



UNIVERSITÉ
DE LORRAINE

BIBLIOTHÈQUES
UNIVERSITAIRES

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact bibliothèque : ddoc-theses-contact@univ-lorraine.fr
(Cette adresse ne permet pas de contacter les auteurs)

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

DOCTEUR DE L'UNIVERSITE DE LORRAINE

Mention : Chimie

par **Clarisse GOSSET-ERARD**

Sous la direction de Pr. Patrick CHAIMBAULT et Dr. Yannis FRANCOIS

**Développement méthodologique en CE-FTICR-MS
pour l'étude de biomolécules à visée thérapeutique**

24 novembre 2022

Membres du jury :

Directeurs de thèse : **Pr. Patrick CHAIMBAULT**

Professeur, LCP-A2MC, Université de Lorraine

Dr. Yannis FRANCOIS

Maître de conférences, LSMIS, Université de Strasbourg

Président de jury : **Pr. Raphaël DUVAL**

Professeur, L2CM, Université de Lorraine

Rapporteurs : **Pr. Reine NEHME**

Professeure, ICOA, Université d'Orléans

Dr. David TOUBOUL

Directeur de recherche, ICSN, Université Paris-Saclay

Examinateur : **Dr. Fanny D'ORLYE**

Maître de conférences, i-CLeHS, Université Paris Sciences & Lettres

Membres invités : **Dr. Emmanuelle LEIZE-WAGNER**

Directrice de recherche, LSMIS, Université de Strasbourg

Pr. Frédéric AUBRIET

Professeur, LCP-A2MC, Université de Lorraine

A ma famille

« Je suis de ceux qui pensent que la science est d'une grande beauté. Un scientifique dans son laboratoire est non seulement un technicien : il est aussi un enfant placé devant des phénomènes naturels qui l'impressionnent comme des contes de fées »

Marie Curie

« Le courage n'est pas l'absence de peur, mais la capacité de la vaincre. »

Nelson Mandela

REMERCIEMENTS

Ça y est ! Le temps, tant attendu, des remerciements est enfin arrivé ! Bien que ce soit une des premières parties lues lorsque l'on ouvre un manuscrit de thèse, c'est souvent la dernière écrite par les doctorants. En effet, les remerciements nous tiennent particulièrement à cœur. Sans toutes les personnes qui y figurent, la belle et surprenante aventure qu'est la thèse ne serait pas la même...

Avant de commencer à remercier toutes les personnes qui ont gravité autour de moi pendant ces trois dernières années, je souhaite remercier sincèrement les membres du jury qui prennent le temps et me font l'honneur d'évaluer mes travaux de thèse : Pr Reine Nehmé, Dr David Touboul, Pr Raphaël Duval, Dr Fanny D'Orlyé, Dr Emmanuelle Leize-Wagner et Pr Frédéric Aubriet.

J'en profite également pour remercier toutes les personnes et institutions impliquées dans le projet RESEX qui a notamment financé ma thèse et l'achat du FTICR.

Passons maintenant à toutes les personnes qui m'entourent depuis trois ans et qui ont rendu cette thèse possible...

Tout d'abord, sans directeur de thèse, pas de doctorant ! J'ai eu la chance d'en avoir deux : Pr Patrick Chaimbault et Dr Yannis François ! C'est donc tout naturellement que vous trouvez votre place au début de ses remerciements. Chacun à votre manière, vous m'avez encadrée, conseillée et poussée vers le haut tout au long de ma thèse. Merci à vous deux de toujours m'avoir mise en avant et d'avoir cru en moi dès le premier jour. Vous avez su m'offrir de l'autonomie, tout en étant présents chaque fois que j'en avais besoin. Vous avez toujours su trouver les mots pour me rassurer face à mes peurs et mes questionnements, qui, il faut l'avouer, n'étaient pas toujours rationnels ! Merci de m'avoir guidée à l'aide de vos nombreux conseils tout au long de cette aventure avec une grande bienveillance. J'ai énormément appris à vos côtés, et si ces trois années m'ont autant fait grandir en tant que chercheuse, mais aussi d'un point de vue personnel, c'est en très grande partie grâce à vous. Pour tout cela, un grand merci !

Associés à mes deux directeurs de thèse, on retrouve mes deux laboratoires d'accueil, le Laboratoire de Chimie et Physique – Approches Multi-échelles des Milieux Complexes (LCP-A2MC) et le Laboratoire de Spectrométrie de Masse et des Systèmes (LSMIS), ainsi que toutes les personnes qui s'y trouvent. Tout d'abord, je remercie Dr Emmanuelle Leize-Wagner et Pr Frédéric Aubriet de m'avoir si gentiment accueillie dans vos équipes respectives, de m'avoir accordé votre confiance et d'avoir initié, par votre amitié, mon projet de thèse. Je vous remercie également pour toutes les discussions scientifiques et non-scientifiques, toujours très intéressantes, que nous avons eues.

L'aventure qu'est la thèse ne serait pas la même si on ne la partageait pas au quotidien avec les collègues doctorants.

Tout d'abord, Mathieu Galmiche, avec qui j'ai commencé ma thèse le même jour. Bien que l'on se connaisse depuis 6 ans maintenant, j'ai réappris à te connaître pendant nos thèses. J'étais déjà impressionnée par ton assurance et ta rapidité de travail, mais j'ai aussi découvert ton sens de l'humour et ta capacité à faire face aux problèmes et aux doutes avec courage et détermination, sans jamais chercher à les éviter. Merci pour ton aide et tes encouragements, tout au long de la thèse mais encore plus lors de ces dernières semaines ! Merci pour tous les bons moments passés ensemble. Je me souviendrai notamment longtemps de notre voyage aux Etats-Unis. Entre les péripéties du trajet, le congrès, les visites de Philadelphie, Washington et New York, et le « quick show time », il y a beaucoup de beaux souvenirs !

Tessa Reinert est ensuite venue compléter cette jolie équipe de doctorants strasbourgeois. D'un labo de doctorants adeptes de café, on a toutes les deux converti petit à petit tous les stagiaires au thé, au point d'avoir autant de thés différents dans la salle de pause que dans un salon de thé ! Tu es toujours pleine d'énergie et tu avances, tout comme Mathieu, à mille à l'heure. Je n'ai aucun doute sur le fait que ta thèse sera incroyable ! Merci pour ta bonne humeur, ton aide, tes encouragements et tous les bons moments passés au laboratoire et en dehors !

À Metz, Anthony Abou Dib a été mon binôme de thèse, mais aussi mon binôme de représentants des doctorants au conseil de labo. Pendant ces trois ans, tu as également été un super gardien de tous mes courriers entre chacune de mes venues. Merci de l'avoir toujours fait de bon cœur ! Merci pour ta bonne humeur contagieuse, pour tes nombreux encouragements et pour tous les bons moments ! Bon courage pour ces derniers mois. C'est promis, il y a bien une lumière au bout du tunnel de la rédaction !

Un grand merci à vous trois, je n'aurais pas pu imaginer meilleurs coéquipiers de thèses ! J'ai beaucoup appris à vos côtés. On a beaucoup ri, on a fêté ensemble chacune de nos réussites et on s'est toujours tirés vers le haut. On a râlé, ensemble ou à tour de rôle, lorsque l'on avait besoin de décompresser. Mais aussi, on s'est serré les coudes dans les moments difficiles. C'était un réel plaisir de travailler avec vous !

Je remercie aussi les doctorants du LCP-A2MC qui sont arrivés plus tard : Farès Slimani et Pierre Pacholski. Merci pour les bons moments partagés ! Et dans quelques jours, ce sera le tour de Théo à Metz et d'Emilie à Strasbourg. Bon courage à vous tous ! Et profitez ! Ça passe si vite trois ans de thèse !

Qui dit doctorants, dit aussi anciens doctorants devenus maintenant docteurs. J'aimerais remercier le super trio qui nous a accueillis, Mathieu et moi, à notre arrivée au laboratoire : Dr Jérémie Giorgetti, Dr Antony Lechner et Dr Christophe Giorgiutti. Merci à vous de nous avoir directement intégrés avec vos jeux de mots les plus pourris et de nous avoir transmis la traditionnelle pause-café que Mathieu ne manque pas de nous rappeler ! Jérémie, merci de m'avoir formée à la CE-MS, à l'analyse des anticorps monoclonaux et à l'attaque des capillaires pendant tes deux derniers mois au laboratoire avant de rejoindre le pays des

caribous ! Antony, merci pour ton calme légendaire à toute épreuve et de m'avoir transmis tes connaissances sur la CE-MS et sur les ARN avec une grande patience ! Christophe, mon co-bureau pendant ma première année de thèse, merci pour le soutien, pour les nombreux encouragements et pour avoir été aussi nul que moi aux fléchettes face à Yannis, Antony et Mathieu ! J'aimerais aussi remercier Dr Yu Min Kiw, l'ancienne doctorante du labo d'à côté. Merci pour nos différentes discussions au pas de la porte ou dans le couloir, c'était toujours très agréable de partager ces moments avec toi !

Je tiens aussi à remercier les (anciens et nouveaux) post-doctorants de Metz : Dr Ali Zaïter, Dr Julien Comel et Dr Marisa Maia. Merci pour votre bonne humeur, votre soutien et votre accueil au laboratoire ! On ne s'est pas beaucoup vu, mais j'ai apprécié chacune de nos petites discussions ! A special part in English for you Marisa : Thank you for your joy, your kindness, your support and all the little discussions we had in the lab or at Analytics ! Thank you also for always making me feel welcome even if I was only there for a few weeks over the last year. It was a real pleasure to get to know you during this last year of my PhD !

Je remercie les autres membres permanents du LSMIS pour leur accueil. Des remerciements particuliers sont adressés à Noëlle : merci pour ta gentillesse, et pour toutes les discussions scientifiques très intéressantes que l'on a eues. C'était toujours un plaisir de discuter avec toi et d'avoir ta vision extérieure sur mes travaux.

Du côté du LCP-A2MC, je souhaite remercier chaleureusement Dr Vincent Carré, Dr Sébastien Schramm, Dr Lionel Vernex-Loset, Dr Gabriel Krier et Dr Jasmine Hertzog. En plus d'être des collègues, les amitiés profondes qui lient tous les membres de l'équipe transparaissent dans votre travail. Merci de m'avoir accueillie les bras ouverts dès mon premier jour à Metz, et de me faire me sentir chez moi, même avec mes affaires dans une valise et tous les équipements du labo dans ma voiture. Un grand merci également pour l'aide précieuse face au FTICR qui était bien souvent capricieux !

Après les doctorants, les post-doctorants et les permanents des deux équipes, il est temps de remercier tous les stagiaires que j'ai eu l'occasion de rencontrer : Florian, Geoffrey, Clémence, Louis, Emilie, Oumayma, Aurore, Théo, Anaïs, Arthur, Tugba, Pierre, Max et Mévie. Merci à tous pour la bonne ambiance et l'air frais que vous avez amené aux laboratoires pendant quelques mois ! Des remerciements particuliers sont destinés à Oumayma et Mévie que j'ai co-encadrées. J'ai beaucoup appris en vous encadrant, notamment à être pédagogue.

Mévie, tu as été notre super stagiaire bio-info ! Merci à toi pour l'immense travail que tu as fourni pendant ton stage, et pour avoir grandement contribué à mes travaux de thèse de dernière année. Merci également d'avoir eu le courage de mettre les pieds dans un laboratoire sans aucun bio-informaticien pour t'aider sur le code, même si Jérôme a ensuite rejoint le projet !

Un petit mot spécial également pour Clémence qui était toujours partante pour une pause jardinage ou pour une soirée burger-jeux de société avec Tessa. Merci pour cette belle rencontre qui s'est transformée en amitié !

Un mot également pour Stéphanie Coutin : même si tu ne fais pas officiellement partie du laboratoire, c'est tout comme ! Merci pour les agréables discussions à la pause de midi, mais aussi pour ta bonne humeur, ta gentillesse, tes conseils et tes encouragements ! Tu as toujours su trouver les mots pour m'aider à y voir plus clair dans les moments de doutes.

Si mes travaux de thèse ont pu se développer sur les ARN à partir de la fin de la première année, c'est grâce à Philippe Wolff et Dr Antony Lechner (Anto, tu as le droit à deux remerciements !), qui sont nommés « les collègues biologistes de l'IBMC » dans ce manuscrit. Merci à vous deux d'avoir répondu à mes innombrables questions avec toujours beaucoup de patience. J'ai tellement fait d'allers-retours entre le labo et l'IBMC, que la dame de l'accueil n'est même plus étonnée de me voir arriver, de temps en temps, plusieurs fois dans la même journée. J'ai beaucoup appris sur les ARN et la biologie grâce à vous. C'était un réel plaisir de travailler ensemble.

À l'IBMC, je souhaite également remercier Lauriane Kuhn. Tu es un véritable puits de connaissances et tu n'hésites pas à prendre le temps pour nous les partager. Merci d'avoir répondu à mes nombreuses questions sur les logiciels de protéomique avec la patience et la gentillesse qui te caractérisent. J'en profite également pour remercier les autres membres de la plateforme protéomique Johana Chicher et Philippe Hammann pour leur gentillesse.

Enfin, pour finir la liste des personnes avec qui j'ai travaillé, il manque Dr Jérôme Pansanel. Merci d'avoir rejoint les projets « Nucleos'ID » et « OligoMod+ » et d'avoir pris le relai sur le travail fait par Mévie. Merci pour ta patience face à nous qui, il faut le dire, ne connaissons pas grand-chose en informatique et en codage.

Ces remerciements ne pourraient pas être complets sans citer Jean Michel Billmann et Dominique Melon. Vous avez sauvé plus d'une fois ma thèse lors de grands moments de désespoirs. Sans vous, je pense que l'on serait encore tournevis ou carte électronique à la main ! J'ai également beaucoup appris grâce à vous. Merci d'avoir répondu avec une grande gentillesse à chacun de nos mails désespérés et à toutes mes questions. Un grand merci de m'avoir transmis une partie de vos connaissances si précieuses pour les analyticiens.

Je remercie également nos voisins de laboratoire : Dr Pierre Adam, Dr Philippe Schaeffer et Estelle Motsch. En plus d'être des voisins très agréables, toujours souriants et prêts à aider en prêtant le tournevis manquant à notre boîte à outils, vous êtes aussi les premières personnes qui m'ont fait réellement découvrir le monde de la recherche académique. Qui l'eut cru que, quelques années plus tard, le destin me ferait poser mes valises quelques bureaux plus loin, chez l'équipe voisine. Mon passage dans votre laboratoire a été de très courte durée, mais c'est à ce moment-là que mon amour pour la recherche a débuté... Alors merci !

Pour continuer mon histoire avec la recherche académique avant la thèse, j'aimerais remercier Pr Saïd Ennahar. Merci d'avoir eu confiance en moi lors de mon stage M2 et de m'avoir poussée à dépasser mes peurs pour me lancer dans le doctorat.

En plus de mes travaux de recherche, j'ai eu l'occasion de réaliser des missions d'enseignements à Metz et à Strasbourg. Cette expérience a été très enrichissante, et si cela s'est aussi bien passé, c'est en grande partie grâce aux personnes suivantes.

À Metz, je souhaite remercier Dr Sophie Legeai et Cathy Aubert de m'avoir fait confiance. Merci également pour votre bonne humeur, votre gentillesse et pour les nombreux conseils que vous m'avez donnés. C'est toujours un plaisir de vous recroiser et de discuter ! Merci également à Vincent de m'avoir accordé ta confiance pour ces TP.

À Strasbourg, mes premiers remerciements vont à Thomas Weissenberger. Merci d'avoir veillé à ce que les TP se passent si bien et que l'on ne manque de rien. Je n'oublierai pas tes encouragements (même à l'autre bout de la planète) dignes du plus grand fan du Racing Club de Strasbourg ! Afin d'assurer les TP, on a fait équipe avec de super doctorantes (devenues pour deux d'entre elles docteures) : Tessa, Dr Anaïs Rodrigues et Dr Supansa Chimjarn (Méry). Avec une telle équipe, les TP ne pouvaient que bien se dérouler, toujours dans la bonne humeur. Merci à vous quatre d'avoir rendu cette expérience d'enseignement si agréable ! Et Yannis, merci de nous avoir fait confiance pour encadrer ces TP !

Enfin, pour la partie enseignement, je souhaite remercier tous les élèves qui ont croisé mon chemin. Merci pour votre enthousiasme. Si j'en retire une si grande satisfaction c'est aussi grâce à vous. Les semaines de TP étaient pour moi de vraies bouffées d'air frais, qui m'ont permis de prendre un peu de recul sur mes problèmes de recherche.

Tout au long de ma vie, j'ai eu la chance de rencontrer d'autres personnes formidables qui m'ont inspirée, aidée, fait grandir ou beaucoup appris... Toutes ces personnes ont contribué de manière indirecte à la réussite de cette thèse, elles méritent donc tout autant une place dans ces remerciements !

Tout d'abord, deux enseignants du lycée m'ont particulièrement marqué : Mr Matthieu Perrin et Mlle Stéphanie Milonet. Vous êtes tous les deux des sources d'inspiration dans ma manière d'enseigner. Merci !

Claudine Weber, professeur de danse qui m'a suivie de mes 9 à mes 19 ans, fait également partie des personnes qui ont marqué ma vie. Grâce à vous, j'ai notamment appris à faire preuve de rigueur, de patience, et de persévérance, et tout cela m'a grandement aidé pendant ma thèse. Finalement, entre la danse et la science, il y a beaucoup de points communs ! Je souhaite remercier par la même occasion toutes les filles avec qui j'ai partagé de magnifiques souvenirs tout au long de ces années : Sarah, Laurine, Elsa, Camille, Amélie, Marion, Elise, Justine C., Justine O., Ophélie, Alexandra, Marie, Sofia...

Pendant mes deux années de prépa, j'ai eu la chance de tomber sur un duo de choc : Emmanuel Bougnol et Dr Jean-Julien Fleck (surnommés, avec beaucoup de respect, Boubou et JJ !), enseignants passionnés de maths et physique de la PCSI1. En plus d'avoir été d'un soutien sans faille pendant les deux ans de prépa, vous m'avez appris à toujours me battre jusqu'au bout pour les choses qui me tiennent à cœur (oui oui, j'ai encore cette petite voix

dans ma tête qui crie « Jusqu'au bout ! Toujours jusqu'au bout ! » même 7 ans plus tard !). Cette leçon m'a grandement aidée pendant les trois dernières années.

Ensuite, même si nos chemins se sont éloignés, j'aimerais remercier Amandine, Élodie, Madeline, Laetitia, Chloë, Marion, Clément, Victor, David, Laurent, Matthieu et Anthony. Vous faites partie des personnes qui ont beaucoup compté à un moment de ma vie, qui m'ont fait grandir et devenir la personne que je suis aujourd'hui.

Enfin, comment ne pas remercier Pr Véronique Hubscher... Je vous l'ai déjà dit, un simple "merci pour tout" ne sera jamais à la hauteur de ce que vous m'avez apporté, bien que ce soit un bon résumé. Je suis extrêmement reconnaissante que nos routes se soient croisées à un moment important de ma vie. Merci pour la petite bulle d'apaisement et de sérénité que m'avez offert lors de nos nombreuses discussions. Un immense merci pour tout le temps que vous m'avez consacré et de m'avoir tendu votre main avec l'empathie, la gentillesse et la bienveillance qui vous caractérisent tant.

Pour la dernière partie de ces remerciements, je souhaite remercier tous mes proches qui ont été d'un soutien sans faille tout au long de cette thèse, mais aussi depuis bien plus longtemps !

En première position, mes amis d'enfance (surnommés « les loulous » ou « les petits loups ») : Flore, Camille, Juliette, Thomas, Philomin et Ioana.

Flore, tu es un vrai rayon de soleil et tu es toujours prête à défendre tes amis contre vents et marées. Nous avions été toutes les deux été surnommées « reines de la chimie » en seconde (on gardera les raisons de ce surnom entre nous), et me voilà des années plus tard en train d'écrire ses remerciements... Il faut croire que ce surnom a enfin trouvé son sens ! Camille, pleine de patience et de bienveillance, tu veilles toujours à ce que l'on se sente bien et que l'on prenne conscience de nos valeurs. Merci pour nos nombreuses discussions à cœur ouvert... Juliette, tu es simple, authentique et sans filtre. De belles qualités que peu de personnes possèdent, mais qu'il faut savoir appréhender. Avec toi, l'amitié c'est tout ou rien ! Ta grande maladresse aura déclenché un bon nombre de fous rires et tes lundis soirs à Strasbourg auront été de vraies bouffées d'air frais pendant ma thèse ! Thomas (surnommé « Sale T » ou « Propre T » en fonction des situations et de ton envie d'être chafouin... je ne voulais pas le dire, mais tu as insisté :p), tu es mon acolyte chimiste de la bande. Merci pour les longues heures passées au téléphone à discuter des joies et difficultés de la thèse, mais aussi du reste de nos vies. Philomin, grand amoureux de la nature et des animaux, merci pour ton immense gentillesse, ta bienveillance et ton écoute toujours très attentive. Et enfin Ioana, fan incontestable des GIFs de loutres, merci pour ta gentillesse et pour ton sens de l'humour !

Tous si différents, et à la fois si complémentaires, vous êtes un des grands piliers de ma vie. Être éparpillés aux quatre coins de la France ne change rien, on est toujours là les uns pour les autres, que ce soit pour partager et fêter nos victoires et nos bonheurs, ou pour se soutenir dans les moments difficiles. Vous ne vous en êtes peut-être pas rendu compte, mais

vous avez eu une grande implication dans la réussite de cette thèse. Merci pour ces amitiés si précieuses à mes yeux, pour votre soutien sans faille, pour vos encouragements, pour nos fous rires et nos moments de folie (le dernier en date étant notre petite activité pédalo au mariage de Juliette !), pour nos weekends jeux de société, pour nos longues discussions et pour tout le reste...

Je ne pouvais pas finir ce (très) long paragraphe qui vous est dédié sans un petit mot spécial « thèse » pour Thomas : On y sera arrivés à la fin de nos thèses ! De notre balade autour du Malsaucy au moment où l'on doutait tous les deux de nos thèses, à la fin de la rédaction de nos manuscrits.... Que de chemin parcouru ! Bon courage à toi pour cette dernière ligne droite ! Acolytes au point de soutenir à 24 h d'intervalle, nous fêterons nos deux thèses comme il se doit le 25 novembre !

Mes prochains remerciements sont adressés à Evolène (ou plutôt Dr Evolène depuis presque un an maintenant !), d'abord marraine à l'école, puis amie ! Merci pour tes petites anecdotes toujours très divertissantes et pour ton soutien ! Et comme pour Mathieu, on se souviendra longtemps de nos péripéties de voyage à Philadelphie !

Un autre grand pilier de ma vie est bien évidemment ma famille, mes derniers remerciements ainsi que ma thèse leur sont donc dédiés. Merci à mes oncles, tantes, cousins, cousines... que j'ai toujours plaisir à revoir ! Un remerciement spécial à Margot qui m'a toujours encouragée dans mes études et dans ma thèse ! Un grand merci à mes grands-parents, Christiane et Cyrille, pour leur amour et leur soutien. Vous m'avez toujours encouragée et avez toujours cru en moi. J'espère vous rendre fiers ! Merci à mes frères et sœur, Flora, Sacha et Arnaud. Plus ou moins âgés que moi, peu importe, j'ai beaucoup appris de vous et j'apprends encore tous les jours. Il m'est impossible de ne pas écrire un petit mot pour nos deux petites boules de poils, Grizzly et Hindi, qui font s'évader tous nos soucis et qui sont toujours d'un grand réconfort. Enfin, probablement le plus important pour la fin, je remercie infiniment mes parents : Jacques, Isabelle et Robert. Merci à vous trois pour votre amour et votre soutien inconditionnel depuis toujours et dans toutes les grandes étapes de ma vie. Vous avez toujours cru en moi, et su me changer les idées dans les moments difficiles, chacun à votre manière. Un remerciement particulier pour toi maman : merci d'avoir relu toute ma thèse avec une grande attention !

Sans vous tous, je n'en serais pas là aujourd'hui... On a tous nos qualités et nos défauts, mais ensemble on est capables de tout affronter ! Pour vous aussi, un simple « merci pour tout » ne suffira jamais...mais...Merci du fond du cœur !

J'aimerais clôturer ces remerciements par une pensée émue pour mes grands-parents, Jean, Germaine, et Mariam, qui, j'en suis sûre, veillent sur moi au quotidien. J'aurais aimé voir vos yeux remplis de fierté en vous montrant ce manuscrit.

Remerciements

TABLE DES MATIERES

Remerciements	3
Table des matières	11
Liste des abréviations.....	15
Liste des figures.....	19
Liste des tableaux.....	19
Introduction générale	31
PARTIE I : Introduction et revue bibliographique	33
CHAPITRE 1 : Electrophorèse capillaire et spectrométrie de masse.....	37
1.1. Introduction	37
1.2. Electrophorèse capillaire	37
1.3. Spectrométrie de masse	42
1.4. Couplage entre l'électrophorèse capillaire et la spectrométrie de masse	55
1.5. Références	57
CHAPITRE 2 : Les couplages entre la FTICR-MS et les méthodes séparatives	61
2.1. Introduction	61
2.1. Publication : « Hyphenation of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FTICR-MS) with separation methods: the art of compromise and the possible - a review »	62
2.2. Références	105
CHAPITRE 3 : Les acides ribonucléiques (ARN).....	119
3.1. Généralités sur la structure des ARN.....	119
3.2. Les différents types d'ARN et leurs rôles.....	120
3.3. Les modifications post-transcriptionnelles	124
3.4. La caractérisation des modifications post-transcriptionnelles des ARN par spectrométrie de masse	128
3.5. L'interprétation des données	137
3.6. Références	139

PARTIE II : Mise en place et étude du couplage CE-FTICR-MS.....	145
CHAPITRE 4 : Mise en place et étude du couplage CE-FTICR-MS.....	149
4.1. Introduction	149
4.2. Optimisation des méthodes de détection des nucléosides	154
4.3. Etude des performances du couplage CE-FTICR-MS	158
4.4. Optimisation du mode automatique pour la MS/MS.....	165
4.5. Comparaison des couplages CE-FTICR-MS et CE-QTOF-MS	167
4.6. Conseils pratiques.....	170
4.7. Conclusion.....	172
4.8. Références bibliographiques	173
CHAPITRE 5 : Evaluation du transfert de méthode vers le couplage CE-FTICR-MS	175
5.1. Introduction	175
5.2. Preuve de concept sur des ARNt purifiés	176
5.3. Preuve de concept sur un extrait total d'ARNt.....	179
5.4. Conclusion.....	183
5.5. Références	184
PARTIE III : Développement méthodologique pour l'étude des modifications post-transcriptionnelles des ARN	185
CHAPITRE 6 : Optimisation de la déphosphorylation des nucléotides pour la caractérisation structurale des ARN.....	189
6.1. Introduction	189
6.2. Publication : Optimization of nucleotides dephosphorylation for RNA structural characterization by tandem mass spectrometry hyphenated with separation methods	189
6.3. Conclusion.....	205
6.4. Références	205
CHAPITRE 7 : Caractérisation des modifications post-transcriptionnelles d'un ARN ribosomique.....	209
7.1. Introduction	209
7.2. Premier niveau d'analyse : identification des nucléosides modifiés.....	210
7.3. Second niveau d'analyse : oligonucléotides	211
7.4. Conclusion.....	214
7.5. Références	215

CHAPITRE 8 : Développement de deux logiciels pour la caractérisation des ARN à partir d'analyses en mode positif	217
8.1. Introduction	217
8.2. Publication : “Nucleos’ID: new search engine enabling the untargeted identification of RNA post-transcriptional modifications from tandem mass spectrometry analyses of nucleosides”	218
8.3. OligoMod+ : nouveau logiciel pour localiser les modifications post-transcriptionnelles à partir d’analyses MS/MS d’oligonucléotides en mode positif	241
8.4. Conclusion.....	243
8.5. Références	244
PARTIE IV : Mise en application du couplage CE-FTICR-MS	249
CHAPITRE 9 : Mise en application du couplage CE-FTICR-MS	251
9.1. Introduction	251
9.2. ARNr 16S de <i>Pseudomonas aeruginosa</i>	252
9.3. Extrait total d’ARNt de <i>Thermus thermophilus</i>	256
9.4. Conclusion.....	261
9.5. Références	262
Conclusion générale et perspectives	265
Liste des communications scientifiques et autres responsabilités	269
Publications à comité de lecture	269
Publications soumises	269
Chapitre de livre	270
Communications orales.....	270
Communication par affiche.....	272
Enseignements délivrés.....	273
Autres responsabilités.....	273
Annexes	275
Annexe 1 : Chapitre de livre sur l’attaque des capillaires.....	275
Annexe 2 : Poster présenté à Analytics 2022 (Nantes).....	285

LISTE DES ABREVIATIONS

A : adénine/adénosine

ADN : acide désoxyribonucléique

APCI : ionisation chimique à pression atmosphérique

API : ionisation à pression atmosphérique

APII : photo-ionisation par laser à pression atmosphérique

APPI : photo-ionisation à pression atmosphérique

ARN : acide ribonucléique

ARNc : acide ribonucléique codant

ARNdb : acide ribonucléique double brin

ARNm : acide ribonucléique messager

ARNmi : micro acide ribonucléique

ARNnc : acide ribonucléique non codant

ARNr : acide ribonucléique ribosomique

ARNs : petit acide ribonucléique

ARNsi : petit acide ribonucléique interférent

ARNsn : petit acide ribonucléique nucléaire

ARNsno : petit acide ribonucléique nucléolaire

ARNt : acide ribonucléique de transfert

ASO : oligonucléotide antisens

BAP : phosphatase alcaline bactérienne

BGE : électrolyte support

C : cytosine/cytidine

CE : électrophorèse capillaire

CEC : électro-chromatographie capillaire

CID : dissociation induite par collision

CZE : électrophorèse capillaire de zon

DDA : acquisition dépendante des données

EC : cellule électrochimique

ECD : dissociation par capture d'électron

EI : impact électronique

ESI : ionisation par électrospray

ETD : dissociation par transfert d'électron

FDR : taux de faux positifs

FN : faux négatif

FP : faux positif

FT : transformées de Fourier

G : guanine/guanosine

GC : chromatographie gazeuse

GCC : colonne de charbon graphite

GCE : électrophorèse capillaire en gel

HFIP : 1,1,1,3,3-hexafluoro-2-propanol

HIC : chromatographie à interaction hydrophobique

ICR : résonance cyclotronique des ions

ID : diamètre interne

IRMPD : dissociation multiphotonique par infrarouge

ISD : dissociation en source

LC : chromatographie liquide

mAb : anticorps monoclonal

MCP : galette de microcanaux

MEKC : électrochromatographie en phase micellaire

MGF : format générique de Mascot

MRM : monitoring de réactions multiples

MS : spectrométrie de masse

MS/MS : spectrométrie de masse en tandem

NMP : nucléotide monophosphate

NPLC : chromatographie liquide en phase normale

OD : diamètre externe

PAGE : électrophorèse sur gel de polyacrylamide

PAH : hydrocarbure aromatique polycyclique

PMS : pH-mediated stacking

PP : polypropylène

PTM : modification post-traductionnelle

Q : quadrupôle

RNase : ribonucléase

RPLC : chromatographie liquide en phase inverse

RT-PCR : transcriptase inverse - réaction en chaîne par polymérase

tITP : isotachophorèse transient

TN : vrai négatif

TNR : taux de vrais négatifs

TOF : temps de vol

TP : vrai positif

TPR : taux de vrais positifs

U : uracile/uridine

UVPD : photodissociation par ultraviolet

VIH : virus de l'immunodéficience

VOC : composé organique volatil

LISTE DES FIGURES

PARTIE I : Introduction et revue bibliographique

Chapitre 1 : Electrophorèse capillaire et spectrométrie de masse

Figure 1.1: Représentation schématique du système d'électrophorèse capillaire.	38
Figure 1.2 : Représentation schématique de la formation de la double couche de charges dans un capillaire de silice vierge.	40
Figure 1.3 : Représentation schématique (A) d'une séparation par CZE sur un capillaire en silice vierge et (B) de l'électrophérogramme obtenu par cette séparation.	41
Figure 1.4 : Processus d'ionisation par électrospray menant à l'apparition d'un cône de Taylor et à la production de gouttelettes chargées.	44
Figure 1.5 : Modèle de Dole (CRM) et d'Iribarne-Thomson (IEM) décrivant la formation des ions en phase gazeuse à partir de gouttelettes chargées.....	45
Figure 1.8 : Représentation schématique du spectromètre de masse Maxis 4G (Bruker Daltonics).....	47
Figure 1.9 : Représentation schématique du spectromètre de masse 7T SolariX 2XR (Bruker Daltonics).....	47
Figure 1.8 : Représentations schématiques d'un quadrupôle sous différents angles.....	48
Figure 1.9 : Diagramme de stabilité de la trajectoire des ions dans un quadrupôle.....	49
Figure 1.10 : Représentation schématique d'un TOF.	50
Figure 1.11 : Représentation schématique du fonctionnement d'un analyseur TOF équipé d'un réflectron.....	51
Figure 1.12 : Représentation schématique d'une cellule ICR cylindrique dans un champ magnétique B_0	51
Figure 1.13 : Représentation schématique (A) des trois mouvements d'un ion confiné dans une cellule ICR et soumis à un champ magnétique B_0 et (B) de leur combinaison pour obtenir la trajectoire globale de l'ion.	53
Figure 1.14 : Représentations schématiques (A) d'une CE utilisant l'interface sheathless et (B) de l'interface sheathless en elle-même.	56
<u>Chapitre 2 : Les couplages entre la FTICR-MS et les méthodes séparatives</u>	
Figure 2.1: Brief principle of ultra-high resolution mass spectrometry using FTICR-MS.	64

Figure 2.2: Schematic representation of (a) the three natural motions of an ion confined in an ICR cell and (b) the resulting ion trajectory shape.....	66
Figure 2.3: (a) Example of the ultra-high-resolution MS analysis using a FT-ICR MS. (b) Resolution of isotope ions using a FT-ICR MS.....	67
Figure 2.4: Schematic view of an HPLC-ESI-FT-ICR MS and its differential pumping stages from the atmospheric pressure ion source to the of 10^{-10} mbar-ultra high vacuum of the ICR cell (Bruker Daltonics BioAPEX -94e fitted with a 9,4T magnet).....	69
Figure 2.5: Design of results of the method optimization experiments from the analysis of PBDE-99, a brominated flame retardant.....	73
Figure 2.6: (Top) Detection set-up for normal detection (1ω) and for detection at twice the cyclotron frequency (2ω). (Bottom) The image current resulting from one revolution of the ion packet for both 1ω and 2ω detection schemes.....	76
Figure 2.7: Capillary LC-FTICR 2-D display for a portion of the analysis of a yeast global soluble protein tryptic digest that contained a total of ~110 000 detected components.....	79
Figure 2.8: FcγRIIIa affinity extracted ion chromatograms of therapeutic mAb detected glycoforms.....	80
Figure 2.9: Chromatogram and mass spectra of proteolytic subunits from IdeS digestion....	82
Figure 2.10: Positive ion mode spectra of <i>A. thaliana</i> leaf metabolites.....	85
Figure 2.11: Comparison of productions of ajmalicine in the MS/MS analysis (upper) and the MS/MS boost analysis (lower).....	88
Figure 2.12: Total ion chromatogram and extracted ions chromatograms of an oil sands processed water sample showing separation of several isomers within O, O ₂ , O ₃ , C, S and SO classes.....	91
Figure 2.13: Total ion chromatogram for one sampling depth (67.5–69.5 cm) and the mass spectra resulting from averaging the acquired scans over 10 min intervals.....	92
Figure 2.14: Extracted ions chromatograms (<i>m/z</i> 135.081539) from a GC-APCI-FT-ICR MS analysis of pyrolysis oil from German brown coal, showing the separation of 9 propylphenol, ethylmethylphenol, and trimethylphenol isomers.	93
Figure 2.15 : Analytical timescales based on the speed of separation of analytical strategies (on the left) and on the total number of potential spectra that can be obtained in a 20-minute chromatogram (on the right).	95

Chapitre 3 : Les acides ribonucléiques

Figure 3.1 : (A) Structure primaire d'une séquence d'ARN composée des quatre nucléosides canoniques et (B) représentation simplifiée de la chaîne de l'ARN.....	119
Figure 3.2 : Motifs créés par appariement des nucléotides des ARN.....	120

Figure 3.3 : (A) Structure secondaire d'un ARNt en forme de feuille de trèfle. (B) Structure 3D d'un ARNt en forme de L inversé.....	121
Figure 3.4 : Reconnaissance par appariement entre un codon d'un ARNm et un anticodon d'un ARNt.....	122
Figure 3.5 : Représentation simplifiée des étapes de la traduction de l'ARNm en protéine.	123
Figure 3.6 : Structure secondaire de l'ARNr 16S de <i>E. coli</i>	123
Figure 3.7 : Exemple d'écriture simplifiée d'un nucléotide modifié (N6,N6,2'-O-triméthyladenosine) et correspondance avec la structure chimique.....	125
Figure 3.8 : Structure 3D d'une interaction codon-anticodon avec ms ² i6A en position 37 de l'ARNt.....	127
Figure 3.9 : Stratégie générale de caractérisation des modifications post-transcriptionnelles des ARN par spectrométrie de masse.....	129
Figure 3.10 : (A) Première dimension dénaturante et (B) seconde dimension semi-dénaturante d'un gel 2D PAGE d'un mélange d'ARNt.....	130
Figure 3.11 : Sites de coupure des liaisons phosphodiesters des ARN et produits formés après coupure..	132
Figure 3.13 : Structures des modifications (A) m ⁶ A et (B) Am.....	134
Figure 3.14 : Nomenclature de McLuckey pour la fragmentation des oligonucléotides et le séquençage De Novo.....	135
Figure 3.15 : Système CE-MS microfluidique « ZipChip » développé par 908devices avec (A) un module qui se positionne directement à l'entrée des spectromètres de masse et (B) la cartouche dans laquelle se glisse (C) la puce en verre où sont gravés les canaux microfluidiques (908devices).	137

PARTIE II : Mise en place et étude du couplage CE-FTICR-MS

Chapitre 4 : Mise en place du couplage CE-FTICR-MS

Figure 4.1 : (A) Structure d'un mAb. (B) Structures des principales N-glycosylations présentes sur les mAbs.	149
Figure 4.2 : Stratégies d'analyses d'un anticorps monoclonal aux niveaux (A) bottom-up, (B) middle-up et (C) middle-down, (D) intact et (E) top-down. DTT : dithiothréitol ; TCEP : tris(2-carboxyéthyl)phosphine.....	150
Figure 4.3 : Electrophérogrammes d'analyses de l'adalimumab aux niveaux (A) intact et (B) middle-up obtenus avec le couplage CE-QTOF-MS, et spectres de masse déconvolus associés.	151

Figure 4.4 : (A) Electrophérogramme d'analyse bottom-up de l'infliximab Remicade obtenu avec le couplage CE-FTICR-MS/MS. et recouvrements de séquence associés à la MS (en rouge) et à la MS/MS (souligné en bleu) (B) de la chaîne lourde et (C) de la chaîne légère.....	152
Figure 4.5 : Spectres de masse déconvolués de l'infliximab Remicade au niveau intact obtenus par infusion avec les couplages (A) CE-QTOF-MS et (B) CE-FTICR-MS.....	153
Figure 4.6 : Spectres de masse déconvolués de l'infliximab Remicade au niveau middle-up obtenus par infusion avec les couplages (A) CE-QTOF-MS et (B) CE-FTICR-MS.	153
Figure 4.7 : Représentation schématique du spectromètre de masse FTICR SolariX 7T 2XR FTMS.....	154
Figure 4.8 : Spectre MS après optimisation de la méthode de détection, issu d'une infusion CE-FTICR-MS d'un échantillon d'ARN ^{Asp GUC} (<i>S. cerevisiae</i>) digéré à la P1 et déphosphorylé à la BAP.	155
Figure 4.9 : Spectres MS/MS de A, U, C, G, m ⁵ C et m ¹ G après optimisation de l'énergie de collision à 15eV, issus d'une infusion CE-FTICR-MS/MS d'un échantillon d'ARN ^{Asp CUC} (<i>S. cerevisiae</i>) digéré à la P1 et à la BAP.....	158
Figure 4.10 : Evolution de (A) l'aire, (B) l'intensité, (C) le rapport signal sur bruit (S/N), (D) le nombre de points pour les quatre pics électrophorétiques extraits, en fonction du temps d'accumulation des ions.....	160
Figure 4.11 : Evolution de (A) l'aire, (B) l'intensité, (C) le rapport signal sur bruit (S/N), (D) le nombre de points, pour les quatre pics électrophorétiques extraits, (E) la résolution MS et (F) le temps d'acquisition dans la cellule ICR en fonction du DataPoint Size.	161
Figure 4.12 : Evolution de (A) l'aire, (B) l'intensité, (C) le rapport signal sur bruit (S/N), (D) le nombre de points, (E) la résolution MS pour les quatre pics électrophorétiques extraits et (F) le temps d'acquisition dans la cellule ICR en fonction du Low <i>m/z</i> du FTICR-MS.	162
Figure 4.13 : Evolution de (A) l'aire, (B) l'intensité, (C) le rapport signal sur bruit (S/N), (D) le nombre de points, (E) la résolution MS pour les quatre pics électrophorétiques extraits et (F) la durée d'un cycle MS en fonction des différentes combinaisons de temps d'accumulation et de DataPoint Size détaillées dans le Tableau 4.5.....	163
Figure 4.14 : Electrophérogramme obtenu par CZE-MS, issu d'une séparation CE-MS des quatre nucléosides canoniques A, C, U et G à 56,5 µM.	164
Figure 4.15 : Evolution de la résolution en fonction du temps d'acquisition dans la cellule ICR. Le temps d'acquisition dans la cellule ICR a été défini en changeant le DataPoint Size.	168
Figure 4.16 : Erreurs moyennes sur la mesure de masse des quatre nucléosides canoniques pour (A) le couplage CE-QTOF-MS et (B) le couplage CE-FTICR-MS (n=9).	169
Figure 4.17 : Position de la lockmass dans l'interface CESI-MS du FTICR SolariX 7T 2XR (Bruker)	
	172

Chapitre 5 : Evaluation du transfert de méthode vers le couplage CE-FTICR-MS

- Figure 5.1 :** Séquences des quatre ARNt étudiés issus de *S. cerevisiae* (ARNt^{Phe GAA}, ARNt^{Trp CCA}, ARNt^{Asp GUC} et ARNt^{Ala AGC}). 176
- Figure 5.2 :** Electrophérogrammes, obtenus par CE-FTICR-MS/MS, des (A) ARNt^{Phe GAA}, (B) ARNt^{Asp GUC}, (C) ARNt^{Ala AGC} et (D) ARNt^{Trp CCA}, issus de *S. cerevisiae*, après digestion à la Nucléase P1 et déphosphorylation à la BAP..... 177
- Figure 5.3 :** Séquences des ARNt de l'extrait total de *B. taurus*. 180
- Figure 5.4 :** Electrophérogramme, obtenu par CE-FTICR-MS/MS, de l'extrait total d'ARNt de *B. taurus*, après digestion à la Nucléase P1 et déphosphorylation à la BAP..... 182

PARTIE III : Développement méthodologique pour l'étude des modifications post-transcriptionnelles des ARN

Chapitre 6 : Optimisation de la déphosphorylation des nucléotides pour la caractérisation structurale des ARN

- Figure 6.1:** Multiple extracted ion electropherograms for the analysis of total tRNA extract from *B. taurus* after digestion with nuclease P1 and dephosphorylation after a 4-hour incubation at 37°C with 0.75 U of BAP per µg of RNA. 195
- Figure 6.2:** Multiple extracted ion electropherograms for the analysis of total tRNA extract from *B. taurus* after digestion with nuclease P1 and dephosphorylation with (A) 0.75 U of BAP for 1 µg of RNA incubated for 4 hours at 37 °C, (B) 7.5 U of BAP for 1 µg of RNA incubated for 4 hours at 37 °C and (C) 3.0 U of BAP for 1 µg of RNA incubated for 4 hours at 37 °C. 196
- Figure 6.3:** Estimated dephosphorylation efficiencies for 29 nucleotides after (A) 4-hour incubation and (B) overnight incubation at 37°C with seven quantities from 0.15 to 7.5 U of BAP for 1 µg of RNA. 197
- Figure 6.4:** Extracted ion electropherogram (*m/z* 378.08) for the analysis total tRNA extract from *B. taurus* after digestion with nuclease P1 and dephosphorylation after a 4-hour incubation at 37°C with (A) 0.15 U of BAP per µg of RNA, (B) 0.75 U of BAP per µg of RNA, (C) 1.5 U of BAP per µg of RNA, (D) 3.0 U of BAP per µg of RNA and (E) 4.5 U of BAP per µg of RNA. (F) Structures of all four isomers of methylated monophosphate guanosine. 198

Chapitre 7 : Caractérisation des modifications post-transcriptionnelles d'un ARN ribosomique

Figure 7.1 : Electrophérogramme des modifications de l'ARNr 16S issu de *P. aeruginosa*, obtenu par CZE-MS/MS après digestion à la Nucléase P1 et déphosphorylation à la BAP... 210

Figure 7.2 : Electrophérogramme des oligonucléotides de l'ARNr 16S issu de *P. aeruginosa*, obtenu par CZE-MS/MS après digestion à la RNase T1 et déphosphorylation à la BAP. 211

Figure 7.3 : Spectre MS/MS de CAAG issu de la fragmentation par CID (15,8 eV) de l'ion *m/z* 624,13 (+2). 212

Figure 7.4 : Recouvrement de la séquence canonique de l'ARNr 16S issu de *P. aeruginosa* digéré à la RNase T1 et à la BAP..... 213

Chapitre 8 : Développement de deux logiciels pour la caractérisation des ARN à partir d'analyses en mode positif

Figure 8.1: Workflow of identification of RNA modifications by CE-MS/MS or LC-MS/MS.. 223

Figure 8.2: Friendly-user interface of Nucleos'ID with its default settings. 225

Figure 8.3: Multiple extracted ion electropherograms for the nucleosides analysis of tRNA^{Phe GAA} from *S. cerevisiae* obtained after manual interpretation. 228

Figure 8.4: Evolution of the numbers of (A) true positives, (B) false positives, (C) true negatives, (D) false negatives, as well as (E) the TPR, the TNR and (F) the FDR, depending on the score threshold used for the Nucleos'ID search for *B. taurus* total tRNA extract..... 232

Figure 8.5 : Interface de la première version du logiciel OligoMod+. 241

Figure 8.6 : Principe expérimental et fonctionnement du logiciel OligoMod+ pour la vérification des positions des modifications post-transcriptionnelles à partir d'analyse CE-MS/MS en mode positive..... 242

Figure 8.7 : Affichage du recouvrement de séquence d'une analyse CE-MS/MS de l'ARNt^{Phe GAA} issu de *S. cerevisiae* et digéré à la RNase T1 et à la BAP. 243

Supplemental information

Figure S8.1: *S. cerevisiae* tRNA^{Phe GAA} sequence referenced in the Modomics database 237

Figure S8.2: *B. taurus* tRNA sequences referenced in the Modomics database. 237

Figure S8.3: MS/MS spectrum with only background noise that can be used to estimate the MS/MS absolute intensity threshold. 238

Figure S8.4: Multiple extracted ion electropherograms for the analysis of *B. taurus* total tRNA extract obtained after manual interpretation. 239

Figure S8.5: Multiple extracted ion electropherograms for the analysis of *P. aeruginosa* 70S ribosome obtained after manual interpretation. 240

PARTIE IV : Mise en application du couplage CE-FTICR-MS

Chapitre 9 : Mise en application du couplage CE-FTICR-M

Figure 9.1 : Electrophérogramme des nucléosides de l'ARNr 16S issu de *P. aeruginosa*, obtenu par CE-QTOF-MS/MS après digestion à la Nucléase P1 et déphosphorylation à la BAP..... 254

Figure 9.2 : Electrophérogramme des modifications de l'ARNr 16S issu de *P. aeruginosa*, obtenu par CE-FTICR-MS/MS à une concentration de 2,0 g/L après digestion à la Nucléase P1 et déphosphorylation à la BAP..... 254

Figure 9.3 : Nombre de nucléosides détectés pour les 4 analyses en CE-QTOF-MS/MS et CE-FTICR-MS/MS de l'ARNr 16S issu de *Pseudomonas aeruginosa*, après vérification manuelle des résultats de Nucleos'ID. 255

Figure 9.4 : Electrophérogramme des nucléosides de l'extrait total d'ARNr issu de *Thermus thermophilus*, obtenu par CE-QTOF-MS/MS après digestion à la Nucléase P1 et déphosphorylation à la BAP..... 258

Figure 9.5 : Electrophérogramme des modifications de l'extrait total d'ARNr issu de *Thermus thermophilus*, obtenu par CE-FTICR-MS/MS après digestion à la Nucléase P1 et déphosphorylation à la BAP..... 258

Figure 9.6 : (A) Spectres MS et (B) spectre MS/MS (fragmentation CID de m/z 314,09166) obtenus lors de l'analyse par CE-FTICR-MS/MS de l'extrait total d'ARNr issu de *Thermus thermophilus*. 260

Figure 9.7 : Structures des modifications (A) 2-methylthio-N6-threonylcarbamoyladenosine (ms^2t^6A), (B) 2-methylthio-N6-(cis-hydroxyisopentenyl) adenosine (ms^2io^6A), (C) 2-methylthio-N6-methyladenosine (ms^2m^6A), (D) 2-methylthio-N6-isopentenyladenosine (ms^2i^6A), (E) 2-methylthio-N6-hydroxynorvalylcarbamoyladenosine (ms^2hn^6A) et (F) 2-methylthio cyclic N6-threonylcarbamoyladenosine (ms^2ct^6A) issues de Modomics. 261

Liste des figures

LISTE DES TABLEAUX

PARTIE I : Introduction et revue bibliographique

Chapitre 1 : Electrophorèse capillaire et spectrométrie de masse

Tableau 1.1 : Récapitulatif des analyseurs de masse utilisés lors de ces travaux de thèse et leurs caractéristiques.	48
--	----

Chapitre 2 : Les couplages entre la FTICR-MS et les méthodes séparatives

Table 2.1: Applications using LC hyphenated to FTICR-MS.	96
--	----

Table 2.2: Applications using CE hyphenated to FTICR-MS	102
--	-----

Table 2.3: Applications using GC hyphenated to FTICR-MS.	104
--	-----

Chapitre 3 : Les acides ribonucléiques

Tableau 3.1 : Principales abréviations utilisées dans l'écriture simplifiée des modifications.	125
--	-----

Tableau 3.2 : Principales enzymes utilisées pour l'hydrolyse totale des ARN.....	132
---	-----

Tableau 3.3 : Principales enzymes utilisées pour digérer spécifiquement les ARN.	133
--	-----

PARTIE II : Mise en place et étude du couplage CE-FTICR-MS

Chapitre 4 : Mise en place du couplage CE-FTICR-MS

Tableau 4.1 : Paramètres du FTICR-MS pouvant être réglés pour optimiser le transfert des ions de la source à la cellule ICR pour la caractérisation des nucléosides.	155
--	-----

Tableau 4.2 : Valeurs du minimum de la gamme de m/z , du DataPoint Size et du temps d'accumulation utilisées pour l'étude de l'influence du temps d'accumulation sur les performances du couplage CE-FTICR-MS.....	159
---	-----

Tableau 4.3 : Valeurs du minimum de la gamme de m/z , du DataPoint Size et du temps d'accumulation utilisées pour l'étude de l'influence du DataPoint Size sur les performances du couplage CE-FTICR-MS.	160
--	-----

Tableau 4.4 : Valeurs du minimum de la gamme de m/z , du DataPoint Size et du temps d'accumulation utilisées pour l'étude de l'influence du minimum de la gamme de m/z sur les performances du couplage CE-FTICR-MS.....	162
---	-----

Tableau 4.5 : Valeurs du minimum de la gamme de m/z , du DataPoint Size et le temps d'accumulation utilisées pour étudier l'influence de différentes combinaisons sur les performances du couplage CE-FTICR-MS.....	163
Tableau 4.6 : Paramètres optimisés de l'AutoMS/MS avec le couplage CE-FTICR-MS pour la caractérisation des nucléosides ainsi que leurs significations respectives.....	165
Tableau 4.7 : Liste des masses références utilisées pour la calibration interne	169
Tableau 4.8 : Nombre de formules brutes possibles données par SmartFormula pour les m/z mesurés des quatre nucléosides canoniques selon l'erreur tolérée	170

Chapitre 5 : Evaluation du transfert de méthode vers le couplage CE-FTICR-MS

Tableau 5.1 : Récapitulatif des nucléosides présents dans les ARNt ^{Phe GAA} , ARNt ^{Trp CCA} , ARNt ^{Asp GUC} et ARNt ^{Ala AGC} , issus de <i>S. cerevisiae</i> , et des données MS et MS/MS utiles à leur identification	176
Tableau 5.2 : Comparaison des données MS et MS/MS obtenues pour l'identification des nucléosides des (A) ARNt ^{Phe GAA} , (B) ARNt ^{Asp GUC} , (C) ARNt ^{Ala AGC} et (D) ARNt ^{Trp CCA} , issus de <i>S. cerevisiae</i> , par les couplages CE-QTOF-MS et CE-FTICR-MS.	178
Tableau 5.3 : Récapitulatif des nucléosides présents dans l'extrait total de tRNA de <i>B. taurus</i> , et des données MS et MS/MS utiles à leur identification.	181
Tableau 5.4 : Comparaison des données MS et MS/MS obtenues pour l'identification des nucléosides de l'extrait total d'ARNt de <i>B. taurus</i> , par les couplages CE-QTOF-MS et CE-FTICR-MS.....	182

PARTIE III : Développement méthodologique pour l'étude des modifications post-transcriptionnelles des ARN

Chapitre 6 : Optimisation de la déphosphorylation des nucléotides pour la caractérisation structurale des ARN

Supplemental information

Table S6.1: List of modifications reported in Modomics database for <i>B. taurus</i> tRNA total extract and theoretical m/z used for their MS and MS/MS identification.	200
Table S6.2: CE separation status of isomeric nucleosides and nucleotides.	201
Table S6.3: Migration times of nucleosides and nucleotides	202
Table S6.4: Estimated dephosphorylation efficiencies for 29 nucleotides after 4-hour incubation at 37°C with seven quantities from 0.15 to 7.5 U of BAP per μg of RNA.	203

Table S6.5: Estimated dephosphorylation efficiencies for 29 nucleotides after overnight incubation at 37°C with seven quantities from 0.15 to 7.5 U of BAP per µg of RNA. 204

Chapitre 8 : Développement de deux logiciels pour la caractérisation des ARN à partir d'analyses en mode positif

Table 8.1 : List of modified and non-modified nucleosides reported in Modomics database for *S. cerevisiae* tRNA^{Phe GAA} and *B. taurus* total tRNA extract. 221

Table 8.2: Confusion matrix for the classification of true positives, true negatives, false positives and false negatives depending on the Nucleos'ID results and the actual nucleosides content of the sample. 226

Table 8.3: Output CSV file open in excel for the Nucleos'ID search of *S. cerevisiae* tRNA^{Phe GAA}. 228

Table 8.4: Identification after manual interpretation and Nucleos'ID search of expected nucleosides *S. cerevisiae* tRNA^{Phe GAA}. 228

Table 8.5: Identification after manual interpretation and Nucleos'ID search of expected nucleosides in *B. taurus* total tRNA extract. 230

Table 8.6: Identification after manual interpretation and Nucleos'ID search of unexpected nucleosides in *B. taurus* total tRNA extract. 230

Table 8.7: Manual verification of all nucleosides found by Nucleos'ID search in *P. aeruginosa* 70S ribosome..... 234

Supplemental information

Table S8.1: Corresponding short names and full names of all nucleosides cited in this study 236

Table S8.2: Output CSV file open in excel for the Nucleos'ID search of *B. taurus* total tRNA extract..... 239

Table S8.3: Output CSV file open in excel for the Nucleos'ID search of *P. aeruginosa* 70S ribosome..... 240

PARTIE IV : Mise en application du couplage CE-FTICR-MS

Chapitre 9 : Mise en application du couplage CE-FTICR-MS

Tableau 9.1 : Nucléosides identifiés, par la recherche Nucleos'ID et après vérification manuelle, dans l'ARNr 16S issu de *Pseudomonas aeruginosa*, analysé par CE-QTOF-MS/MS et CE-FTICR-MS/MS. 253

Tableau 9.2 : Nucléosides identifiés, par la recherche Nucleos'ID et après vérification manuelle, dans l'extrait total d'ARNT issu de <i>Thermus thermophilus</i> , analysé par CE-QTOF-MS/MS et CE-FTICR-MS/MS.....	257
---	-----

INTRODUCTION GENERALE

La spectrométrie de masse (MS) s'est imposée depuis le début des années 2000 comme une méthode essentielle pour la caractérisation des molécules. En effet, grâce à ses grandes sensibilité et sélectivité, il est devenu possible de détecter, identifier et quantifier la majorité d'entre elles. Il est également possible d'obtenir des informations structurales sur ces dernières grâce aux nombreuses méthodes de fragmentation. Parmi les différents spectromètres de masse disponibles sur le marché, les spectromètres de masse à transformées de Fourier et résonance cyclotronique des ions (FTICR-MS) sont les instruments de référence connus pour leur ultra-haute résolution et leur très grande précision de mesure de masse. Pour obtenir de telles performances, de longs temps d'acquisition sont nécessaires.

En parallèle, les méthodes séparatives (chromatographiques ou électrophorétiques) sont régulièrement couplées à la MS afin de faciliter la caractérisation d'échantillons complexes. Parmi les méthodes séparatives, l'électrophorèse capillaire (CE) est une des méthodes offrant les meilleures résolution, efficacité de séparation et capacité de pics tout en permettant des analyses rapides. La CE repose sur la séparation des composés selon leurs rayons hydrodynamiques et leurs charges sous l'effet d'un champ électrique. Du fait des différences de sélectivité, la CE montre une grande complémentarité avec la chromatographie liquide qui est souvent la méthode de référence pour un grand nombre d'applications. Enfin, la CE est une méthode miniaturisée. Elle permet donc d'utiliser peu de quantité d'échantillon et d'obtenir une sensibilité optimale en MS grâce à l'utilisation de nano-débits.

De prime abord, le couplage de la CE à la FTICR-MS peut sembler impossible du fait des longs temps d'acquisition du FTICR et de l'efficacité de séparation de la CE. Cependant, en trouvant les bons compromis entre les performances de séparation et la sensibilité apportées par la CE, et les temps d'acquisition du FTICR-MS, il est possible de combiner les avantages des deux méthodes. D'une part, un tel couplage permet notamment de séparer les isomères, de réduire la complexité spectrale et les phénomènes de suppression d'ions, ou encore d'augmenter la gamme dynamique et la sensibilité de l'analyse. D'autre part il permet d'avoir accès à une résolution et une précision de mesure de masse supérieures à ce qui est atteignable avec d'autres couplages CE-MS.

C'est dans ce contexte que se positionnent les travaux détaillés dans ce manuscrit. Le choix de l'application pour mettre en valeur le couplage CE-FTICR-MS s'est porté sur les biomolécules. La caractérisation des biomolécules est en effet de grand intérêt pour tenter de comprendre en profondeur les phénomènes biologiques qui leur sont associés. Ces dernières années, de nombreuses biomolécules ont également suscité un grand intérêt dans le développement de nouvelles stratégies thérapeutiques.

Parmi ces biomolécules, les acides ribonucléiques (ARN) jouent un rôle central dans la vie des cellules. Ils sont le support de l'information génétique contenue dans l'acide désoxyribonucléique (ADN), permettent de traduire l'information génétique en protéines, ou encore régulent l'expression génétique. De nombreuses modifications post-transcriptionnelles sont présentes sur les nucléosides des ARN et notamment dans les ARN de transfert (ARNt) et les ARN ribosomiques (ANRr). Ces modifications sont impliquées dans de nombreuses fonctions biologiques et caractéristiques structurales de l'ARN. Certaines modifications sont également impliquées dans la résistance aux antibiotiques ou encore dans certaines maladies, et d'autres sont à la base du développement de nouveaux traitements. La nature de ces modifications ainsi que leurs localisations sur la séquence de l'ARN ont une importance cruciale dans toutes ces activités biologiques. L'étude des modifications post-transcriptionnelles des ARN est donc de grand intérêt.

Ainsi, l'objectif principal de ces travaux était de mettre en place et développer le couplage CE-FTICR-MS au laboratoire en s'appuyant sur des analyses de biomolécules à visée thérapeutique, et notamment sur la caractérisation des modifications post-transcriptionnelles des ARN.

Ce manuscrit est composé de 4 parties :

- Une première partie est consacrée à la présentation théorique de l'électrophorèse capillaire et de la spectrométrie de masse, ainsi que de leur couplage. Elle comporte également un état de l'art sur les acides ribonucléiques et plus particulièrement sur la caractérisation de leurs modifications post-transcriptionnelles, application choisie pour la mise en place et le développement du couplage CE-FTICR-MS.
- Une seconde partie est dédiée à la mise en place du couplage CE-FTICR-MS, nouveau au laboratoire, et au premier transfert de méthode sur des échantillons biologiques bien connus et décrits dans la littérature.
- Une troisième partie présente les divers développements de méthodes réalisés pour approfondir les connaissances actuelles des modifications post-transcriptionnelles des ARN. Dans un premier temps, l'optimisation d'une étape de la préparation d'échantillon est décrite. Ensuite, l'évaluation de la méthode CE-MS sur un échantillon s'enchaîne avec le développement de deux nouveaux logiciels facilitant l'interprétation des données.
- Une dernière partie propose la mise en application du couplage CE-FTICR-MS ainsi que des derniers développements de méthode, sur des échantillons plus complexes et dont les modifications post-transcriptionnelles ne sont pas décrites dans la littérature.

PARTIE I : INTRODUCTION ET REVUE BIBLIOGRAPHIQUE

Cette première partie, divisée en trois chapitres, est consacrée à la présentation théorique des deux techniques, et de leur couplage, au centre de ces travaux : l'électrophorèse capillaire et la spectrométrie de masse. Elle comporte également un état de l'art sur la caractérisation des modifications post-transcriptionnelles des acides ribonucléiques, application choisie pour la mise en place et le développement du couplage CE-FTICR-MS.

Le premier chapitre présente les bases théoriques de l'électrophorèse capillaire, de la spectrométrie de masse, ainsi que du couplage des deux techniques. Les systèmes utilisés lors cette thèse y sont particulièrement détaillés.

Le second chapitre comporte une revue bibliographique sur les couplages entre une méthode séparative et un spectromètre de masse à transformées de Fourier et résonance cyclotronique des ions (FTICR). L'intérêt d'un tel couplage, ainsi que ses défis et les diverses applications possibles sont présentés.

Enfin, le dernier chapitre porte sur l'application choisie pour mettre en place et développer, au laboratoire, le couplage de l'électrophorèse capillaire à un spectromètre de masse à transformées de Fourier et résonance cyclotronique des ions : les modifications post-transcriptionnelles des acides ribonucléiques.

CHAPITRE 1 : ELECTROPHORESE CAPILLAIRE ET SPECTROMETRIE DE MASSE

1.1. Introduction

La spectrométrie de masse (MS) s'est imposée depuis quelques années comme une méthode essentielle pour la caractérisation de molécules. Afin de faciliter la caractérisation d'échantillons complexes dans de nombreux domaines, la MS a été couplée à des méthodes séparatives comme la chromatographie liquide, la chromatographie gazeuse ou encore l'électrophorèse capillaire. Ces couplages permettent de réduire la complexité des spectres, de discriminer des isomères, mais également de limiter les phénomènes de suppression d'ions. Les travaux de cette thèse portent plus particulièrement sur le couplage de l'électrophorèse capillaire à la spectrométrie de masse à transformées de Fourier et résonance cyclotronique des ions pour l'étude de biomolécules.

Dans ce chapitre, les bases théoriques de l'électrophorèse capillaire et de la spectrométrie de masse seront détaillées, en se focalisant sur les techniques utilisées lors de ces travaux. Le type d'interface utilisé pour réaliser le couplage sera également présenté.

1.2. Electrophorèse capillaire

L'électrophorèse capillaire (CE) est une méthode séparative basée sur la séparation rapide et efficace d'espèces chargées en solution, soumises à un champ électrique. La séparation par électrophorèse a été introduite pour la première fois en 1930 par Tiselius [1], lauréat du prix Nobel de chimie en 1948. Depuis, de nombreuses applications variées ont été développées, telles que l'analyse de protéines [2], de peptides [3], de molécules chirales [4], d'acides désoxyribonucléiques (ADN) [5] et ribonucléiques (ARN) [6], ou encore de métabolites [7].

Un système d'électrophorèse capillaire (Figure 1.1) est composé d'un capillaire, rempli d'un électrolyte support (BGE) conducteur. Les deux extrémités du capillaire sont plongées chacune dans un compartiment d'électrolyte support dans lequel plonge également une électrode afin d'appliquer une différence de potentiel entre les deux extrémités. Sous l'effet du champ électrique, les analytes migrent à des vitesses différentes, selon leurs charges et leurs rayons hydrodynamiques, et ainsi se séparent. Les phénomènes physiques impliqués dans la séparation sont à l'origine de la complémentarité de la CE par rapport aux autres méthodes séparatives telles que la chromatographie liquide.

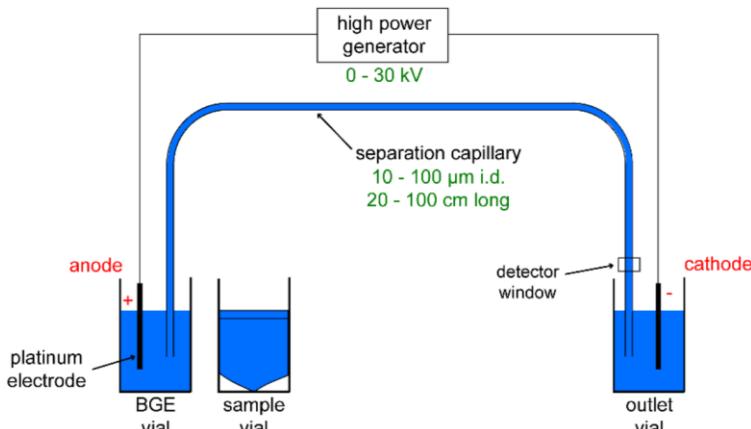


Figure 1.1: Représentation schématique du système d'électrophorèse capillaire. Extrait de [8].

1.2.1. Phénomènes impliqués dans la séparation par électrophorèse capillaire

Lors d'une séparation par CE, les analytes sont soumis à deux phénomènes physiques induits par le champ électrique appliqué aux extrémités du capillaire : l'électrophorèse et l'electroosmose.

1.2.1.1. Mobilité électrophorétique

Le principal phénomène qui explique la séparation des analytes en électrophorèse capillaire est l'électrophorèse. Sous l'effet d'un champ électrique, les analytes chargés migrent. La vitesse de migration électrophorétique d'un analyte dépend de sa charge, de son rayon hydrodynamique, mais aussi du champ électrique appliqué pour réaliser la séparation. Chaque analyte possède donc sa propre mobilité électrophorétique.

La mobilité électrophorétique μ_{ep} d'un analyte est définie par l'équation (1.1) et sa vitesse électrophorétique v_{ep} est définie par l'équation (1.2) :

$$\mu_{ep} = \frac{q}{6\pi\eta r} \quad (1.1)$$

$$v_{ep} = \mu_{ep} * E \quad (1.2)$$

Avec μ_{ep} la mobilité électrophorétique ($\text{cm}^2.\text{V}^{-1}.\text{s}^{-1}$), η la viscosité du BGE ($\text{kg}.\text{cm}^{-1}.\text{s}^{-1}$), q la charge globale de l'analyte (C), r son rayon hydrodynamique (cm), v_{ep} la vitesse électrophorétique ($\text{cm}.\text{s}^{-1}$), et E l'intensité du champ électrique ($\text{V}.\text{cm}^{-1}$).

Les analytes hautement chargés possèdent une mobilité électrophorétique plus importante que les analytes faiblement chargés. À charge équivalente, l'analyte ayant la plus grande densité de charge, donc le rayon hydrodynamique le plus faible, possède une mobilité électrophorétique plus importante. Les anions migrent en direction de l'anode, alors que les cations migrent vers la cathode. Les analytes neutres, quant à eux, possèdent une mobilité électrophorétique nulle.

La mobilité électrophorétique est donc caractéristique de chaque analyte, mais elle dépend également de l'électrolyte support utilisé (composition et pH) ainsi que de la température à laquelle est réalisée la séparation.

1.2.1.2. Flux électroosmotique

L'électroosmose est le deuxième phénomène induit par l'application d'un champ électrique lors de séparation par électrophorèse capillaire. Sous l'application d'un champ électrique, un écoulement dans le capillaire, appelé flux électroosmotique, est engendré par l'interaction des charges contenues dans l'électrolyte support avec celles situées à la surface du capillaire. Les charges contenues dans l'électrolyte support forment, à la surface du capillaire, une double couche de charges. La première, appelée couche de Stern, est compacte et la deuxième, appelée couche de Gouy-Chapman, est diffuse. Le potentiel de surface de cette double couche, appelé potentiel zéta, est défini par l'équation (1.3) :

$$\zeta = \frac{\delta\sigma}{\varepsilon} \quad (1.3)$$

Avec ζ le potentiel zéta de la double couche (V), δ son épaisseur (cm), σ la densité de charge par unité de surface de la double couche ($C.cm^{-2}$) et ε la constante diélectrique de l'électrolyte support ($kg.m.V^{-2}.s^{-2}$).

Dans le cas d'un capillaire de silice vierge, la surface de capillaire est chargée négativement et la double couche est composée de cations du BGE (Figure 1.2). Sous l'effet du champ électrique, les charges de la double couche migrent vers la cathode et créent ainsi le flux électroosmotique. La vitesse du flux électroosmotique v_{eo} est définie par l'équation (1.4) et le flux électroosmotique μ_{eo} qui lui est associé est défini par l'équation (1.5) de Helmholtz-Smoluchowski :

$$v_{eo} = \frac{\varepsilon\zeta}{4\pi\eta} * E \quad (1.4)$$

$$\mu_{eo} = \frac{\varepsilon\zeta}{4\pi\eta} = \frac{\delta\sigma}{4\pi\eta} \quad (1.5)$$

Avec v_{eo} la vitesse électroosmotique ($cm.s^{-1}$), μ_{eo} le flux électroosmotique ($cm^2.V^{-1}.s^{-1}$), η la viscosité du BGE ($kg.cm^{-1}.s^{-1}$), ε sa constante diélectrique du BGE ($kg.cm.V^{-2}.s^{-2}$), ζ le potentiel zéta de la double couche (V), δ l'épaisseur de la double couche (cm), σ la densité de charge par unité de surface de la double couche ($C.cm^{-2}$) et E l'intensité du champ électrique ($V.cm^{-1}$).

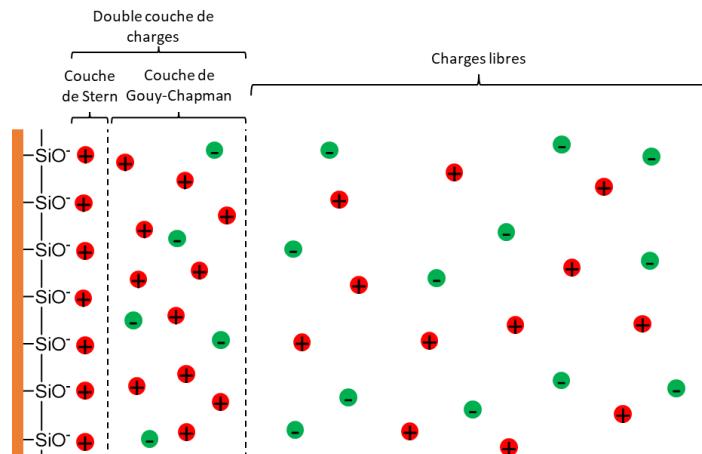


Figure 1.2 : Représentation schématique de la formation de la double couche de charges dans un capillaire de silice vierge.
Adapté à partir de [8].

Le flux électroosmotique dépend donc du BGE (composition, concentration et pH), de la température à laquelle est réalisée la séparation, mais aussi de la nature des charges à la surface du capillaire utilisé (type de capillaire).

Le front d'écoulement plat du flux électroosmotique est à l'origine de la grande efficacité de séparation obtenue par électrophorèse capillaire, en comparaison avec d'autres techniques de séparation comme la chromatographie liquide dont l'écoulement à un front parabolique dû aux phénomènes de diffusion.

1.2.1.3. Mobilité apparente

Pour obtenir la vitesse de migration des analytes dans un capillaire soumis à un champ électrique, il est donc essentiel de prendre en compte la mobilité électrophorétique, mais également le flux électroosmotique. La vitesse de migration apparente et la mobilité apparente de l'analyte sont respectivement définies par les équations (1.6) et (1.7) :

$$v_{app} = \mu_{app} * E \quad (1.6)$$

$$\mu_{app} = \mu_{ep} + \mu_{eo} \quad (1.7)$$

Avec v_{app} la vitesse de migration apparente (cm.s^{-1}), μ_{app} la mobilité apparente ($\text{cm}^2.\text{V}^{-1}.\text{s}^{-1}$), μ_{ep} la mobilité électrophorétique ($\text{cm}^2.\text{V}^{-1}.\text{s}^{-1}$), μ_{eo} le flux électroosmotique ($\text{cm}^2.\text{V}^{-1}.\text{s}^{-1}$) et E l'intensité du champ électrique (V.cm^{-1}).

Il est possible de déterminer la mobilité apparente des analytes dans les conditions expérimentales choisies à partir des électrophérogrammes en utilisant l'équation (1.8) :

$$\mu_{app} = \frac{l}{E * t_m} = \frac{l * L}{V * t_m} \quad (1.8)$$

Avec μ_{app} la mobilité apparente ($\text{cm}^2 \cdot \text{V}^{-1} \cdot \text{s}^{-1}$), l la longueur effective de l'entrée du capillaire au détecteur du capillaire (cm), E l'intensité du champ électrique ($\text{V} \cdot \text{cm}^{-1}$), t_m le temps de migration (s), L la longueur totale du capillaire (cm) et V la tension appliquée pour réalisée la séparation (V).

Les mobilités électrophorétiques μ_{ep} , intrinsèques aux caractéristiques de chaque analyte, peuvent être déduites de la mobilité apparente μ_{app} en utilisant un marqueur de flux neutre dont la mobilité apparente sera égale au flux électroosmotique μ_{eo} .

1.2.2. Electrophorèse capillaire de zone (CZE)

Différents modes de séparation par électrophorèse capillaire existent tels que l'électrophorèse capillaire en gel (GCE), l'électrochromatographie capillaire (CEC), l'isoélectrofocalisation capillaire (CIEF) ou encore la chromatographie électrocinétique micellaire (MEKC). Cependant, le plus simple et le plus utilisé, est l'électrophorèse capillaire de zone (CZE). C'est également le mode de séparation le plus compatible avec la spectrométrie de masse, grâce à la possibilité de constituer des tampons suffisamment volatiles et à faible contenance en sel du BGE. La séparation par CZE se base uniquement sur les différences de mobilités apparentes des composés.

Dans le cas d'un capillaire de silice vierge, le flux électrophorétique est dirigé vers la cathode. Les cations sont donc les premiers à être détectés. Plus le cation est chargé et de grande densité de charge, plus il sera détecté rapidement. Les espèces neutres, uniquement soumises au flux électroosmotique ne peuvent pas être séparées et sont donc détectées en même temps. Enfin, les anions sont les derniers à être détectés. Plus l'anion est chargé et de faible densité de charge, plus il sera détecté tardivement. Si le flux électroosmotique n'est pas assez important par rapport à la mobilité électrophorétique d'un anion, celui-ci ne sera pas détecté (Figure 1.3).

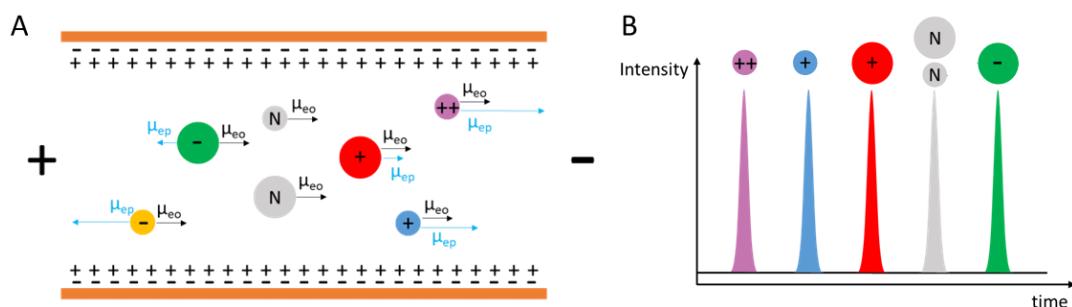


Figure 1.3 : Représentation schématique (A) d'une séparation par CZE sur un capillaire en silice vierge et (B) de l'électrophérogramme obtenu par cette séparation. Adapté à partir de [8].

Pour étudier les anions, optimiser les conditions de séparation ou encore lutter contre les phénomènes d'absorption entre les analytes et la surface du capillaire en silice vierge, il est possible de greffer le capillaire et ainsi de supprimer (greffage neutre) ou d'inverser (greffage positif) le flux électroosmotique [9]. Dans le cadre de cette thèse, la quasi-totalité des travaux ont été réalisés sur un capillaire de silice vierge, non greffé.

1.2.3. Efficacité de séparation

La CE est connue pour être une des techniques séparatives les plus efficaces avec des largeurs de pics pouvant descendre jusqu'à quelques secondes.

L'équation de Van Deemter (équation (1.9) permet de modéliser l'efficacité de séparation de ces méthodes. Plus la hauteur équivalente d'un plateau théorique H est faible, plus la séparation est efficace, c'est-à-dire que la largeur des pics est faible.

$$H = A + \frac{B}{v_x} + Cv_x \quad (1.9)$$

Avec H la hauteur équivalente d'un plateau théorique, v_x la vitesse de la phase mobile, la viscosité de l'électrolyte support ($\text{kg} \cdot \text{m}^{-1} \cdot \text{s}^{-1}$), q la charge globale de l'analyte, A la constante reliée à la diffusion turbulente, B la constante reliée à la diffusion longitudinale et C la constante reliée à la résistance au transfert de charge.

La diffusion turbulente A représente les nombreux « chemins » que peut emprunter un analyte pour parcourir une phase stationnaire. Ces différents « chemins » sont créés par l'hétérogénéité de la distribution des particules au sein de la phase stationnaire. La résistance au transfert de charge C représente la capacité de la phase stationnaire à s'opposer à l'entraînement des analytes par la phase mobile. Dans le cadre de l'électrophorèse capillaire, qui ne possède pas de phase stationnaire (hormis la CEC), les termes A et C sont donc nuls. Concernant la diffusion longitudinale B , elle est limitée par le front plat du flux électroosmotique caractéristique de la CE (cf. paragraphe 1.2.1.2). Tout cela implique une faible hauteur équivalente d'un plateau théorique H en CE, ce qui explique la grande efficacité de séparation de cette technique.

1.3. Spectrométrie de masse

La spectrométrie de masse (MS) est basée sur la détermination du rapport masse-surcharge (m/z) de molécules préalablement ionisées. Cette technique est devenue, grâce à sa spécificité et sa sensibilité, une référence pour détecter, identifier et quantifier les biomolécules [10]. Il est également possible d'obtenir des informations structurales.

Une analyse par spectrométrie de masse se déroule en plusieurs étapes. Les analytes sont d'abord ionisés et volatilisés au niveau de la source, puis les optiques transmettent les ions à l'analyseur qui détermine les rapports m/z des ions. Enfin, le signal électrique produit par les ions est mesuré au niveau du détecteur et les spectres de masse sont obtenus après traitement de ce signal.

Il existe de nombreux spectromètres de masse composés de sources, d'optiques de transfert des ions, d'analyseurs et de détecteurs variés. Chaque type de source, d'interface, d'analyseur et de détecteur possède ses propres caractéristiques et performances. En effet,

les différentes sources sont plus ou moins adaptées pour ioniser les composés en fonction de leurs poids moléculaires et de leurs polarités. Certains analyseurs sont caractérisés par leur ultra-haute résolution et leur grande précision de masse, d'autres par leur grande vitesse d'acquisition, ou encore par leur capacité en quantification. Et enfin, les différents détecteurs peuvent avoir, quant à eux, des limites de détection différentes. Le choix du type d'instrument est donc adapté selon les applications et les besoins de chaque étude. Les différents éléments qui composent les deux instruments utilisés pour réaliser ces travaux sont détaillés dans les paragraphes suivants. Il s'agit des spectromètres de masse Maxis 4G (Bruker Daltonics) et FTICR 7T SolariX 2XR (Bruker Daltonics).

1.3.1. Source d'ionisation par électrospray (ESI)

L'utilisation de l'ionisation par électrospray (ESI) en tant que source pour la MS a été démontrée pour la première fois en 1984 par Fenn *et al.* [11], lauréat du prix Nobel de chimie en 2002. Depuis, elle est communément utilisée en MS pour de nombreuses applications, notamment biologiques.

L'ESI est une méthode d'ionisation douce à pression atmosphérique, elle n'engendre donc généralement pas de fragmentation lors de l'ionisation. Elle permet également de produire des ions multichargés, en mode positif ou négatif. Tout cela fait de l'ESI une source de choix pour l'étude de biomolécules telles que les protéines ou les acides ribonucléiques (ARN).

Le processus d'ionisation se compose de trois étapes. Dans un premier temps, des gouttelettes chargées vont se former, puis chacune va se diviser en de multiples gouttelettes plus fines. Pour finir, des ions en phase gazeuse sont produits à partir de ses fines gouttelettes.

1.3.1.1. Formation des gouttelettes chargées

La formation des gouttelettes chargées est due au champ électrique intense appliqué entre le capillaire et la contre-électrode. En effet, ce champ électrique provoque une accumulation des charges à l'extrémité du capillaire qui déstabilise la surface du liquide. Ce phénomène engendre la formation d'un cône, appelé cône de Taylor (Figure 1.4) [12,13].

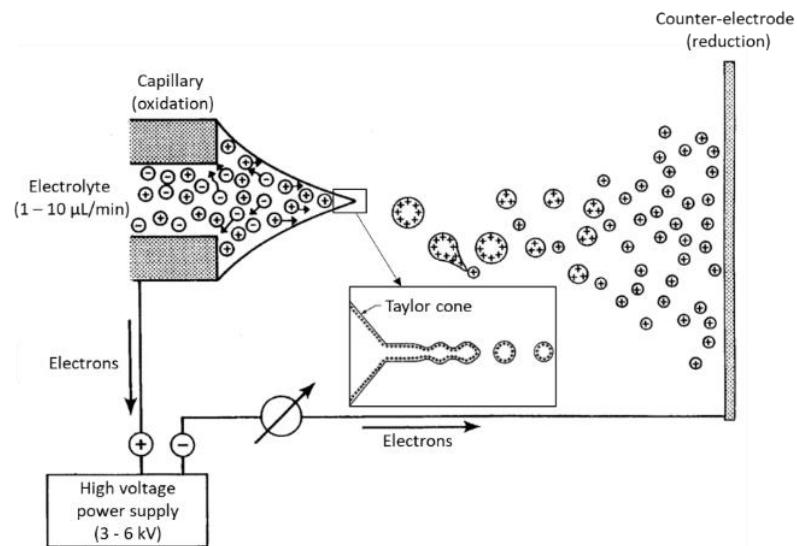


Figure 1.4 : Processus d'ionisation par électrospray menant à l'apparition d'un cône de Taylor et à la production de gouttelettes chargées. Adapté de [14].

1.3.1.2. Fission des gouttelettes chargées

Avec la chaleur émise par la source, le solvant contenu dans les gouttelettes générées par le cône de Taylor s'évapore et leurs rayons diminuent, à mesure que celles-ci se rapprochent de l'entrée du spectromètre de masse. Le rayon de gouttelettes diminue ainsi jusqu'à atteindre la « limite de Rayleigh ». Ce rayon critique est défini par l'équation de Rayleigh (équation (1.10)) et dépend notamment de la charge de la gouttelette [15].

$$q = 8\pi\sqrt{\varepsilon_0\gamma R^3} \quad (1.10)$$

Avec q la charge de la gouttelette (C), ε_0 la permittivité du vide ($F.m^{-1}$), γ la tension de surface ($N.m^{-1}$) et R le rayon critique de Rayleigh (m).

À mesure que la taille de la gouttelette se rapproche du rayon critique de Rayleigh, les répulsions coulombiennes entre les charges augmentent jusqu'à dépasser la tension de surface. La gouttelette explose alors pour former plusieurs gouttelettes plus fines et de densité de charge plus élevée, qui sont à leur tour soumises à une succession d'explosions coulombiennes [12,14].

1.3.1.3. Production des ions en phase gazeuse

Bien que le mécanisme de formation des ions en phase gazeuse à partir de fines gouttelettes ne soit, à ce jour, pas complètement élucidé, deux principaux modèles ont été proposés (Figure 1.5).

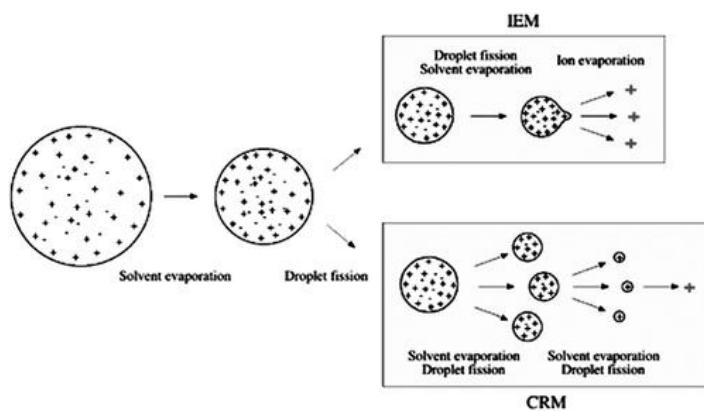


Figure 1.5 : Modèle de Dole (CRM) et d'Iribarne-Thomson (IEM) décrivant la formation des ions en phase gazeuse à partir de gouttelettes chargées. Extrait de [16].

Le premier, appelé modèle de charge résiduelle (CRM) et proposé par Dole *et al.* en 1968 [17] et 1970 [18], suggère que les gouttelettes issues du cône de Taylor sont soumises à des évaporations et des explosions coulombiennes successives jusqu'à l'obtention de gouttelettes contenant un ion unique.

Le second, appelé modèle d'évaporation d'ion (IEM) et proposé par Iribarne et Thomson en 1976 [19], suggère que les ions sont directement émis depuis la surface des gouttelettes hautement chargées soumises au fort champ électrique, lorsque leur densité de charge devient trop importante. Ce phénomène se produirait avant même que le rayon de la gouttelette n'atteigne le rayon critique de Rayleigh.

Bien que le mécanisme de formation des ions en phase gazeuse soit encore sujet de débat, il semblerait que le modèle de Dole soit plus adapté pour décrire la formation d'espèces multichargées de grosses molécules, telles que les protéines, et que le modèle de Iribarne-Thomson soit, quant à lui, plus adapté pour décrire la formation de petits ions monochargés [20].

1.3.1.4. Miniaturisation des sources nanoESI

La miniaturisation de l'ESI a débuté en 1994, avec la source dite microESI de Emmet et Caprioli qui a permis d'obtenir un spray stable avec des débits compris entre 300 et 800 nL/min. À titre de comparaison, les débits classiquement utilisés en ESI sont compris entre 1 et 10 µL/min. Ces travaux ont montré, dans le cadre d'études de protéines, que la sensibilité et le rapport signal/bruit ont significativement augmenté avec cette nouvelle source [21].

Dans la même année, la source nanoESI de Mann et Wilm a permis d'atteindre des débits inférieurs à 25 nL/min. Ces travaux ont montré que la diminution du débit permet, en plus d'améliorer l'efficacité d'ionisation des analytes et la sensibilité de la détection par MS, de diminuer les effets de suppression d'ions causés par les sels contenus dans l'échantillon et de travailler avec de plus faibles quantités d'échantillons [22].

L'impact des sources nanoESI sur l'efficacité d'ionisation et sur la sensibilité peut être expliqué par la réduction du diamètre du capillaire et du débit d'injection. En effet, les diamètres des capillaires sont compris entre 1 et 10 µm, ce qui permet de produire, à partir du cône de Taylor, de plus fines gouttelettes par rapport à celles émises en ESI classique. De plus, Mann et Wilm ont montré, à partir de l'équation de Bernoulli, que le rayon d'émission des gouttelettes est proportionnel à $d^{2/3}$ (d étant le débit d'injection) [22]. De plus faibles débits permettent donc également de produire de plus fines gouttelettes. Plus les gouttelettes sont fines et plus elles sont soumises aux explosions coulombiennes, ce qui augmente donc le rendement d'ionisation.

Grâce à leur impact sur l'efficacité de l'ionisation et la sensibilité du signal, le développement des sources nanoESI a été bénéfique pour les infusions directes, mais aussi pour les couplages entre la MS et les méthodes séparatives.

1.3.2. Optiques et configuration

Les optiques d'un spectromètre de masse qualifient l'ensemble des éléments qui permettent le transfert des ions entre la source et l'analyseur. Ils permettent également de relier la source, qui est à pression atmosphérique, et l'analyseur, qui est dans un vide poussé sous atmosphère inerte, en augmentant graduellement le vide. Par un jeu de lentilles et de gradients de potentiel électrique, les ions sont guidés tout au long du spectromètre de masse. Les optiques jouent un rôle essentiel dans l'analyse par MS puisque l'efficacité du transfert des ions impacte directement la sensibilité de la détection.

Les configurations des deux spectromètres de masse utilisés pour réaliser ces travaux de thèse vont être détaillées dans les prochains paragraphes.

1.3.2.1. Spectromètre de masse Maxis 4G (Bruker Daltonics)

Un schéma du spectromètre de masse Maxis 4G (Bruker Daltonics) est présenté en Figure 1.6. Les ions produits par la source traversent tout d'abord un capillaire de transfert en verre où un gaz séchant chaud (généralement de l'azote) circule en contre-sens afin d'évaporer le solvant résiduel. Les ions sont ensuite entraînés et focalisés par une série de lentilles électrostatiques (funnels) où un gradient de potentiel est appliqué. Afin d'éliminer les dernières molécules neutres présentes dans le flux d'ions, le centre du premier funnel est légèrement désaxé par rapport au capillaire de transfert. Le flux d'ions est ensuite focalisé par un hexapole jusqu'à leur entrée dans le quadrupôle qui est le premier analyseur. À la sortie du quadrupôle, les ions traversent une cellule de collision pour arriver au deuxième analyseur, le temps de vol.

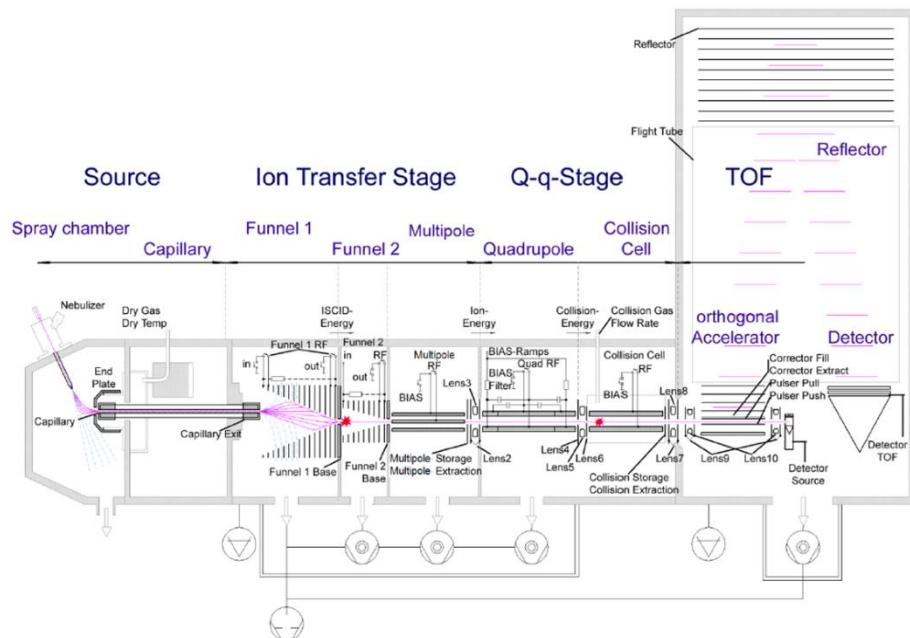


Figure 1.6 : Représentation schématique du spectromètre de masse Maxis 4G (Bruker Daltonics)

1.3.2.2. Spectromètre de masse FTICR 7T SolariX 2XR (Bruker Daltonics)

Un schéma du spectromètre de masse FTICR 7T SolariX 2XR (Bruker Daltonics) est présenté en Figure 1.7. La première partie, entre la source et le quadrupôle, ressemble fortement à celle du Maxis 4G, mis à part que le capillaire de transfert est plus long et perpendiculaire au premier funnel. Cette configuration a permis au constructeur d'ajouter une source MALDI dans l'alignement des funnels. À la sortie du quadrupôle, les ions sont dirigés vers la cellule de collision où ils sont accumulés. Puis, ils sont envoyés par paquets, en appliquant un potentiel à la sortie de la cellule, dans un hexapôle qui permet de guider et de focaliser le flux d'ions jusqu'à atteindre l'analyseur positionné au centre de l'aimant, qui est une cellule à résonance cyclotronique (ICR).

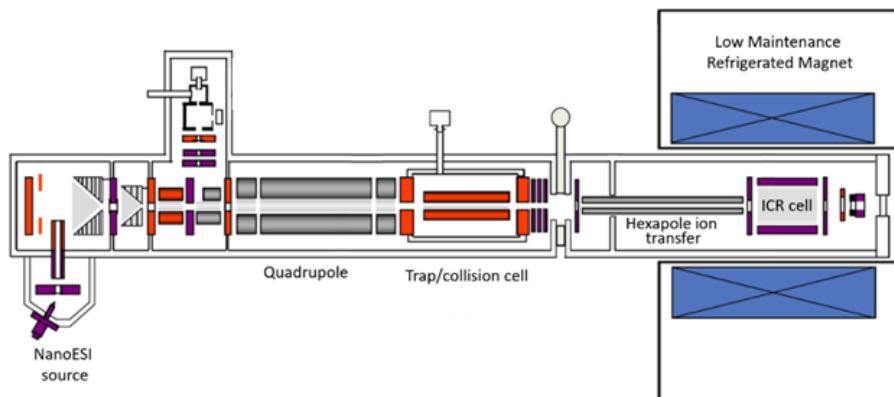


Figure 1.7 : Représentation schématique du spectromètre de masse 7T SolariX 2XR (Bruker Daltonics)

1.3.3. Analyseurs de masse

L'analyseur d'un spectromètre de masse a pour but de séparer et de déterminer les rapports m/z des ions détectés. De nos jours, différents types d'analyseurs sont disponibles

sur le marché. Chacun d'eux faisant appel à une technologie particulière, leurs performances en ce qui concerne la résolution, la précision de mesure, la vitesse d'acquisition ou encore la gamme de m/z , sont donc bien différentes. Les analyseurs utilisés dans le cadre de cette thèse, ainsi que leurs caractéristiques, sont récapitulés dans le Tableau 1.1 et décrits dans les paragraphes suivants.

Tableau 1.1 : Récapitulatif des analyseurs de masse utilisés lors de ces travaux de thèse et leurs caractéristiques. Adapté à partir de [23].

Type d'analyseur	Gamme de m/z	Résolution	Précision de mesure (ppm)	Vitesse d'acquisition (Hz)
Quadrupôle (Q)	0 – 3 000	3 000 – 5 000	Faible	2 – 10
Temps de vol (TOF)	0 – 20 000	10 000 – 60 000	1 – 5	10 – 50
Résonance cyclotronique (ICR)	50 – 10 000	750 000 – 2 500 000	0,3 – 1	0,5 – 2

1.3.3.1. Quadrupôle (Q)

Le quadrupôle est un filtre de masse composé de quatre électrodes, cylindriques ou hyperboliques, disposées parallèlement autour d'un axe (Figure 1.8.A). Un potentiel continu U et un potentiel alternatif de radiofréquence $V \cos(\omega t)$ sont appliqués sur les électrodes de telle façon que les électrodes opposées aient le même potentiel, et que les électrodes adjacentes aient un potentiel de polarité opposée (Figure 1.8.B).

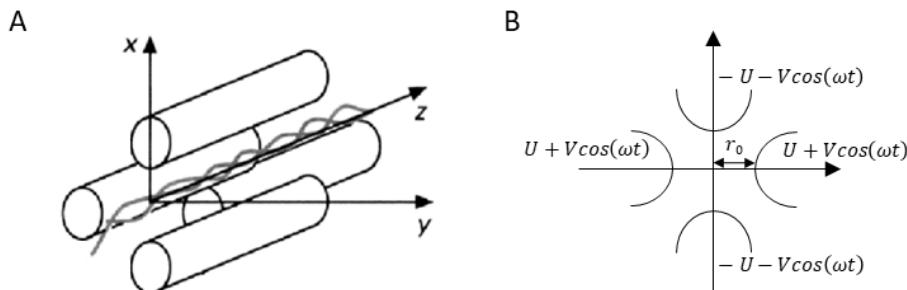


Figure 1.8 : Représentations schématiques d'un quadrupôle sous différents angles. Adapté à partir de [8].

Le potentiel électrique ϕ résultant dans le plan xOy est défini par l'équation (1.11) :

$$\phi(x, y, z, t) = Q_0 \frac{(x^2 - y^2)}{r_0^2} \quad \text{où} \quad Q_0 = U + V \cos(\omega t) \quad (1.11)$$

Avec ϕ le potentiel électrique dans le plan xOy, U le potentiel constant, V l'amplitude du potentiel alternatif, ω la fréquence du potentiel alternatif et r_0 la distance entre l'axe et les électrodes.

Sous l'application de ce potentiel électrique, la trajectoire des ions à l'intérieur de ce quadrupôle est définie par les équations de Mathieu (équations (1.12) à (1.15)).

$$\frac{d^2x}{d\zeta^2} + x(a_x - 2q_x \cos(2\zeta)) = 0 \quad (1.12)$$

$$\frac{d^2y}{d\zeta^2} - y(a_y - 2q_y \cos(2\zeta)) = 0 \quad (1.13)$$

$$\frac{d^2z}{d\zeta^2} = 0 \quad (1.14)$$

Où $\zeta = \frac{\omega t}{2}$; $a_x = -a_y = \frac{8zeU}{mr_0^2\omega^2}$ et $q_x = -q_y = \frac{4zeV}{mr_0^2\omega^2}$ (1.15)

Avec ze la charge de l'ion, m sa masse, U le potentiel constant, V l'amplitude du potentiel alternatif, ω la fréquence du potentiel alternatif et r_0 la distance entre l'axe et les électrodes.

Si l'amplitude de la trajectoire de l'ion est stable et inférieure à r_0 , l'ion passera à travers le filtre qu'est le quadrupôle. Au contraire, si la trajectoire de l'ion n'est pas stable et que son amplitude finit par dépasser r_0 , soit l'ion entre en collision avec l'une des électrodes, soit il s'échappe entre deux électrodes. Il ne passera donc pas à travers le quadrupôle. Pour un rapport U/V fixé, uniquement les ions de masses définies par l'intersection de la zone de stabilité et de la ligne de scan sur le diagramme de stabilité (Figure 1.9) passent à travers le quadrupôle. Le rapport m/z des ions est mesuré en faisant varier les potentiels U et V tout en maintenant le rapport des deux constant.

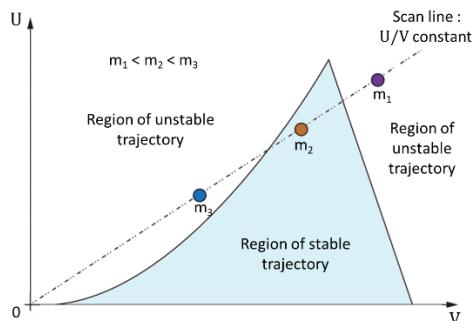


Figure 1.9 : Diagramme de stabilité de la trajectoire des ions dans un quadrupôle. Pour un rapport U/V fixé, les ions de masses m_1 et m_3 ne passent pas à travers le filtre quadrupolaire alors que les ions situés dans la zone de stabilité (comme ceux de masse m_2) passent à travers. Adapté à partir de [24].

De nos jours, le quadrupôle est souvent combiné à un second analyseur. Dans cette configuration, il permet de réaliser des expériences de spectrométrie de masse en tandem (MS/MS) en servant de filtre pour isoler les ions précurseurs d'intérêt.

1.3.3.2. Temps de vol (TOF)

L'analyseur à temps de vol (TOF) est composé d'un tube de vol sous vide poussé (10^{-7} mbar). En arrivant à la base du tube de vol, les ions sont accélérés par un « pusher » qui

applique une différence de potentiel. Cette dernière permet aux ions d'acquérir une énergie cinétique (équation (1.16)) qui leur permet de parcourir le tube de vol.

$$E_C = \frac{mv^2}{2} = zE \quad (1.16)$$

Avec E la différence de potentiel, E_C l'énergie cinétique de l'ion, m sa masse, z sa charge et v sa vitesse dans le tube de vol.

À partir de la mesure du temps nécessaire à l'ion pour traverser le tube de vol, il est possible d'en déduire son rapport m/z selon l'équation (1.17).

$$t = \frac{L}{v} = L \sqrt{\frac{m}{2zE}} \quad \text{d'où} \quad \frac{m}{z} = 2E \left(\frac{t}{L}\right)^2 \quad (1.17)$$

Avec E la différence de potentiel, m la masse de l'ion, z sa charge, v sa vitesse, t son temps de vol et L la longueur du tube de vol.

Ainsi, les ions de faibles m/z atteignent le détecteur avant les ions de plus grands rapports de m/z (Figure 1.10).

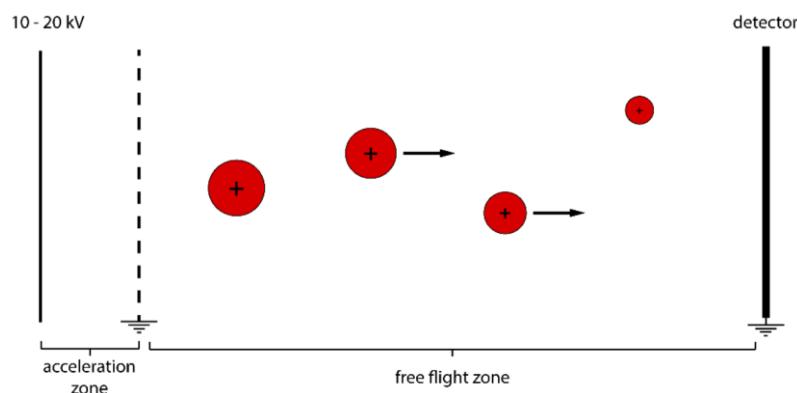


Figure 1.10 : Représentation schématique d'un TOF. Extrait de [8].

Les TOF relativement récents, comme celui utilisé dans le spectromètre Maxis 4G, sont équipés d'un réflecteur électrostatique, aussi appelé réflectron. Ce dernier est composé d'une série de lentilles électrostatiques de potentiels croissants. Le gradient de potentiel agit comme un miroir électrostatique. Plus les ions ont une énergie cinétique importante, plus ils vont avancer au sein de ce gradient avant d'être envoyés dans le sens inverse. Son utilisation permet donc de refocaliser les ions de même m/z et de compenser les légères différences d'énergies cinétiques pour des ions de même m/z (Figure 1.11). Il permet également d'augmenter la distance totale parcourue par les ions dans le tube de vol et ainsi d'augmenter la résolution et la précision de masse du spectromètre. Dans cette configuration, l'utilisateur peut choisir de travailler soit en mode linéaire, sans utiliser le réflecteur électrostatique, soit en mode réflectron.

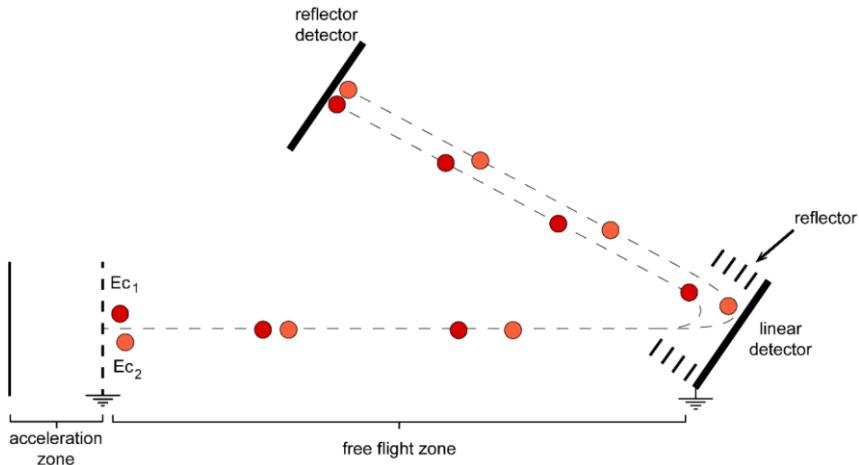


Figure 1.11 : Représentation schématique du fonctionnement d'un analyseur TOF équipé d'un réflecteur. Les ions oranges et rouges possèdent le même m/z , mais des énergies cinétiques légèrement différentes ($E_{c1} < E_{c2}$). Extrait de [8].

1.3.3.3. Résonance cyclotronique des ions (ICR)

La cellule à résonance cyclotronique des ions (ICR), sous vide encore plus poussé que dans le TOF ($10^{-9} - 10^{-10}$ mbar), est placée au centre d'un aimant supraconducteur qui crée un puissant champ magnétique B_0 . A titre d'exemple, le SolariX utilisé lors de ces travaux de thèse, possède un aimant de 7 Tesla. À noter que le champ magnétique le plus fort sur un appareil commercial est obtenu avec un aimant de 21 Tesla. La cellule ICR est composée de deux plaques d'excitation, de deux plaques de piégeage et de deux plaques de détection (Figure 1.12). Au sein de cette cellule, la trajectoire des ions est composée de plusieurs mouvements qui seront détaillés dans les prochains paragraphes.

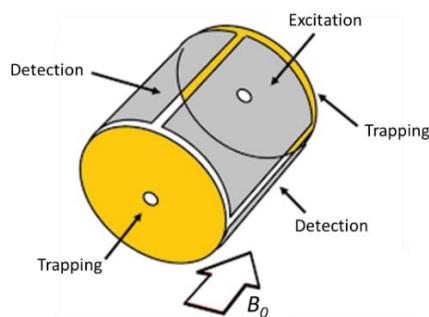


Figure 1.12 : Représentation schématique d'une cellule ICR cylindrique dans un champ magnétique B_0 . Adapté à partir de[25].

- Mouvement cyclotronique

Lorsque les ions arrivent dans la cellule ICR, ils sont excités jusqu'à entrer en résonance et sont envoyés en orbite, perpendiculairement au champ magnétique B_0 , grâce à un champ électrique alternatif créé par les plaques d'excitation. Ce phénomène est appelé mouvement cyclotronique (Figure 1.13.A). Le fréquence cyclotronique w_c des ions en orbite dépend du champ magnétique B_0 et est inversement proportionnelle au rapport m/z des ions. Elle est définie par l'équation (1.18) [26] :

$$w_c = \frac{eB_0}{m/z} \quad (1.18)$$

Avec w_c la fréquence cyclotronique, m/z le rapport masse-sur-charge des ions, e la charge élémentaire et B_0 le champ magnétique.

- Mouvement de piégeage

Une oscillation harmonique entre les deux plaques de piégeage, appelée mouvement de piégeage (Figure 1.13.A), est engendrée par un puits de potentiel électrique. Ce mouvement permet de maintenir les ions au sein de la cellule ICR. La fréquence de piégeage w_T est définie par l'équation (1.19) [26] :

$$w_T = \sqrt{\frac{2e \propto V_T}{m/z}} \quad (1.19)$$

Avec w_T la fréquence de piégeage, \propto un facteur dépendant de la géométrie de la cellule ICR, V_T la tension appliquée aux plaques de piégeage, m/z le rapport masse-sur-charge des ions et e la charge élémentaire.

La fréquence cyclotronique théorique, détaillée dans le paragraphe précédent, est légèrement altérée par le mouvement de piégeage. On parle alors de fréquence cyclotronique réduite w_+ . Cette dernière est définie par l'équation (1.20) [26] :

$$w_+ = \frac{w_c}{2} + \sqrt{\frac{w_c^2}{4} - \frac{w_T^2}{2}} \quad (1.20)$$

Avec w_+ la fréquence cyclotronique réduite, w_T la fréquence de piégeage and w_c la fréquence cyclotronique.

- Mouvement magnetron

Enfin, la combinaison du champ magnétique B_0 et du champ électrique, produit par les plaques de piégeage, induit une rotation des ions autour de l'axe du champ magnétique. On parle de mouvement magnetron (Figure 1.13.A). La fréquence magnetron w_- est définie par l'équation (1.21) [26] :

$$w_- = \frac{w_c}{2} - \sqrt{\frac{w_c^2}{4} - \frac{w_T^2}{2}} \quad (1.21)$$

Avec w_- la fréquence magnetron, w_T la fréquence de piégeage and w_c la fréquence cyclotronique.

Le mouvement magnetron engendre des signaux non attribuables à des ions sur le spectre. Ces pics « fantômes » peuvent être minimisés en *shimming* l'appareil. Le *shimming* est un réglage des différentes tensions appliquées au sein de la cellule ICR afin d'optimiser la trajectoire des ions.

- Trajectoire globale des ions

La trajectoire globale des ions dans la cellule ICR, représentée en Figure 1.13.B, est issue de la superposition de ces trois mouvements.

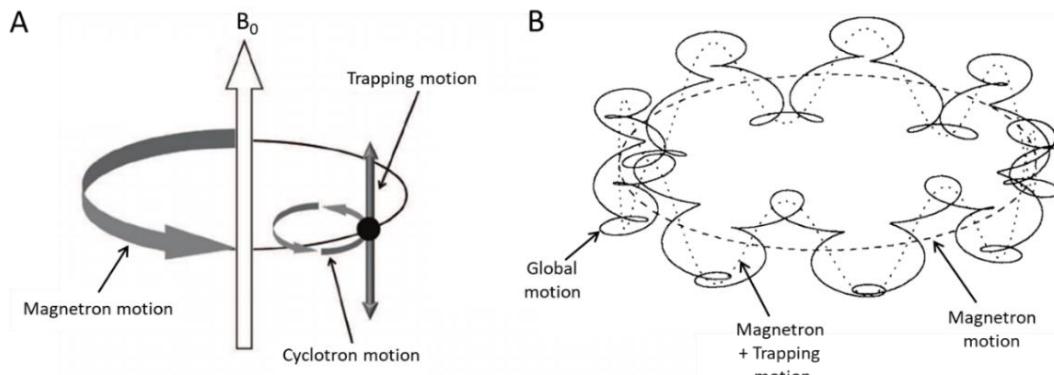


Figure 1.13 : Représentation schématique (A) des trois mouvements d'un ion confiné dans une cellule ICR et soumis à un champ magnétique B_0 et (B) de leur combinaison pour obtenir la trajectoire globale de l'ion. Adapté à partir de [27,28].

En fonction de leur rapport m/z , les ions ont une trajectoire et une fréquence particulières. C'est à partir de la mesure des fréquences globales de rotation des ions, que leurs rapports m/z peuvent être déduits.

1.3.4. DéTECTEURS

Le rôle du détecteur est de convertir le flux d'ions en signal électrique afin de générer le spectre de masse après traitement du signal.

Le multiplicateur d'électrons à galette micro-canaux (MCP) est un détecteur communément utilisé dans de nombreux spectromètres de masse, dont le Maxis 4G. Il est composé d'une multitude de petits canaux incurvés et recouverts d'un dépôt métallique. Cette fine couche conductrice est souvent un alliage de nickel et de chrome, mais d'autres matériaux peuvent également être utilisés [29]. Lorsqu'un ion entre en collision avec la surface d'un micro-canal, un électron secondaire est généré par transmission de son énergie cinétique. Lorsque cet électron secondaire entre à son tour en collision avec la surface du micro-canal, il est multiplié en plusieurs électrons de plus faible énergie. Ces électrons sont accélérés par une différence de potentiel et entrent à leur tour en collision avec la surface du micro-canal. Ainsi de suite, les électrons sont multipliés tout au long du micro-canal. Le courant induit par ses électrons est mesuré.

Dans un spectromètre de masse FTICR, la détection est effectuée au sein de l'analyseur puisque celle-ci est réalisée par les plaques de détection de la cellule ICR. La rotation des ions

dans la cellule induit un courant électrique à partir duquel les fréquences sont déterminées par transformées de Fourier, et les rapports m/z en sont déduits. Dans le cadre du Solarix 7T 2XR, équipé d'une ParaCell, les plaques d'excitation de la cellule ICR servent également de plaques de détection. Le courant mesuré donne donc le double des fréquences. Cela permet, soit de multiplier par deux la résolution obtenue pour un même temps d'acquisition, soit de doubler la vitesse d'acquisition pour une même résolution, par rapport à une détection avec uniquement deux plaques.

1.3.5. Spectrométrie de masse en tandem (MS/MS)

La spectrométrie de masse en tandem (MS/MS) permet de confirmer l'identification d'une molécule, mais aussi d'obtenir des informations structurales. À l'exception de la trappe ionique et la cellule ICR, la MS/MS requiert une combinaison de deux analyseurs, positionnés de part et d'autre d'une cellule de collision. Lorsqu'une expérience de MS/MS est réalisée, un premier scan permet de déterminer le rapport m/z de l'ion qui est sélectionné pour être ensuite fragmenté dans la cellule de collision. Les rapports m/z des fragments sont ensuite déterminés dans le second analyseur.

Différents modes de fragmentation ont été développés et peuvent fournir des informations structurales complémentaires. Les principaux modes de fragmentation sont :

- La fragmentation en source (ISD) [30] ;
- La dissociation induite par collision (CID) [31];
- La dissociation par transfert d'électrons (ETD) [32];
- La dissociation par capture d'électrons (ECD) [33] ;
- La photodissociation multi-photonique par infrarouge (IRMPD) [34] ;
- La photodissociation par rayonnement ultra-violet (UVPD) [35].

L'UVPD, l'IRMPD, l'ECD sont uniquement disponibles avec certains spectromètres de masse FTICR puisque la fragmentation a lieu directement au sein la cellule ICR. La CID, utilisée lors de ces travaux de thèse, est détaillée dans le prochain paragraphe.

1.3.5.1. Dissociation induise par collision (CID)

La CID est le mode de fragmentation le plus communément utilisé aujourd'hui. Un flux de gaz inerte (souvent de l'azote) est présent dans la cellule de collision. La collision entre ce gaz et l'ion provoque la fragmentation. Des fragments différents peuvent être obtenus en fonction de l'énergie de collision utilisée. Cette dernière, de l'ordre de quelques dizaines de V ou eV, correspond à l'énergie cinétique de l'ion qui est déterminée en fonction des tensions au niveau des lentilles d'entrée et de sortie de la cellule de collision. Le débit de gaz peut également être adapté pour augmenter ou limiter la fragmentation.

1.4. Couplage entre l'électrophorèse capillaire et la spectrométrie de masse

Depuis les débuts du couplage CE-MS en 1987 [36], de nombreuses applications ont été développées, notamment pour l'analyse de mélanges complexes. En plus des avantages classiques d'un couplage entre une méthode séparative et la MS, tels que la réduction de la complexité des spectres et des effets de suppression d'ions ainsi que la discrimination de certains isomères, la CE apporte une grande efficacité de séparation et une sélectivité complémentaire aux autres méthodes séparatives.

Cependant, lors d'un couplage entre la CE et la MS, deux contraintes doivent être respectées. La première concerne la stabilité du spray ESI malgré les faibles débits générés par le flux électroosmotique. La seconde est la nécessité de maintenir le contact électrique entre les deux électrodes de la CE, alors que la sortie du capillaire est positionnée devant l'entrée du spectromètre de masse. Le maintien du contact électrique est essentiel afin d'éviter la formation d'un arc électrique entre la sortie de capillaire et l'entrée du spectromètre de masse.

Pour répondre à ces contraintes, de nombreuses interfaces ont été développées depuis les débuts du couplage CE-MS. Elles peuvent être regroupées selon trois catégories : les interfaces *sheath-liquid*, les interfaces *liquid junction* et les interfaces *sheathless*.

Les interfaces dites « *sheath-liquid* » sont les plus utilisées grâce à leur robustesse. La première a été décrite pour la première fois en 1988 par Smith *et al.* [37]. Cette interface se compose d'un nébuliseur dans lequel est placé le capillaire de la CE. Le maintien du contact électrique est assuré par l'ajout d'un capillaire co-axial rempli d'un liquide additionnel conducteur relié à l'électrode de sortie. Ce liquide additionnel se mélange avec le BGE à la sortie du capillaire de la CE et garantit ainsi le maintien du contact électrique et la stabilité du spray, mais une diminution de la sensibilité en détection MS est généralement causée par la dilution des analytes.

Les interfaces dites « *liquid junction* », introduites pour la première fois en 1988 par Lee *et al.* [38], utilisent également un liquide additionnel pour maintenir le contact, mais la jonction entre celui-ci et le BGE est différente. Les designs de *liquid junction* ont pour but de réduire les effets de dilution des analytes observés avec les interfaces *sheath-liquid*.

Enfin, les interfaces dites « *sheathless* », décrites lors du tout premier couplage CE-MS en 1987 par Olivares *et al.* [36], n'utilisent pas de liquide additionnel et suppriment donc totalement les effets de dilution observés avec les autres interfaces. Le contact électrique est maintenu par différents systèmes de connexion directement reliés au capillaire de la CE.

Dans le paragraphe suivant, uniquement l'interface *sheathless* utilisée lors de ces travaux sera détaillée.

1.4.1. Interface sheathless

Il existe plusieurs design d'interfaces *sheathless*, dont celui utilisé lors de ces travaux de thèse : l'interface *sheathless* qui utilise un capillaire dont l'extrémité devant la source MS est poreuse. Elle a été développée par Moini *et al.* En 2007 [39], puis reprise par Beckmann Coulter devenu Sciex Separation qui la commercialise sous le nom *CESI-MS* avec les systèmes CESI 8000. Cette interface est devenue la plus utilisée parmi les interfaces *sheathless* pour de nombreuses applications (telles que les analyses de protéines [40] ou d'ARN [6], ou encore les analyses métabolomiques [41]).

Cette interface est constituée de deux capillaires (Figure 1.14). Le premier, appelé capillaire de séparation, est celui où a lieu la séparation électrophorétique. Il possède une extrémité poreuse passant dans une canule en acier inoxydable et dirigée vers la source MS. Pour obtenir cette extrémité poreuse, la silice a subi, sur cette partie, une attaque à l'acide fluorhydrique. L'autre extrémité, ainsi que l'électrode d'entrée de la CE, sont plongées dans le BGE. Le second, appelé capillaire de contact, sert à maintenir le contact électrique au niveau de l'extrémité poreuse du capillaire de séparation. Pour cela, une de ces extrémités est plongée dans le BGE au niveau de l'électrode de sortie. Son autre extrémité est positionnée dans la canule afin de la remplir de BGE. Ce dernier est donc en contact avec la partie poreuse du capillaire de séparation. La porosité de la silice à l'extrémité du capillaire de séparation laisse passer, à travers la paroi, les ions de faibles rayons, ce qui permet de maintenir le contact électrique. Les faibles débits présents dans le capillaire de séparation (< 20 – 30 nL/min) permettent de générer un spray nanoESI directement à l'extrémité poreuse du capillaire de séparation.

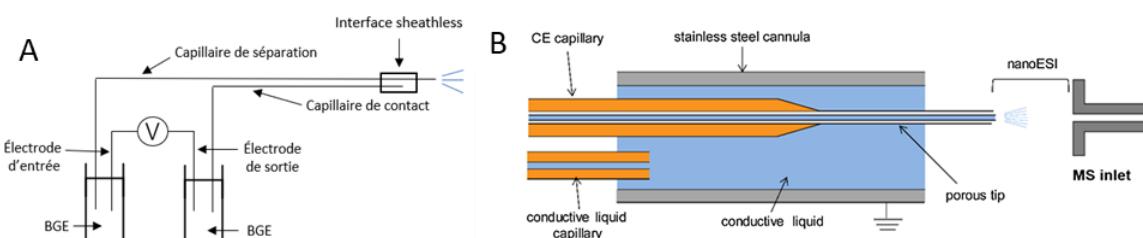


Figure 1.14 : Représentations schématiques (A) d'une CE utilisant l'interface sheathless et (B) de l'interface sheathless en elle-même. Extrait et adapté de [42].

Au laboratoire, un protocole pour l'attaque des capillaires à l'acide fluorhydrique a été développé il y a plusieurs années afin de faire nos propres capillaires. Ce protocole a fait l'objet, lors de ma thèse, d'un chapitre de livre disponible en Annexe 1 : Chapitre de livre sur l'attaque des capillaires[43].

Bien que plus fragile que les autres à cause de la porosité du capillaire, cette interface a été choisie pour réaliser ces travaux de thèse, car elle favorise une ionisation efficace ainsi qu'une bonne sensibilité. En effet, les faibles débits issus du flux électroosmotique permettent d'obtenir de très fines gouttelettes, plus facilement soumises aux explosions coulombiennes

lors de l'ionisation, et l'absence de liquide additionnel permet d'éviter les problèmes liés à la dilution des analytes.

1.5. Références

- [1] Tiselius, A., *The Moving Boundary Method of Studying the Electrophoresis of Proteins*; Nova acta Regiae Societatis Scientiarum Upsaliensis; Almqvist & Wiksell, 1930.
- [2] Haselberg, R. *et al.*, “Capillary Electrophoresis–Mass Spectrometry for the Analysis of Intact Proteins.”, *J. Chromatogr. A*, **2007**, 1159, 81–109.
- [3] Scriba, G. K. E. *et al.*, “Separation of Peptides by Capillary Electrophoresis.”, *Methods Mol. Biol. Clifton NJ*, **2008**, 384, 483–506.
- [4] Chankvetadze, B., “Separation of Enantiomers with Charged Chiral Selectors in CE.”, *ELECTROPHORESIS*, **2009**, 30 (S1), S211–S221.
- [5] Durney, B. C. *et al.*, “Capillary Electrophoresis Applied to DNA: Determining and Harnessing Sequence and Structure to Advance Bioanalyses (2009–2014).”, *Anal. Bioanal. Chem.*, **2015**, 407, 6923–6938.
- [6] Lechner, A. *et al.*, “Characterization of Post-Transcriptional RNA Modifications by Sheathless Capillary Electrophoresis–High Resolution Mass Spectrometry.”, *Anal. Chem.*, **2020**, 92 (10), 7363–7370.
- [7] Ramautar, R., “Capillary Electrophoresis–Mass Spectrometry for Clinical Metabolomics.” in: *Advances in Clinical Chemistry*; Elsevier, 2016; Vol. 74, pp 1–34.
- [8] Gahoual, R., “Développement Du Couplage Électrophorèse Capillaire - Spectrométrie de Masse Haute Sensibilité : Application à La Caractérisation Fine de Protéines,” Strasbourg - Ecole Doctorale des Sciences Chimiques, 2014.
- [9] Huhn, C. *et al.*, “Relevance and Use of Capillary Coatings in Capillary Electrophoresis–Mass Spectrometry.”, *Anal. Bioanal. Chem.*, **2010**, 396 (1), 297–314.
- [10] Lössl, P. *et al.*, “The Diverse and Expanding Role of Mass Spectrometry in Structural and Molecular Biology.”, *EMBO J.*, **2016**, 35 (24), 2634–2657.
- [11] Yamashita, M. *et al.*, “Electrospray Ion Source. Another Variation on the Free-Jet Theme.”, *J. Phys. Chem.*, **1984**, 88 (20), 4451–4459.
- [12] Taylor, G. I., “Disintegration of Water Drops in an Electric Field.”, *Proc. R. Soc. Lond. Ser. Math. Phys. Sci.*, **1964**, 280 (1382), 383–397.
- [13] Kebarle, P. *et al.*, “From Ions in Solution to Ions in the Gas Phase - the Mechanism of Electrospray Mass Spectrometry.”, *Anal. Chem.*, **1993**, 65 (22), 972A–986A.
- [14] Kebarle, P., “A Brief Overview of the Present Status of the Mechanisms Involved in Electrospray Mass Spectrometry.”, *J. Mass Spectrom.*, **2000**, 35 (7), 804–817.
- [15] Rayleigh, Lord, “XX. On the Equilibrium of Liquid Conducting Masses Charged with Electricity.”, *Lond. Edinb. Dublin Philos. Mag. J. Sci.*, **1882**, 14 (87), 184–186.
- [16] Rohner, T. C. *et al.*, “Electrochemical and Theoretical Aspects of Electrospray Ionisation.”, *Phys. Chem. Chem. Phys.*, **2004**, 6 (12), 3056–3068.
- [17] Dole, M. *et al.*, “Molecular Beams of Macroions.”, *J. Chem. Phys.*, **1968**, 49 (5), 2240–2249.
- [18] Mack, L. L. *et al.*, “Molecular Beams of Macroions. II.”, *J. Chem. Phys.*, **1970**, 52 (10), 4977–4986.
- [19] Iribarne, J. V. *et al.*, “On the Evaporation of Small Ions from Charged Droplets.”, *J. Chem. Phys.*, **1976**, 64 (6), 2287–2294.

- [20] Konermann, L. *et al.*, “Unraveling the Mechanism of Electrospray Ionization.”, *Anal. Chem.*, **2013**, 85 (1), 2–9.
- [21] Emmett, M. R. *et al.*, “Micro-Electrospray Mass Spectrometry: Ultra-High-Sensitivity Analysis of Peptides and Proteins.”, *J. Am. Soc. Mass Spectrom.*, **1994**, 5 (7), 605–613.
- [22] Wilm, M. S. *et al.*, “Electrospray and Taylor-Cone Theory, Dole’s Beam of Macromolecules at Last?”, *Int. J. Mass Spectrom. Ion Process.*, **1994**, 136 (2), 167–180.
- [23] Holčapek, M. *et al.*, “Recent Developments in Liquid Chromatography-Mass Spectrometry and Related Techniques.”, *J. Chromatogr. A*, **2012**, 1259, 3–15.
- [24] “Introduction To LC-MS Technology.” in: *Interpretation of MS-MS Mass Spectra of Drugs and Pesticides*; John Wiley & Sons, Ltd, 2017; pp 1–53.
- [25] Qi, Y. *et al.*, “Data Processing in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry.”, *Mass Spectrom. Rev.*, **2014**, 33 (5), 333–352.
- [26] Boldin, I. A. *et al.*, “Fourier Transform Ion Cyclotron Resonance Cell with Dynamic Harmonization of the Electric Field in the Whole Volume by Shaping of the Excitation and Detection Electrode Assembly.”, *Rapid Commun. Mass Spectrom.*, **2011**, 25 (1), 122–126.
- [27] Amorim Madeira, P. J. *et al.*, “High Resolution Mass Spectrometry Using FTICR and Orbitrap Instruments.” in: *Fourier Transform - Materials Analysis*; Salih, S., Ed.; InTech, 2012.
- [28] Schweikhard, L. *et al.*, “The Trapping Condition and a New Instability of the Ion Motion in the Ion Cyclotron Resonance Trap.”, *Int. J. Mass Spectrom. Ion Process.*, **1995**, 141 (1), 77–90.
- [29] Ladislas Wiza, J., “Microchannel Plate Detectors.”, *Nucl. Instrum. Methods*, **1979**, 162 (1), 587–601.
- [30] He, Z. *et al.*, “The Application of In-Source Fragmentation in Ultra-High Performance Liquid Chromatography-Electrospray Ionization -Tandem Mass Spectrometry for Pesticide Residue Analysis.”, *J. Chromatogr. A*, **2020**, 1633, 461637.
- [31] Mitchell Wells, J. *et al.*, “Collision-Induced Dissociation (CID) of Peptides and Proteins.” in: *Methods in Enzymology*; Biological Mass Spectrometry; Academic Press, 2005; Vol. 402, pp 148–185.
- [32] Syka, J. E. P. *et al.*, “Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry.”, *Proc. Natl. Acad. Sci. U. S. A.*, **2004**, 101 (26), 9528–9533.
- [33] Xie, Y. *et al.*, “Top-Down ESI-ECD-FT-ICR Mass Spectrometry Localizes Noncovalent Protein-Ligand Binding Sites.”, *J. Am. Chem. Soc.*, **2006**, 128 (45), 14432–14433.
- [34] Little, D. P. *et al.*, “Infrared Multiphoton Dissociation of Large Multiply Charged Ions for Biomolecule Sequencing.”, *Anal. Chem.*, **1994**, 66 (18), 2809–2815.
- [35] Becher, S. *et al.*, “Influence of Protein Ion Charge State on 213 Nm Top-down UV-PD.”, *Analyst*, **2021**, 146 (12), 3977–3987.
- [36] Olivares, J. A. *et al.*, “On-Line Mass Spectrometric Detection for Capillary Zone Electrophoresis.”, *Anal. Chem.*, **1987**, 59 (8), 1230–1232.
- [37] Smith, R. D. *et al.*, “Improved Electrospray Ionization Interface for Capillary Zone Electrophoresis-Mass Spectrometry.”, *Anal. Chem.*, **1988**, 60 (18), 1948–1952.
- [38] Lee, E. D. *et al.*, “On-Line Capillary Zone Electrophoresis-Ion Spray Tandem Mass Spectrometry for the Determination of Dynorphins.”, *J. Chromatogr. A*, **1988**, 458, 313–321.

- [39] Cao, P. *et al.*, “A Novel Sheathless Interface for Capillary Electrophoresis/Electrospray Ionization Mass Spectrometry Using an in-Capillary Electrode.”, *J. Am. Soc. Mass Spectrom.*, **1997**, 8 (5), 561–564.
- [40] Gahoual, R. *et al.*, “Novel Sheathless CE-MS Interface as an Original and Powerful Infusion Platform for NanoESI Study: From Intact Proteins to High Molecular Mass Noncovalent Complexes.”, *Anal. Bioanal. Chem.*, **2014**, 406 (4), 1029–1038.
- [41] Sánchez-López, E. *et al.*, “Sheathless CE-MS Based Metabolic Profiling of Kidney Tissue Section Samples from a Mouse Model of Polycystic Kidney Disease.”, *Sci. Rep.*, **2019**, 9 (1), 806.
- [42] Gahoual, R. *et al.*, “Revealing the Potential of Capillary Electrophoresis/Mass Spectrometry: The Tipping Point.”, *Rapid Commun. Mass Spectrom.*, **2019**, 33 (S1), 11–19.
- [43] Gosset-Erard, C. *et al.*, “Protocol for Etching Bare-Fused Silica Capillaries for Sheathless Capillary Electrophoresis–Mass Spectrometry Coupling.” in: *Capillary Electrophoresis-Mass Spectrometry : Methods and Protocols*; Neusüß, C. *et al.*, Eds.; Methods in Molecular Biology; Springer US: New York, NY, 2022; pp 49–59.

CHAPITRE 2 : LES COUPLAGES ENTRE LA FTICR-MS ET LES METHODES SEPARATIVES

2.1. Introduction

Depuis l'obtention du premier spectre en 1974 [1], la spectrométrie de masse à résonance cyclotronique des ions et à transformées de Fourier (FTICR-MS) a connu de nombreux progrès technologiques qui l'ont amenée à être un instrument de référence connu pour son ultra-haute résolution et sa très grande précision de mesure de masse. La FTICR-MS offre donc, en infusion directe, des avantages indéniables pour le profilage moléculaire d'une grande variété de mélanges complexes. Pour atteindre de telles capacités, la FTICR-MS requiert des temps d'acquisition relativement élevés et qui peuvent sembler, à première vue, incompatibles avec les fenêtres de temps disponibles lors d'une séparation par chromatographie liquide (LC), chromatographie gazeuse (GC) ou électrophorèse capillaire (CE). Afin de rendre possible le couplage de la FTICR-MS avec une méthode séparative et combiner leurs différents avantages, il est nécessaire de faire certains compromis, notamment sur le temps d'acquisition dans la cellule ICR. Ces compromis ont un impact direct sur la résolution et la précision de mesure de masse accessibles en couplage.

Les couplages on-line entre la FTICR-MS et les méthodes de séparation soulèvent deux interrogations principales. D'une part, du point de vue des utilisateurs de méthodes séparatives, pourquoi un instrument aussi coûteux devrait-il être utilisé comme détecteur alors que des alternatives moins chères et plus compatibles sont disponibles ? D'autre part, du point de vue des massistes, quels sont les avantages à utiliser des méthodes de séparation chronophages en amont de la FTICR-MS, alors que de plus grandes résolution et précision de mesure de masse peuvent être obtenues en infusion directe ?

Pour discuter de ces points de vue, une revue bibliographique a été rédigée afin d'expliquer les défis à relever pour un tel couplage et en préciser les intérêts. Cette revue s'est voulue être la plus exhaustive possible (sans être à l'abri d'une ou deux références qui aurait pu échapper au crible) et montre la variété du champ des possibles en couplage LC, GC et CE avec la FTICR-MS illustrée par de nombreuses applications puisées au sein du répertoire. Cette revue a été soumise dans le journal à comité de lecture *Talanta*.

2.1. Publication : « Hyphenation of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FTICR-MS) with separation methods: the art of compromise and the possible - a review »

Clarisse Gosset-Erard, Frédéric Aubriet, Emmanuelle Leize-Wagner, Yannis-Nicolas François and Patrick Chaimbault

Submitted in *Talanta*.

2.1.1. *Introduction*

There is quite a long time now since, in 1974, the first spectrum obtained by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR MS) was published by Comisarow and Marshall [1]. With the relentless progress in this technology, nowadays, evoking FT-ICR MS is immediately synonymous of extreme resolution mass spectrometry (with typical resolving powers higher than 1 000 000 and up to 24 000 000 [2]) and very high mass measurement accuracy (sub parts-per-million enabling the determination of the molecular formula). FT-ICR MS provides incontestable benefits for the molecular profiling of all types of complex mixtures. Direct flow injection FT-ICR MS is widely used in many “omic” studies such as proteomic [3], metabolomics [4], foodomic [5] petroleomic [6] and environmental analyses [7]. This non-exhaustive list of applications deserves to be completed by Mass Spectrometry Imaging for which, the unique FT-ICR MS resolving power enables to distinguish the specific localization of isobaric compounds in human [8] or plant tissues [9]. At last, different fragmentation modes, such as sustained off-resonance irradiation collision-induced dissociation (SORI-CID), infrared multiple photon dissociation (IRMPD) or electron-capture dissociation (ECD), are often implemented to complete structural information obtained in these FT-ICR MS profiles by direct infusion [10] and to confirm the signal attributions to specific analytes.

If the FT-ICR MS story began in the middle of 1970's, the chapter concerning its hyphenation with separation methods started being written a little bit later. Indeed, very high-resolution mass measurement requires a low acquisition time frequency, which has long been an obstacle to the implementation of coupling of FT-ICR MS to separation methods. Ledford *et al.* were the first to attempt the adventure of the hyphenation with gas chromatography (GC) in 1980 [11]. Getting a high-resolution ICR mass spectrum from a GC peak has not been easy to reach at that time, mainly due to the vacuum problems when GC is interfaced with FT-ICR MS. However, since 1983, Sack and Gross succeeded in reaching a resolution of 63 000 (full width at half maximum, FWHM) at m/z 128.062 (for naphthalene) in GC-FT-ICR MS [12]. The first coupling of FT-ICR MS with liquid chromatography (LC) is achieved in the early 1990's by Stockston *et al.* [13] using electrospray (ESI) as ion source. In 1994, the LC-ESI-FT-ICR MS analysis of peptides carried out by Stacey *et al.* reached a resolution of 35 000 (FWHM) [14] even if it appears to be very low compared with actual performance in LC-FT-ICR MS. The sensitivity was also not entirely successful but, thanks to the progress in ion trapping, this

drawback was slowly overcome. Since 2004, a review in this field claimed that “The coupling of liquid chromatography (LC) to mass spectrometry (MS) is nowadays well established in routine analysis” [15]. “Routine” is still appearing a little bit excessive today for LC-FT-ICR MS, but this is probably the view in a limited number of specialized research groups. Capillary electrophoresis (CE) was the last online separation method coupled to FT-ICR MS. Even if CE is still today less widespread than LC, curiously their hyphenations with FT-ICR MS were developed almost simultaneously and Hofstadler *et al.* published the first papers in CE-FT-ICR MS in 1993 [16].

Online hyphenation of FT-ICR MS with separation methods raises at least two questions. On one hand, from a chromatographer point of view, why would it be necessary to use a so expensive instrument as detector when cheaper alternatives are available? On the other hand, from a mass spectrometrist point of view, why would it be necessary to use a time-consuming separation method when the resolving power of FT-ICR mass spectrometer is doing the job? This review will try to answer both questions before moving on to the matter of applications of online hyphenation.

2.1.2. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FTICR-MS)

The first article dealing with gas chromatography coupled with FT-ICR MS was published more than 25 years now. However, at the exception of the reviews published by Schrader *et al.* in 2004 [15] and Hertzog *et al.* in 2021 [17], the published reviews in the field of FT-ICR MS were generally focused on recent research and technological developments in the field of the MS instrumentation rather than on the hyphenation with separation methods. The first paragraph of this review is not the umpteenth version of theoretical aspects of FT-ICR MS. Nevertheless, to understand the challenges, the instrumental specificities, and the benefits of online FT-ICR MS hyphenation with separation methods, it is essential to dive into few simplified theoretical aspects of FT-ICR MS.

2.1.2.1. FTICR-MS principle

As depicted in Figure 2.1, after their ionization in the source, the ions are accumulated and transferred by packets into the ion cyclotron resonance (ICR) cell. The ion transfer (including their accumulation) between the source and the FT-ICR MS cell is a crucial step to understand the challenge of coupling FT-ICR MS with separation methods, this issue will be, however, discussed later in section 2.1.3.

The ICR cell is placed in an ultra-high vacuum region (typically lower than 10^{-9} Torr) located itself inside a superconducting magnet producing high-field intense magnetic field (B_0). At its simplest, the cell is composed of two trapping plates, two excitation plates and two detection plates. After their transfer, the ions are localized close to the center of the ICR cell, on relatively small orbits around B_0 axis. The ions, which circular motion is uniformly distributed (non-coherent motion), are then excited due to a swept RF pulse orthogonal to

the magnetic field. As a result, ions having the same mass-to-charge (m/z) ratio acquire a coherent motion and bundles together on a higher gyration orbit allowing their detection by the detection plates.

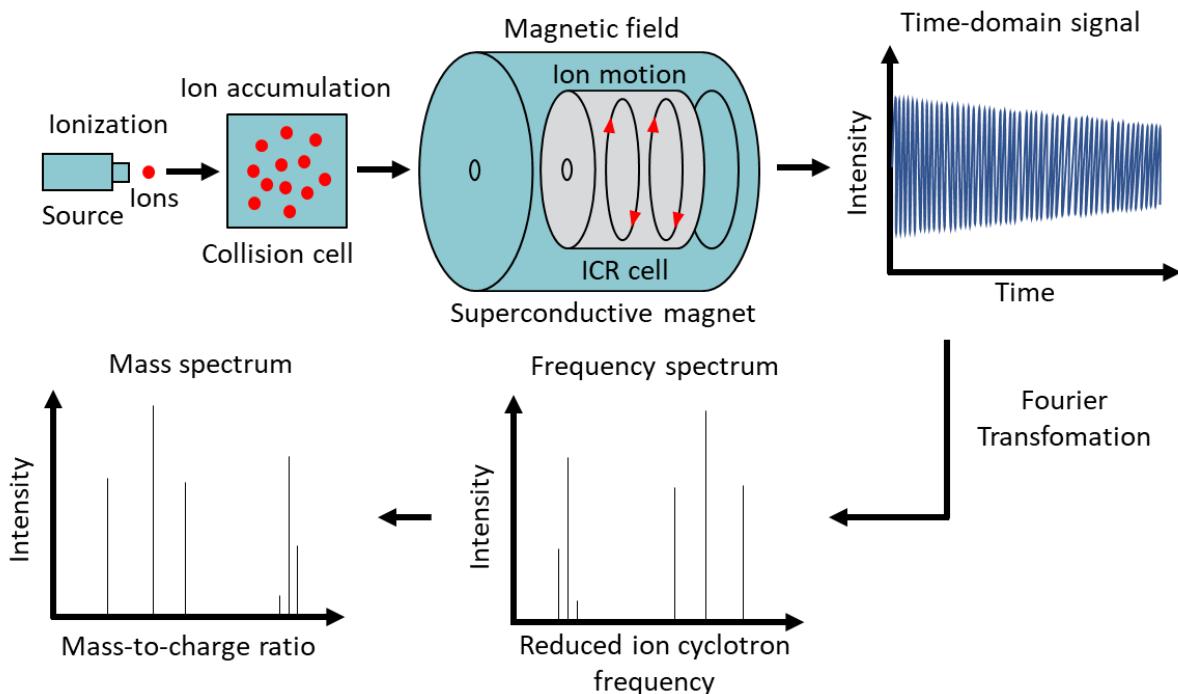


Figure 2.1: Brief principle of ultra-high resolution mass spectrometry using FTICR-MS. Ionized analytes are accumulated in the collision cell and sent by packet in the ICR cell. Ions are excited and the frequencies are detected for the deconvolution by Fourier-transform and converted from a frequency spectrum to a mass spectrum. Reproduced and adapted from [18].

This motion, called the cyclotron motion, is the basis of the m/z measurement in FT-ICR MS. The ion packets are rotating at their own cyclotron frequencies w_c which are proportional to the inverse of their m/z ratios as presented in Equation (2.1) [19]:

$$w_c = \frac{eB_0}{m/z} \quad (2.1)$$

With w_c the ion cyclotron frequency e the elementary charge and B_0 the magnetic field value.

The recorded sinusoidal signal induced by the ion motion near the detection plates is called time domain signal or transient. FTMS transient is curved with a rapidly decreasing frequency. This transient acquisition is the longest step in FT-ICR MS analysis since the resolution and the mass measurement accuracy increase with low acquisition rate. The time domain signal is then converted into cyclotron frequencies by Fourier transformation and finally, the mass spectrum is deduced from cyclotron frequencies.

In practice, the overall motion of the ions in the cell is much more complex due to the combined action of an electrostatic field (application of potential on the trapping plates) and the magnetic field. Indeed, harmonic oscillations, called the trapping motion, are created by the electrical potential applied between the two trapping plates, which is required to confine the ions in the ICR cell. The trapping frequency is defined by Equation (2.2) [19]:

$$w_T = \sqrt{\frac{2e \propto V_T}{m/z}} \quad (2.2)$$

With w_T the ion trapping frequency, $\propto V_T$ a coefficient proportional to the trapping voltage V_T , e the elemental charge constant, V_T the trapping voltage and m/z the ion mass-to-charge ratio.

Because of the trapping motion, the real frequency of ion cyclotron motion is modified from the theoretical one. The reduced ion cyclotron frequency w_+ is actually the detected frequency (from which the mass spectrum is deduced) and is defined by Equation (2.3) [19]:

$$w_+ = \frac{w_c}{2} + \sqrt{\frac{w_c^2}{4} - \frac{w_T^2}{2}} \quad (2.3)$$

With w_+ the reduced ion cyclotron frequency, w_T the ion trapping frequency and w_c the ion cyclotron frequency.

Finally, the combination of the magnetic field B_0 and the electrical potential created by the trapping plates induces the magnetron motion of the ions around the magnetic field axis. The magnetron frequency w_- is described by Equation (2.4) [19]:

$$w_- = \frac{w_c}{2} - \sqrt{\frac{w_c^2}{4} - \frac{w_T^2}{2}} \quad (2.4)$$

With w_- the ion magnetron frequency, w_T the ion trapping frequency and w_c the ion cyclotron frequency.

The consequence of the superposition of these three motions is a very complex trajectory (represented in the Figure 2.2). More details on the deep theory of FT-ICR MS can be found in Marshall *et al.* paper [10].

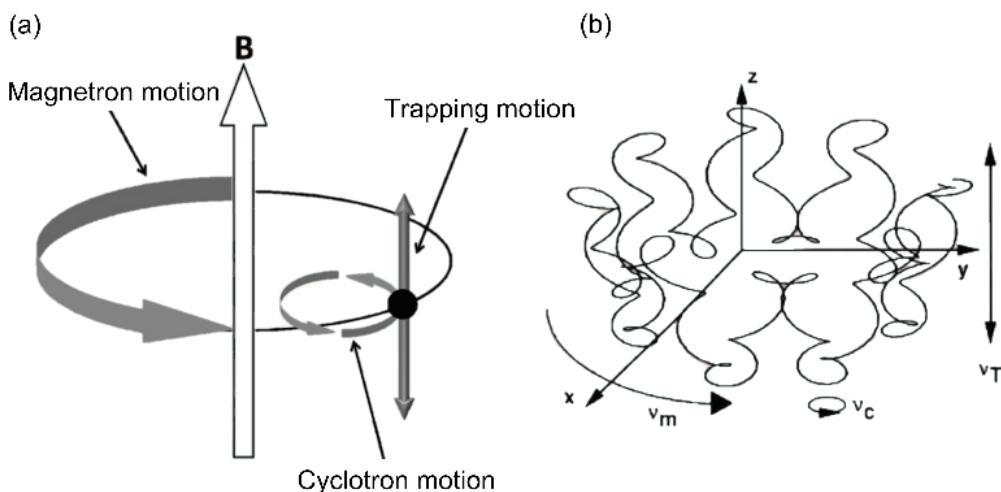


Figure 2.2: Schematic representation of (a) the three natural motions of an ion confined in an ICR cell and (b) the resulting ion trajectory shape. Adapted from [20].

2.1.2.2. Advantages of FTICR-MS

Everyone will understand that, even if FT-ICR mass spectrometers became more user-friendly than in the past, they are still more complicated to manage than other mass spectrometers. This is largely compensated by the provided advantages. FT-ICR MS is currently the mass spectrometer providing the highest resolution and the best mass measurement accuracy at least in low m/z range [21,22]. Indeed, the use of a strong magnetic field enables FT-ICR MS to provide an ultra-high resolution and a high mass measurement accuracy. Already in the beginning of the 2010's, the resolving power of a 7 T FT-ICR MS equipped with a dynamically harmonized ICR cell reported by Nikolaev *et al.* reached 24 000 000 at m/z 609 (protonated reserpine) with a mass measurement accuracy of 0.9 ppm for a detection time of 180 s [2]. Concerning higher mass compounds such as proteins, Kelleher *et al.* achieved in 1997, with custom build 9.4 T FT-ICR mass spectrometer and a detection time of 100-150 s, the unit mass resolution with a 3 Da-accuracy for a 112 kDa-protein (Chondroitinase II), which is corresponding to a resolution power of 170 000 [23]. Later on, Nikolaev *et al.* achieved isotopic resolution for a 186 kDa tetrameric enolase with a 7 T FT-ICR MS [24]. Some of the keys to reach such ultra-high resolutions are undoubtedly the progress in very high vacuum technologies (a typical vacuum around $10^{-10} - 10^{-11}$ bar is required in the ICR cell to enable such long acquisition times) and more recent instrumental and data treatment developments. Those will be discussed later in this review.

The high mass measurement accuracy achieved by the FT-ICR enables the reliable assignment of a molecular formula to each detected feature, and the high resolution ensures the distinction of very close isobaric species. Indeed, only high-resolution mass spectrometry enables the separation of overlapping isotopologue peaks containing ^{13}C , ^2H , ^{18}O , ^{15}N , ^{31}P and ^{34}S , as shown in Figure 2.3. The isotopic peak pattern gives crucial information on the elemental composition. Therefore, high resolution is helping to narrow down the number of candidates for elemental compositions until the assignment of an unambiguous elemental

formula especially for small molecules [18,25,26]. Another advantage of the FT-ICR MS is the availability of numerous fragmentation techniques. Indeed, besides the common collision induced dissociation (CID) and electron-transfer dissociation (ETD) fragmentation techniques, ultraviolet photodissociation (UVPD), infrared multiphoton dissociation (IRMPD) and electron-capture dissociation (ECD) fragmentation techniques are available in the ICR cell. All those techniques produce complementary MS^n data, enabling easier and deeper structural elucidation of analytes [27].

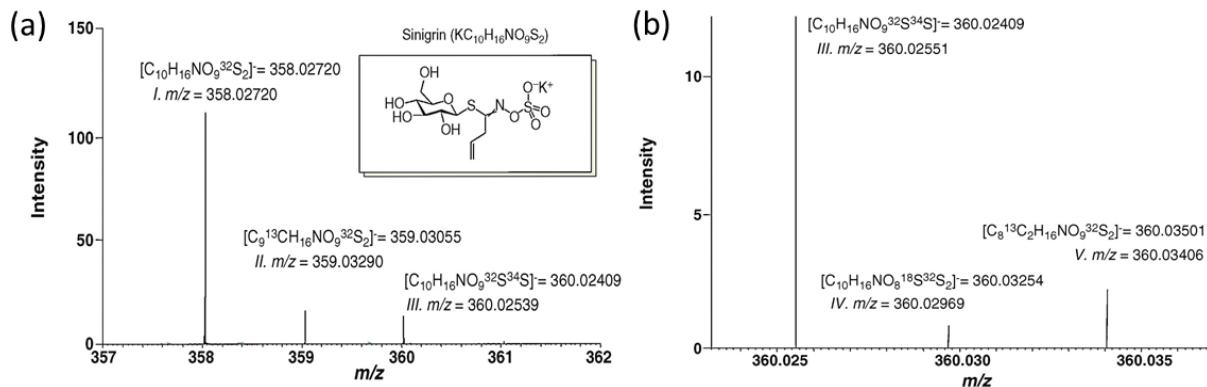


Figure 2.3: (a) Example of the ultra-high-resolution MS analysis using a FT-ICR MS. The MS spectra were obtained using an LTQ-FT-ICR MS (Thermo Scientific). A glucosinolate produced by the Brassicaceae plants, sinigrin ($KC_{10}H_{16}NO_9S_2$), was subjected to the separation of isotope ion peaks in an ESI-negative mode. In a broadband spectrum, the detected ions of I ($m/z = 358.02890$), II ($m/z = 359.03290$), and III ($m/z = 360.02539$) corresponded to $[C_{10}H_{16}NO_9S_2]^-$, $[C_9^{13}CH_{16}NO_9^{32}S_2]^-$, and $[C_{10}H_{16}NO_9^{32}S^{34}S]^-$, respectively. (b) Resolution of isotope ions using a FT-ICR MS. Magnification of the m/z window near the ion III ($C_{10}H_{16}NO_9^{32}S^{34}S$, $m/z = 360.02551$) further revealed the presence of ions of V ($m/z = 360.02969$) and VI ($m/z = 360.03406$) corresponding to $[C_{10}H_{16}NO_8^{18}O^{32}S_2]^-$ and $[C_8^{13}C_2H_{16}NO_9^{32}S_2]^-$, respectively. Figure and caption extracted from [18].

2.1.2.3. Limits of FTICR-MS overcome by hyphenation with separation methods

FT-ICR MS is a powerful tool for elemental composition assignment and structural elucidation thanks to its ultra-high resolution, its sub-ppm mass measurement accuracy, and its multiple fragmentation modes. However, this technique still has some limitations when used in direct infusion.

There are two essential “prerequisites” to the specific detection and identification of a given analytes in mass spectrometry and FT-ICR MS is no exception to this rule.

The first rule concerns the charge “z” itself: no charge means no ion, making the m/z measurement impossible. The direct infusion of a complex mixtures or matrices is often subject to the well-known ion suppression effect. Indeed, during the ionization step, the analytes of interest can compete with other compounds to become electrically charged, leading to reduction of their ionization efficiency and thus a poor detection sensitivity. Such ionization suppression was nicely demonstrated by Guillemant *et al.* in a recent publication dealing with the analysis of N and S hydrocarbons in gasoline [28]. A separation prior MS detection limits the ion suppression phenomenon since fewer molecules are present simultaneously in the ion source [29–33] and in spite of the chromatographic dilution, the number of detected compounds is higher than in direct infusion, including miniaturized

techniques such as microchip mass spectrometry. For example, during their proteomic study of cerebrospinal fluid tryptic digest, Wetterhal *et al.* succeeded in identifying by CE-FT-ICR MS twice the number of identified compounds by direct infusion [34]. The authors attributed this difference to a better selectivity but also to a lower ion suppression in CE-MS.

Precisely, the above-mentioned conclusion leads to the second issue which concerns the discrimination of isomers that cannot be achieved only by MS detection in spite of the high resolving power of the mass spectrometer. Unless ion-mobility is used, structural isomers cannot be distinguished since they have the same exact mass. The solution is their chromatographic or electrophoretic separation before their MS detection. Then, isomers can be discriminated and identified on the base of their retention/migration time. This approach may be associated to MS^2 and MS^n data, and eventually comparison to commercially available standard [35–38]. Mass spectra can be very complex, particularly with extreme resolution mass spectrometer such as FT-ICR MS. The additional dimension brought by the selectivity of a separation method, regardless of the obvious question of isomers, reduces the complexity of the individual mass spectra obtained along the run and thus facilitating the assignment of a molecular formula to each detected ion. Despite this, Zacs *et al.* identified more than 5 000 m/z values in every GC-FT-ICR MS data point of their untargeted analysis of brominated and chlorinated flame retardants in food [39], which still represents a huge work of data interpretation. In their studies, the assignment of the elemental compositions was facilitated by the specific isotope patterns of $^{79/81}Br$ and $^{35/37}Cl$ atoms. However, it is easy to imagine how tremendous would have been this data treatment in case of compounds overlapping.

Apart from these two critical points, there are still few additional advantages in the hyphenation. As already mentioned, the separation technique is reducing the number of compounds simultaneously present in the ion source. Consequently, it can also help limiting the peak coalescence of ions having close m/z values which is observed when the charge load is too high in the ICR cell. When the ion clouds are too dense, the MS peaks tend to merge together but the chromatographic dilution decreases the amount of ions in the analyzer cell and the peak coalescence phenomenon is less likely to happen [40]. In the same idea, the hyphenation of ultra-high resolution mass spectrometry with separation methods enables the increase of the dynamic range since the number of ions in the cell is reduced [31,41–44]. Limiting the number of ions in the ICR cell allows reducing the charge-space effect which can cause frequency shifts, and therefore mass-to-charge ratio shifts, due to Coulombic interactions between ions [45]. Direct infusion is thus less accurate in the quantification of the analytes, and relative and absolute quantifications are rather performed by hyphenation with a separation method [39,41,46–50]. At last, direct infusion is limited for low sample concentration and small volume contrary to hyphenated methods [34]. Sensitivity can be even improved using miniaturized and low flow rates separation techniques [51]. Such hyphenations are essential when working with samples in very restricted/limited available quantities, which is a common issue with biological or environmental samples.

In conclusion to this part, the hyphenation of a separation method to an FT-ICR mass spectrometer enables separation of isomers, extended dynamic range, reduced complexity for each spectrum obtained along the analytical run, easier and deeper data interpretation, and quantification. Hence, hyphenations of separation methods with FT-ICR mass spectrometers have been increasingly developing over the past few years for various applications. Hyphenation of liquid chromatography (LC) [30,31,43,49,52–54], gas chromatography (GC) [39,55,56] and capillary electrophoresis (CE) [16,41,57–59] with FT-ICR MS are used in multiple studies. The choice of the separation techniques is based on the analytes of interest, the matrix of the sample and the sample preparation needed, to provide the best selectivity and sensitivity possible.

2.1.3. Challenges of FTICR-MS and separation method hyphenation

As mentioned in the introduction, the FT-ICR MS chapter concerning its hyphenation with separation techniques started in 1980 with gas chromatography (GC) [15]. The main problem was to maintain a sufficient vacuum state (typically less than 10^{-8} Torr) for the optimal mass analysis when GC is interfaced with FT-ICR MS. In the immediately following years, there has been an explosion for (home-made most of the time) tips dedicated to circumvent this challenge. Even if the ionization step was still sometimes carried out in the ICR cell (internal ionization) in few studies, the external ionization technique is nowadays dominating even making possible the use of an atmospheric pressure ionization (API) source and its coupling with LC and CE. With the external ionization techniques, the ion source is separated from the ICR cell by several stages of differential pumping in order to reach the required extreme low pressure (Figure 2.4). However, this solution implies to guide ions towards the ICR cell. If such interfaces procure advantages, the ion transfer from the source to the ICR cell, which can be long, is also bringing drawbacks concerning the hyphenation of separation techniques with FT-ICR MS. This will be discussed later in section 2.1.3.2.

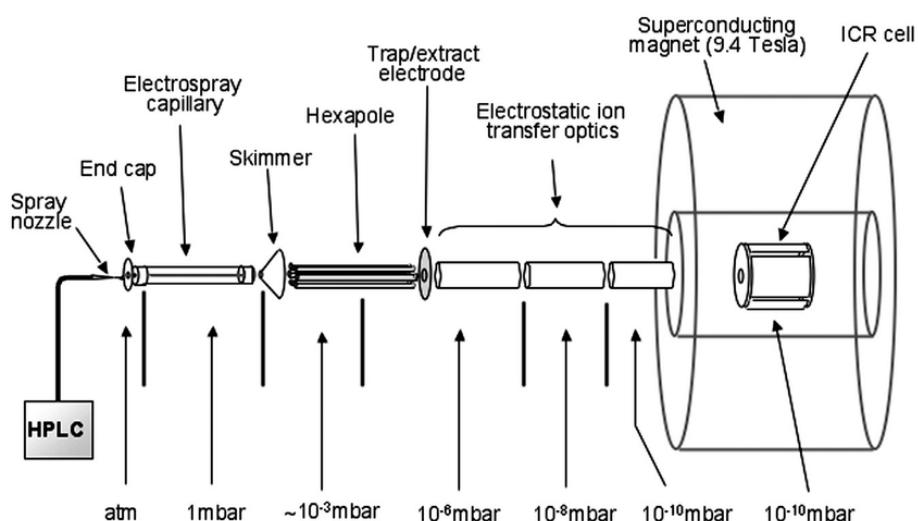


Figure 2.4: Schematic view of an HPLC-ESI-FT-ICR MS and its differential pumping stages from the atmospheric pressure ion source to the of 10^{-10} mbar-ultra high vacuum of the ICR cell (Bruker Daltonics BioAPEX -94e fitted with a 9,4T magnet). Ions formed at atmospheric pressure are transferred through a glass capillary and then through a skimmer into a hexapole ion trap where they are stored before they are further transferred to the ion cyclotron resonance cell. Extracted from [60].

2.1.3.1. Compatibility of ion sources

Solving the problem of vacuum, there are no technical barriers to the coupling. Each separation techniques can be used with its traditional ionization sources. Thus, many studies have been carried out in gas chromatography with electron impact (EI) whether in external [55,56], or, even still in internal ionization [61]. The chemical ionization, which is “softer” than EI and consequently leads more easily to the molecular weight, is less widespread in GC-FT-ICR MS. However, Szulejko *et al.* [61] developed the introduction of ethanol vapors as chemical ionization reagent and Luo *et al.* [35] employed this strategy to characterize multicomponent gasoline samples. Among the atmospheric pressure ionization (API) sources, atmospheric pressure chemical ionization (APCI) is likely the most used ionization source in GC-FT-ICR MS [39,62,63]. As expected with a soft ionization technique, APCI reduces the fragmentation compared to EI and preserves the intact molecule, whether as protonated [62] or deprotonated ions [39] depending on the chosen ion mode. APCI is rather used to analyze mid polar compounds comprising few heteroatoms such as phenols, terpenoids, flame retardants, long chain alkyl and alkylene carboxylic acids. Atmospheric photoionization laser ionization (APLI) is the latest developed among atmospheric ionization methods. In the world of ion sources dedicated to apolar compounds, APLI is the immediate rival of Atmospheric Pressure Chemical Ionization (APCI) and yet APLI, with its stepwise two-photon process, is much more efficient and selective for the ionization of non-polar compounds such as polycyclic aromatic hydrocarbons (PAHs) [64]. Not surprisingly, GC-APLI-FT-ICR MS has been successfully employed to improve the molecular coverage in PAHs in fossil oil [65].

For its part, the hyphenation of mass spectrometry with liquid chromatography (including UPLC or nanoLC) takes “the lion’s share” with around 70 % of the published papers and requires exclusively API sources among which electrospray ionization (ESI) and its declination (e.g. ultrasonic assisted nebulization ESI). This can be easily explained by the involvement of FT-ICR MS in the omics molecules of biological interest such as proteomics or metabolomics (targeted and untargeted, including its specific declination like lipidomics) but also foodomics where highly polar molecules play essential roles. This will be illustrated in the last part of this review. Decreasing the polarity of the analytes compound, APCI may also be employed in LC-FT-ICR MS. Thus, for example, UPLC-APCI-FT-ICR MS was used for the qualitative/quantitative characterization of *Spirulina* pigments dietary supplements [66]. Another example is the analysis by LC-APCI-FT-ICR MS of toxicants (e.g. hydroxyl or amino PAH) in complex environmental mixtures [67]. The coupling of LC with APPI- and APLI-FT-ICR MS is rare and to the best of our knowledge, there is very few studies published showing this hyphenation. In the paper from Lababidi and Schrader [68], the authors showed normal-phase LC-FT-ICR MS ion chromatograms of complex crude oil mixtures obtained with ESI, APPI, APCI and APLI. Their study highlights the complementarity of each source regarding the analyte polarity and aromaticity.

Due to its principle, capillary zone electrophoresis (CZE) is dedicated to the separation of ionic and ionizable compounds. It is then easy to understand that ESI is the source of predilection for its hyphenation with mass spectrometry. However, CE is probably the separation technique with the least obvious material compatibility for online hyphenation with mass spectrometry. Indeed, the principle itself of the two techniques is based on electrical charge circulation and the first challenge in CE-MS is to combine both electrical circuits, while maintaining both voltages (for the CE separation and for the MS ionization) independently. The history of CE-FT-ICR MS illustrates the host of ideas and the progresses in this matter. In the early phases, Hofstadler *et al.* published the first CE-FT-ICR MS analyses of proteins using sheathless CE interfaces [16,69,70]. Few centimeters of the external protecting polyacrylamide coating were removed from the outlet of separation capillary and the nude fused silica was etched using hydrofluoric acid to taper it. Then, a gold conductive coating was deposited allowing the electrical contact with the buffer solution and ensuring the electrospray process. Much less frequently, CE-ESI-FT-ICR MS was also implemented using the liquid junction interface [48]. However, the majority of the currently published method in CE-ESI-FT-ICR MS have mainly resorted to the coaxial liquid junction interface (sheath flow or sheath liquid) [71,42,72–74,58,75]. The make-up (or sheath-liquid) is delivered independently of the CE effluent flow rate providing a robust and reproducible spray [76] compared to the liquid junction and the composition of this make-up can be carefully optimized to improve significantly the detection sensitivity [77]. The main drawback of sheath liquid interface is the dilution effect brought by the addition of the make-up. The porous tip interface developed by Whitt and Moini is the most recently developed one [78]. There are many similarities in the capillary preparation. The external protecting polyimide coating was removed from the outlet, and the nude fused silica was also etched using hydrofluoric acid until it reaches a porosity allowing a contact closure with a conductive liquid. No conducting coating is then required and only nitrogen was introduced into the capillary during the etching process to preserve the capillary inner wall. The commercial version of porous tip interfaces from Sciex Separations (CESI 8000) is now currently available and Gstöttner *et al.* published recently a paper [59] in CE-FT-ICR MS for the analysis of bispecific antibodies subunits.

2.1.3.2. Mass accuracy, resolving power and sensitivity: the art of compromise in hyphenation

If there is no apparent technical limitation, hyphenating separation techniques with FT-ICR MS is somehow like squaring the circle and means some compromises.

Let us go back to ultra-high mass measurement accuracy and resolving power, which are the ultimate performances expected from an FTICR mass spectrometer. Mass measurement accuracy requires that the peak of interest must be distinguished from others. Consequently, accurate measurement is strongly depending on the mass resolving power R (Equation (2.5)).

$$R = \frac{m/z}{\Delta m/z} = \frac{w_c}{\Delta w_c} \quad (2.5)$$

where m/z is the mass-to-charge ratio of the peak of interest and $\Delta m/z$ the peak width at half maximum (FWHM).

The relation is obviously also true when expressed as cyclotron frequency. Combining Equation (2.1) with Equation (2.5), Equation (2.6) shows the well-known dependence for an ion of a given m/z (m and q are fixed) on the magnetic field strength. As an example, Hendrickson *et al.* obtained a resolving power about 150 000 for intact BSA (66 kDa) with detection period of 0.38 s with a 21 T FT-ICR MS. The resolution went up to more than 2 000 000 for a detection time of 12 s [79].

$$R = \frac{eB_0}{m/z \cdot \Delta w_c} \quad (2.6)$$

However, without getting too specific, Marshall *et al.* have shown that the resolving power of any peak is also depending on the time-domain ICR signal acquisition period T [80]. The relationship between resolution and the time-domain ICR signal acquisition period can be expressed as Equation (2.7):

$$R = \frac{k z B_0 T}{m} \quad (2.7)$$

where k is a peak width constant (e.g. $k = 1,274 \cdot 10^{-7}$ for the resolution at FWHM and at low pressure).

From Equation (2.7), it can be deduced that the highest time-domain ICR signal acquisition period, the highest resolution. That is all the paradox of coupling FT-ICR MS to separation methods. High resolution is time demanding but the continuous flow of the separation technique required speed analyses and therefore some compromises. There are only few articles involving FT-ICR MS hyphenation illustrating this dichotomy. The effect of transient duration on resolution is well known in direct infusion [10]. Nevertheless, Zacs *et al.* published an interesting paper in 2019 showing how transient length also influences many other key parameters in hyphenation. The study was carried out in GC-APCI-FT-ICR MS in their case [39], which is reputed for its high separation efficiency (narrow peak width of only a few seconds), and resolving power. Drawing an accurate chromatogram requires commonly at least around 10 to 15 points per peaks, below which the chromatographic peak often looks angular (rather than Gaussian shaped) and most of the time truncated at the top, reducing the chromatographic signal-to-noise ratio (S/N). The Figure 2.5 shows very well that such number of points per peak can only be achieved by drastically reducing the transient length (only from 256 thousand data points and lower), at the expense of the resolving power. In their study, the mass resolving power was only around 40 000 (FWHM) for 256 thousand data points at m/z 486 whereas it reached around 100 000 (FWHM) for 1 million data points. Still

on this particular subject, whatever the separation technique (GC, LC including UPLC, and CE), the number of points per peak is a key parameter to monitor very low concentrated compounds and when a study required automated MS/MS experiments to achieve structural elucidation, for instance in non-targeted “omics” approaches, particularly in case of partial coelution of analytes.

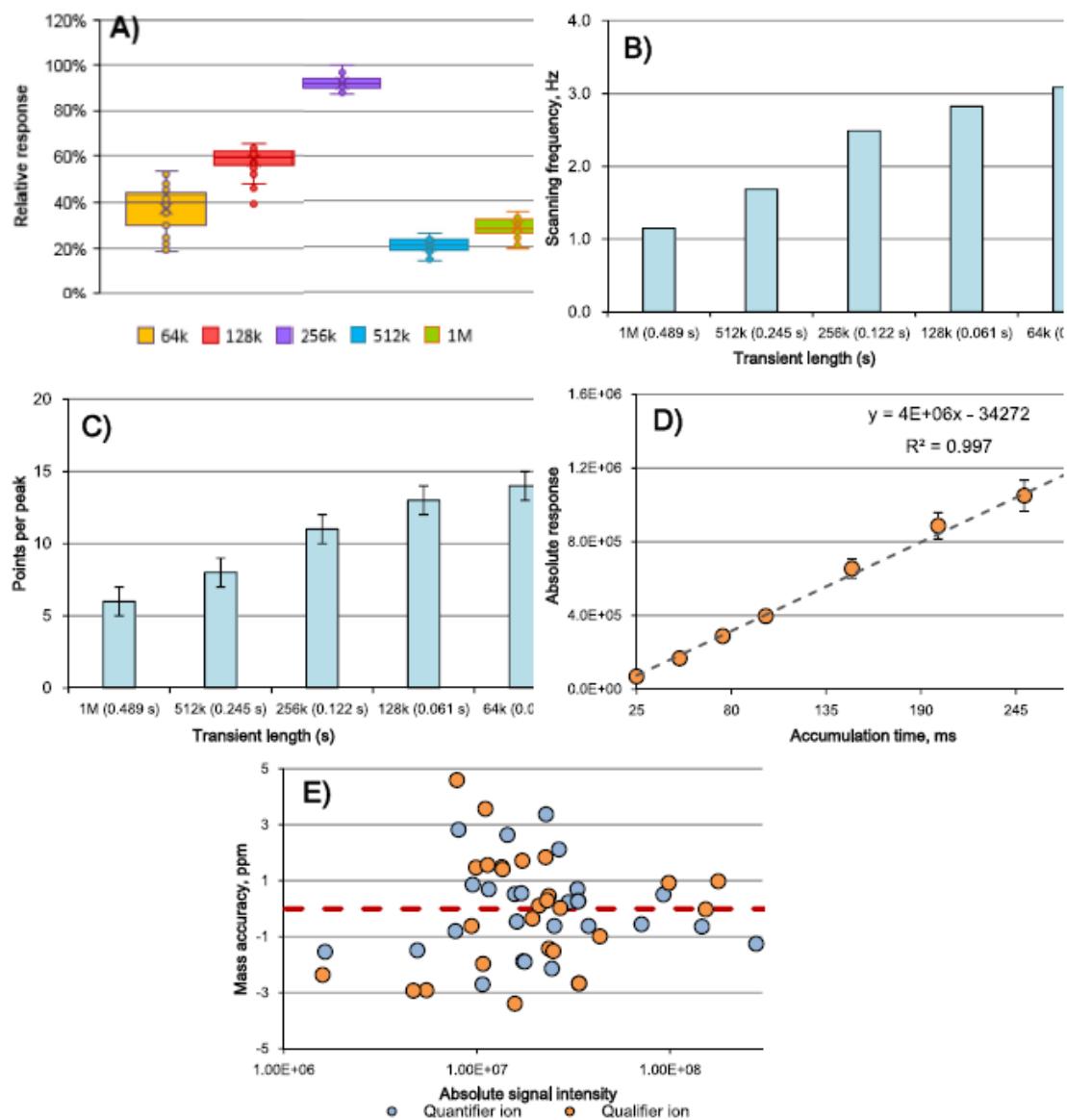


Figure 2.5: Design of results of the method optimization experiments from the analysis of PBDE-99, a brominated flame retardant, (50 pg on-column): (A) instrumental response of the system obtained by applying different transient lengths; (B) relationship between the scanning speed and transient length; (C) relationship between the number of data points per chromatographic peak and different transient lengths; (D) relationship between the observed instrumental response and accumulation time; (E) mass measurement accuracy for targeted analyte fragment ions observed with the optimized method. Extracted from [39].

Another issue is the detection sensitivity in qualitative and, above all, in quantitative studies. The signal intensity and more specifically the limit of detection or quantification, which are linked to the signal-to-noise ratio, are the key values. In FT-ICR MS, the signal-to-noise ratio is a function of the data acquisition time T and a collisional damping constant τ

(depending on the frequency of collision between the measured ions with other ions and gas molecules in the cell) as shown by Equation (2.8).

$$S/N \propto \frac{\tau}{\sqrt{T}} \cdot (1 - e^{-T/\tau}) \quad (2.8)$$

Whereas a high acquisition time T significantly improves the resolving power R , a too high value of T impairs the S/N of the mass spectra [81]. The sampling rate is also a reality in separation sciences [82]. With high sampling frequencies, chromatographic peaks are well defined but the background noise is increased. On the contrary, small sampling frequencies reduce the noise on the chromatographic baseline but when a too small number of points per peak is reached, the signal is truncated and the signal-to-noise ratio also decreases. Both for mass spectrometry and chromatographic reasons, the “optimal T value” must be considered as a compromise between the resolving power in one hand and, the signal-to-noise ratio and the number of points per peak in the other hand. Of course, in many commercial FT-ICR mass spectrometers, the signal itself can be increased using ion accumulation (e.g. in a hexapole ion trap as proposed in Figure 2.4) along their way towards the ICR cell as depicted in Fig. 5D. Nevertheless, the reader should be aware that ion accumulation takes time at the expense of the acquisition time T and thus at the expense of the resolving power as well as the number of points per chromatographic peak: once more the art of compromise.

2.1.4. Recent technical developments of FTICR-MS improving higher performance in hyphenation

Efficient separation methods do not allow low MS acquisition rates because, contrary to direct infusion, they provide small retention or migration time windows for the detection of analytes. With very few data points, chromatographic or electrophoretic peaks can be distorted, and their intensities can be biased. This can also lead to low concentrated compounds being not detected at all. The question is then: how to improve the resolving power without increasing acquisition time too much.

Obviously, during the implementation of the hyphenation with FT-ICR MS, it is critical to spend time on the optimization of all available parameters in order to make the best of both techniques. However, from the initial concept in 1974 [1] to nowadays, many technical developments in FT-ICR MS have been made, making the hyphenation with separation methods more friendly, even if it is still challenging, and reducing the loss of resolution.

All the historic instrumentation developments of FT-ICR MS were well described by Marshall *et al.* [83] and Wanczel *et al.* [84] in their respective reviews. In this review, the focus is rather put on the technical advances benefiting to the implementation of the hyphenation of FT-ICR MS with separation techniques in order to increase the MS resolving power, the mass measurement accuracy or/and acquisition rate.

From Equation (2.7), the increase of the magnetic B_0 field strength is one of the first ways to improve the resolving power R independently of the transient duration T . The implementation of superconductive magnets of higher strength is thus beneficial for the hyphenation of FT-ICR MS with a separation method. Indeed, higher magnetic field strength enables higher resolution and higher mass measurement accuracy, but also increases the acquisition speed, the dynamic range and the resistance to peak coalescence [85]. The first FT-ICR system was home built with a magnet of 0.32 T [1]. Nowadays, the highest magnetic field commercially available is 15 T, however custom-built FT-ICR MS with 21 T magnet exists. For instance, Smith *et al.* analyzed samples containing dissolved organic matter by 9.4 T FT-ICR MS and 21 T FT-ICR MS, and results showed that the 21 T superconductive magnet enabled to increase by 2.2 the resolution, by 2.6 the mass measurement accuracy and by 1.3 the number of assigned molecular formulae compared to the 9.4 T magnet [86]. To increase the acquisition rate and the mass measurement accuracy while keeping the same resolution, higher magnetic field strength has a significant advantage for the implementation of the hyphenation of FT-ICR MS with a separation method. Indeed, magnets of higher strength limit the sacrifice of resolution required for the hyphenation. The only drawback is the tremendous cost of such systems.

Another way to bypass the problem of short time-domains “T” requested by the hyphenation is the design of the ICR cell itself. Several ICR cells have been designed over the time and the last one, the dynamically harmonized cell, called the ParaCell, was first design in 2011 [19] and is now commercially available (Bruker Daltonics, Bremen, Germany). This ICR cell offers the best ion stability and coherence in ion cyclotron motion than any other cells by dynamically harmonizing the potential. This harmonization is performed by averaging the hyperbolic electrical field over the cyclotron motion. Unharmonized potential influences the ion cyclotron motion and causes phase shift. During the phase shift of ion cyclotron motion, MS peaks become wider and distorted, leading to lower resolution. As the design of this cell enables the dynamic harmonization of the potential, it enables higher resolving power [19]. An adaptation of the dynamically harmonized cell was proposed by Kostyukovich *et al.* in 2012. It is based on the addition of supplemental shaped electrodes adapting the electrical field (flat-bottom potential well) in order to compensate the inhomogeneity of the magnetic field inside the ICR cell. This adaptation enables reducing even more the phase shift of ion cyclotron motion [87].

Besides the usual detection with a single pair of plates, multiple detection is enabled by the modification of the ICR cell design. Quadrupole detection is based on an ICR cell using two pairs of opposite detection plates instead of a single pair. The set-up of the detection plates and the resulting image current for an ion cloud of both traditional (1ω) and 2ω ICR cells are presented in Figure 2.6. In a conventional ICR cell (1ω), the reduced cyclotron frequency is measured by the pair of detection plates, whereas the second harmonic of the reduced cyclotron frequency is measured by the two pairs of detection plates in quadrupole detection. This second harmonic is the double of the reduced cyclotron frequency and enables

to double either the acquisition rate while conserving the same resolution or the resolution while conserving the same acquisition rate [88–90]. The only commercially available instrument using this 2ω detection technology is the 7 T SolariX 2xR FT-ICR MS (Bruker Daltonics, Bremen, Germany). However, Valadóres Tose *et al.* used a 3ω detection (hexapolar detection) during LC-FT-ICR MS 14.5 T, resolution higher than 1 000 000 (FWHM) at m/z 800 and mass measurement accuracy lower than 0.5 ppm [91]. Shaw *et al.* also published a study using 4ω detection with a home-build analyzer cell and obtained a resolution of 150 000 at m/z 200 with a transient of 200 ms [92]. Therefore, using multiple detection is a strong asset for the hyphenation of FT-ICR MS with separation methods in order to enable high resolution and mass accuracy while using short acquisition times. Nevertheless, the rise of undesired harmonics of the reduced cyclotron frequencies is more likely with FT-ICR MS instruments using multiple detection [88]. Hence, to reduce the intensity of the undesired harmonic signals, the ICR cell must be frequently and finely shimmed and gated. It enables to control and reduce the magnetron motion and to recenter the ion cyclotron motion in the ICR cell. Moreover, with the relentless progress of signal treatments, m/z signals corresponding to harmonics are indicated on the mass spectra by software and these peaks can be automatically removed from the lists of the peak of interests. Finally, with 3ω or 4ω detection, loss of sensitivity can also occur [89].

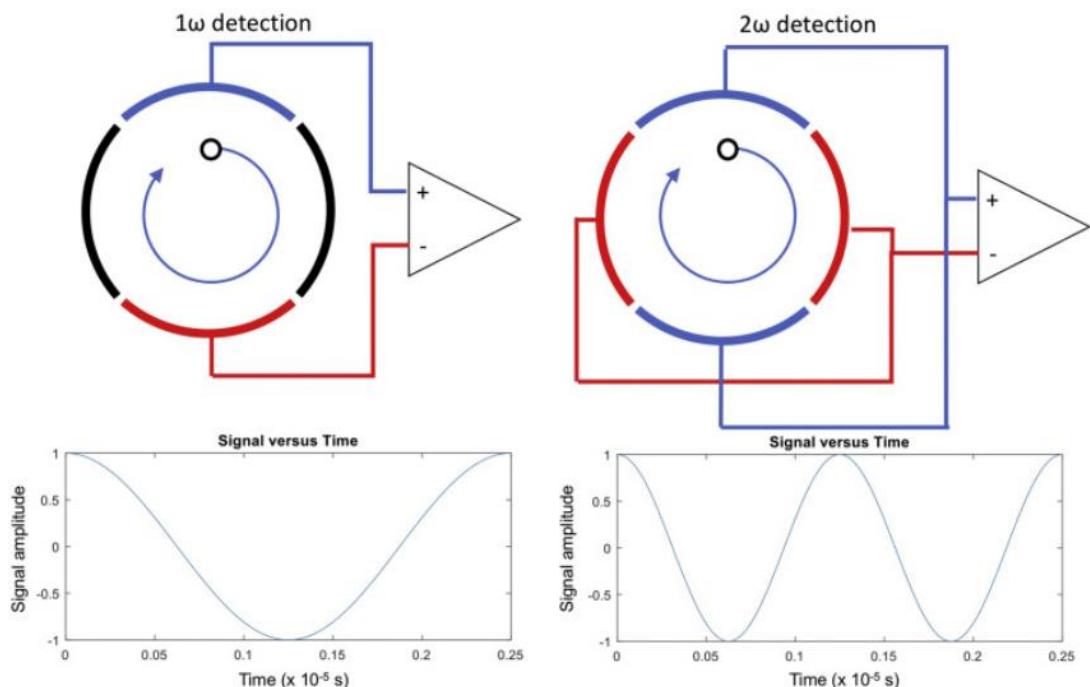


Figure 2.6: (Top) Detection set-up for normal detection (1ω) and for detection at twice the cyclotron frequency (2ω). (Bottom) The image current resulting from one revolution of the ion packet for both 1ω and 2ω detection schemes. Extracted from [89].

Advances may also result from the signal processing. Besides the classical data processing method called “magnitude mode”, another method called “absorption mode” is based on the complex impedance of the measured signal [31]. The real part of the signal is extracted and plotted to obtain frequency and mass spectra. This processing method enables

to increase the resolution by a factor up to 2, the signal-to-noise by a factor of 1.4. and the mass measurement accuracy by a factor of 1.3. However, the signal extracted in absorption mode needs to be phased with the signal of magnitude mode to prevent peak distortion. Therefore, phase correction of the absorption signal must be performed and can make the reprocessing more complex. It should be noted that the ParaCell is extremely compatible with absorption mode as the phasing of the ion cyclotron motion is more maintained than with other cell designs [90]. Therefore, the combination of the absorption mode and the quadrupole detection enables to increase by a factor of nearly four, either the acquisition rate while conserving the same resolution, or the resolution while conserving the same acquisition rate. [90] This can benefit the performances of the hyphenation of FT-ICR MS to a separation method by increasing the compatibility of both techniques.

Another strategic development is the possibility to accumulate ions during detection [93]. Indeed, as the acquisition time is the longest in FT-ICR MS, no ion accumulation during detection with a separation prior to MS can lead to a loss of information. This parameter is often called “LC capture” or “Accumulate During Detect” in the software. The user can choose how the accumulation time is distributed during the acquisition. However, this feature is more useful in MS mode only. If some MS/MS fragmentation is done after a full scan MS, the next full scan MS will not be able to accumulate ions during the MS/MS analysis.

2.1.5. The application field: the art of the possible in hyphenation

By way of introduction to this last section, it is interesting to keep in mind that in coupling, the separation methods and mass spectrometry are supporting each other to compensate their weaknesses and add their strengths. Thus, it is not always necessary to reach the ultimate resolving power that the FT-ICR MS handles and in the same way, a full chromatographic separation is not always required to explore in-depth the molecular content of complex samples. However, it is essential to ensure that the minimal MS resolution, sensitivity and separation are reached to achieve the objectives. Over the past years, the use of LC, GC or CE hyphenated to FT-ICR MS has been expanded in various application domains from proteomics, metabolomics and other “omics” subdivision (e.g. lipidomics) to petroleomics and environmental studies. The last part of this review is dedicated to applications showing the art of the possible in coupling FT-ICR MS with separation techniques.

2.1.6. Proteomics applications

Proteomics application are not only the first but also the most common applications of hyphenation with FT-ICR MS. This is likely due to the need of high-resolving power and mass measurement accuracy which, in the early days of MS instrumentation, could not be achieved otherwise than by a FT-ICR mass spectrometer. Because of their high polarity and molecular weight, peptides and proteins are commonly ionized by electrospray implying LC (Table 2.1) or CE-FTICR-MS (Table 2.2) analyses.

Proteomic analyses can be categorized according to chosen strategies. In the following paragraphs, some illustrations of the bottom-up, middle-up, intact, middle-down and top-down strategies will be presented.

2.1.6.1.1. Bottom-up analyses

The bottom-up strategy consists in digesting the protein into peptides, typically with trypsin, before analyzing them by MS and MS/MS fragmentation. In bottom-up analyses, the digestion step results in a complex mixture of peptides. Thus, the analytical challenge of the final peptide mixture is rapidly increasing with the number of proteins in the initial sample. Separation methods are especially useful to reduce the complexity of the mass spectra recorded along the analytical run.

The bottom-up strategy is, for instance, useful to identify the modifications present on a protein. Irungu *et al.* [94] characterized the 31 potential glycosylation sites of the HIV-1 envelope glycoprotein (*CON-S gp140ΔCFI*) by analyzing its tryptic digest by LC-FT-ICR MS/MS. In this application, the LC separation helped to reduce ion suppression effects due to non-glycosylated peptides, enabling the detection of various low abundance glycoforms. Two hundred fifteen glycans were detected and identified, corresponding to 18 of 31 potential sites. Combined with an offline LC/MALDI-TOF/TOF approach, a total of 341 glycans were identified. Other bottom-up analyses of glycopeptides hyphenating a separation method to FT-ICR MS have been carried out [95,96]. Another example is Boulanger *et al* [75] who characterized by CZE-FT-ICR MS the modifications on muropeptides from muramidase digestion of *Bacillus licheniformis* cell wall. The high resolution (50 000 at *m/z* 400) and mass measurement accuracy (0.01 - 2.26 ppm) provided by the FT-ICR MS enabled confident identification of 49 peptidoglycans separated in 25 electrophoretic peaks. Several modifications such as deacetylations and particularly isomer muropeptides differing by their amidation sites were identified after their separation by CZE. FT-ICR MS/MS (CID) and FT-ICR MS³ (ECD or ETD) enabled the final discrimination of these isomers.

The bottom-up strategy also enables the proteome identification of a complex mixture such as body fluids [34,97,98] or cell lysate [43,44,99–104]. For example, to overcome the impact of environmental conditions or the development state of the organism on its proteome, multiple analyses can be performed to obtain comprehensive results. This requires high throughput methods. In 2001, Shen *et al.* [44] developed a new strategy for high throughput bottom-up analysis based on capillary LC-FT-ICR MS in which two LC columns were used in parallel (one column was regenerating while the other one was employed for the analysis). This new strategy was tested on a trypsic digest of *Saccharomyces cerevisiae* cell lysate. The additional dimension provided by LC increased the dynamic range and simplified the mass spectra. Around 110 000 peptides were identified in this cell lysate (Figure 2.7).

On their own, Calrsohn *et al.* [52] performed a nanoLC-FT-ICR MS/MS profiling of the outer membrane proteome of several strains of *Helicobacter pylori* to find potential markers

or target proteins for vaccine development. The different strains of *H. pylori* were collected from asymptomatic patients and others suffering from duodenal ulcers or gastric cancer. A total of 60 membrane proteins, including the low-abundant ones, were identified by database searching, MS and MS/MS matching. One of these proteins was found as a potential marker for all strains. Furthermore, the strains causing duodenal ulcers seem to be characterized by the presence of one specific protein and the absence of another one.

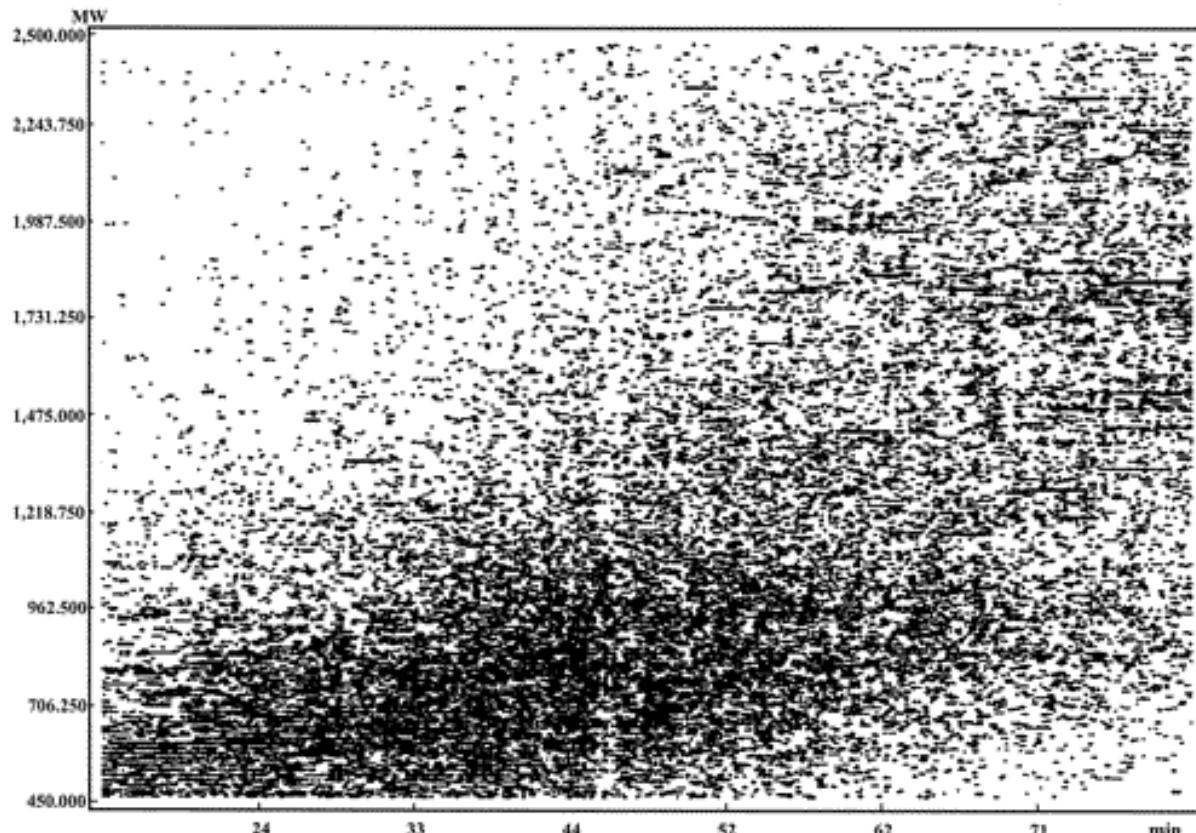


Figure 2.7: Capillary LC-FTICR 2-D display for a portion of the analysis of a yeast global soluble protein tryptic digest that contained a total of ~110 000 detected components. Figure and caption reprinted with permission from [44]. Copyright 2001 American Chemical Society.

2.1.6.1.2. Middle-up analyses

Middle-up analyses are based on the partial digestion of proteins into large subunits followed by MS analysis.

To the best of our knowledge, the only use of separation methods hyphenated to FT-ICR MS for middle-up analyses is the structural characterization of bispecific antibodies subunits. In 2020, Gstöttner *et al.* [59] used CE-FT-ICR MS to study the heterogeneity of subunits. All 25 kDa subunits (Fd_1 , Fd_2 , Lc_1 , Lc_2 , $Fc/2_1$ and $Fc/2_2$) were partially separated by CE, detected, and identified with great mass measurement accuracy (2 - 13 ppm) and high resolution (145 000 at m/z 1100). Several modifications were identified such as pyroglutamate formation on Lc and Fd subunits, or K-clips and glycosylations on both $Fc/2$ subunits.

2.1.6.1.3. Intact protein analyses

The intact protein strategy consists of analyzing by MS the whole protein without any digestion step and without MS/MS fragmentation. This kind of approach is usually investigating the various proteoforms. For instance, in 2009, Thakur *et al.* [48] analyzed the various glycoforms of the α subunit of intact recombinant human chorionic gonadotropin from murine cell line by CE-LTQ-FT-ICR MS. Initially the analysis was performed by CE-LTQ-MS without the use of FT-ICR to refine the m/z range for CE-FT-ICR MS analysis, which enabled a faster acquisition speed and a better sensitivity. FT-ICR enabled to reach the high resolution of 55 000 at m/z 1800. More than 60 glycoforms have been detected and identified. The relative abundance of the variants was achieved thanks to LC separation. The authors reported that the 20 most abundant glycoforms represented 90% of the whole protein content. Another example of protein heterogeneity studies is Lippold *et al.* [105] who hyphenated the Fc γ RIIIa affinity chromatography to FT-ICR MS to study the influence of glycosylation heterogeneity on the Fc γ RIIIa affinity of a therapeutic monoclonal antibody (mAb). The differences in Fc γ RIIIa affinity of the mAb according to the glycoform is displayed in Figure 2.8.

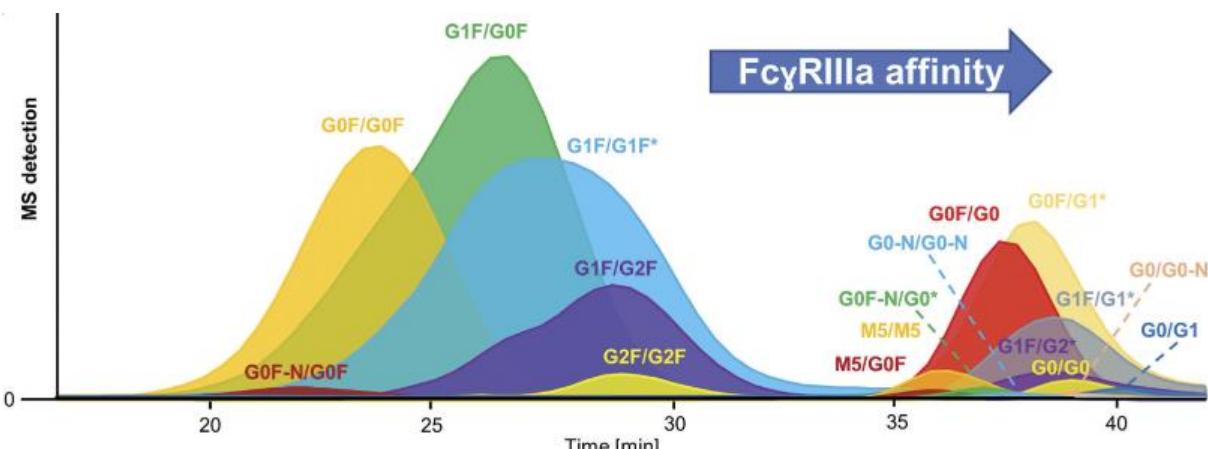


Figure 2.8: Fc γ RIIIa affinity extracted ion chromatograms of therapeutic mAb detected glycoforms. * indicates that there are multiple possibilities. The most probable glycoform is annotated. Extracted and adapted from [105].

Glycosylations are not the only modifications leading to various proteoforms. There are also smaller modifications such as methylations or oxidations. A good illustration is the work of Lee *et al.* [106] who characterized by LC-FT-ICR MS the proteins from the large ribosomal subunit of yeast. Due to the high mass measurement accuracy and resolution of FT-ICR, as well as the LC separation of isoforms with high structural similarity, the identification of 42 proteins and 58 isoforms, as well as their various modifications (mostly methylations, phosphorylations, N-terminal modifications), were enabled. Once more, high sensitivity and broad dynamic range were achieved thank to the hyphenation.

The analyses of intact proteins are also carried out to identify the protein content in complex mixtures such as cell lysates [41,69,71,107–110]. Jensen *et al.* [41] hyphenated a capillary isoelectric focusing separation (CIEF) to FT-ICR MS to study complex protein mixtures. Their method was based on an isotope depletion strategy and to demonstrate the usefulness

of their method, the response to a cadmium (Cd^{2+}) stress of the proteomes from *E. coli* and *D. radiodurans* cell lysates were evaluated. Stressed and unstressed cell lysates were mixed in equal proportions and the relative abundances of proteins were measured after 0 min, 45 min and 150 min of exposure to Cd^{2+} . From the stressed/unstressed ratio, it was possible to deduce which proteins are induced or suppressed by the Cd^{2+} stress. As shown by Marshall *et al.* in 1997, the isotope depletion strategy associated with FT-ICR MS detection can improve the mass measurement accuracy [111]. CIEF prior to FT-ICR MS enabled relative quantification and extended dynamic range. CIEF-FT-ICR MS has also been used by Hofstadler *et al.* [69,110] to analyze whole single cells (human erythrocytes) without preliminary lysis. The whole cells were directly lysed in the CE capillary due to an osmotic shock. By injecting several cells, partial CIEF separation of proteins enabled the detection of low abundance proteins like carbonic anhydrase (approx. 7 amol/cell), compared to hemoglobin which is the most abundant in the sample (approx. 450 amol/cell) [69]. By injecting one single cell, both α and β chains of hemoglobin could be detected [110].

The intact protein analysis strategy is also implemented to study protein complexes. In 2000, Martinovic *et al.* [72,112] analyzed non-covalent protein complexes by CIEF-FT-ICR MS. The coupling of FT-ICR MS with a separation method in native condition is a very useful tool to characterize non-covalent complexes. In native condition, CIEF enabled the separation of the different non-covalent complexes according to their isoelectric point. Depending on the composition of the sheath liquid at the CE-MS interface, it was possible to preserve the intact complexes or to dissociate the complex into subunits allowing to detect and identify non-covalent protein complexes and their subunits.

2.1.6.1.4. Middle-down analyses

Middle-down analyses are based on the partial digestion of proteins into large subunits followed by MS analysis and MS/MS fragmentation.

To our knowledge, the only use of separation methods hyphenated to FT-ICR MS for middle-down analyses is a structural characterization of monoclonal antibodies (mAbs) subunits. In 2019, Jin *et al.* [113] characterized the 25 kDa subunits (Fc/2, Fd and Lc) of an IgG1 mAb by LC-FT-ICR MS/MS. The middle-down analyses (Figure 2.9) highlighted and enabled the identification of various glycoforms and other micro-variants (oxidation, loss of water, C-terminal glycine clipping and proline amidation). Additional top-down analysis with direct infusion and offline middle-down analysis were carried out to increase bond cleavage and confidence in identification. He *et al.* [114] also used LC-FT-ICR MS for middle-down analysis of 25k Da subunits of adalimumab spiked into human serum.

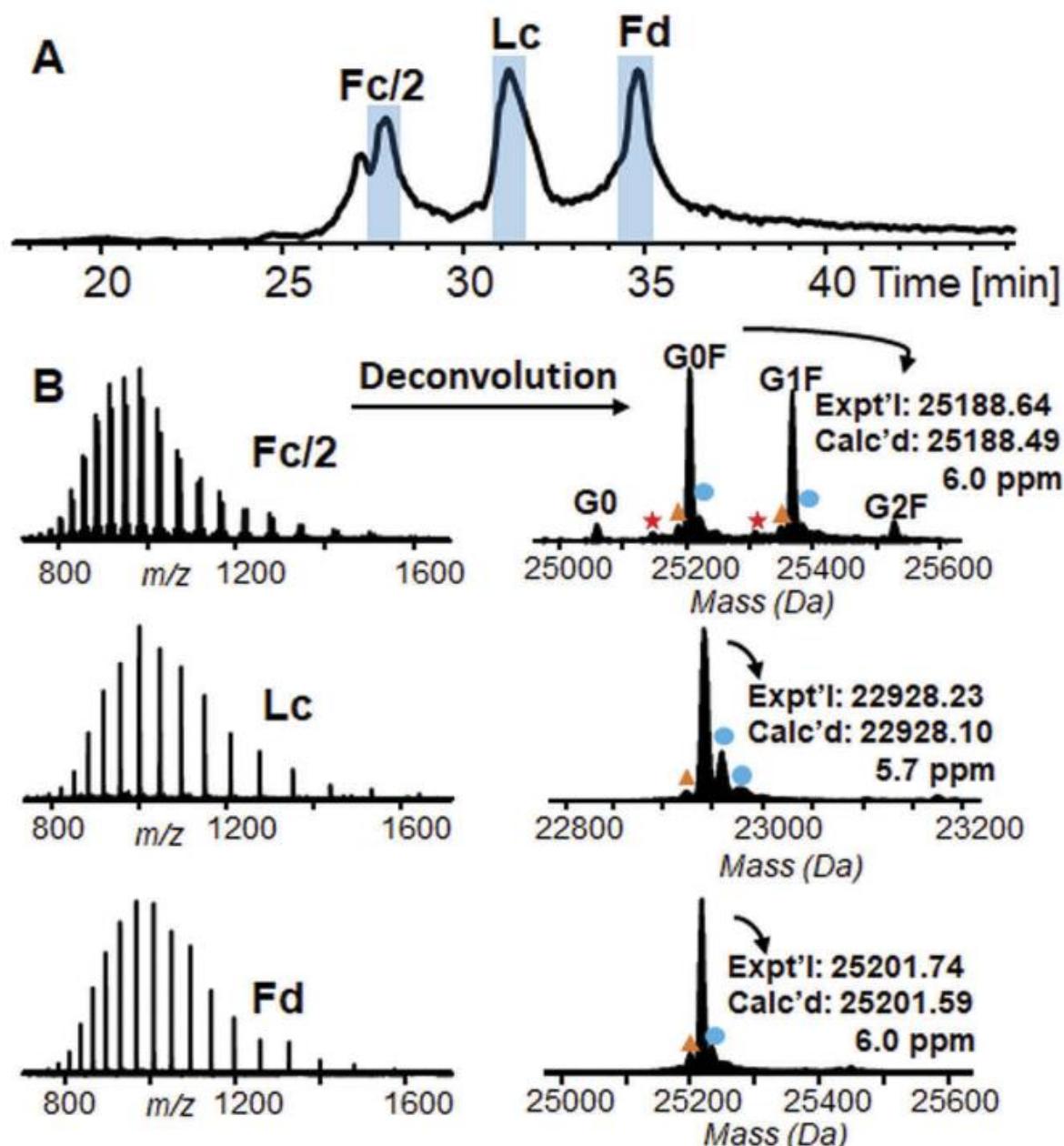


Figure 2.9: Chromatogram and mass spectra of proteolytic subunits from IdeS digestion. (A) TIC chromatogram of three proteolytic subunits, Fc/2, Lc and Fd, from online LC/MS analysis. (B) Left, experimental mass spectra of Fc/2, Lc, and Fd. Right, deconvoluted mass spectra of Fc/2, Lc, and Fd. Asteroid, -58 Da from C-terminal glycine clipping and proline amidation; triangle, -18 Da from neutral loss of water; dot, $+16\text{ Da}$ from oxidation. Expt'l, experimental monoisotopic mass based on data from MS experiments. Calc'd, calculated monoisotopic mass based on amino acid sequences. Figure and captions extracted from [113].

2.1.6.1.5. Top-down analyses

The top-down strategy is based on the MS/MS fragmentation of intact protein enabling the structural characterization of protein proteoforms. This analytical strategy is harder to implement than previous ones due to the high sample complexity and the size of intact proteins.

The work of Anderson *et al.* [115] on a human colorectal cancer cell lysate is a perfect illustration of the relevance of LC-FT-ICR MS/MS for the top-down characterization of proteoforms. Before the online LC-FT-ICR MS analysis, an electrophoretic gel of the lysate was run to simplify the original protein mixture into 8 fractions according to their molecular weight. Collision-induced dissociation (CID) and electron-transfer dissociation (ETD) fragmentation modes were both used for LC-FT-ICR MS/MS top-down analyses. With a total of 40 LC-FT-ICR MS/MS analyses, the authors detected and identified in the lysate 684 unique proteins, including proteins larger than 30 kDa, and 3238 proteoforms, with a false discovery rate of 1%. For their part, in 2012, Brinkmalm *et al.* [116] hyphenated the FT-ICR MS with nanoLC to highlight potential biomarkers of Alzheimer's disease in human cerebrospinal fluids. The abnormal degradation of amyloid β and amyloid precursor protein in cerebrospinal fluid could be characteristic of Alzheimer's disease. Top-down analyses were carried out with CID, ECD and IRMPD fragmentation and the ultra-high resolution provided by FT-ICR enabled isotopic resolution of $> 3\,000$ Da peptides, the extra dimension provided by LC enabling their semi-quantification.

Top-down analyses can also be used to get information about 3D structures of proteins. Indeed, in 2018, Chen *et al.* [117] hyphenated hydrophobic interaction chromatography (HIC) with FT-ICR MS/MS using ECD broadband fragmentation to analyze a mixture of two deglycosylated mAbs. The conditions for the separation of the two studied mAbs were adapted to be compatible with mass spectrometry while maintaining their native conditions. ECD fragmentation preserved disulfide bonds and noncovalent interactions limiting the fragmentation, which in turn could provide information about the localization of binding sites of complementary-determining regions (CDR). When comparing ECD results with crystallographic 3D structure, hypotheses have been put forward concerning the role of N- and C-termini of light and heavy chains in interaction interface.

Other top-down analyses using a separation method hyphenated to FT-ICR MS have been carried out on cell lysates [42,118–120] and human ribosome [121]. Additional information on those is available in Table 2.1 for LC and Table 2.2 for CE.

2.1.6.2. Metabolomics and other related omics disciplines

Metabolomics are the study of all metabolites in various samples derived from living organisms, from plants to animals, from bacteria to fungi, etc. Two strategies are usually implemented in this kind of study: the untargeted approach, which generally explores the organism metabolome as comprehensively as possible, and the targeted approach, which focus on already known specific metabolites. Both approaches aim at understanding metabolic pathways on the basis of the collected information (molecular formulae, structure).

Untargeted and targeted applications for metabolomics and other related omics disciplines carried out using hyphenation of a separation method to FT-ICR MS are gathered

Table 2.1 for LC and Table 2.2 for CE. Some of them have been taken as illustrations in the following paragraphs.

2.1.6.2.1. Untargeted metabolites profiling

Compared to the targeted approach, untargeted analyses highlight hundreds of known but, more interestingly, of unknown compounds as well. Chromatography is especially useful to reduce the complexity of each mass spectrum recorded along the analytical run. The data interpretation remains a huge work and it often requires specific identification procedures and the assistance of dedicated software.

In 2008, Iijima *et al.* [36] developed an annotation procedure for untargeted metabolites profiling using LC-FT-ICR MS in *Solanum lycopersicum* cv. Micro-Tom. Tomato was used as a model but the procedure can be used in any other untargeted metabolites analysis. The annotation of each metabolite included mass spectra, predicted molecular formulae and structures, and each metabolite received a confidence grade (A, B or C) based on the eventual ambiguity of its identification. This procedure enabled the annotation of 869 metabolites including 496 new ones after comparison with public databases. The authors also performed the semi-quantification of flavonoids and glycoalkaloids after having compared their relative signal intensities in several samples.

Another example is relative to Giavalisco *et al.* [50] who developed a new strategy for untargeted qualitative and semi-quantitative analysis of metabolites in *Arabidopsis thaliana* leaves using LC-FT-ICR MS. Their approach is based on the $^{12}\text{C}/^{13}\text{C}$ ratio to determine the number of carbon atoms for each component. Knowing the number of carbon atoms in an elemental composition reduced the number of potential molecular formulae to a unique one or at least a few ones. Thanks to the LC separation and the $^{12}\text{C}/^{13}\text{C}$ sample labeling, reliable relative quantification was demonstrated for 20 compounds. Moreover, the LC separation of isomers allowed discriminating biological from non-biological signal, and increased the number of annotated non-redundant elemental compositions by two compared to FT-ICR MS direct infusion (2304 vs 1024). As shown in Figure 2.10, MS spectra for ^{12}C and ^{13}C samples were compared at similar retention time, and the number of carbons for each compound was deduced from the difference in mass-to-charge ratio between both spectra (Figure 2.10a). If there is no difference between both spectra (Figure 2.10b), the signal is determined as non-biological.

In 2013, Nakabayashi *et al.* [122] used a similar approach to study the sulfur-containing metabolites in onion bulb (*Allium cepa*). The chemical assignment of these metabolites was performed based on a ^{34}S labeling. The high mass measurement accuracy ($> 1 \text{ ppm}$) and the ultra-high resolution ($> 250\,000 \text{ FWHM}$) enabled the distinction of S-monoisotopic ions from other species. Based on the natural abundance of ^{32}S and ^{34}S and their mass difference, the elemental compositions of various S-monoisotopic ions were assigned. The use of the

hyphenation with LC enabled the assignment of 67 S-containing monoisotopic ions among 4693 chromatographic compounds.

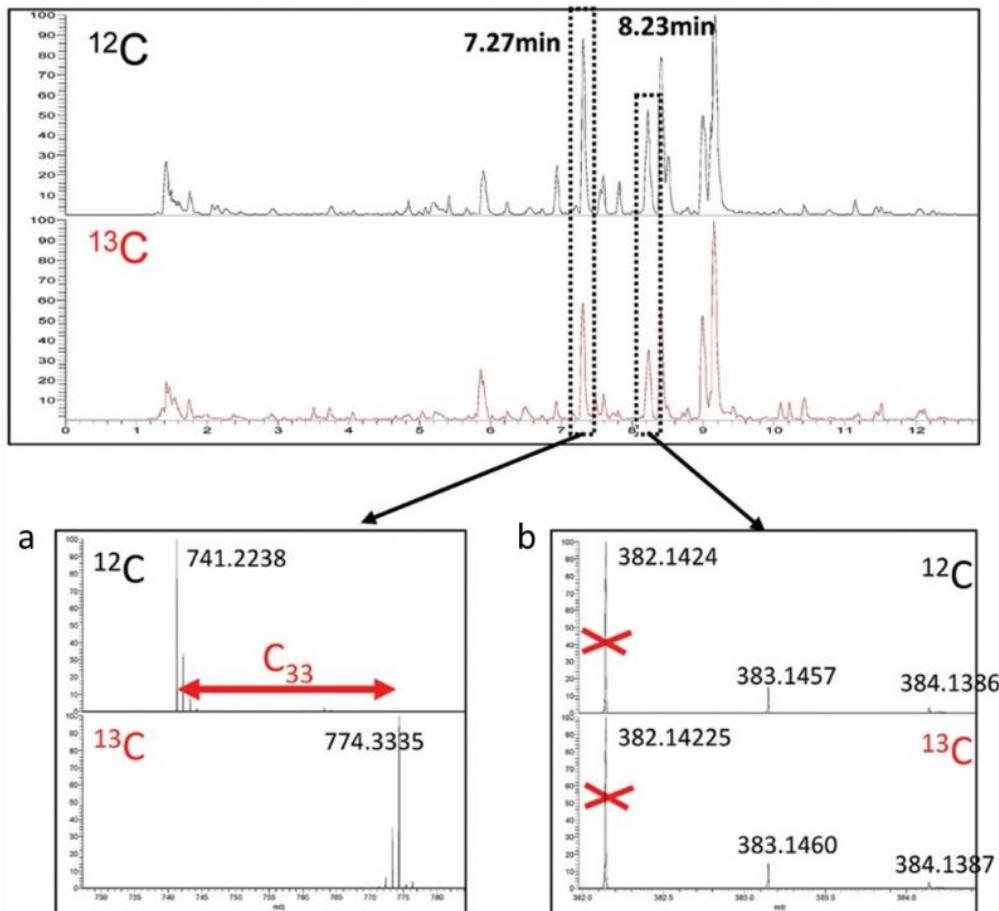


Figure 2.10: Positive ion mode spectra of *A. thaliana* leaf metabolites. The upper chromatogram represents the total ion chromatogram (TIC) of the ^{12}C (black) and ^{13}C (red) sample. The lower left spectra show the ^{12}C and ^{13}C mass spectra at the apex of the chromatographic peak at 7.27 min, while the right-hand spectrum shows the ^{12}C and ^{13}C mass spectra of the peak at 8.31 min. The presence/absence of a mass shift between the displayed masses in the upper (^{12}C) and lower (^{13}C) row of each mass spectrum indicates the biological (a) / non-biological (b) origin of the compound, and in case a mass shift is detectable, the number of carbon atoms of the measured compound. Figure reprinted with permission from [50]. Copyright 2009 American Chemical Society.

If the high mass measurement accuracy and the ultra-high resolution of FT-ICR MS are crucial to assign raw molecular formulae, the chromatographic separation is essential to highlight the presence of isomers.

For instance, Kong *et al.* [37] investigated by LC-FT-ICR MS the degradation and bioconversion of ginsenosides from red ginseng extract in *in vitro* gastrointestinal tract in order to have a better understanding of its pharmacological activity. Based on the mass measurement accuracy and MS/MS data from FT-ICR MS as well as their retention time in LC, 42 compounds were tentatively identified. The additional LC dimension was primordial to discriminate isomers with similar fragmentation pattern. All the identified compounds were also relatively quantified over time using some ginsenoside standards. Ren *et al.* [123]

investigated the metabolism of flavonoids from *Ziziphi spinosae* semen extracts in urine and feces samples after oral administration to rats and dogs by using for the first time LC-FT-ICR MS. *Ziziphi spinosae* semen is used in traditional Chinese medicine to treat insomnia and anxiety. The use of FT-ICR MS enabled the prediction of elemental composition and structures thanks to both the mass measurement accuracy and the multistage fragmentation. A total of 15 compounds were detected and 9 were identified, including some never reported before in *Ziziphi spinosae* semen extracts. Among them, several isomers have been discriminated with the LC separation prior to MS detection and flavonoids metabolomic pathways could have been suggested.

Guan *et al.* [124] profiled with UPLC-FT-ICR MS *in vivo* metabolites of 2-(2-hydroxypropanamido)benzoic acid, a drug isolated from marine fungus *Penicillium chrysogenum*, in rat plasma, urine, feces and bile samples after oral administration. The additional dimension brought by UPLC to MS and MS/MS detection was once more essential to discriminate isomers with close fragmentation pattern. A total of 13 metabolites were detected and tentatively identified thanks to the high level of confidence brought by FT-ICR MS high mass measurement accuracy, contributing to a better understanding of its safety and mechanism of action.

Metabolomic studies were performed on living organism samples to investigate the composition of their metabolome to improve the knowledge or to deduce metabolic pathways.

For example, Sun *et al.* [38] studied the *in vitro* metabolism of paeoniflorin in human intestinal microflora by using LC-FT-ICR MS. *In vitro* model was general anaerobic media broth mixed with fresh human feces. Thanks to the accurate mass measurements and retention times obtained in LC-FT-ICR MS 31 metabolites of paeoniflorin were highlighted and completed with the fragmentation from LC-QTOF-MS/MS experiments, 16 metabolites were identified. Thirteen metabolites were discovered for the first time in human intestinal microflora *in vitro* study extending the known metabolic reactions in the paeoniflorin pathway.

Pollier *et al.* [125] hyphenated LC to FT-ICR MS to profile triterpene saponins in *Medicago truncatula* hairy roots. Hairy roots which are the product of an infection of *M. truncatula* make their cultivation easier and thus could be used as an alternative to phytochemicals to prevent the destruction of natural roots. In order to reduce the number of potential molecular formulae for each detected triterpene saponins, the authors compared the elemental compositions with mass errors lower than 5 ppm to the different isotopic pattern present in the FTICR mass spectra. The combination of accurate mass measurement and investigation of the isotopic pattern improved the confidence in elemental composition annotation. A total of 79 compounds in *M. truncatula* hairy roots were annotated, including 61 ones not previously detected in this plant.

The list of the untargeted metabolomic studies carried out in LC-FT-ICR MS detailed here is obviously not exhaustive, but the readers will find some complements in Table 2.1 on what has been done up to now in this field. This complements includes other close related “omics” such as foodomics [47,53,126,127] and lipidomics [29,74].

2.1.6.2.2. Targeted metabolites analysis

The targeted analysis focuses only on known compounds. Therefore, the data interpretation and the quantification are most of the time easier. Usually, targeted analyses are performed after untargeted analyses to have a better sight on the content of the sample in order to adapt the quantification methodology. All targeted applications of the hyphenation of a separation method to FT-ICR MS are recorded in the Table 2.1 (for LC-FTICR-MS) and Table 2.2 (for CE-FTICR-MS). Two of them have been chosen to illustrate this paragraph.

In 2015, Nakabayashi *et al.* [128] developed a method to boost the sensitivity of the products ion of monoterpene indole alkaloids during LC-FT-ICR MS/MS analysis. The study focuses on the biosynthetic pathways of monoterpene indole alkaloids in *Catharanthus roseus* and *Uncaria rhynchophylla*, since those metabolites are known for their pharmaceutical activity. To boost the sensitivity of the product ions, MS/MS was performed with an increase of the product ion accumulation by 20 only for the MS/MS compared to the ion accumulation for the MS analysis. The gain of sensitivity with the MS/MS boost is illustrated in Figure 2.11. This MS/MS boost strategy was then applied to targeted analyses of *Catharanthus roseus* and *Uncaria rhynchophylla*. It enabled the identification of major and minor monoterpene indole alkaloids. Moreover, the use of the LC prior to the FT-ICR MS analysis simplified the mass spectrum interpretation by distributing the total ion population, including isomers, over the analysis time and enabled the fragmentation of fewer compounds at the same time.

The second chosen example was the only one involving the hyphenation of FT-ICR MS with capillary electrophoresis in the area of metabolomic analysis. In 2008, the targeted analysis of cationic metabolites from *Desulfovibrio vulgaris* Hildenborough lysate was performed by Baidoo *et al.* [58]. *D. vulgaris* is a Gram-negative anaerobic bacterium, easily cultivable and able to oxide soluble heavy metals into insoluble forms. Therefore, its metabolome study is of great interest for preventing soil and water contamination. The authors developed an original method using pH-mediated stacking and transient isotachophoresis (PMS-tITP) to improve the detection of the targeted metabolites by injecting higher sample volume than in classical CE. Thanks to the separation and the reduction of ion suppression, the absolute quantification of 28 metabolites, including leucine and isoleucine which are structural isomers, was achieved with only 124 nL injected. Despite the little sacrifice in resolution and mass measurement accuracy, the hyphenation of such system was therefore crucial compared to FT-ICR MS direct infusion.

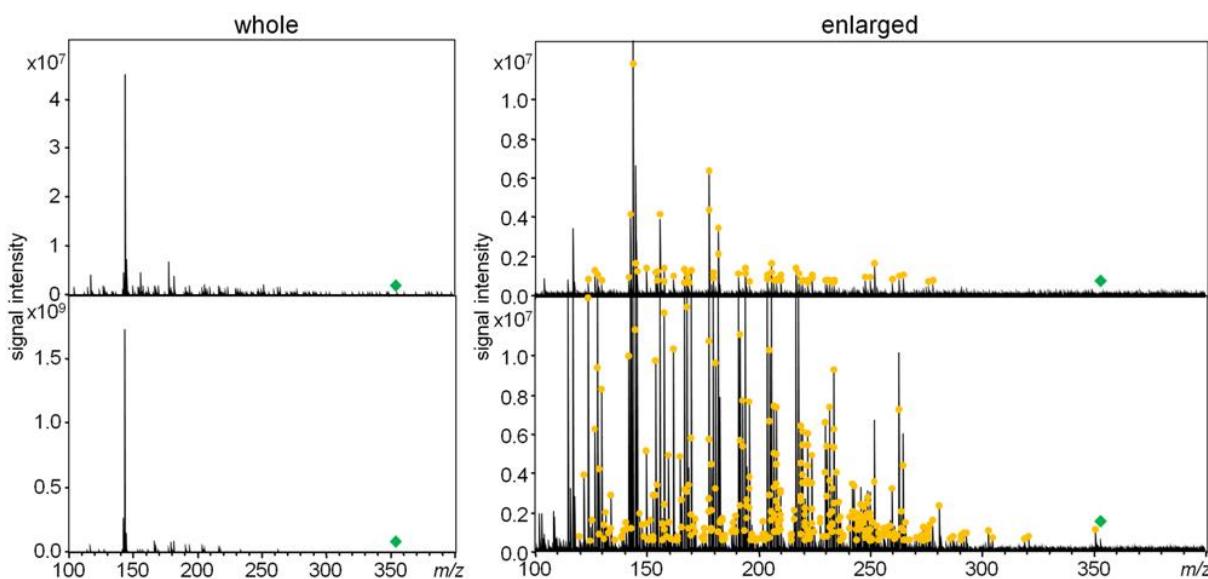


Figure 2.11: Comparison of productions of ajmalicine in the MS/MS analysis (upper) and the MS/MS boost analysis (lower). The yellow circles indicate chemically assigned monoisotopic ions. The green diamond indicates the precursor ion. Figure reprinted with permission from [128]. Copyright 2009 American Chemical Society.

Many other studies were focused on the targeted analysis of biological samples from various matrices. As evoked, the class of metabolites involved in these studies as well as the type of matrix can be found in Table 2.1 and Table 2.2.

2.1.6.3. Environmental applications

Environmental analyses cover a wide range of environments and purposes. This includes the analysis of aqueous environments, soils, fauna and flora, but also different atmospheric environments. The aims of those analyses can be various, from a comprehensive study of compounds naturally present in the environment to the monitoring of an environmental pollution. The chemical and physical properties of the analyzed molecules in environmental application are wide as well. Most of the time these analyses concern volatile organic compounds (VOC) making GC the most useful separation method hyphenated to FT-ICR MS. However, LC is also widely used, particularly when the molecule polarity is increasing.

Environmental applications using separation method coupled to FT-ICR MS are summarized in Table 2.3 for GC, and Table 2.1 for LC. Some of them have been detailed in the following paragraph.

2.1.6.3.1. Food contaminants analyses

The analyses of food contaminants are a good link between the foodomic evoked in the previous paragraph and the environmental analyses as most of food pollutants may come from anthropogenic sources. The hyphenation of GC and APCI-FT-ICR MS was used by Zacs *et al.* [39] to assess food contamination by halogenated flame retardants. The method was developed on various food samples (meat, whole milk, eggs and cereals) and then applied on cod liver, fish liver oil and soybean meal. The study was focused on halogenated flame

retardants accumulés dans les chaînes alimentaires représentant des risques cumulatifs et de santé. La haute précision de mesure de la masse et la haute résolution ultra-haute de la FTICR ont permis de résoudre l'overlapping des isotopologues contenant ^{13}C , ^2H , ^{18}O , ^{15}N , $^{79}\text{Br}/^{81}\text{Br}$ et $^{35}\text{Cl}/^{37}\text{Cl}$. L'interprétation de l'ensemble isotopique a réduit le nombre d'éléments potentiels pour la composition élémentaire des composés inconnus. La connaissance du nombre de bromine et de chlore dans chaque composé assigné a aidé à discriminer les retardants de flamme brominés des retardants de flamme chlorinés et a assuré une classification basée sur le nombre d'atomes de Br et Cl. La dimension GC non seulement a mis en évidence les isomères pendant les analyses mais a également facilité l'interprétation des données comparée à la FT-ICR MS direct infusion. En effet, dans cette étude, le spectre de masse de chaque pic GC contenait environ 5 000 m/z values, ce qui permettait d'imager à quel point il aurait été un cauchemar pour l'interprétation des données sans la dimension supplémentaire GC. Finalement, la quantification de 18 retardants de flamme brominés et chlorinés a été effectuée.

2.1.6.3.2. Biosphère environmental studies

Évoquer les problèmes environnementaux est le plus souvent lié aux polluants et à l'évaluation de la qualité de la biosphère. Les milieux aquatiques sont très souvent sujets à la pollution, et la plupart du temps la LC-FT-ICR MS est employée en raison de la forte polarité des molécules solubles dans l'eau. La caractérisation de la matière organique dissoute dans l'eau peut être utilisée pour surveiller la dégradation naturelle des substances organiques et la pollution organique dans les milieux aquatiques. Par exemple, Qi *et al.* [49] ont développé une méthode LC-FT-ICR MS pour caractériser la matière organique dissoute dans les échantillons d'eau de surface du fleuve Yongding (Chine) et de l'eau de pluie de Shanghai. Comparé à l'infusion directe, la LC a aidé à réduire le phénomène d'atténuation ionique, à discriminer les composés d'intérêt de la matrice et à limiter les artefacts causés par la formation d'adduits métalliques. Quatre classes moléculaires de composés ont été séparées : les sels métalliques, les molécules alicycliques riches en carboxyles, les organosulfates et les lignines-like. L'utilisation de la LC était essentielle pour la caractérisation prédictive de la matière organique dissoute en raison de leur grande complexité et de leur grande variété de structures chimiques. La LC a également permis de détecter de nombreux composés organiques dissous qui étaient difficilement détectables par la FT-ICR MS directe. La semi-quantification des composés de la classe O_x a été effectuée et les diagrammes de van Krevelen avec les rapports d'abondance relatives ont été utilisés pour comparer la composition de la matière organique dissoute dans les différentes échantillonées d'eau. Sandron *et al.* [54] ont également utilisé la hybride LC-FT-ICR MS pour caractériser la matière organique dissoute dans l'eau de mer et l'eau douce du fleuve Suwannee, mais ils ont ajouté une étape de fractionnement en ligne. La combinaison de la fractionnement en ligne et de la LC a permis une séparation remarquable des espèces isomères. En 2019, Kim *et al.* [129] ont utilisé une méthode LC-FT-ICR MS équipée d'une détection quadrupolaire (autrement dit, une détection 2ω) pour caractériser la matière organique dissoute dans les cours d'eau. La détection quadrupolaire a permis d'augmenter le nombre de composés détectés par 1,4 fois par rapport à la détection dipolaire (détection 1ω) et par 2 fois par rapport à l'infusion directe, puisque le temps d'acquisition nécessaire pour obtenir la même résolution que la détection dipolaire a été divisé par deux et la complexité du spectre de masse a été réduite par la séparation LC.

water can also be contaminated with polar molecules when extracting bitumen from oil sands. Among these contaminations, naphthenic acids are known for their toxicity. In this case, GC is appropriate. In 2014, Ortiz *et al.* [130] used the GC-FT-ICR MS hyphenation to characterize naphthenic acids from oil sand processed water and groundwater from the Athabasca river region. Between 1 300 and 2 200 compounds, including many isomers, were detected in each sample and assigned thanks to the high mass measurement accuracy of the FT-ICR MS and to the Kendrick mass defect. Those compounds are distributed into 11 classes based on their heteroatom content. Thanks to GC, isomers were separated within a same class of compounds (Figure 2.12). The high resolution and mass measurement accuracy of the FT-ICR MS were essential to characterize naphthenic acids due to their very close structures.

Soil and sediments are other environments potentially exposed to pollutants. In 2010, Bataineh *et al.* [67] developed a LC-APCI-FT-ICR MS/MS method to characterize polar mutagenic compounds in aquatic sediment. The method was developed on 55 compounds representing all classes of polycyclic aromatic compounds basically found in sediment samples. Their tentative identification was based on accurate mass measurement, MS/MS fragmentation but also on LC retention behavior which, in addition to reveal the presence of isomers, reduced the number of candidates. Soils and sediments are witnesses of contamination over time. The different soil layers may contain different contaminations caused along the history of the site. In 2019, Thomas *et al.* [131] used several methods, including direct infusion on a 12 T FT-ICR MS and a 7 T FT-ICR MS coupled to GC using a 2ω -detection, for the analysis of Salt marsh soil from New York/New Jersey (NY/NJ) estuary at different depths. The 2ω -detection allowed reaching very high resolutions (300 000 for GC-APCI-FT-ICR MS against 650 000 for DI-APPI-FT-ICR MS at m/z 200) while maintaining cycle times suitable for GC hyphenation. Soils are complex matrices and every mass spectrum acquired and summed on 10-minute segments (Figure 2.13) remains of great complexity despite the GC fractionation, which requires high resolution for their interpretation. Once again, the separation of analytes prior their detection enables reduction of ionization competition, and several compounds containing heteroatoms were detected with higher intensities compared to direct infusion. The FT-ICR MS high resolution and mass measurement accuracy allow the classification of the different detected compounds according to their elemental composition revealing that the NY/NJ estuary may have been contaminated by petroleum, agricultural effluent, or industrial chemicals.

Flora and fauna (including humans through the foodstuffs mentioned in section 2.1.6.2.1) is the ultimate destination of the biosphere pollutants where they are bioaccumulated causing health damages. Several studies monitoring these potential contaminations have been carried out in GC-FT-ICR MS like Solouki *et al.* [132] on volatile organic compounds (terpenes), Taguchi *et al.* [133] on dioxins in extracts of plants exposed to industrial fire fallout, or Sühring *et al.* [134] who controlled the presence of chemical pollution in different parts of European eels (including eggs and gonads) to assess the maternal transfer of these contaminants.

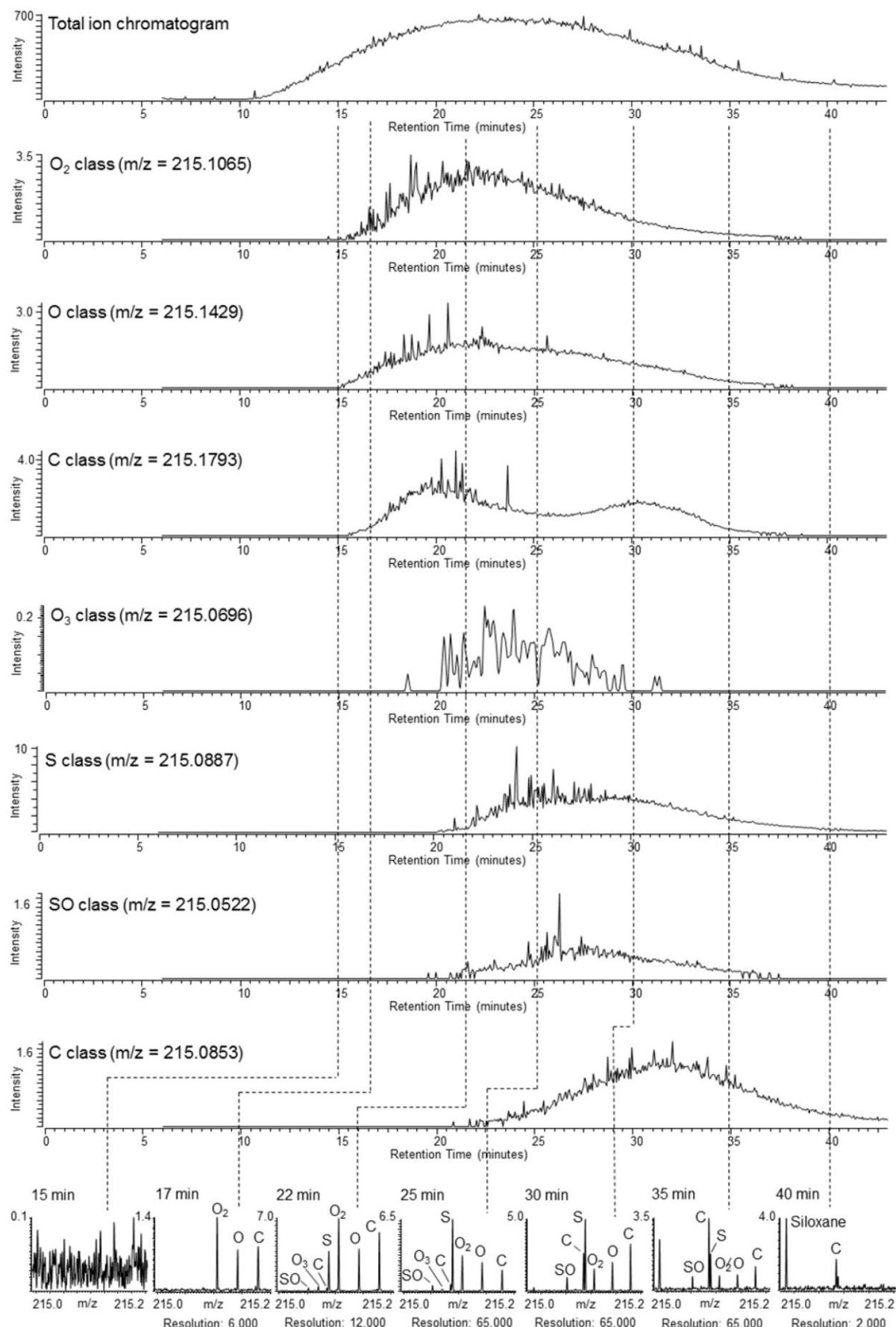


Figure 2.12: Total ion chromatogram and extracted ions chromatograms of an oil sands processed water sample showing separation of several isomers within O, O₂, O₃, C, S and SO classes. Figure reprinted with permission from [130]. Copyright 2014 American Chemical Society.

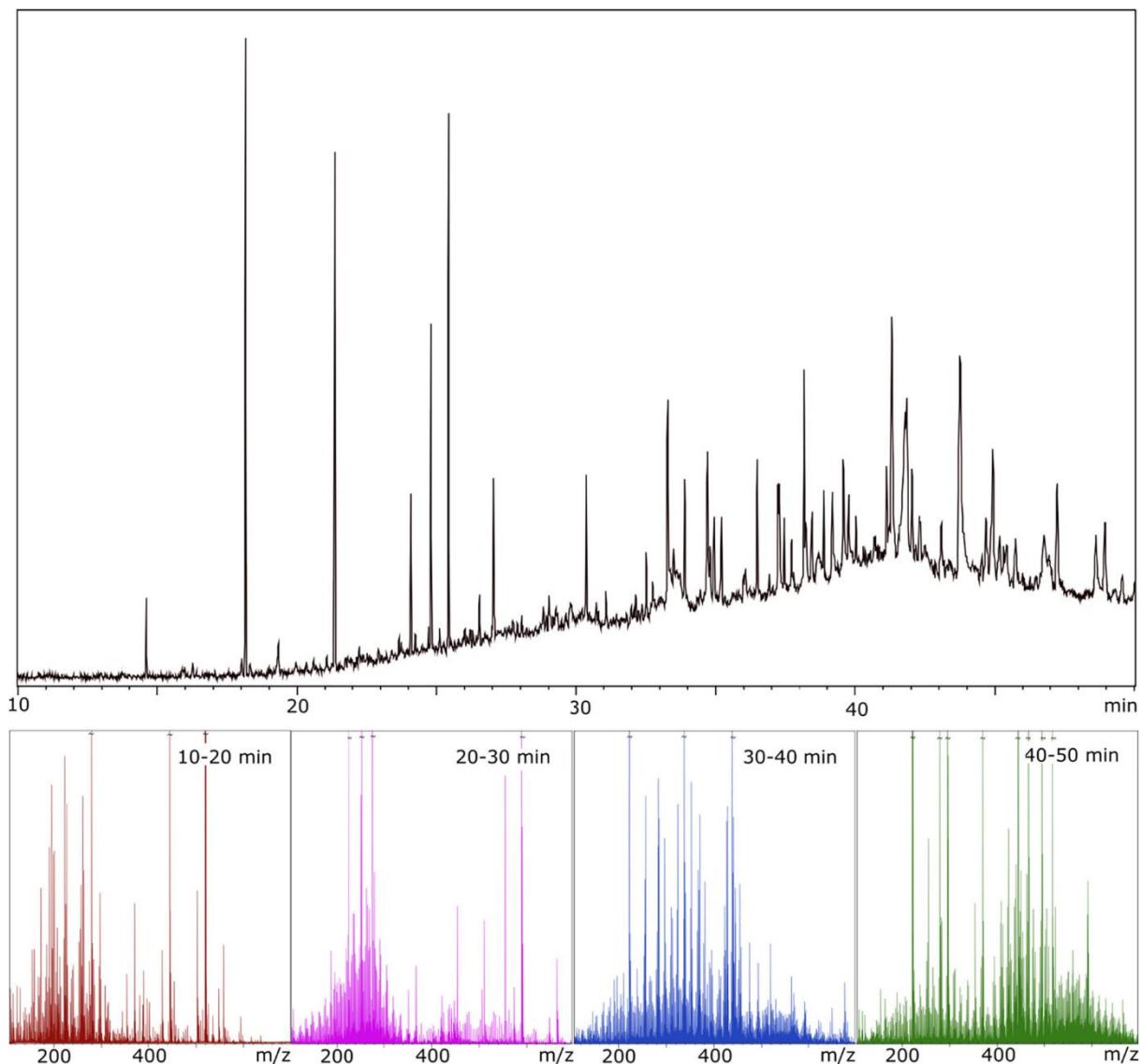


Figure 2.13: Total ion chromatogram for one sampling depth (67.5–69.5 cm) and the mass spectra resulting from averaging the acquired scans over 10 min intervals. Figure extracted and caption adapted from [131].

2.1.6.4. Petroleomics applications

As every omics, petroleomics studies face very high sample complexity requiring mass spectrometers with ultra-high-resolution power and very high mass measurement accuracy such as FT-ICR MS. Only few of these studies involved a separation method, although isomer discrimination cannot be achieved without an additional dimension (separation method or ion mobility). Petroleomic applications using separation methods hyphenated to FT-ICR MS found in the literature are gathered in Table 2.1 for LC but the majority have involved GC collected in Table 2.3.

GC-EI-FT-ICR MS is often used to explore gasoline and fuel content [35,55]. For instance, Luo *et al.* [35] associated GC with a 7 T FT-ICR MS to study low concentrated gasoline samples from four retail gas stations. Over 50 compounds were chromatographically resolved among which some C₈H₁₀ isomers, 9 hexene and 10 heptane isomers. Molecular formulae were confidently assigned within mass errors mostly below 5 ppm and some identifications

were also confirmed with standards. Additional experiments carried out in CI succeeded in the discrimination of some coeluted isomers thanks to their different proton affinities. When samples are containing aromatic polar molecules, APCI is then involved. Zuber *et al.* [63] analyzed pyrolysis oils from German brown coal by GC-APCI-FTICR MS. Once again, the GC separation allowed the separation and detection of many isomers among which the nine propylphenols, ethylmethylphenols and trimethylphenols are displayed in Figure 2.14.

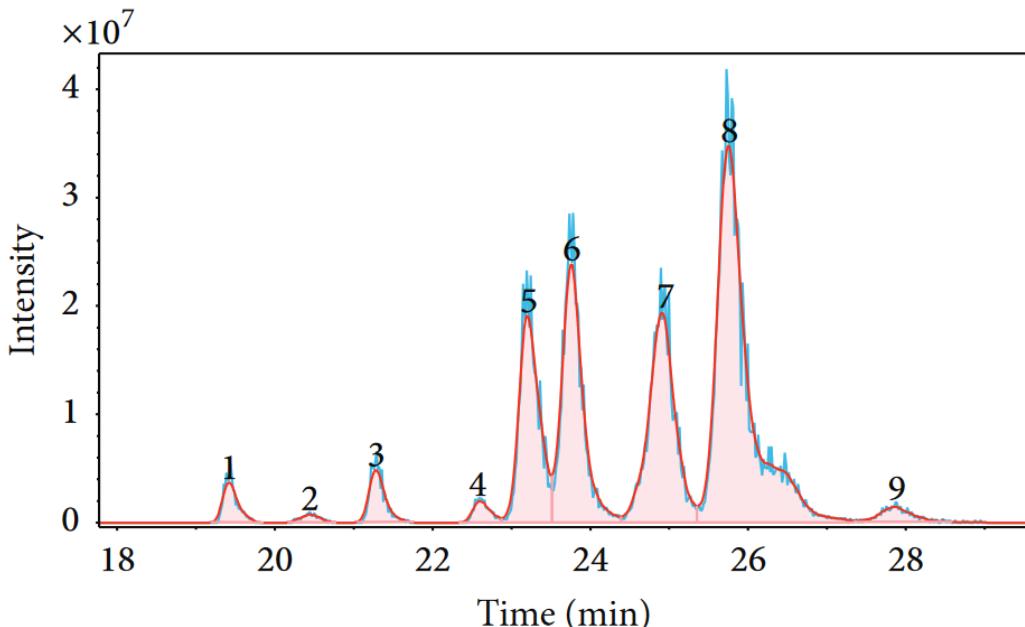


Figure 2.14: Extracted ions chromatograms (m/z 135.081539) from a GC-APCI-FT-ICR MS analysis of pyrolysis oil from German brown coal, showing the separation of 9 propylphenol, ethylmethylphenol, and trimethylphenol isomers. Real signal is represented in blue, and Gaussian-fitted signal is represented in red. Extracted from [63].

API is also an interesting ionization source for PAH and Benigni *et al.* [65] analyzed their content in fossil oil by GC-APLI-FT-ICR MS. Various reference fossil oils standards such as shale oil, petroleum crude oil and heavy sweet crude oil were analyzed to illustrate the benefits of using GC-FT-ICR MS for the characterization of PAH. Due to the intrinsic properties of GC separations, a high mass cut off at m/z 450 was observed. Compared to direct APLI-FT-ICR MS infusion, more compounds were detected by GC-FT-ICR MS in the m/z range below m/z 450 (up to 25% are only identified with GC-FT-ICR MS). Furthermore, the hyphenation enabled isomers separation and increased sensitivity and dynamic range due to reduced complexity.

If very few LC-FT-ICR MS hyphenated systems were used for petroleomics studies, in 2013 Lababidi *et al.* [135] developed the first online LC-APLI-FT-ICR MS method for crude oil analysis. The online LC-FT-ICR MS was more efficient in the assignment than the offline mode (480 assigned formula offline and 1438 assigned formula online) despite a degradation of the FT-ICR MS resolving power to keep compatible the acquisition in LC (resolution of 200 000 at m/z 400 online instead of 750 000 offline). Online LC-APPI-FT-ICR MS analyses of crude oil were carried out later by Rowland *et al.* [31] and the same statement was presented.

2.1.7. What about the future of hyphenation with FT-ICR MS?

Hyphenations of separation methods are undoubtedly a little bit trickier to implement with FT-ICR MS than with other analyzers (quadrupoles, ion traps, time of flight and even orbitraps). This issue raises questions about their future confronted to the relentless progress of MS technologies. No one can predict and plan with precision what will be the dates of the apparition and particularly the diffusion of the new technologies in development in the research laboratories. However, from the recent story, the progress goes ever faster, ever higher and ever further. This leitmotiv is in line with the facilitation of hyphenation. Indeed, from recent articles [89,136–138], it is possible to envisage future commercial ICR cells enabling the detection in 3ω , 4ω , 5ω mode (and even more?) even associated with 21 T magnetic field [92]. Such progress will allow faster acquisitions in phases with the requisites of online separations. Thus, maybe in a few years, the burdensome compromises between FT-ICR MS performance and acquisition time will no longer be needed to get enough acquisition points per chromatographic (from GC to UPLC) or electrophoretic peak.

The effective competition for online separation methods is ion mobility mass spectrometry. It was already evoked in this manuscript as a solution to overcome the separation of isomers (isobars do not present any real problems with FT-ICR MS). As shown in Figure 2.15 [139], chromatography is extremely time-consuming compared to all other MS steps. In the full-time equivalent of the 20-minute chromatogram, it is possible to collect around 100 000 spectra using time dispersive IM. The same performance is only reached at a frequency of around 84 spectra per second in hyphenation (regardless the molecular information and its precision which decreases with the acquisition time). Even if already used for the separation of phosphopeptides [140] in FT-ICR MS analysis, the coupling of time dispersive ion mobility is not suitable for the slower mass analyzers. However, more recent IMS techniques such trapped ion mobility spectrometry (TIMS) and particularly oversampling selective accumulation TIMS became effective enough (IMS resolving up to 250) to be hyphenated to FT-ICR MS without prejudice to the achievement of mass accuracy < 1 ppm and mass resolution > 1 million [141].

Nevertheless, achieving the detection of such a large number of m/z signals does not mean that the analysis a complex mixture is exhaustive. Indeed, mass spectrometers, no matter how powerful (resolution and accuracy), only detect m/z ratios and the analyte ions are all produced upstream from the IMS cells. This crucial step is well known to be subjected to ion suppression effects due to competition during the ionization process. It was clearly highlighted in the application section of this review that there is often a huge difference between the number of analytes detected by direct infusion FT-ICR MS and the number of analytes revealed in very complex mixture by online separation hyphenated with FT-ICR MS. This is not only due to the presence of isomers but above all because of ion suppression. With the constant need to explore deeper and deeper complex samples, hyphenation of separation methods to FT-ICR MS will probably remain essential for years. At last, instead of confronting

chromatographic methods (including capillary electrophoresis) to ion mobility, it will undoubtedly be more fruitful to associate them to go a step further in the determination of the molecular content of complex mixtures.

Nesting of Analytical Timescales

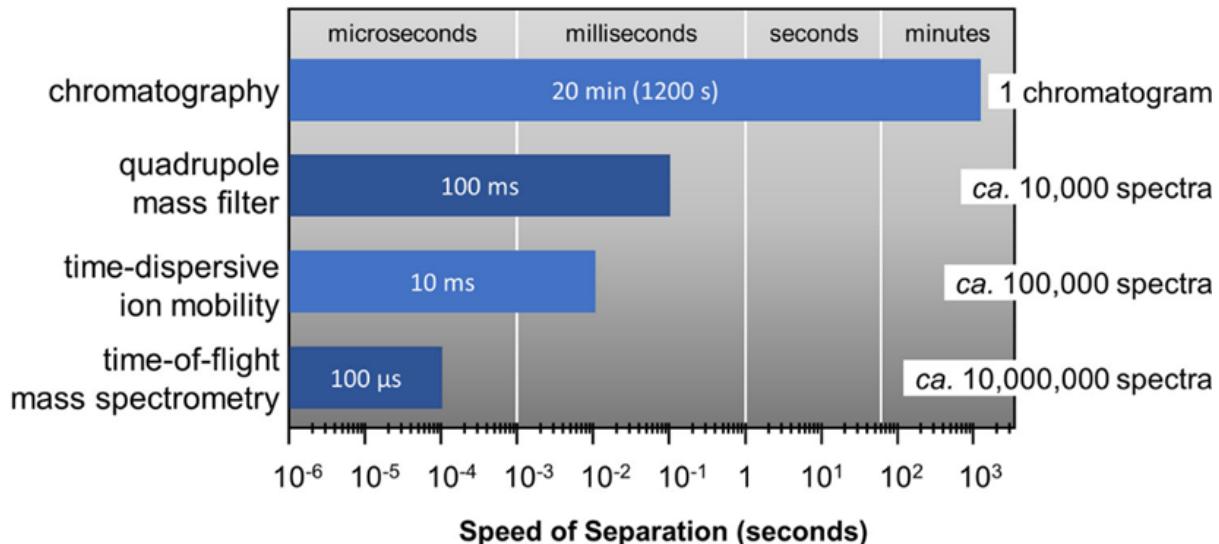


Figure 2.15 : Analytical timescales based on the speed of separation of analytical strategies (on the left) and on the total number of potential spectra that can be obtained in a 20-minute chromatogram (on the right). Figure extracted from [139].

Chapitre 2 : Les couplages entre la FTICR-MS et les méthodes séparatives

Table 2.1: Applications using LC hyphenated to FTICR-MS. (RPLC: reversed-phase liquid chromatography, NPLC: normal-phase liquid chromatography, HIC: hydrophobic interaction chromatography; GCC: graphited carbon column; EC: electrochemical cell; CZE: capillary zone electrophoresis; A: isomers separation, B: quantification, C: high sensitivity; D: high dynamic range, E: easier data interpretation; n.d.:not disclosed)

Applications			Instrumental parameters							Advantages brought by the LC hyphenation				References		
Domain	Application	Sample	LC mode	Ionization source (polarity)	FTICR magnetic field (T)	MS mode	Other parameters	MS resolution (FWHM)	Mass measurement accuracy [ppm]	Advantages brought by LC hyphenation	Number of detected compounds	Number of unambiguous formula assignments	Number of identifications	Authors	Year	Reference
Proteomics	Peptides analysis	Standard peptides in artificial cerebrospinal fluid	RPLC	ESI (+)	n.d.	MS	Accumulation during 0.5 - 3 s, up to 100% duty cycle	n.d.	n.d.	C	n.d.	n.d.	n.d.	Senko <i>et al.</i>	1997	[142]
Proteomics	Peptides identification	Neuroactive peptides mixture	RPLC	ESI (+)	9.4	MS, MS ² (IRMPD)	64 k - 128 k data points	8000 - 1100	n.d.	C	n.d.	n.d.	n.d.	Emmet <i>et al.</i>	1998	[143]
Proteomics	Proteins identification	Mixtures of standard proteins	RPLC	ESI (n.d.)	9.4	MS, MS ² (IRMPD)	256 k data points, 5 s/scan	n.d.	n.d.	n.d.	5	n.d.	5	Li <i>et al.</i>	1999	[144]
Proteomics	Proteome analysis	Tryptic digests of fractionated <i>S. cerevisiae</i> cell lysate	RPLC	ESI (n.d.)	11.5	MS, MS ² (n.d.)	128 k data points, 5 s/spectrum	n.d.	0.3 - 4.5	B, C	n.d.	n.d.	n.d.	Gao <i>et al.</i>	2000	[46]
Proteomics	Proteins identification	Digest of protein mixture (β -lactoglobulin, β -casein, bovine serum albumin, carbonic anhydrase II, cytochrome c, glyceraldehyde-3-phosphate dehydrogenase)	RPLC	ESI (+)	n.d.	MS, MS ² (IRMPD)	64 k data points, 2 s/spectrum	n.d.	<0.01 Da	C	n.d.	n.d.	n.d.	Martin <i>et al.</i>	2000	[145]
Proteomics	Bottom-up identification of proteins	Cytosolic tryptic digest of <i>S. cerevisiae</i>	RPLC	ESI (n.d.)	11.5	MS, MS ² (n.d.)	5.7 s/ spectrum	n.d.	<1	C, D, E	110000	n.d.	~1000	Shen <i>et al.</i>	2001	[44]
Proteomics	Proteome analysis	Tryptic digest of <i>D. radiodurans</i> cell lysate	RPLC	ESI (+)	3.5	MS, MS ² (n.d.)	2.5 s/spectrum	n.d.	n.d.	C, D	48664	n.d.	n.d.	Shen <i>et al.</i>	2001	[43]
Proteomics	Peptides and proteins identification	Tryptic digests of <i>D. radiodurans</i> cell lysate and bovine serum albumine	RPLC	ESI (+)	7	MS, MS ² (IRMPD, SORI-CID)	256 k data points	n.d.	1.1 - 35.1	D	44	n.d.	44	Li <i>et al.</i>	2001	[104]
Proteomics	Peptides analysis	Biological peptide mixtures	RPLC	ESI (+)	7	MS	64 k data points, 2 s/spectrum	n.d.	n.d.	C	n.d.	n.d.	n.d.	Quenzer <i>et al.</i>	2001	[146]
Proteomics	Proteins identification	Tryptic digest of diacetyl reductases in the human liver cytosol	RPLC	ESI (+)	n.d.	MS	m/z range 400 - 2400, 256 k data points, 6 s/2 spectra, accumulation during 0.5 - 4.8 s	n.d.	0.0 - 3;5	n.d.	n.d.	n.d.	n.d.	Tanaka <i>et al.</i>	2001	[147]
Proteomics	Peptides characterization	BSA tryptic digest	RPLC	ESI (n.d.)	3.5	MS	Accumulation during 0.105 s	n.d.	0.059 - 4.997	C, D	83	n.d.	83	Belov <i>et al.</i>	2002	[148]
Proteomics	Peptides characterization	Mixture of standard peptides and bovine serum albumin tryptic digest	RPLC	ESI (n.d.)	9.4	MS, MS ² (ECD)	128 k - 512 k data points, 17.5 - 28.1 s/spectrum	n.d.	n.d.	n.d.	75	n.d.	n.d.	PalmbLad <i>et al.</i>	2002	[149]
Proteomics	Identification of intact proteins	Yeast large ribosomal subunits	RPLC	ESI (n.d.)	7	MS	n.d.	n.d.	0.5 - 20.3	n.d.	71	n.d.	62	Lee <i>et al.</i>	2002	[106]
Proteomics	Proteome analysis	Tryptic digests of <i>D. radiodurans</i> cell lysate	RPLC	ESI (+)	11.4	MS, MS ² (n.d.)	n.d.	n.d.	≤ 1	C, D	n.d.	n.d.	n.d.	Lipton <i>et al.</i>	2002	[102]

Chapitre 2 : Les couplages entre la FTICR-MS et les méthodes séparatives

Proteomics	Proteome characterization	Tryptic digest of <i>Deinococcus radiodurans</i> cell lysate	RPLC	ESI (+)	11.5	MS, MS ² (n.d.)	5 s/spectrum	100000	< 1	B, C	n.d.	n.d.	n.d.	Smith <i>et al.</i>	2002	[150]
Proteomics	Analysis of protein digest	Cytochrome c digest	RPLC	ESI (+)	7	MS (broadband ECD)	<i>m/z</i> range 292 – 2000, 512 k data points, accumulation during 0.5	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	Davidson <i>et al.</i>	2002	[151]
Proteomics	Proteins analysis	Protein mixture from COS-6 membranes	RPLC	ESI (+)	7	MS	<i>m/z</i> range 400 – 2000	n.d.	0.4 – 4.0	E	n.d.	n.d.	n.d.	Ihling <i>et al.</i>	2003	[152]
Proteomics	Proteins identification	Tryptic digests of cerebrospinal fluid samples	RPLC	n.d.	9.4	n.d.	n.d.	n.d.	~ 2	C, D	6551	n.d.	n.d.	Ramström <i>et al.</i>	2003	[97]
Proteomics	Proteins identification	BSA tryptic digest	RPLC	ESI (+)	7	MS	Infinity ICR cell, <i>m/z</i> 790 – 3000, 512 k data points.	n.d.	≤ 5 (external calibration), ≤ 2 (internal calibration ^a)	n.d.	n.d.	n.d.	n.d.	Witt <i>et al.</i>	2003	[153]
Proteomics	Bottom-up analysis	<i>Deinococcus radiodurans</i> cell lysate	RPLC	ESI (n.d.)	9.4	MS, MS ² (n.d.)	Infinity ICR Cell, 512 k data points, transient of 0.8 s, accumulation during 0.6 s, 1.6 s/scan	n.d.	≤ 10	C, D	n.d.	n.d.	n.d.	Belov <i>et al.</i>	2004	[100]
Proteomics	Analysis of complex biological samples	Tryptic digests of BSA and cerebrospinal fluid samples	RPLC-CZE	ESI (n.d.)	9.4	n.d.	7.5 s/spectrum	n.d.	≤ 10	n.d.	10 - 3126	n.d.	10 n.d. 3126	Bergström <i>et al.</i>	2006	[154]
Proteomics	Characterization of proteins (potential biomarkers or vaccine antigens)	Tryptic digest of <i>Helicobacter pylori</i> outer membrane proteins	RPLC	ESI (+)	7	MS, MS ² (n.d.)	<i>m/z</i> range 400 – 1600	60000 at <i>m/z</i> 600	≤ 5	C, D, E	n.d.	n.d.	60	Carlsohn <i>et al.</i>	2006	[52]
Proteomics	Bottom-up characterization of the proteome	Tryptic digest of <i>S. oneidensis</i> MR-1 cell lysates	RPLC	ESI (n.d.)	7	MS, MS ² (CID)	Open cylindrical ICR cell, time of flight of 0.150 s	n.d.	MS: ≤ 10 MS ² : ≤ 25	n.d.	94	n.d.	n.d.	Kang <i>et al.</i>	2007	[99]
Proteomics	Datasets processing	Tryptic digests of savinase, a protein mixture and cerebrospinal fluid	RPLC	ESI (+)	7	n.d.	Infinity ICR Cell, time of flight of 0.4 s, 1.3 s/scan	n.d.	≤ 20	n.d.	n.d.	n.d.	34	Burgt <i>et al.</i>	2007	[155]
Proteomics	Characterization of oxidized forms of a protein	Calmodulin from activated macrophage lysates	RPLC	ESI (n.d.)	12	MS, MS ² (n.d.)	<i>m/z</i> range 400 – 2000, 2 s/spectrum	10 000	< 15	n.d.	n.d.	n.d.	n.d.	Smallwood <i>et al.</i>	2007	[156]
Proteomics	Analysis of glycopeptides	Tryptic digest of a HIV-1 envelope protein	RPLC	ESI (+)	7	MS, MS ² (CID)	<i>m/z</i> range 800 - 2000	50000 at <i>m/z</i> 400	n.d.	n.d.	215	n.d.	215	Irungu <i>et al.</i>	2008	[94]
Proteomics	Top-down identification of proteins	<i>A. flavus</i> cell lysate	RPLC	ESI (n.d.)	7	MS, MS ² (CID)	n.d.	100000 at <i>m/z</i> 400	0.44 - 2.34	B	659	n.d.	22	Collier <i>et al.</i>	2008	[118]
Proteomics	Bottom-up characterization of the proteome	Tryptic digest of proteins in mouse bronchoalveolar lavage fluid	RPLC	n.d.	9.4	MS, MS ² (n.d.)	n.d.	n.d.	≤ 5	n.d.	n.d.	n.d.	1797	Pounds <i>et al.</i>	2008	[98]
Proteomics	Top-down characterization of the degree of nitration and oxidation of intact calmodulin	Calmodulin in macrophage lysates	RPLC	ESI (+)	12	MS, MS ² (CID)	Infinity ICR Cell, accumulation during 0.2 s, 2 s/spectrum	n.d.	0.31 - 23.98	B	266	n.d.	144	Lourette <i>et al.</i>	2010	[120]
Proteomics	Characterization of glycopeptides	Tryptic digest of fetal bovine serum	RPLC	ESI (+)	14.5	MS, MS ² (CID)	Accumulation during 0.2 s	100000 at <i>m/z</i> 400	0.01 - 2.08	n.d.	n.d.	137	n.d.	Wang <i>et al.</i>	2010	[96]
Proteomics	Proteoforms identification	α-synuclein from human brain tissue	RPLC	ESI (+)	7	MS, MS ² (CID)	<i>m/z</i> range 600 - 1500 or 350 - 1500	MS: 100000 MS ² : 50000	2 - 5	n.d.	3	n.d.	3	Örhfelt <i>et al.</i>	2011	[157]
Proteomics	Top-down analyses of endogenous Amyloid β peptide species and other proteolytic cleavage products of amyloid precursor protein	Human and cat cerebrospinal fluid samples	RPLC	ESI (+)	7	MS, MS ² (CID, IRMPD, ECD)	<i>m/z</i> range 350 - 1500	MS: 100000 MS ² : 50000	0.0 - 9.3	B	n.d.	n.d.	n.d.	Brinkmalm <i>et al.</i>	2012	[116]
Proteomics	Top-down identification of proteins	HeLa cell lysate	n.d.	ESI (+)	14.5	MS, MS ² (CID)	Transient of 1.8 s	n.d.	≤ 10	n.d.	n.d.	n.d.	n.d.	Tipton <i>et al.</i>	2012	[119]

Chapitre 2 : Les couplages entre la FTICR-MS et les méthodes séparatives

Proteomics	Top-down identification of proteins	Human 2OS and 19S proteasome complexes	RPLC	ESI (+)	15	MS, MS ² (FS-CAD, CASI-CAD)	<i>m/z</i> range 150 – 3000, 1 M data points, transient of 0.34 s, accumulation during 0.2 s, 0.63 s/spectrum	45000 – 50000	0.4 - 17	n.d.	n.d.	n.d.	n.d.	Lakshmanan <i>et al.</i>	2014	[121]
Proteomics	Glycan structural analysis of a glycoprotein	Ovalbumin	GCC-LC	ESI (+)	n.d.	MS, MS ² (CID)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	Xin <i>et al.</i>	2015	[95]
Proteomics	Characterization of protein disulfide bonds	β -lactoglobulin and ribonuclease B	RPLC-EC	ESI (+)	15	MS, MS ² (CID)	ParaCell, <i>m/z</i> range 307 – 3000, 1 M data points, accumulation during 3 s	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	Switzer <i>et al.</i>	2016	[158]
Proteomics	Top-down analysis of protein	Human Mitochondrial Branched-Chain Amino Acid Aminotransferase 2	RPLC	miESI (+)	21	MS, MS ² (ETD, CID)	Dynamically harmonized ICR cell, transient of 0.76 s	300000 at <i>m/z</i> 400 110000 at <i>m/z</i> 920	0.1 - 3.6	n.d.	n.d.	n.d.	n.d.	Anderson <i>et al.</i>	2017	[159]
Proteomics	Top down/middle down analysis of mAbs	mAbs in human serum	RPLC	ESI (+)	21	MS, MS ² (ETD, CID)	<i>m/z</i> range 650 - 2000 or 300 – 2000, transient of 0.76 s or 0.38 s	150000 at <i>m/z</i> 400	MS: 0.2 - 0.4 MS ² : 2.9 - 4.7	B	n.d.	n.d.	n.d.	He <i>et al.</i>	2017	[114]
Proteomics	Proteome top-down analysis	Fractionated human colorectal cancer cells lysate	RPLC	ESI (+)	21	MS, MS ² (ETD, CID)	Transient of 0.381 - 0.762 s, duty cycle (1MS + 2MS/MS) of 3 - 6 s	150000 - 300000 at <i>m/z</i> 400	n.d.	n.d.	n.d.	n.d.	141 n.d. 593	Anderson <i>et al.</i>	2017	[115]
Proteomics	Top-down analysis of mAbs	Deglycosylated IgG1	HIC	n.d. (+)	12	MS, MS ² (ECD)	ParaCell, <i>m/z</i> range 200 - 10000 or 500 – 10000, 1 M data points, accumulation during 1.3 s, time of flight of 2 ms	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	Chen <i>et al.</i>	2018	[117]
Proteomics	Middle-down analysis of mAbs	universal antibody standard human IgG1	RPLC	nanoESI (+)	12	MS, MS ² (ECD/CAD)	<i>m/z</i> range 100 – 3000, 1 M data points, accumulation during 0.04 s	Isotopic resolution	n.d.	n.d.	6	n.d.	6	Jin <i>et al.</i>	2019	[113]
Proteomics	Analysis of mAbs at intact level	4 mAbs intacts	FcyRIIIa affinity	ESI (+)	15	MS	<i>m/z</i> range 506 – 20000, 128 k data points, accumulation during 1 s, 2.6 data points/min	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	Lippold <i>et al.</i>	2019	[105]
Targeted metabolomics	Profiling of N-acyl-homoserine lactones	Pseudomonas aeruginosa isolates	RPLC	ESI (+)	7	MS, MS ² (IRMPD)	<i>m/z</i> range 100 – 500	50000 at <i>m/z</i> 400	0.10 - 4.92	A, B, C, D	12	n.d.	12	Cataldi <i>et al.</i>	2007	[30]
Targeted metabolomics	Targeted and untargeted analysis of amine containing metabolites	Human urine	RPLC	ESI (+)	9.4	MS	n.d.	n.d.	0.00 - 1.95	A, B, D	438	n.d.	18	Guo <i>et al.</i>	2007	[160]
Targeted metabolomics	Targeted bile acids profiling in bodyfluids and tissues	Human urine, plasma, liver tissue methanol extracts	RPLC	ESI (-)	n.d.	MS	n.d.	100000 at <i>m/z</i> 400	0.10 - 0.37	A, B	20	n.d.	17	Bobeldijk <i>et al.</i>	2008	[161]
Targeted metabolomics	Analysis of N-acyl-homoserine-lactones and cognate lactone-opened compounds	Pseudomonas aeruginosa isolates	RPLC	ESI (+)	7	MS, MS ² (CID)	<i>m/z</i> range 150 - 500	50000 at <i>m/z</i> 400	0.07 - 1.79	A, B	13	n.d.	n.d.	Cataldi <i>et al.</i>	2009	[162]
Targeted metabolomics	Quantitative profiling of amine and phenol metabolites using 12C/13C isotope dansylation labeling	Human urine samples	RPLC	ESI (+)	9.4	MS	n.d.	n.d.	< 1.5	A, B, D	672	n.d.	121	Guo <i>et al.</i>	2009	[163]
Targeted metabolomics	Metabolites profiling using 13C/12C isotope	Human cerebrospinal fluid	RPLC	ESI (+)	9.4	MS	n.d.	n.d.	< 2	A, B	1132	n.d.	85	Guo <i>et al.</i>	2011	[164]
Targeted metabolomics	Targeted analysis of monoterpane indole alkaloids	Extracts of <i>C. roseus</i> and <i>U. rhynchophylla</i>	RPLC	ESI (+)	7	MS ² (QCID)	Infinity ICR Cell, <i>m/z</i> range 100 - 600, 0.4893 s, accumulation during 0.1 s for MS and 2 s for MS/MS, time of flight of 0.4 s	66000 at <i>m/z</i> 400	n.d.	A, C, D, E	n.d.	19	n.d.	Nakabayashi <i>et al.</i>	2015	[128]
Targeted metabolomics	Cannabinoids characterization	Extracts of <i>Cannabis sativa</i>	RPLC	ESI (+)	7	MS, MS ² (CID, IRMPD)	<i>m/z</i> range 50–1000	100000 at <i>m/z</i> 400	0.02 – 1.22	A, B	6	n.d.	6	Lelario <i>et al.</i>	2021	
Untargeted metabolomics	Profiling of N-acyl-homoserine lactones	Pseudomonas aeruginosa isolates	RPLC	ESI (+)	7	MS, MS ² (IRMPD)	<i>m/z</i> range 100 – 500	50000 at <i>m/z</i> 400	0.10 - 4.92	A, B, C, D	12	n.d.	12	Cataldi <i>et al.</i>	2007	[30]

Chapitre 2 : Les couplages entre la FTICR-MS et les méthodes séparatives

Untargeted metabolomics	Profiling of metabolites	Peel and flesh extracts of tomato	RPLC	ESI (+/-)	n.d.	MS, MS ² (CID)	<i>m/z</i> range 100 - 1500	100000 at <i>m/z</i> 400	0.00 - 3.86	A, B	869	493	31	Iijima <i>et al.</i>	2008	[36]
Untargeted metabolomics	Profiling of flavonoids	Lotus japonicus extracts	RPLC	ESI (+)	n.d.	MS, MS ² (CID)	<i>m/z</i> range 200 - 1500	100000 at <i>m/z</i> 400	0.006 - 0.435	A, B	61	59	14	Suzuki <i>et al.</i>	2008	[165]
Untargeted metabolomics	Analysis of N-acyl-homoserine-lactones and cognate lactone-opened compounds	Pseudomonas aeruginosa isolates	RPLC	ESI (+)	7	MS, MS ² (CID)	<i>m/z</i> range 150 - 500	50000 at <i>m/z</i> 400	0.07 - 1.79	A, B	13	n.d.	n.d.	Cataldi <i>et al.</i>	2009	[162]
Untargeted metabolomics	Metabolites profiling	Leaf of Arabidopsis thaliana	RPLC	ESI (+/-)	n.d.	MS	<i>m/z</i> range 100 - 1300, transient ≤ 0.5 s	Set at 50000	≤ 1	A, B, C, D	643	94 over 98 most abundant	n.d.	Giavalisco <i>et al.</i>	2009	[50]
Untargeted metabolomics	Metabolites profiling of ginsenosides <i>in vitro</i> models of the gastrointestinal tract	Red ginseng extracts	RPLC	ESI (-)	n.d.	MS, MS ² (CID)	<i>m/z</i> range 300 - 1500	n.d.	≤ 1.2	A, B	n.d.	n.d.	42	Kong <i>et al.</i>	2009	[37]
Untargeted metabolomics	Profiling of metabolites triterpene saponins	Hairy roots of <i>Medicago truncatula</i> extracts	RPLC	ESI (-)	n.d.	MS, MS ² , MS ³ (CID)	<i>m/z</i> range 120 - 1400	Set at 100000	1.352 - 6.539	A	79	n.d.	79	Pollier <i>et al.</i>	2011	[125]
Untargeted metabolomics	Analysis of sulfur-containing metabolites	Allium cepa extracts	RPLC	ESI (+)	7	MS, MS ² (n.d.)	<i>m/z</i> range 50 - 1500, 4 M MS data points (transient 1.9573 s), 1 M MS/MS data points, MS transient 1.9573 s, MS/MS transient of 0.2621 s, accumulation during 0.001 s, MS time of flight of 0.5 ms, MS/MS time of flight of 0.7 ms	MS: 260000 at <i>m/z</i> 400 MS ² : 35000 at <i>m/z</i> 400	< 1	n.d.	67	22	6	Nakabayashi <i>et al.</i>	2013	[122]
Untargeted metabolomics	Profiling of <i>in vivo</i> metabolites of flavonoids after oral admission of extracts of ziziphi spinosae semen to rats and dogs	Feces and urine from rats and dogs	RPLC	ESI (+)	n.d.	MS, MS ² , MS ³ (CID)	<i>m/z</i> range 100 - 1000	Set at 50000	0.03 - 3.83	A	15	15	9	Ren <i>et al.</i>	2013	[123]
Untargeted metabolomics	Profiling of <i>in vivo</i> metabolites of 2-(2 hydroxypropanamido) benzoic acid	Plasma, urine, feces, bile from rats	RPLC	ESI (+)	7	MS, MS ² (CID)	<i>m/z</i> range 50 - 1000	n.d.	0.07 - 0.94	A, C	13	13	n.d.	Guan <i>et al.</i>	2015	[124]
Untargeted metabolomics	Chemical constituent profiling	Rhodiola crenulata extract	RPLC	ESI (+/-)	7	MS, MS ² (CID)	<i>m/z</i> range 50 - 3 000	n.d.	0.01 - 1.43	D	48	n.d.	48	Han <i>et al.</i>	2016	[166]
Untargeted metabolomics	Analysis of sulfur-containing metabolites	Allium cepa, Allium fistulosum and Allium sativum extracts	RPLC	ESI (+)	7	MS, MS ² (n.d.)	<i>m/z</i> range 50 - 1500, 4 M MS data points (transient 1.9573 s), 1 M MS/MS data points, MS transient 1.9573 s, MS/MS transient of 0.2621 s, accumulation during 0.001 s, MS time of flight of 0.5 ms, MS/MS time of flight of 0.7 ms	MS: 260000 at <i>m/z</i> 400 MS ² : 35000 at <i>m/z</i> 400	< 1	A	69	n.d.	11	Nakabayashi <i>et al.</i>	2016	[167]
Untargeted metabolomics	Profiling of <i>in vivo</i> metabolites of esculin after oral administration	Plasma, urine, feces and bile from rats	RPLC	ESI (+)	7	MS, MS ² (CID)	<i>m/z</i> range 100 - 1000	n.d.	0.03 - 3.00	A	19	19	n.d.	Wang <i>et al.</i>	2016	[168]
Untargeted metabolomics	Metabolites profiling of pyrexia for biomarkers identification	Plasma of rats after subcutaneous administration of pyrexia	RPLC	ESI (+)	7	MS	<i>m/z</i> range 50 - 1000, accumulation during 0.15 s	n.d.	0.01 - 2.83	n.d.	52	n.d.	52	Liu <i>et al.</i>	2017	[169]
Untargeted metabolomics	Profiling of metabolites of erysolin after intragastric administration	Plasma, urine, feces and bile from rats	RPLC	ESI (+)	7	MS, MS ² (CID)	<i>m/z</i> range 50 - 1500	n.d.	0.08 - 0.45	A, E	6	6	n.d.	Li <i>et al.</i>	2018	[170]
Untargeted metabolomics	Profiling of metabolites of Gegenginlian decoction after oral administration	Plasma, urine, bile and feces from rats	RPLC	ESI (+)	7	MS, MS ² (CID)	<i>m/z</i> range 50 - 3000	n.d.	0.00 - 3.87	A, C, D	174	n.d.	≥ 19	Liu <i>et al.</i>	2018	[171]
Untargeted metabolomics	Metabolites profiling of paeoniflorin	Human intestinal microflora	RPLC	ESI (+)	7	MS	<i>m/z</i> range 100 - 1000	n.d.	0.00 - 2.28	A, C	31	n.d.	16	Sun <i>et al.</i>	2018	[38]

Chapitre 2 : Les couplages entre la FTICR-MS et les méthodes séparatives

Untargeted metabolomics	Metabolites profiling of Palbociclib after oral administration	Urine, feces and bile from rats	RPLC	ESI (+)	7	MS, MS ² (n.d.)	<i>m/z</i> range 150 - 1000	n.d.	0.04 - 1.64	A	29	n.d.	29	Yao et al.	2019	[172]
Targeted lipidomics	Study of the triglyceride dynamics during the ovarian previtellogenesis stage	Ovaries of <i>Aedes aegypti</i>	RPLC	ESI (+)	14.5	MS	Hexapolar detection ICR cell	> 1000000 at <i>m/z</i> 800	0.065 - 0.440	A, B	19	n.d.	19	Valadares Tose et al.	2021	[91]
Untargeted lipidomics	Lipids profiling	Murine primary hepatocytes	RPLC	ESI (+/-)	7	MS, MS ² (CID)	<i>m/z</i> range 400/350 - 1050	200000 at <i>m/z</i> 400	< 2	A, B, C, D	242	n.d.	n.d.	Fauland et al.	2011	[29]
Foodomics	Analysis of glucosinolates	Extracts of wild bud flowers of <i>Capparis spinosa</i>	RPLC	ESI (-)	7	MS, MS ² (IRMPD)	<i>m/z</i> range 50 - 1000	100000 at <i>m/z</i> 400	0.23 - 1.56	B, C	n.d.	8	n.d.	Bianco et al.	2012	[126]
Foodomics	Analysis of intact glucosinolates	Brassicaceae (broccoli, cauliflower, rocket salad) extracts	RPLC	ESI (-)	n.d.	MS, MS ² (CID)	<i>m/z</i> range 70 - 1000	100000 at <i>m/z</i> 400	0.1 - 2.0	A, B, C	n.d.	n.d.	24	Lelario et al.	2012	[53]
Foodomics	Analysis of flavan-3-ols	Extracts of seeds of grape pomace	RPLC	ESI (-)	7	MS	<i>m/z</i> range 200 - 2000 or 150 - 1500	100000 at <i>m/z</i> 400	0.01 - 2.00	A	n.d.	251	n.d.	Rockenbach et al.	2012	[127]
Foodomics	Profiling of glycoalkaloids	Eggplants extracts	RPLC	ESI (+)	n.d.	MS, MS ² (IRMPD)	<i>m/z</i> range 50 - 1300	n.d.	0.2 - 1.3	A, B	n.d.	19	n.d.	Lelario et al.	2019	[47]
Environment	Organic aerosol characterization	VOCs from gas-phase ozonolysis of α -pinene, sabinene and cyclohexane	RPLC	ESI (-)	7	MS, MS ² (CID)	<i>m/z</i> range 100 - 500	100000 at <i>m/z</i> 400	0.1 - 2.1	A	58	58	n.d.	Müller et al.	2009	[173]
Environment	Mutagenic polycyclic aromatic compounds analysis	Fractions of aquatic sediment extract	RPLC	APCI (+/-)	n.d.	MS (SIM), MS ²	<i>m/z</i> range 100 - 400	25000	\leq 1	A	17	n.d.	17	Bataineh et al.	2010	[67]
Environment	Dissolved and natural organic matter analysis	Coastal seawater from the Tasman Peninsula and Suwannee River natural organic matter	RPLC	ESI (-)	n.d.	MS	<i>m/z</i> range 50 - 1000	Target at 30000	\leq 2	A	2200	n.d.	n.d.	Sandon et al.	2018	[54]
Environment	Natural organic matter analysis	Suwannee River fluvic acid and Upper Mississippi River natural organic matter	RPLC	ESI (-)	7 (2XR)	MS	2 ω detection, <i>m/z</i> range 100 - 1500, 2 M data points, transient of 0.524 s, accumulation during 0.2 s, 0.803 s/scan	170000 at <i>m/z</i> 400	\leq 1	B	n.d.	n.d.	n.d.	Kim et al.	2019	[129]
Environment	Soil organic matter characterization	Soil extracts	RPLC	ESI (-)	12	MS	<i>m/z</i> range 200 - 1200	n.d.	n.d.	C, D	n.d.	n.d.	n.d.	Shen et al.	2019	[174]
Environment	Siderophore profiling of soil bacteria	Soil extract after enrichment culture	RPLC	ESI (+)	21	MS, MS ² (CID)	<i>m/z</i> range 220 - 1400, transient of 1-2 s	600000 at <i>m/z</i> 400	n.d.	D	46	n.d.	n.d.	Boiteau et al.	2019	[32]
Environment	Geographical origin determination	Raw propolis samples	UPLC	ESI (+)	15	MS	<i>m/z</i> range 150 - 1000, 4 M data points, transient of 1.398 s, accumulation during 50 ms	n.d.	0.0 - 1.098	B	n.d.	n.d.	43	Kim et al.	2019	[175]
Environment	Photodegradation products of benzisothiazolinone monitoring	Delaware River to the Sargasso Sea	RPLC	ESI (+/-)	9.4	MS, MS ² (CID)	<i>m/z</i> range 57.7 - 1000, 4 M data points, accumulation during 0.02s	n.d.	n.d.	A	15	n.d.	15	Varga et al.	2020	[176]
Environment	Dissolved organic matter analysis	Surface water from Yongding River and rainwater from Shanghai	RPLC	ESI (-)	7 (2XR)	MS, MS ² (CID)	<i>m/z</i> range 200 - 800	n.d.	\leq 0.5	A, B, D, E	n.d.	n.d.	n.d.	Qi et al.	2021	[49]
Petroleomics	Component analysis of crude oil	Deasphalted crude oil rich in nitrogen	NPLC	APLI (+)	12	MS	n.d.	200000 at <i>m/z</i> 400	n.d.	C	1438	n.d.	n.d.	Lababidi et al.	2013	[135]
Petroleomics	Component analysis of crude oil	Deasphalted crude oil rich in nitrogen	NPLC	ESI, APPI, APCI, APLI (+)	12	MS	<i>m/z</i> range 150 - 800, 3 s/spectrum	n.d.	n.d.	A	n.d.	150 - 939	n.d.	Lababidi et al.	2014	[68]
Petroleomics	Component analysis of crude oil	Heavy gas oil distillate and bitumen distillate	NPLC-SCX	APPI (+)	21	MS	Dynamically harmonized ICR cell, <i>m/z</i> range 150 - 2000, transient of 6.2 s, absorption mode	3200000 at <i>m/z</i> 400	RMS : 0.120	A, C, D	49356	n.d.	n.d.	Rowland et al.	2021	[31]

Chapitre 2 : Les couplages entre la FTICR-MS et les méthode séparatives

Miscellaneous	Assignment of compounds produced by combinatorial chemistry	Mixture of pyrazole carboxylic acids	RPLC	ESI (+)	4.7	MS	<i>m/z</i> range 250 – 2500, 64 k data points, 3 s/scan	95 000 at <i>m/z</i> 361	n.d.	n.d.	120	n.d.	n.d.	Schmid <i>et al.</i>	2001	[33]
Miscellaneous	Oligosaccharides characterization	Rat brain	GCC-LC	ESI (+/-)	7	MS, MS ² , MS ³ , MS ⁴ (n.d.)	<i>m/z</i> range 700 - 2000	n.d.	n.d.	A	103	n.d.	n.d.	Itoh <i>et al.</i>	2006	[177]
Miscellaneous	Identification of microcystin toxins produced by a cyanobacteria	Extract of <i>Microcystis aeruginosa</i> cell lysate	RPLC	n.d.	7	MS, MS ² (CAD)	<i>m/z</i> range 400 – 1200, MS: 0.2 s/scan, MS/MS: 0.4 s/scan	MS : 100000 at <i>m/z</i> 400 MS/MS: 25000	MS: 0.11 - 0.76 MS ² : 0.00 - 7.8	n.d.	8	n.d.	8	Diehnelt <i>et al.</i>	2006	[178]
Miscellaneous	Cell therapy products quality control	N-glycolylneurameric acid incorporated into human cells	RPLC	ESI (+)	n.d.	MS, MS ² (n.d.)	<i>m/z</i> range 400 - 450	n.d.	n.d.	B, C	n.d.	n.d.	n.d.	Hashii <i>et al.</i>	2007	[179]
Miscellaneous	Criteria determination for hormone and veterinary drug residue analyses	Stanozolol and designer β-agonist "Clenbuterol-R"	RPLC	ESI (+)	7	MS, MS ² (CID)	<i>m/z</i> range 75 - 400	250000 at <i>m/z</i> 161	MS ² : 0.1 - 0.5	n.d.	n.d.	n.d.	n.d.	Nielen <i>et al.</i>	2007	[180]
Miscellaneous	Ruthenium complexes characterization	Ruthenium arene anticancer complexes	RPLC	ESI (+)	9.4	MS	<i>m/z</i> range 90 – 3000, 512 k data points	n.d.	n.d.	n.d.	2	n.d.	2	Wang <i>et al.</i>	2008	[181]
Miscellaneous	Drug impurities identification	Moxifloxacin (drug substance)	RPLC	ESI (+)	n.d.	MS, MS ² , MS ³ (n.d.)	<i>m/z</i> range 100 - 1000	50000	n.d.	n.d.	11	n.d.	11	Wu <i>et al.</i>	2012	[182]

Chapitre 2 : Les couplages entre la FTICR-MS et les méthodes séparatives

Table 2.2: Applications using CE hyphenated to FTICR-MS. (CZE: capillary zone electrophoresis; CIEF: capillary isoelectric focusing; CEC: capillary electrochromatography; tITP: transient isotachophoresis; PMS: pH-mediated stacking; RPLC: reversed-phase liquid chromatography; A: isomers separation, B: quantification, C: high sensitivity; D: high dynamic range, E: easier data interpretation; n.d.: not disclosed)

Domain	Application	Sample	Applications		Instrumental parameters					Advantages brought by the LC hyphenation				References		
			LC mode	Ionization source (polarity)	FTICR magnetic field (T)	MS mode	Other parameters	MS resolution (FWHM)	Mass measurement accuracy (ppm)	Advantages brought by LC hyphenation	Number of detected compounds	Number of unambiguous formula assignments	Number of identifications	Authors	Year	Reference
Proteomics	analysis of intact proteins	mixture of standard proteins	CZE	ESI (+)	7	MS	<i>m/z</i> range 800 – 1800, transient of 0.3 – 0.4 s, 6 s/spectrum	> 30000	n.d.	C	6	n.d.	6	Hofstadler <i>et al.</i>	1993	[16]
Proteomics	Proteins analysis	Mixtures of standard proteins and peptides	CZE	ESI (+)	7	MS, MS ² (SORI-CAD)	5 s/spectrum	50000	n.d.	n.d.	n.d.	n.d.	n.d.	Hofstadler <i>et al.</i>	1994	[70]
Proteomics	Cells analysis	Human erythrocyte single cell	CZE	ESI (+)	7	MS	<i>m/z</i> range ≥ 690, 256 k data points, transient of 0.836 s	≥ 100000	n.d.	n.d.	n.d.	n.d.	n.d.	Hofstadler <i>et al.</i>	1995	[69]
Proteomics	Intact proteins analysis	Mixture of standard peptides and proteins	CZE	ESI (+)	7	MS	64 k data points, 6 s/spectrum	≥ 45000	n.d.	n.d.	n.d.	n.d.	n.d.	Wahl <i>et al.</i>	1995	[183]
Proteomics	Proteins top-down analysis	Mixture of standard proteins	CZE	ESI (+)	6	MS, MS ² (n.d.)	<i>m/z</i> range 500 - >2500	50000 – 60000	n.d.	C	n.d.	n.d.	n.d.	Valaskovic <i>et al.</i>	1996	[184]
Proteomics	Single cell analysis	Human erythrocyte cells	CZE	ESI (+)	7	MS	256 k data points, pulse sequence of 12.6 s	≥ 45000	n.d.	C	n.d.	n.d.	n.d.	Hofstadler <i>et al.</i>	1996	[110]
Proteomics	Intact proteins analysis	Mixture of standard proteins and <i>E. coli</i> cell lysate	CIEF	ESI (+)	7	MS	<i>m/z</i> range 700 – 3000, 128 k – 256 k data points, transients of 0.435 - 0.835 s, accumulation during 0.4 s, 2.24 s/scan	≥ 45000 (256 k)	1 - 5	C, D	112	n.d.	n.d.	Yang <i>et al.</i>	1998	[71]
Proteomics	Proteome top-down analysis	<i>E. coli</i> cell lysate	CIEF	ESI (+)	7	MS, MS ² (SORI CID)	256 k data points, 3.5 s/spectrum	n.d.	n.d.	C, D	1500	n.d.	n.d.	Jensen <i>et al.</i>	1999	[42]
Proteomics	Intact proteins analysis	<i>E. coli</i> cell lysate	CIEF	n.d.	7	MS	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	Pasa-Tolic <i>et al.</i>	1999	[109]
Proteomics	Intact proteins analysis	Mixture of standard proteins and <i>E. coli</i> cell lysate	CIEF	ESI (+)	7	MS	<i>m/z</i> range 150 – 2000, 16 k data points, 3 s/spectrum	n.d.	n.d.	n.d.	100	n.d.	n.d.	Zhang <i>et al.</i>	2000	[108]
Proteomics	Proteome analysis	<i>E. coli</i> cell lysate	CIEF	ESI (+)	7	MS	4 s/spectrum	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	Veenstra <i>et al.</i>	2000	[185]
Proteomics	Analysis of intact protein complexes	Protein complexes mixture	CIEF	ESI (+)	7	MS	4 s/spectrum	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	Martinovic <i>et al.</i>	2000	[112]
Proteomics	Analysis of intact enzymes	Human liver alcohol dehydrogenase isoenzymes	CIEF	ESI (+)	7	MS	4 s/spectrum	n.d.	n.d.	n.d.	n.d.	n.d.	9	Martinovic <i>et al.</i>	2000	[72]
Proteomics	Intact proteins analysis	<i>E. coli</i> and <i>D. radiodurans</i> cell lysates	CIEF	ESI (+)	7	MS	256 k data points, 3.5 s/spectrum	n.d.	n.d.	B, C, D	400 - 1000	n.d.	n.d.	Jensen <i>et al.</i>	2000	[41]

Chapitre 2 : Les couplages entre la FTICR-MS et les méthodes séparatives

Proteomics	Peptide characterization	Mixture of standard peptides and human serum albumine tryptic digest	n.d.	ESI (+)	9.4	MS, MS ² (ECD)	128 k data points, transient of 0.2 s, accumulation during 0.6 - 1.0 s, time of flight < 0.01 s, 0.4 - 0.5 s/scan, 2 - 4 s/spectrum	n.d.	n.d.	E	n.d.	n.d.	n.d.	Tsybin <i>et al.</i>	2002	[103]
Proteomics	Protein identification	Tryptic digests of cerebrospinal fluid samples	CZE	ESI (n.d.)	9.4	n.d.	256 k data points, 4.5 s/spectrum	n.d.	≤ 5	C, D	1497	n.d.	n.d.	Wetterhall <i>et al.</i>	2002	[34]
Proteomics	Identification of intact proteins	Protein extract of <i>E. coli</i> and <i>S. cerevisiae</i>	CIEF	ESI (+)	7	n.d.	4 s/spectrum	n.d.	≤ 10 (resolved peaks), ≤ 100 (unresolved peaks)	n.d.	n.d.	n.d.	n.d.	Martinovic <i>et al.</i>	2002	[107]
Proteomics	Protein identification	Tryptic digest of <i>S. oneidensis</i> cell lysate	tITP/CZE	ESI (n.d.)	7	MS	256 k data points	n.d.	n.d.	C, D	n.d.	n.d.	3071	Mohan <i>et al.</i>	2003	[101]
Proteomics	Characterization of oligosaccharides from glycoproteins	Mixture oligosaccharides from bovin mucin and bile salt-stimulated lipase	CEC	ESI (+)	7	MS, MS ² (SORI-CID)	Infinity ICR Cell, 256 k data points, 1.6 - 1.8 s/scan	30 000	0.00 - 9.13	n.d.	27	n.d.	27	Que <i>et al.</i>	2003	[186]
Proteomics	Identification of biomarkers for renal disease (bottom-up and top-down)	Urine samples	CZE	ESI (+)	9.4	MS, MS ² (ECD, IRMPD)	512 k data points, accumulation during 2 - 30 s, time of flight of 0.012 s	n.d.	3 - 6.5	n.d.	4	n.d.	n.d.	Chalmers <i>et al.</i>	2005	[73]
Proteomics	Analysis of complex biological samples	Tryptic digests of BSA and cerebrospinal fluid samples	RPLC-CZE	ESI (+)	9.4	n.d.	7.5 s/spectrum	n.d.	≤ 10	n.d.	10 - 3126	n.d.	10 - 3126	Bergström <i>et al.</i>	2006	[154]
Proteomics	Polypeptides identification	Polypeptides from rat urine	n.d.	ESI (n.d.)	12	MS	m/z range 300 - 2000, 1 M data points, 0.5 s of accumulation, 5 s/scan	n.d.	≤ 0.5	n.d.	100	n.d.	n.d.	Frommberger <i>et al.</i>	2007	[57]
Proteomics	Characterization of intact glycoproteins	Intact α subunit of recombinant human chorionic gonadotrophin	CZE	ESI (+)	n.d.	MS	m/z range 1400 - 2000, accumulation during 0.2 s	55000 at m/z 1800	n.d.	B	n.d.	n.d.	n.d.	Thakur <i>et al.</i>	2009	[48]
Proteomics	Bottom-up characterization of muropeptide variants	Bacillus licheniformis peptidoglycan	CZE	ESI (+)	7	MS, MS ² (CID), MS ³ (CID, ETD)	m/z range 400 - 2000 or 200 - 2000	50000 at m/z 400	0.01 - 2.26	A	25	n.d.	25	Boulanger <i>et al.</i>	2019	[75]
Proteomics	Middle up and intact analyses of bispecific antibodies	2 bispecific antibodies	CZE	ESI (+)	12	MS	ParaCell, m/z range 202.70 - 3000 or 405.41 - 3000 or 589.68 - 3000, 1 M data points, transient of 0.144 s or 1.25 s or 1.78 s, accumulation during 0.1 s, time of flight of 1.2 ms	140000 at m/z 1100	n.d.	n.d.	18 - 19	n.d.	18 - 19	Gstötter <i>et al.</i>	2020	[59]
Targeted metabolomics	Targeted cationic metabolites profiling	D. vulgaris lysate	PMS-tITP	ESI (+)	9.4	MS	m/z range 65 - 1000, 131 k data points, transient of 0.0328 s, 1 s of accumulation, 1.3 s/spectrum	15000 at m/z 250, 30000 at m/z 125	0.08 - 3.08	A, B	n.d.	n.d.	28	Baidoo <i>et al.</i>	2008	[58]
Lipidomics	Characterization of lipopolysaccharides and lipids	Isolated lipopolysaccharides from <i>Escherichia coli</i> and <i>Salmonella enterica</i>	CZE	ESI (-)	7	MS	512 k data points, accumulation during 2 s, 3.3 s/spectrum	n.d.	0.11 - 5.30	A, C	15	n.d.	n.d.	Hübner <i>et al.</i>	2009	[74]

Chapitre 2 : Les couplages entre la FTICR-MS et les méthodes séparatives

Table 2.3: Applications using GC hyphenated to FTICR-MS. (A: isomers separation, B: quantification, C: high sensitivity; D: high dynamic range, E: easier data interpretation; n.d.: not disclosed).

Applications			Instrumental parameters							Advantages brought by the LC hyphenation				References		
Domain	Application	Sample	Ionization source (polarity)	FTICR magnetic field (T)	MS mode	Other parameters	MS resolution (FWHM)	Mass measurement accuracy (ppm)	Advantages brought by GC hyphenation	Number of detected compounds	Number of unambiguous formula assignments	Number of identifications	Authors	Year	Reference	
Foodomics	Analysis of flame retardants as food contaminants	Food extracts (meat, whole milk, eggs, cereals, cod liver, fish liver oil, soybean meal)	APCI (+)	7	MS	<i>m/z</i> range 100 – 1000, 256 k data points, transient of 0.122 s, accumulation during 0.2 s, time of flight of 0.7 s	40000 at <i>m/z</i> 560	0.23 - 3.81	A, B, C, D	n.d.	n.d.	18	Zacs <i>et al.</i>	2019	[39]	
Environment	Naphthenic acids analysis	Oil sand processed water and groundwater from the Athabasca River region	EI/Cl (n.d.)	9.4	MS	<i>m/z</i> range 75 – 650, transient of 0.524 s, cycle time of 1.5	100000 – 150000	≤ 5	A	1300 – 2200	n.d.	n.d.	Ortiz <i>et al.</i>	2014	[130]	
Environment	Flora VOC analysis	Headspace from pine tree twig samples	EI/SCI (+)	7	MS	Duty cycle of 0.06 – 29 s	140000 - 185000	0.07 - 0.88	A, C	7	n.d.	7	Solouki <i>et al.</i>	2004	[132]	
Environment	Control of maternal transfer of contaminants in European eels	Eggs and gonads extracts from eels	EI (+)	n.d.	MS	<i>m/z</i> range 75 – 650, transient of 0.524 s	100000 – 150000	n.d.	n.d.	n.d.	n.d.	n.d.	Sürhing <i>et al.</i>	2016	[134]	
Environment	Dioxin analysis	Industrial fire fallout exposed vegetation extract	n.d.	9.4	MS	<i>m/z</i> range 250 – 850, transient of 0.262 s, cycle time of 1.5 s	50000 – 100000	≤ 1	n.d.	n.d.	n.d.	n.d.	Taguchi <i>et al.</i>	2010	[56]	
Environment	Petroleopic profiling of soil extracts	Extracts of Salt marsh soil from Staten Island in New York	APCI (+)	7 (2XR)	MS	ParaCell, <i>m/z</i> range 107 - 3000, 2 M data point	300000 at <i>m/z</i> 200	n.d.	C	n.d.	n.d.	n.d.	Thomas <i>et al.</i>	2019	[131]	
Environment/Petroleomics	Characterization of oil sands process water and groundwater compounds	Oil sands process water and groundwater from Athabasca oil sands area	APCI (+)	12	MS	<i>m/z</i> range 98 – 3000, transient of 0.56 s, 1 s/scan	330000 at <i>m/z</i> 200, 220000 at <i>m/z</i> 300	n.d.	A	n.d.	n.d.	n.d.	Barrow <i>et al.</i>	2014	[62]	
Environment/Petroleomics	Data processing	Oil sands process water and groundwater from Athabasca oil sands area	APCI (+)	7 (2XR)	MS (2w)	ParaCell	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	Gavard <i>et al.</i>	2020	[187]	
Petroleomics	Multicomponent characterization of gasoline samples	Gasoline	Internal CI, EI (+)	7	MS	128 k - 512 k data points	n.d.	< 6	A, C	≥ 50	≥ 25	≥ 5	Luo <i>et al.</i>	2009	[35]	
Petroleomics	Characterization of pyrolysis oil	Pyrolysis oil from German brown coal	APCI (-)	15	MS	<i>m/z</i> range 46 – 500, 2 M data points, accumulation during 0.3 s, 1.003 s/spectrum	n.d.	0.021 - 0.427	A	n.d.	n.d.	32	Zuber <i>et al.</i>	2016	[63]	
Petroleomics	Polyaromatic hydrocarbons (PAH) analysis	Fossil oils (organics in shale oil, petroleum crude oil, and heavy sweet crude oil)	API (+)	7	MS	<i>m/z</i> range 100 – 900, 2 M data points, transient of 1 s, accumulation during 0.1 s	264000 at <i>m/z</i> 400	≤ 0.5	A, C	1655-1734	n.d.	n.d.	Benigni <i>et al.</i>	2016	[65]	
Petroleomics	Fingerprinting of complex mixtures	Automobile gasoline and aviation jet fuel samples	Internal EI (+)	7	MS	Cylindrical ICR cell, 128 k data points	30000 at <i>m/z</i> 58	0.0 - 18.1	A	n.d.	37	n.d.	Szulejko <i>et al.</i>	2002	[55]	
Miscellaneous	Low-mass ions detection	Grob sample (mixture of 12 organic compounds)	EI (n.d.)	9.4	n.d.	256 k data point	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	Zekavat <i>et al.</i>	2014	[188]	

2.1.8. Conclusion

The purpose of this review is to provide an overview of the hyphenation of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR MS) with separation methods to date. As with all the coupling between separation techniques and mass spectrometry, the question is essentially one of compromises. This is even truer for FT-ICR MS which coupling with GC, LC or LC may appear counter-intuitive and scarcely compatible at the first sight. If the relentless progress has suppressed some apparent technical limitations (e.g. vacuum compatibilities with ion sources), coupling separation techniques with FT-ICR MS still means some compromises and is not as straightforward as with other analyzers (such as QTOF-MS for instance). In a few words, FT-ICR MS needs time (ion transfer, accumulation and especially acquisition times) to reach its highest resolving power and mass measurement accuracy capabilities whereas separation methods have no time to lose to remain effective (sharp peaks for an optimal chromatographic resolving power). Fortunately, in many applications, the highest performances of both techniques are not required to achieve the level of success envisioned, the strength of the first one balancing the weakness of the second (and vice versa). “Omics” is undoubtedly the main application area of FT-ICR MS combined to online separation techniques. This can be easily understood by the huge complexity of the explored samples in this field, particularly in untargeted studies. FT-ICR MS enables to obtain the required high mass measurement accuracy to determine accurate molecular formulae and resolution for isobar distinction, while separation techniques highlight isomers. However, the separation of isomers is not the only significant advantage of online chromatography or capillary electrophoresis to FT-ICR MS. They also help to reduce ion suppression effects, enable to extend the dynamic range of detection (detection of minor compounds and quantification) and reduce the complexity of mass spectra acquired along the run, enabling easier and deeper data interpretation.

2.2. Références

- [1] Comisarow, M. B. *et al.*, “Fourier Transform Ion Cyclotron Resonance Spectroscopy”, *Chem. Phys. Lett.*, **1974**, 25 (2), 282–283.
- [2] Nikolaev, E. N. *et al.*, “Initial Experimental Characterization of a New Ultra-High Resolution FTICR Cell with Dynamic Harmonization”, *J. Am. Soc. Mass Spectrom.*, **2011**, 22 (7), 1125–1133.
- [3] Bogdanov, B. *et al.*, “Proteomics by FTICR Mass Spectrometry: Top down and Bottom Up”, *Mass Spectrom. Rev.*, **2005**, 24 (2), 168–200.
- [4] Maia, M. *et al.*, “FT-ICR-MS-Based Metabolomics: A Deep Dive into Plant Metabolism”, *Mass Spectrometry Reviews*, **2021**, 1–22.
- [5] Rychlik, M. *et al.*, Chapter 21 - Foodomics Assessed by Fourier Transform Mass Spectrometry. In *Fundamentals and Applications of Fourier Transform Mass Spectrometry*; Kanawati, B. *et al.*, Eds.; Elsevier, 2019; pp 651–677.
- [6] Rodgers, R. P. *et al.*, Petroleomics: Advanced Characterization of Petroleum-Derived Materials by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR

- MS). In *Asphaltenes, Heavy Oils, and Petroleomics*; Mullins, O. C. et al., Eds.; Springer: New York, NY, 2007; pp 63–93.
- [7] Kujawinski, E. B., “Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (ESI FT-ICR MS): Characterization of Complex Environmental Mixtures”, *Environmental Forensics*, **2002**, 3 (3–4), 207–216.
 - [8] Buck, A. et al., “High-Resolution MALDI-FT-ICR MS Imaging for the Analysis of Metabolites from Formalin-Fixed, Paraffin-Embedded Clinical Tissue Samples”, *The Journal of Pathology*, **2015**, 237 (1), 123–132.
 - [9] Lohse, M. et al., “Direct Imaging of Plant Metabolites in the Rhizosphere Using Laser Desorption Ionization Ultra-High Resolution Mass Spectrometry”, *Front Plant Sci*, **2021**, 12, 753812.
 - [10] Marshall, A. G. et al., “Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer”, *Mass Spectrometry Reviews*, **1998**, 17 (1), 1–35.
 - [11] Ledford, E. B. et al., “Coupling of Capillary Gas Chromatograph and Fourier Transform Mass Spectrometer”, *Anal. Chem.*, **1980**, 52, 2450–2451.
 - [12] Sack, T. M. et al., “Pulsed Valve Interface for Gas Chromatography/Fourier Transform Mass Spectrometry”, *Anal. Chem.*, **1983**, 55 (14), 2419–2421.
 - [13] Stockton, G. W. et al., An External Ion Source FT-ICR Mass Spectrometer for Ultrahigh-mass Accuracy Measurements and High-performance Liquid Chromatographic Sample Introduction. In *FT-ICR/MS: Analytical Applications of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry*; Asamoto, B., Ed.; VCH: New York, 1991.
 - [14] Stacey, C. C. et al., “Reverse-Phase Liquid Chromatography/Electrospray-Ionization Fourier-Transform Mass Spectrometry in the Analysis of Peptides”, *Rapid Communications in Mass Spectrometry*, **1994**, 8 (7), 513–516.
 - [15] Schrader, W. et al., “Liquid Chromatography/Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (LC-FTICR MS): An Early Overview”, *Anal Bioanal Chem*, **2004**, 379 (7), 1013–1024.
 - [16] Hofstadler, S. A. et al., “On-Line Capillary Electrophoresis with Fourier Transform Ion Cyclotron Resonance Mass Spectrometry”, *J. Am. Chem. Soc.*, **1993**, 115 (15), 6983–6984.
 - [17] Hertzog, J. et al., “Characterization of Heavy Products from Lignocellulosic Biomass Pyrolysis by Chromatography and Fourier Transform Mass Spectrometry: A Review”, *Energy Fuels*, **2021**, 35 (22), 17979–18007.
 - [18] Ohta, D. et al., “Application of Fourier-Transform Ion Cyclotron Resonance Mass Spectrometry to Metabolic Profiling and Metabolite Identification”, *Current Opinion in Biotechnology*, **2010**, 21 (1), 35–44.
 - [19] Boldin, I. A. et al., “Fourier Transform Ion Cyclotron Resonance Cell with Dynamic Harmonization of the Electric Field in the Whole Volume by Shaping of the Excitation and Detection Electrode Assembly”, *Rapid Communications in Mass Spectrometry*, **2011**, 25 (1), 122–126.
 - [20] Amorim Madeira, P. J. et al., High Resolution Mass Spectrometry Using FTICR and Orbitrap Instruments. In *Fourier Transform - Materials Analysis*; Salih, S., Ed.; InTech, 2012.
 - [21] Nikolaev, E. N. et al., Chapter 4 - Fundamentals and Simulations in FT-ICR-MS. In *Fundamentals and Applications of Fourier Transform Mass Spectrometry*; Kanawati, B. et al., Eds.; Elsevier, 2019; pp 89–111.

- [22] Xian, F. *et al.*, "High Resolution Mass Spectrometry", *Anal. Chem.*, **2012**, 84 (2), 708–719.
- [23] Kelleher, N. L. *et al.*, "Unit Resolution Mass Spectra of 112 kDa Molecules with 3 Da Accuracy", *Journal of the American Society for Mass Spectrometry*, **1997**, 8 (4), 380–383.
- [24] N. Nikolaev, E. *et al.*, "From Supercomputer Modeling to Highest Mass Resolution in FT-ICR", *Mass Spectrom (Tokyo)*, **2013**, 2 (Spec Iss), S0010.
- [25] Huang, D. *et al.*, "Comparison of High-Resolution Fourier Transform Mass Spectrometry Platforms for Putative Metabolite Annotation", *Anal. Chem.*, **2021**.
- [26] Moseley, H. N., "Correcting for the Effects of Natural Abundance in Stable Isotope Resolved Metabolomics Experiments Involving Ultra-High Resolution Mass Spectrometry", *BMC Bioinformatics*, **2010**, 11 (1), 139.
- [27] Stingl, C. *et al.*, "Application of Different Fragmentation Techniques for the Analysis of Phosphopeptides Using a Hybrid Linear Ion Trap-FTICR Mass Spectrometer", *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, **2006**, 1764 (12), 1842–1852.
- [28] Guillemant, J. *et al.*, "Towards a New Pseudo-Quantitative Approach to Evaluate the Ionization Response of Nitrogen Compounds in Complex Matrices", *Sci Rep*, **2021**, 11 (1), 6417.
- [29] Fauland, A. *et al.*, "A Comprehensive Method for Lipid Profiling by Liquid Chromatography-Ion Cyclotron Resonance Mass Spectrometry", *J. Lipid Res.*, **2011**, 52 (12), 2314–2322.
- [30] Cataldi, T. R. I. *et al.*, "Profiling of N-Acyl-Homoserine Lactones by Liquid Chromatography Coupled with Electrospray Ionization and a Hybrid Quadrupole Linear Ion-Trap and Fourier-Transform Ion-Cyclotron-Resonance Mass Spectrometry (LC-ESI-LTQ-FTICR-MS)", *J. Mass Spectrom.*, **2007**, 43 (1), 82–96.
- [31] Rowland, S. M. *et al.*, "Online Coupling of Liquid Chromatography with Fourier Transform Ion Cyclotron Resonance Mass Spectrometry at 21 T Provides Fast and Unique Insight into Crude Oil Composition", *Anal. Chem.*, **2021**, 93 (41), 13749–13754.
- [32] Boiteau, R. M. *et al.*, "Siderophore Profiling of Co-Habitating Soil Bacteria by Ultra-High Resolution Mass Spectrometry", *Metalomics*, **2019**, 11 (1), 166–175.
- [33] Schmid, D. G. *et al.*, "High-Resolution Analysis of a 144-Membered Pyrazole Library from Combinatorial Solid Phase Synthesis by Using Electrospray Ionisation Fourier Transform Ion Cyclotron Resonance Mass Spectrometry", *Rapid Communications in Mass Spectrometry*, **2001**, 15 (5), 341–347.
- [34] Wetterhall, M. *et al.*, "Rapid Analysis of Tryptically Digested Cerebrospinal Fluid Using Capillary Electrophoresis-Electrospray Ionization-Fourier Transform Ion Cyclotron Resonance-Mass Spectrometry", *J. Proteome Res.*, **2002**, 1 (4), 361–366.
- [35] Luo, Z. *et al.*, "Multidimensional GC-Fourier Transform Ion Cyclotron Resonance MS Analyses: Utilizing Gas-Phase Basicities to Characterize Multicomponent Gasoline Samples", *Journal of Chromatographic Science*, **2009**, 47 (1), 75–82.
- [36] Iijima, Y. *et al.*, "Metabolite Annotations Based on the Integration of Mass Spectral Information", *The Plant Journal*, **2008**, 54 (5), 949–962.
- [37] Kong, H. *et al.*, "Bioconversion of Red Ginseng Saponins in the Gastro-Intestinal Tract in Vitro Model Studied by High-Performance Liquid Chromatography-High Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry", *Journal of Chromatography A*, **2009**, 1216 (11), 2195–2203.

- [38] Sun, S. *et al.*, "Studies on the Metabolism of Paeoniflorin in Human Intestinal Microflora by High Performance Liquid Chromatography/Electrospray Ionization/Fourier Transform Ion Cyclotron Resonance Mass Spectrometry and Quadrupole Time-of-Flight Mass Spectrometry", *Journal of Chromatography B*, **2018**, 1085, 63–71.
- [39] Zacs, D. *et al.*, "Evaluation of Analytical Performance of Gas Chromatography Coupled with Atmospheric Pressure Chemical Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (GC-APCI-FT-ICR-MS) in the Target and Non-Targeted Analysis of Brominated and Chlorinated Flame Retardants in Food", *Chemosphere*, **2019**, 225, 368–377.
- [40] Boldin, I. A. *et al.*, "Theory of Peak Coalescence in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry", *Rapid Commun Mass Spectrom*, **2009**, 23 (19), 3213–3219.
- [41] Jensen, P. K. *et al.*, "Mass Spectrometric Detection for Capillary Isoelectric Focusing Separations of Complex Protein Mixtures", *Electrophoresis*, **2000**, 21, 1372–1380.
- [42] Jensen, P. K. *et al.*, "Probing Proteomes Using Capillary Isoelectric Focusing-Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry", *Anal. Chem.*, **1999**, 71 (11), 2076–2084.
- [43] Shen, Y. *et al.*, "Packed Capillary Reversed-Phase Liquid Chromatography with High-Performance Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Proteomics", *Anal. Chem.*, **2001**, 73 (8), 1766–1775.
- [44] Shen, Y. *et al.*, "High-Throughput Proteomics Using High-Efficiency Multiple-Capillary Liquid Chromatography with On-Line High-Performance ESI FTICR Mass Spectrometry", *Anal. Chem.*, **2001**, 73 (13), 3011–3021.
- [45] Adamson, J. T. *et al.*, Chapter 14 - Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Lectin Analysis. In *Lectins*; Nilsson, C. L., Ed.; Elsevier Science B.V.: Amsterdam, 2007; pp 343–371.
- [46] Gao, H. *et al.*, "Two-Dimensional Electrophoretic/Chromatographic Separations Combined with Electrospray Ionization FTICR Mass Spectrometry for High Throughput Proteome Analysis", *Journal of Microcolumn Separations*, **2000**, 12 (7), 383–390.
- [47] Lelario, F. *et al.*, "A Complete Survey of Glycoalkaloids Using LC-FTICR-MS and IRMPD in a Commercial Variety and a Local Landrace of Eggplant (*Solanum Melongena L.*) and Their Anticholinesterase and Antioxidant Activities", *Toxins*, **2019**, 11 (4), 230.
- [48] Thakur, D. *et al.*, "Profiling the Glycoforms of the Intact α Subunit of Recombinant Human Chorionic Gonadotropin by High-Resolution Capillary Electrophoresis-Mass Spectrometry", *Anal. Chem.*, **2009**, 81 (21), 8900–8907.
- [49] Qi, Y. *et al.*, "Online Liquid Chromatography and FT-ICR MS Enable Advanced Separation and Profiling of Organosulfates in Dissolved Organic Matter", *ACS EST Water*, **2021**, 1 (8), 1975–1982.
- [50] Giavalisco, P. *et al.*, "13C Isotope-Labeled Metabolomes Allowing for Improved Compound Annotation and Relative Quantification in Liquid Chromatography-Mass Spectrometry-Based Metabolomic Research", *Anal. Chem.*, **2009**, 81 (15), 6546–6551.
- [51] Karas, M. *et al.*, "Nano-Electrospray Ionization Mass Spectrometry: Addressing Analytical Problems beyond Routine", *Fresenius' Journal of Analytical Chemistry*, **2000**, 366 (6–7), 669–676.
- [52] Carlsohn, E. *et al.*, "Characterization of the Outer Membrane Protein Profile from Disease-Related Helicobacter Pylori Isolates by Subcellular Fractionation and Nano-LC FT-ICR MS Analysis", *J. Proteome Res.*, **2006**, 5 (11), 3197–3204.

- [53] Lelario, F. *et al.*, "Establishing the Occurrence of Major and Minor Glucosinolates in Brassicaceae by LC–ESI-Hybrid Linear Ion-Trap and Fourier-Transform Ion Cyclotron Resonance Mass Spectrometry", *Phytochemistry*, **2012**, 73, 74–83.
- [54] Sandron, S. *et al.*, "Fractionation of Dissolved Organic Matter on Coupled Reversed-Phase Monolithic Columns and Characterisation Using Reversed-Phase Liquid Chromatography-High Resolution Mass Spectrometry", *Chromatographia*, **2018**, 81 (2), 203–213.
- [55] Szulejko, J. E. *et al.*, "Potential Analytical Applications of Interfacing a GC to an FT-ICR MS: Fingerprinting Complex Sample Matrixes", *Anal. Chem.*, **2002**, 74 (14), 3434–3442.
- [56] Taguchi, V. Y. *et al.*, "Dioxin Analysis by Gas Chromatography-Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (GC-FTICRMS)", *J Am Soc Mass Spectrom.*, **2010**, 21 (11), 1918–1921.
- [57] Frommberger, M. *et al.*, "Peptidomic Analysis of Rat Urine Using Capillary Electrophoresis Coupled to Mass Spectrometry", *Prot. Clin. Appl.*, **2007**, 1 (7), 650–660.
- [58] Baidoo, E. E. K. *et al.*, "Capillary Electrophoresis-Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for the Identification of Cationic Metabolites via a PH-Mediated Stacking-Transient Isotachophoretic Method", *Anal. Chem.*, **2008**, 80, 3112–3122.
- [59] Gstöttner, C. *et al.*, "Intact and Subunit-Specific Analysis of Bispecific Antibodies by Sheathless CE-MS", *Analytica Chimica Acta*, **2020**, 1134, 18–27.
- [60] Baykut, D. *et al.*, "Molecular Differentiation of Ischemic and Valvular Heart Disease by Liquid Chromatography/Fourier Transform Ion Cyclotron Resonance Mass Spectrometry", *Eur J Med Res*, **2006**, 11 (6), 221–226.
- [61] Szulejko, J. E. *et al.*, "Simultaneous Determination of Analyte Concentrations, Gas-Phase Basicities, and Proton Transfer Kinetics Using Gas Chromatography/Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (GC/FT-ICR MS)", *International Journal of Mass Spectrometry*, **2006**, 257 (1–3), 16–26.
- [62] Barrow, M. P. *et al.*, "An Added Dimension: GC Atmospheric Pressure Chemical Ionization FTICR MS and the Athabasca Oil Sands", *Anal. Chem.*, **2014**, 86 (16), 8281–8288.
- [63] Zuber, J. *et al.*, "Gas Chromatography/Atmospheric Pressure Chemical Ionization-Fourier Transform Ion Cyclotron Resonance Mass Spectrometry of Pyrolysis Oil from German Brown Coal", *International Journal of Analytical Chemistry*, **2016**, 2016, 1–12.
- [64] Constapel, M. *et al.*, "Atmospheric-Pressure Laser Ionization: A Novel Ionization Method for Liquid Chromatography/Mass Spectrometry", *Rapid Communications in Mass Spectrometry*, **2005**, 19 (3), 326–336.
- [65] Benigni, P. *et al.*, "Increasing Polyaromatic Hydrocarbon (PAH) Molecular Coverage during Fossil Oil Analysis by Combining Gas Chromatography and Atmospheric-Pressure Laser Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FT-ICR MS)", *Energy Fuels*, **2016**, 30 (1), 196–203.
- [66] Sommella, E. *et al.*, "Fast Profiling of Natural Pigments in Different Spirulina (Arthrospira Platensis) Dietary Supplements by DI-FT-ICR and Evaluation of Their Antioxidant Potential by Pre-Column DPPH-UHPLC Assay", *Molecules*, **2018**, 23 (5), 1152.
- [67] Bataineh, M. *et al.*, "HPLC/APCI-FTICR-MS as a Tool for Identification of Partial Polar Mutagenic Compounds in Effect-Directed Analysis", *J Am Soc Mass Spectrom.*, **2010**, 21 (6), 1016–1027.

- [68] Lababidi, S. et al., "Online Normal-Phase High-Performance Liquid Chromatography/Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: Effects of Different Ionization Methods on the Characterization of Highly Complex Crude Oil Mixtures: Normal-Phase HPLC/FT-ICR MS of Crude Oil", *Rapid Commun. Mass Spectrom.*, **2014**, 28 (12), 1345–1352.
- [69] Hofstadler, S. A. et al., "Capillary Electrophoresis-Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Direct Analysis of Cellular Proteins", *Anal. Chem.*, **1995**, 67, 1477–1480.
- [70] Hofstadler, S. A. et al., "Capillary Electrophoresis Fourier Transform Ion Cyclotron Resonance Mass Spectrometry with Sustained Off-Resonance Irradiation for the Characterization of Protein and Peptide Mixtures", *J. Am. Soc. Mass Spectrom.*, **1994**, 5, 894–899.
- [71] Yang, L. et al., "Capillary Isoelectric Focusing Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Protein Characterization", *Anal. Chem.*, **1998**, 70, 3235–3241.
- [72] Martinovic, S. et al., "Characterization of Human Alcohol Dehydrogenase Isoenzymes by Capillary Isoelectric Focusing-Mass Spectrometry", *Electrophoresis*, **2000**, 21, 2368–2375.
- [73] Chalmers, M. J. et al., "Combined Top-Down and Bottom-Up Mass Spectrometric Approach to Characterization of Biomarkers for Renal Disease", *Analytical Chemistry*, **2005**, 77, 7163–7171.
- [74] Hübner, G. et al., "Separation of R-Form Lipopolysaccharide and Lipid A by CE-Fourier-Transform Ion Cyclotron Resonance MS", *Electrophoresis*, **2009**, 30, 1808–1816.
- [75] Boulanger, M. et al., "Bacillus Licheniformis Peptidoglycan Characterization by CZE-MS: Assessment with the Benchmark RP-HPLC-MS Method", *ELECTROPHORESIS*, **2019**, 40 (20), 2672–2682.
- [76] Pleasance, S. et al., "Comparison of Liquid-Junction and Coaxial Interfaces for Capillary Electrophoresis-Mass Spectrometry with Application to Compounds of Concern to the Aquaculture Industry", *Journal of Chromatography A*, **1992**, 591 (1), 325–339.
- [77] Bezy, V. et al., "Analysis and Validation of the Phosphorylated Metabolites of Two Anti-Human Immunodeficiency Virus Nucleotides (Stavudine and Didanosine) by Pressure-Assisted CE-ESI-MS/MS in Cell Extracts: Sensitivity Enhancement by the Use of Perfluorinated Acids and Alcohols as Coaxial Sheath-Liquid Make-up Constituents", *ELECTROPHORESIS*, **2006**, 27 (12), 2464–2476.
- [78] Whitt, J. T. et al., "Capillary Electrophoresis to Mass Spectrometry Interface Using a Porous Junction", *Anal. Chem.*, **2003**, 75 (9), 2188–2191.
- [79] Hendrickson, C. L. et al., "21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: A National Resource for Ultrahigh Resolution Mass Analysis", *J. Am. Soc. Mass Spectrom.*, **2015**, 26 (9), 1626–1632.
- [80] Marshall, A. G. et al., "Relaxation and Spectral Line Shape in Fourier Transform Ion Resonance Spectroscopy", *J. Chem. Phys.*, **1979**, 71 (11), 4434–4444.
- [81] Marshall, A. G., "Theoretical Signal-to-Noise Ratio and Mass Resolution in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry", *Anal. Chem.*, **1979**, 51 (11), 1710–1714.
- [82] Wahab, M. F. et al., "Sampling Frequency, Response Times and Embedded Signal Filtration in Fast, High Efficiency Liquid Chromatography: A Tutorial", *Anal Chim Acta*, **2016**, 907, 31–44.

- [83] Marshall, A. G. *et al.*, “40 Years of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry”, *International Journal of Mass Spectrometry*, **2015**, 377, 410–420.
- [84] Wanczek, K. P. *et al.*, Chapter 1 - Historical Developments in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. In *Fundamentals and Applications of Fourier Transform Mass Spectrometry*; Kanawati, B. *et al.*, Eds.; Elsevier, 2019; pp 3–33.
- [85] Shaw, J. B. *et al.*, “21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer Greatly Expands Mass Spectrometry Toolbox”, *J. Am. Soc. Mass Spectrom.*, **2016**, 27 (12), 1929–1936.
- [86] Smith, D. F. *et al.*, “21 Tesla FT-ICR Mass Spectrometer for Ultrahigh-Resolution Analysis of Complex Organic Mixtures”, *Anal. Chem.*, **2018**, 90 (3), 2041–2047.
- [87] Kostyukovich, Y. I. *et al.*, “Dynamically Harmonized FT-ICR Cell with Specially Shaped Electrodes for Compensation of Inhomogeneity of the Magnetic Field. Computer Simulations of the Electric Field and Ion Motion Dynamics”, *J. Am. Soc. Mass Spectrom.*, **2012**, 23 (12), 2198–2207.
- [88] Cho, E. *et al.*, “Application of FT-ICR MS Equipped with Quadrupole Detection for Analysis of Crude Oil”, *Anal. Chem.*, **2017**, 89 (22), 12101–12107.
- [89] Driver, J. A. *et al.*, “Simulations of Nw Measurement Using Multiple Detection Electrodes in FTICR Mass Spectrometry”, *International Journal of Mass Spectrometry*, **2020**, 455, 116372.
- [90] Easterling, M. L. *et al.*, Chapter 3 - Fundamentals, Strengths, and Future Directions for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. In *Fundamentals and Applications of Fourier Transform Mass Spectrometry*; Kanawati, B. *et al.*, Eds.; Elsevier, 2019; pp 63–88.
- [91] Tose, L. V. *et al.*, “Following de Novo Triglyceride Dynamics in Ovaries of Aedes Aegypti during the Previtellogenic Stage”, *Sci Rep*, **2021**, 11 (1), 9636.
- [92] Shaw, J. B. *et al.*, “High Speed Intact Protein Characterization Using 4X Frequency Multiplication, Ion Trap Harmonization, and 21 Tesla FTICR-MS”, *Anal. Chem.*, **2018**, 90 (9), 5557–5562.
- [93] Hu, P. *et al.*, “Routine Absorption Mode FTMS Data Display with an Ethoxylated Anionic Detergent as a Dual-Role (Mass and Phase) Calibrant”, *J. Am. Soc. Mass Spectrom.*, **2019**, 30 (3), 468–475.
- [94] Irungu, J. *et al.*, “Comparison of HPLC/ESI-FTICR MS versus MALDI-TOF/TOF MS for Glycopeptide Analysis of a Highly Glycosylated HIV Envelope Glycoprotein”, *J Am Soc Mass Spectrom*, **2008**, 19 (8), 1209–1220.
- [95] Xin, L. *et al.*, “Online Effective Identification of Glycopeptide Using Liquid Chromatography Combined with Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FTICR-MS)*”, *Advances in Bioscience and Biotechnology*, **2015**, 6, 624–628.
- [96] Wang, X. *et al.*, “Liquid Chromatography Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometric Characterization of N -Linked Glycans and Glycopeptides”, *Anal. Chem.*, **2010**, 82 (15), 6542–6548.
- [97] Ramström, M. *et al.*, “Protein Identification in Cerebrospinal Fluid Using Packed Capillary Liquid Chromatography Fourier Transform Ion Cyclotron Resonance Mass Spectrometry”, *Proteomics*, **2003**, 3 (2), 184–190.
- [98] Pounds, J. G. *et al.*, “Characterization of the Mouse Bronchoalveolar Lavage Proteome by Micro-Capillary LC-FTICR Mass Spectrometry”, *Journal of Chromatography B*, **2008**, 864 (1–2), 95–101.

- [99] Kang, H. *et al.*, “Targeted Tandem Mass Spectrometry for High-Throughput Comparative Proteomics Employing NanoLC-FTICR MS with External Ion Dissociation”, *J Am Soc Mass Spectrom.*, **2007**, 18 (7), 1332–1343.
- [100] Belov, M. E. *et al.*, “An Automated High Performance Capillary Liquid Chromatography-Fourier Transform Ion Cyclotron Resonance Mass Spectrometer for High-Throughput Proteomics”, *J Am Soc Mass Spectrom.*, **2004**, 15 (2), 212–232.
- [101] Mohan, D. *et al.*, “Integration of Electrokinetic-Based Multidimensional Separation/Concentration Platform with Electrospray Ionization-Fourier Transform Ion Cyclotron Resonance-Mass Spectrometry for Proteome Analysis of *Shewanella Oneidensis*”, *Anal. Chem.*, **2003**, 75 (17), 4432–4440.
- [102] Lipton, M. S. *et al.*, “Global Analysis of the *Deinococcus Radiodurans* Proteome by Using Accurate Mass Tags”, *Proceedings of the National Academy of Sciences*, **2002**, 99 (17), 11049–11054.
- [103] Tsybin, Y. O. *et al.*, “Capillary Electrophoresis and Electron Capture Dissociation Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Peptide Mixture and Protein Digest Analysis”, *Eur J Mass Spectrom (Chichester)*, **2002**, 8 (5), 389–395.
- [104] Li, L. *et al.*, “High-Throughput Peptide Identification from Protein Digests Using Data-Dependent Multiplexed Tandem FTICR Mass Spectrometry Coupled with Capillary Liquid Chromatography”, *Anal. Chem.*, **2001**, 73 (14), 3312–3322.
- [105] Lippold, S. *et al.*, “Glycoform-Resolved Fc_YRIIIa Affinity Chromatography–Mass Spectrometry”, *mAbs*, **2019**, 11 (7), 1191–1196.
- [106] Lee, S.-W. *et al.*, “Direct Mass Spectrometric Analysis of Intact Proteins of the Yeast Large Ribosomal Subunit Using Capillary LC/FTICR”, *Proceedings of the National Academy of Sciences*, **2002**, 99 (9), 5942–5947.
- [107] Martinović, S. *et al.*, “Selective Incorporation of Isotopically Labeled Amino Acids for Identification of Intact Proteins on a Proteome-Wide Level”, *J Mass Spectrom.*, **2002**, 37 (1), 99–107.
- [108] Zhang, C.-X. *et al.*, “Stepwise Mobilization of Focused Proteins in Capillary Isoelectric Focusing Mass Spectrometry”, *Anal. Chem.*, **2000**, 72, 1462–1468.
- [109] Pasa-Tolic, L. *et al.*, “High Throughput Proteome-Wide Precision Measurements of Protein Expression Using Mass Spectrometry”, *J. Am. Chem. Soc.*, **1999**, 121, 7949–7950.
- [110] Hofstadler, S. A. *et al.*, “Analysis of Single Cells with Capillary Electrophoresis Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry”, *Rapid Communications in Mass Spectrometry*, **1996**, 10, 919–922.
- [111] Marshall, A. G. *et al.*, “Protein Molecular Mass to 1 Da by ¹³C, ¹⁵N Double-Depletion and FT-ICR Mass Spectrometry”, *J. Am. Chem. Soc.*, **1997**, 119 (2), 433–434.
- [112] Martinovic, S. *et al.*, “Separation and Detection of Intact Noncovalent Protein Complexes from Mixtures by On-Line Capillary Isoelectric Focusing-Mass Spectrometry”, *Analytical Chemistry*, **2000**, 72, 5356–5360.
- [113] Jin, Y. *et al.*, “Comprehensive Characterization of Monoclonal Antibody by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry”, *mAbs*, **2019**, 11 (1), 106–115.
- [114] He, L. *et al.*, “Analysis of Monoclonal Antibodies in Human Serum as a Model for Clinical Monoclonal Gammopathy by Use of 21 Tesla FT-ICR Top-Down and Middle-Down MS/MS”, *J. Am. Soc. Mass Spectrom.*, **2017**, 28 (5), 827–838.

- [115] Anderson, L. C. *et al.*, “Identification and Characterization of Human Proteoforms by Top-Down LC-21 Tesla FT-ICR Mass Spectrometry”, *J. Proteome Res.*, **2017**, *16* (2), 1087–1096.
- [116] Brinkmalm, G., “An Online Nano-LC-ESI-FTICR-MS Method for Comprehensive Characterization Of Endogenous Fragments from Amyloid and Amyloid Precursor Protein in Human and Cat Cerebrospinal Fluid”, *Journal of Mass Spectrometry*, **2012**, *47*, 591–603.
- [117] Chen, B. *et al.*, “Online Hydrophobic Interaction Chromatography–Mass Spectrometry for the Analysis of Intact Monoclonal Antibodies”, *Anal. Chem.*, **2018**, *90* (12), 7135–7138.
- [118] Collier, T. S. *et al.*, “Top-Down Identification and Quantification of Stable Isotope Labeled Proteins from *Aspergillus Flavus* Using Online Nano-Flow Reversed-Phase Liquid Chromatography Coupled to a LTQ-FTICR Mass Spectrometer”, *Anal. Chem.*, **2008**, *80* (13), 4994–5001.
- [119] Tipton, J. D. *et al.*, “Nano-LC FT-ICR Tandem Mass Spectrometry for Top-Down Proteomics: Routine Baseline Unit Mass Resolution of Whole Cell Lysate Proteins up to 72 kDa”, *Anal. Chem.*, **2012**, *84* (5), 2111–2117.
- [120] Lourette, N. *et al.*, “A Top-down LC-FTICR MS-Based Strategy for Characterizing Oxidized Calmodulin in Activated Macrophages”, *J Am Soc Mass Spectrom*, **2010**, *21* (6), 930–939.
- [121] Lakshmanan, R. *et al.*, “Top-down Protein Identification of Proteasome Proteins with NanoLC-FT-ICR-MS Employing Data-Independent Fragmentation Methods”, *Proteomics*, **2014**, *14* (10), 1271–1282.
- [122] Nakabayashi, R. *et al.*, “Combination of Liquid Chromatography–Fourier Transform Ion Cyclotron Resonance–Mass Spectrometry with ¹³C-Labeling for Chemical Assignment of Sulfur-Containing Metabolites in Onion Bulbs”, *Anal. Chem.*, **2013**, *85* (3), 1310–1315.
- [123] Ren, Y. *et al.*, “Identification of the Metabolites after Oral Administration of Extract of Ziziphi Spinosae Semen to Rats or Dogs by High-Performance Liquid Chromatography/Linear Ion Trap FTICR Hybrid Mass Spectrometry: Metabolites of Ziziphi Spinosae Semen in Rats or Dogs”, *Biomed. Chromatogr.*, **2013**, *27* (1), 17–26.
- [124] Guan, J. *et al.*, “Metabolic Profile of 2-(2-Hydroxypropanamido) Benzoic Acid in Rats by Ultra High Performance Liquid Chromatography Combined with Fourier Transform Ion Cyclotron Resonance Mass Spectrometry”, *Journal of Chromatography B*, **2015**, 993–994, 60–68.
- [125] Pollier, J. *et al.*, “Metabolite Profiling of Triterpene Saponins in *Medicago Truncatula* Hairy Roots by Liquid Chromatography Fourier Transform Ion Cyclotron Resonance Mass Spectrometry”, *J. Nat. Prod.*, **2011**, *74* (6), 1462–1476.
- [126] Bianco, G. *et al.*, “Identification of Glucosinolates in Capers by LC-ESI-Hybrid Linear Ion Trap with Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (LC-ESI-LTQ-FTICR MS) and Infrared Multiphoton Dissociation”, *Journal of Mass Spectrometry*, **2012**, *47* (9), 1160–1169.
- [127] Rockenbach, I. I. *et al.*, “Characterization of Flavan-3-Ols in Seeds of Grape Pomace by CE, HPLC-DAD-MSn and LC-ESI-FTICR-MS”, *Food Research International*, **2012**, *48* (2), 848–855.
- [128] Nakabayashi, R. *et al.*, “Boosting Sensitivity in Liquid Chromatography–Fourier Transform Ion Cyclotron Resonance–Tandem Mass Spectrometry for Product Ion Analysis of Monoterpene Indole Alkaloids”, *Front. Plant Sci.*, **2015**, *6*.

- [129] Kim, D. *et al.*, "Application of Online Liquid Chromatography 7 T FT-ICR Mass Spectrometer Equipped with Quadrupolar Detection for Analysis of Natural Organic Matter", *Analytical Chemistry*, **2019**.
- [130] Ortiz, X. *et al.*, "Characterization of Naphthenic Acids by Gas Chromatography-Fourier Transform Ion Cyclotron Resonance Mass Spectrometry", *Anal. Chem.*, **2014**, 86 (15), 7666–7673.
- [131] Thomas, M. J. *et al.*, "Petroleum Depth Profiling of Staten Island Salt Marsh Soil: 2ω Detection FTICR MS Offers a New Solution for the Analysis of Environmental Contaminants", *Science of The Total Environment*, **2019**, 662, 852–862.
- [132] Solouki, T. *et al.*, "A Preconcentrator Coupled to a GC/FTMS: Advantages of Self-Chemical Ionization, Mass Measurement Accuracy, and High Mass Resolving Power for GC Applications", *Journal of the American Society for Mass Spectrometry*, **2004**, 15 (8), 1191–1200.
- [133] Taguchi, V. Y. *et al.*, "Dioxin Analysis by Gas Chromatography-Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (GC-FTICRMS)", *Journal of the American Society for Mass Spectrometry*, **2010**, S1044030510005118.
- [134] Sühring, R. *et al.*, "Evidence for High Concentrations and Maternal Transfer of Substituted Diphenylamines in European Eels Analyzed by Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry and Gas Chromatography-Fourier Transform Ion Cyclotron Resonance Mass Spectrometry", *Environ. Sci. Technol.*, **2016**, 50 (23), 12678–12685.
- [135] Lababidi, S. *et al.*, "Direct Coupling of Normal-Phase High-Performance Liquid Chromatography to Atmospheric Pressure Laser Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for the Characterization of Crude Oil", *Anal. Chem.*, **2013**, 85 (20), 9478–9485.
- [136] Park, S.-G. *et al.*, "Characterization of Harmonic Signal Acquisition with Parallel Dipole and Multipole Detectors", *J. Am. Soc. Mass Spectrom.*, **2018**, 29 (7), 1394–1402.
- [137] Park, S.-G. *et al.*, "Application of Frequency Multiple FT-ICR MS Signal Acquisition for Improved Proteome Research", *Int. J. Mass Spectrom.*, **2021**, 465, 116578.
- [138] Park, S.-G. *et al.*, "Parallel Detection of Fundamental and Sixth Harmonic Signals Using an ICR Cell with Dipole and Sixth Harmonic Detectors", *J. Am. Soc. Mass Spectrom.*, **2020**, 31 (3), 719–726.
- [139] May, J. C. *et al.*, "Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation", *Anal. Chem.*, **2015**, 87 (3), 1422–1436.
- [140] Tang, X. *et al.*, "Design and Performance of an Atmospheric Pressure Ion Mobility Fourier Transform Ion Cyclotron Resonance Mass Spectrometer", *Rapid Commun. Mass Spectrom.*, **2007**, 21 (7), 1115–1122.
- [141] Benigni, P. *et al.*, "Oversampling Selective Accumulation Trapped Ion Mobility Spectrometry Coupled to FT-ICR MS: Fundamentals and Applications", *Anal. Chem.*, **2016**, 88 (14), 7404–7412.
- [142] Senko, M. W. *et al.*, "External Accumulation of Ions for Enhanced Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry", *J Am Soc Mass Spectrom.*, **1997**, 8 (9), 970–976.
- [143] Emmett, M. R. *et al.*, "Application of Micro-Electrospray Liquid Chromatography Techniques to FT-ICR MS to Enable High-Sensitivity Biological Analysis", *J Am Soc Mass Spectrom.*, **1998**, 9 (4), 333–340.

- [144] Li, W. *et al.*, "Identification of Intact Proteins in Mixtures by Alternated Capillary Liquid Chromatography Electrospray Ionization and LC ESI Infrared Multiphoton Dissociation Fourier Transform Ion Cyclotron Resonance Mass Spectrometry", *Anal. Chem.*, **1999**, 71 (19), 4397–4402.
- [145] Martin, S. E. *et al.*, "Subfemtomole MS and MS/MS Peptide Sequence Analysis Using Nano-HPLC Micro-ESI Fourier Transform Ion Cyclotron Resonance Mass Spectrometry", *Anal. Chem.*, **2000**, 72 (18), 4266–4274.
- [146] Quenzer, T. L. *et al.*, "High Sensitivity Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Biological Analysis with Nano-LC and Microelectrospray Ionization", *Anal. Chem.*, **2001**, 73 (8), 1721–1725.
- [147] Tanaka, Y. *et al.*, "Identification of Human Liver Diacetyl Reductases by Nano-Liquid Chromatography/Fourier Transform Ion Cyclotron Resonance Mass Spectrometry", *Analytical Biochemistry*, **2001**, 293 (2), 157–168.
- [148] Belov, M. E. *et al.*, "Higher-Resolution Data-Dependent Selective External Ion Accumulation for Capillary LC-FTICR", *International Journal of Mass Spectrometry*, **2002**, 218 (3), 265–279.
- [149] Palmblad, M. *et al.*, "Liquid Chromatography and Electron-Capture Dissociation in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: LC and ECD with FTICRMS", *Rapid Commun. Mass Spectrom.*, **2002**, 16 (10), 988–992.
- [150] Smith, R. D. *et al.*, "An Accurate Mass Tag Strategy for Quantitative and High-Throughput Proteome Measurements", *PROTEOMICS*, **2002**, 2 (5), 513–523.
- [151] Davidson, W. *et al.*, "Micro-High-Performance Liquid Chromatography/Fourier Transform Mass Spectrometry with Electron-Capture Dissociation for the Analysis of Protein Enzymatic Digests", *Rapid Communications in Mass Spectrometry*, **2002**, 16 (10), 993–998.
- [152] Ihling, C. *et al.*, "Nano-High-Performance Liquid Chromatography in Combination with Nano-Electrospray Ionization Fourier Transform Ion-Cyclotron Resonance Mass Spectrometry for Proteome Analysis", *Rapid Communications in Mass Spectrometry*, **2003**, 17 (12), 1240–1246.
- [153] Witt, M. *et al.*, "Fourier Transform Ion Cyclotron Resonance Mass Spectrometry with NanoLC/Microelectrospray Ionization and Matrix-Assisted Laser Desorption/Ionization: Analytical Performance in Peptide Mass Fingerprint Analysis", *Journal of the American Society for Mass Spectrometry*, **2003**.
- [154] Bergström, S. K. *et al.*, "A Simplified Multidimensional Approach for Analysis of Complex Biological Samples : On-Line LC-CE-MS", *Analyst*, **2006**, 131 (15), 791–798.
- [155] Burgt, Y. E. M. *et al.*, "Parallel Processing of Large Datasets from NanoLC-FTICR-MS Measurements", *J Am Soc Mass Spectrom*, **2007**, 18 (1), 152–161.
- [156] Smallwood, H. S. *et al.*, "Identification of a Denitrase Activity against Calmodulin in Activated Macrophages Using High-Field Liquid Chromatography-FTICR Mass Spectrometry †", *Biochemistry*, **2007**, 46 (37), 10498–10505.
- [157] Öhrfelt, A. *et al.*, "Identification of Novel α-Synuclein Isoforms in Human Brain Tissue by Using an Online NanoLC-ESI-FTICR-MS Method", *Neurochem Res*, **2011**, 36 (11), 2029–2042.
- [158] Switzer, L. *et al.*, "In-Depth Characterization of Protein Disulfide Bonds by Online Liquid Chromatography-Electrochemistry-Mass Spectrometry", *J. Am. Soc. Mass Spectrom.*, **2016**, 27 (1), 50–58.

- [159] Anderson, L. C. *et al.*, "Intact Protein Analysis at 21 Tesla and X-Ray Crystallography Define Structural Differences in Single Amino Acid Variants of Human Mitochondrial Branched-Chain Amino Acid Aminotransferase 2 (BCAT2)", *J. Am. Soc. Mass Spectrom.*, **2017**, *28* (9), 1796–1804.
- [160] Guo, K. *et al.*, "Stable-Isotope Dimethylation Labeling Combined with LC-ESI MS for Quantification of Amine-Containing Metabolites in Biological Samples", *Anal. Chem.*, **2007**, *79* (22), 8631–8638.
- [161] Bobeldijk, I. *et al.*, "Quantitative Profiling of Bile Acids in Biofluids and Tissues Based on Accurate Mass High Resolution LC-FT-MS: Compound Class Targeting in a Metabolomics Workflow", *Journal of Chromatography B*, **2008**, *871* (2), 306–313.
- [162] Cataldi, T. R. I. *et al.*, "Accurate Mass Analysis of N-Acyl-Homoserine-Lactones and Cognate Lactone-Opened Compounds in Bacterial Isolates of *Pseudomonas Aeruginosa* PAO1 by LC-ESI-LTQ-FTICR-MS", *J. Mass Spectrom.*, **2009**, *44* (2), 182–192.
- [163] Guo, K. *et al.*, "Differential 12C-/13C-Isotope Dansylation Labeling and Fast Liquid Chromatography/Mass Spectrometry for Absolute and Relative Quantification of the Metabolome", *Anal. Chem.*, **2009**, *81* (10), 3919–3932.
- [164] Guo, K. *et al.*, "Qualitative Metabolome Analysis of Human Cerebrospinal Fluid by 13C-/12C-Isotope Dansylation Labeling Combined with Liquid Chromatography Fourier Transform Ion Cyclotron Resonance Mass Spectrometry", *J Am Soc Mass Spectrom*, **2011**, *22* (2), 339–347.
- [165] Suzuki, H. *et al.*, "Metabolic Profiling of Flavonoids in Lotus Japonicus Using Liquid Chromatography Fourier Transform Ion Cyclotron Resonance Mass Spectrometry", *Phytochemistry*, **2008**, *69* (1), 99–111.
- [166] Han, F. *et al.*, "Characterization of Chemical Constituents in Rhodiola Crenulata by High-Performance Liquid Chromatography Coupled with Fourier-Transform Ion Cyclotron Resonance Mass Spectrometer (HPLC-FT-ICR MS): Constituents in R. Crenulata by HPLC-FTICR-MS", *J. Mass Spectrom.*, **2016**, *51* (5), 363–368.
- [167] Nakabayashi, R. *et al.*, "Chemical Assignment of Structural Isomers of Sulfur-Containing Metabolites in Garlic by Liquid Chromatography–Fourier Transform Ion Cyclotron Resonance–Mass Spectrometry", *The Journal of Nutrition*, **2016**, *146* (2), 397S-402S.
- [168] Wang, Y. *et al.*, "Metabolic Profile of Esculetin in Rats by Ultra High Performance Liquid Chromatography Combined with Fourier Transform Ion Cyclotron Resonance Mass Spectrometry", *Journal of Chromatography B*, **2016**, *1020*, 120–128.
- [169] Liu, T. *et al.*, "A Plasma Metabonomic Analysis on Potential Biomarker in Pyrexia Induced by Three Methods Using Ultra High Performance Liquid Chromatography Coupled with Fourier Transform Ion Cyclotron Resonance Mass Spectrometry", *Journal of Chromatography B*, **2017**, *1063*, 214–225.
- [170] Li, B. *et al.*, "In Vivo Study of Erysolin Metabolic Profile by Ultra High Performance Liquid Chromatography Coupled to Fourier Transform Ion Cyclotron Resonance Mass Spectrometry", *Journal of Chromatography B*, **2018**, *1072*, 173–181.
- [171] Liu, T. *et al.*, "Metabolic Profiling of Gegenqinlian Decoction in Rat Plasma, Urine, Bile and Feces after Oral Administration by Ultra High Performance Liquid Chromatography Coupled with Fourier Transform Ion Cyclotron Resonance Mass Spectrometry", *Journal of Chromatography B*, **2018**, *1079*, 69–84.
- [172] Yao, J. *et al.*, "An Available Strategy Based on Accurate Mass by Ultra High Performance Liquid Chromatography Coupled to Fourier Transform Ion Cyclotron Resonance Mass

- Spectrometry Technology to Characterization of Metabolic Profile of Palbociclib in Rat Urine, Feces and Bile”, *Journal of Chromatography B*, **2019**, 1124, 37–46.
- [173] Müller, L. *et al.*, “Characterization of Oligomeric Compounds in Secondary Organic Aerosol Using Liquid Chromatography Coupled to Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry”, *Rapid Commun. Mass Spectrom.*, **2009**, 23 (7), 971–979.
- [174] Shen, Y. *et al.*, “Online Supercritical Fluid Extraction Mass Spectrometry (SFE-LC-FTMS) for Sensitive Characterization of Soil Organic Matter”, *Faraday Discuss.*, **2019**, 218 (0), 157–171.
- [175] Kim, C. H. *et al.*, “UPLC/FT-ICR MS-Based High-Resolution Platform for Determining the Geographical Origins of Raw Propolis Samples”, *J Anal Sci Technol*, **2019**, 10 (1), 8.
- [176] Varga, Z. *et al.*, “Photodegradation of Benzisothiazolinone: Identification and Biological Activity of Degradation Products”, *Chemosphere*, **2020**, 240, 124862.
- [177] Itoh, S. *et al.*, “N-Linked Oligosaccharide Analysis of Rat Brain Thy-1 by Liquid Chromatography with Graphitized Carbon Column/Ion Trap-Fourier Transform Ion Cyclotron Resonance Mass Spectrometry in Positive and Negative Ion Modes”, *Journal of Chromatography A*, **2006**, 1103 (2), 296–306.
- [178] Diehnelt, C. W. *et al.*, “Identification of Microcystin Toxins from a Strain of *Microcystis Aeruginosa* by Liquid Chromatography Introduction into a Hybrid Linear Ion Trap-Fourier Transform Ion Cyclotron Resonance Mass Spectrometer”, *Anal. Chem.*, **2006**, 78 (2), 501–512.
- [179] Hashii, N. *et al.*, “Study on the Quality Control of Cell Therapy Products”, *Journal of Chromatography A*, **2007**, 1160 (1–2), 263–269.
- [180] Nielsen, M. W. F. *et al.*, “Screening and Confirmation Criteria for Hormone Residue Analysis Using Liquid Chromatography Accurate Mass Time-of-Flight, Fourier Transform Ion Cyclotron Resonance and Orbitrap Mass Spectrometry Techniques”, *Analytica Chimica Acta*, **2007**, 586 (1–2), 122–129.
- [181] Wang, F. *et al.*, “Identification of Clusters from Reactions of Ruthenium Arene Anticancer Complex with Glutathione Using Nanoscale Liquid Chromatography Fourier Transform Ion Cyclotron Mass Spectrometry Combined with ¹⁸⁰O-Labeling”, *J Am Soc Mass Spectrom*, **2008**, 19 (4), 544–549.
- [182] Wu, C. S. *et al.*, “Separation and Identification of Moxifloxacin Impurities in Drug Substance by High-Performance Liquid Chromatography Coupled with Ultraviolet Detection and Fourier Transform Ion Cyclotron Resonance Mass Spectrometry”, *Chinese Chemical Letters*, **2012**, 23 (10), 1185–1188.
- [183] Wahl, J. H. *et al.*, “Direct Electrospray Ion Current Monitoring Detection and Its Use with On-Line Capillary Electrophoresis Mass Spectrometry”, *Anal. Chem.*, **1995**, 67, 462–465.
- [184] Valaskovic, G. A. *et al.*, “Attomole Protein Characterization by Capillary Electrophoresis-Mass Spectrometry”, *Science*, **1996**, 273, 1199–1202.
- [185] Veenstra, T. D. *et al.*, “Proteome Analysis Using Selective Incorporation of Isotopically Labeled Amino Acids”, *J. Am. Soc. Mass Spectrom*, **2000**, 11, 78–82.
- [186] Que, A. H. *et al.*, “Coupling Capillary Electrochromatography with Electrospray Fourier Transform Mass Spectrometry for Characterizing Complex Oligosaccharide Pools”, *Analytical Chemistry*, **2003**, 75, 1684–1680.
- [187] Gavard, R. *et al.*, “KairosMS: A New Solution for the Processing of Hyphenated Ultrahigh Resolution Mass Spectrometry Data”, *Anal. Chem.*, **2020**, 92 (5), 3775–3786.

- [188] Zekavat, B. *et al.*, “Efficient Injection of Low-Mass Ions into High Magnetic Field Fourier Transform Ion Cyclotron Resonance Mass Spectrometers: Injection of Low-Mass Ions for High Magnetic Field FTICRMS”, *Rapid Commun. Mass Spectrom.*, **2014**, 28 (2), 230–238.

CHAPITRE 3 : LES ACIDES RIBONUCLÉIQUES (ARN)

3.1. Généralités sur la structure des ARN

Les ARN sont des chaînes de nucléotides liés par des liaisons phosphodiesters (Figure 3.1). Chaque nucléotide est composé d'un groupement phosphate, d'un ribose et d'une des quatre bases azotées canoniques. Parmi ces quatre bases, il y a l'adénine (A) et la guanine (G) qui sont des purines, la cytidine (C) et l'uracil (U) qui sont pyrimidines. À chaque extrémité (3' et 5'), différentes terminaisons peuvent être présentes. Du côté 5', il est possible de retrouver un groupe hydroxyle (noté 5'-OH) ou un groupe phosphate (noté 5'-p). Du côté 3', en plus du groupe hydroxyle (noté 3'-OH) et du groupe de phosphate (noté 3'-p), un groupe phosphate cyclique (noté 3'->p) peut également être présent. Par convention, la séquence d'un ARN se lit de l'extrémité 5' à l'extrémité 3', et si les extrémités ne sont pas précisées, il s'agit de groupes hydroxyles.

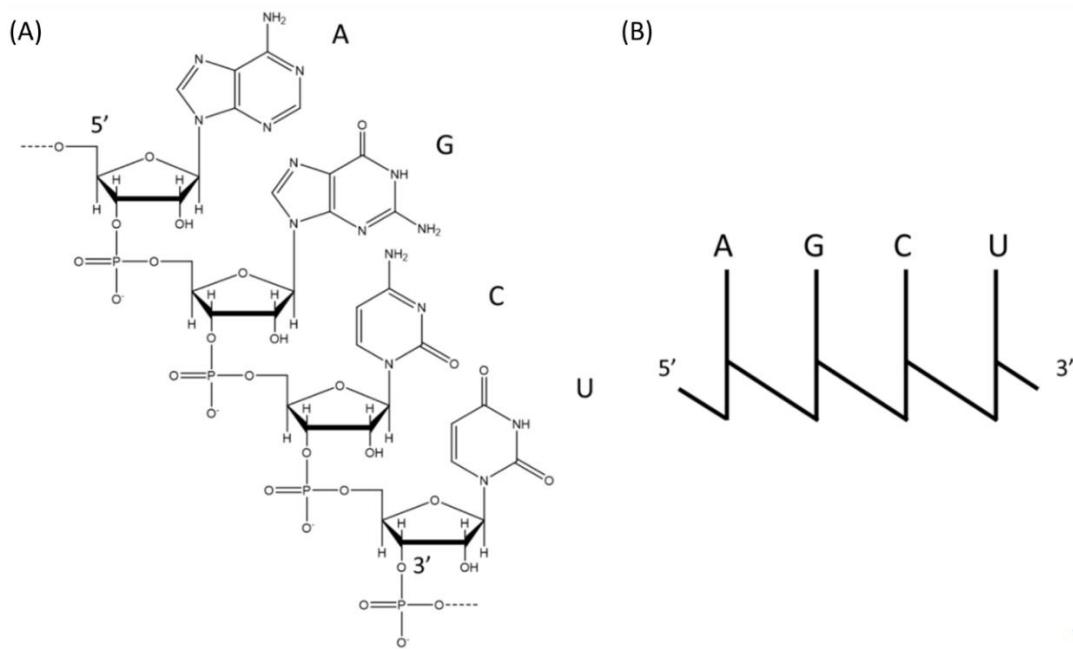


Figure 3.1 : (A) Structure primaire d'une séquence d'ARN composée des quatre nucléosides canoniques et (B) représentation simplifiée de la chaîne de l'ARN. Extrait de [1].

3.1.1. Appariements des bases et motifs

Les bases azotées des ARN peuvent s'apparier deux à deux, en formant des liaisons hydrogènes, en suivant le principe de Watson et Crick qui a été énoncé pour l'ADN en 1953 [2]. Ainsi, l'adénosine et l'uridine s'apparent avec deux liaisons hydrogènes (A=U), et la guanosine et la cytidine s'apparent avec trois liaisons (G≡C). En plus de ces deux paires de bases principales, d'autres paires de bases bancales, aussi appelées « wobble », peuvent être présentes. Par exemple, un appariement « wobble » bien connu chez les ARN et l'appariement de la guanine et de l'uracil à l'aide de deux liaisons hydrogènes (G=U) [3].

Ces appariements impliquent un repliement des ARN sur eux-mêmes en formant des doubles hélices compactes et stabilisées par empilement des bases. En cas d'absence de complémentarité et donc d'appariements, des motifs sont créés au niveau de la structure secondaire des ARN. Les principaux motifs sont représentés en Figure 3.2.

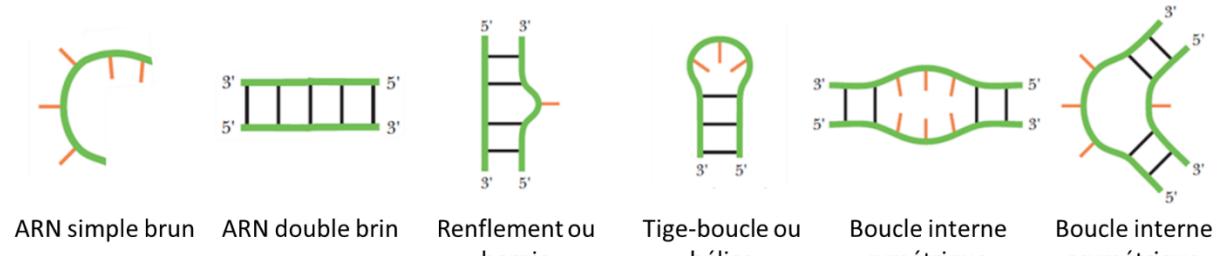


Figure 3.2 : Motifs créés par appariement des nucléotides des ARN. Les bases non appariées sont représentées en orange alors que les appariements sont représentés en noir. Extrait de [1] et adapté de [4].

L'apparition de motifs dans la structure secondaire des ARN affecte leurs structures tertiaires, aussi appelées structures 3D, qui jouent un rôle dans le fonctionnement des ARN au sein de la cellule [5,6].

3.2. Les différents types d'ARN et leurs rôles

Les principales missions des ARN au sein d'une cellule sont :

- La transmission de l'information génétique lors de la production de protéines (processus de traduction) [7],
- La catalyse de certaines réactions biochimiques [8],
- La régulation de l'expression des gènes [9].

Pour effectuer ces différentes missions, il existe plusieurs types d'ARN dans les cellules : certains sont dits codants (ARNc), et d'autres non-codants (ARNnc). Les principaux types d'ARN seront détaillés dans les paragraphes suivants.

3.2.1. Les ARN messagers (ARNm)

Les ARN messagers (ARNm) servent comme supports de l'information génétique et sont donc codants. En effet, leurs séquences sont les complémentaires et antiparallèles de celles de l'ADN, telle une copie. Du fait de leur implication dans l'apport des informations utiles lors de la production des protéines, les ARNm sont appelés des ARN codants. Chaque triplet de nucléotides, aussi appelé codon, dans la séquence de l'ARNm, correspond à un unique acide aminé (à contrario, plusieurs codons peuvent correspondre au même acide aminé). Cette règle de correspondance est appelée le code génétique [7]. Ainsi, le ribosome est capable, en lisant le message encodé de l'ARNm, de traduire une protéine spécifique et nécessaire à la cellule. L'ARNm permet donc de coder la séquence en acides aminés d'une

protéine spécifique. Entre différentes cellules, les ARNm sont très hétérogènes. La majorité se présente sous la forme d'une chaîne individuelle de taille très variable.

3.2.2. Les ARN de transfert (ARNt)

Les ARN de transfert (ARNt) reconnaissent de manière spécifique les ARNm afin d'apporter les acides aminés lors de la synthèse des protéines. Bien que ces ARN de 70 à 100 nucléobases [10] soient non-codants, ils sont considérés comme fonctionnels puisqu'ils participent directement à la traduction des protéines. Les ARNt sont généralement nommés par l'acide aminé qu'ils portent et occasionnellement, les séquences de leurs anticodons (nucléotides en positions 34, 35 et 36) y sont ajoutées. À titre d'exemple, l'ARNt de levure portant une phénylalanine est nommé ARNt^{Phe GAA}.

Les ARNt ont une structure secondaire bien particulière en forme de feuille de trèfle (Figure 3.3.A) [6]. Cette structure est induite par les appariements intramoléculaires qui forment des motifs tiges-boucles. Par exemple, la boucle D est engendrée par des dihydrouridines (D) qui ne s'apparent pas. La boucle TψC (T : thymine, ψ : pseudouridine) s'appelle de cette manière, car sa séquence de sept nucléotides non appariés est toujours constituée de l'enchaînement de nucléotides TψC. A contrario, la boucle variable diffère d'un ARNt à un autre. Du fait d'autres appariements de certains nucléotides, les différentes boucles interagissent entre elles et se rapprochent, ce qui crée un repliement de l'ARNt. La structure 3D en L inversé (Figure 3.3.B) est très stable et permet aux ARNt d'assurer leur rôle lors de la synthèse des protéines [6].

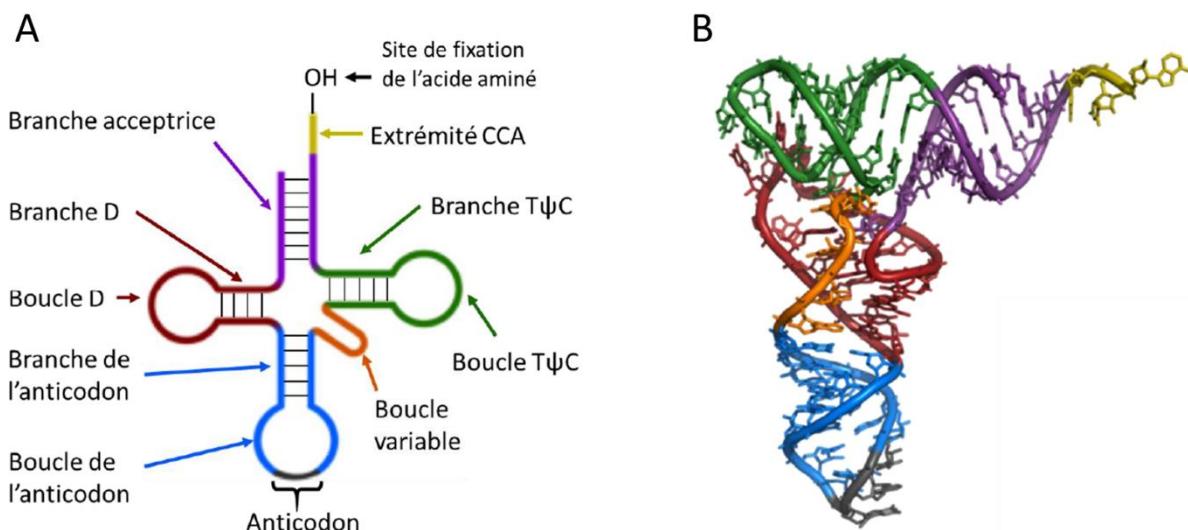


Figure 3.3 : (A) Structure secondaire d'un ARNt en forme de feuille de trèfle. (B) Structure 3D d'un ARNt en forme de L inversé. Chaque région est distinguée par une couleur. Extrait de [1].

L'extrémité CCA-OH en 3' sur la branche acceptrice (portant l'acide aminé spécifique) est identique à tous les ARNt. En effet, cette extrémité permet de former une liaison ester, aussi appelée liaison ester d'aminoacyle, entre l'adénine et l'acide aminé spécifique porté par l'ARNt.

La boucle anticodon joue également un rôle crucial dans la traduction des protéines. Il s'agit d'une région à double-hélice composée de sept nucléotides non appariés, dont trois d'entre eux, en positions 34, 35 et 36, constituent l'anticodon. Lors de la traduction, les codons de l'ARNm sont reconnus spécifiquement par les ARNt afin d'apporter le bon acide aminé à la séquence de la protéine. La reconnaissance des codons de l'ARNm s'effectue par appariement complémentaire et antiparallèle avec les anticodons des ARNt. On parle alors d'interactions codon-anticodon (Figure 3.4).

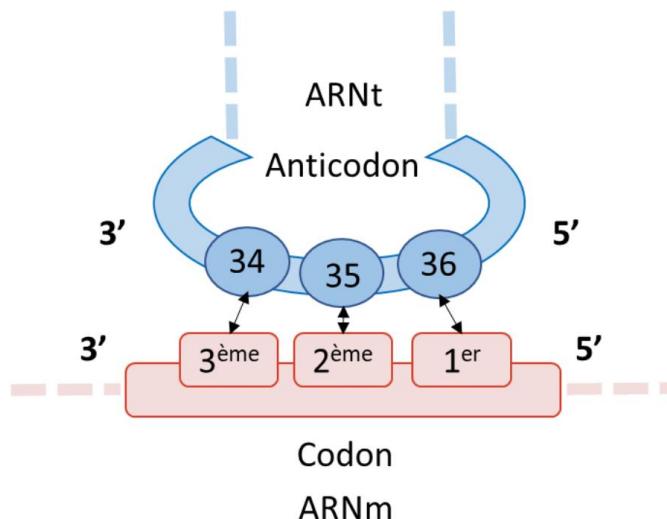


Figure 3.4 : Reconnaissance par appariement entre un codon d'un ARNm et un anticodon d'un ARNt. Extrait de [1].

3.2.3. Les ARN ribosomiques (ARNr)

Les ARN ribosomiques (ARNr) appartiennent à un grand complexe ribonucléoprotéique, appelé ribosome, d'une masse totale supérieure à 10^6 Da. Ce dernier est constitué de plusieurs ARNr et de plusieurs protéines formant des sous-unités. Les ARNr sont les ARN les plus abondants dans les cellules et sont généralement très longs (entre 120 et 1300 nucléotides).

Ce sont des ARN non-codants mais ils jouent un rôle essentiel dans la traduction des protéines. En effet, la traduction des protéines a lieu au sein de cet assemblage supramoléculaire à l'aide de multiples enzymes et cofacteurs [11]. Il agit alors comme un catalyseur. En se déplaçant le long de l'ARNm dans le sens 5' → 3', le ribosome lit la suite de codons. Une première sous-unité du ribosome assiste l'apport de l'acide aminé à la protéine par l'ARNt. Une seconde sous-unité du ribosome est chargée de réaliser le transfert de l'acide aminé de l'ARNt à la séquence de la protéine. Le ribosome se déplace ensuite au codon suivant et libère l'ARNt. On parle de translocation. Ainsi de suite, la séquence de la protéine est complétée jusqu'à ce que le ribosome lise le codon stop de l'ARNm. Des protéines spécifiques reconnaissant ce codon stop permettent alors de libérer du ribosome l'ARNm, la protéine et le dernier ARNt utilisé. Le processus de traduction d'une protéine est représenté en Figure 3.5.

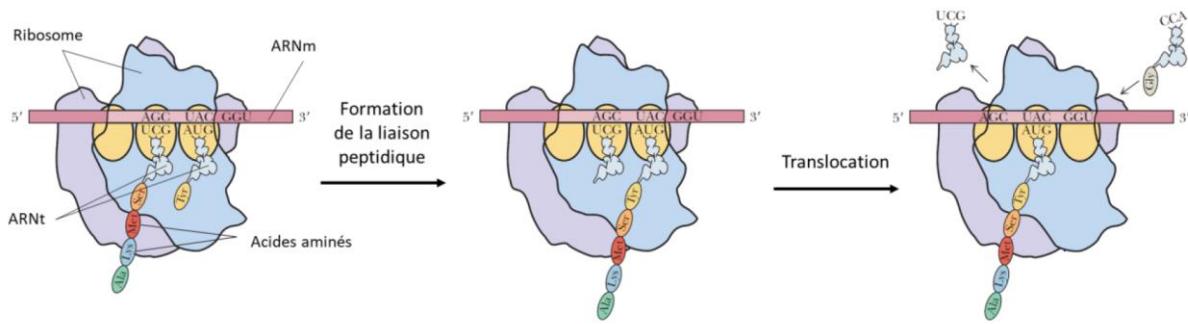


Figure 3.5 : Représentation simplifiée des étapes de la traduction de l'ARNm en protéine. Extrait de [1] et adapté de [4].

Pour différencier les différents types d'ARN, ils sont communément nommés par leur coefficient de sédimentation en Svedberg (S). Les eucaryotes possèdent un ribosome 80 S constitué des sous-unités 60 S et 40 S. La grande sous-unité 60 S contient les ARNr 5 S, 5.8 S et 28 S. La petite sous-unité 40 S contient quant-à-elle, le ribosome 18S . Enfin, les eubactéries ainsi que les archées possèdent un ribosome 70 S constitué des sous-unités 50 S et 30 S. La grande sous-unité 50 S contient les ARNr 5 S et 23 S. La petite sous-unité 30 S contient quant à elle, le ribosome 16 S. Enfin, les archées possèdent un ribosome 70 S constitué des sous-unités 50 S et 30 S. La grande sous-unité 50 S contient les ARNr 5 S et 23 S. Du fait de leur longueur, les ARNr comportent de nombreux motifs dans leur structure secondaire [4]. À titre d'exemple, la structure secondaire d'un ARNr 16 S est présenté en Figure 3.6.

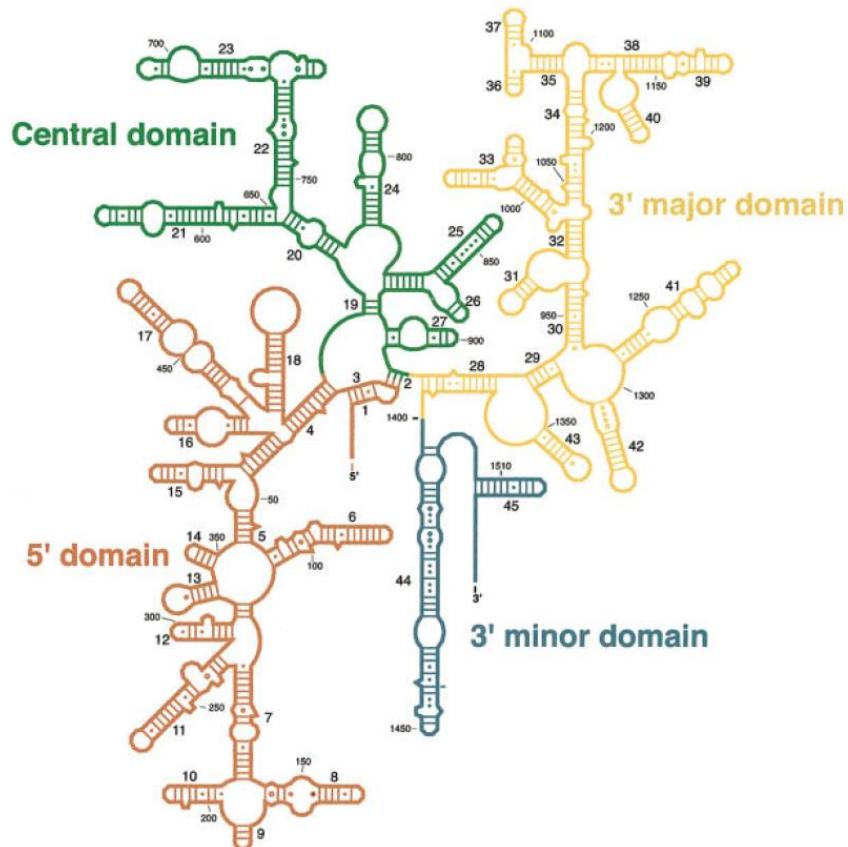


Figure 3.6 : Structure secondaire de l'ARNr 16S de *E. coli*. Extrait de [12].

3.2.4. Autres ARN

Plusieurs autres classes d'ARN sont impliquées dans de nombreuses fonctions biologiques et sont indispensables pour réguler l'expression génétique, même si elles ne participent pas directement au processus de traduction des protéines [4,13]. Parmi elles, on retrouve les petits ARN (ARNs). Ces ARNs sont généralement composés d'une vingtaine de nucléotides, et capables de cibler un ARN spécifique par appariement complémentaire [14]. Selon leurs rôles dans la cellule, ils ont été divisés en différentes sous-classes :

- Les petits ARN interférents (ARNsi) sont capables d'inhiber la production de protéines spécifiques. En effet, en s'appariant avec un ARNm cible, des régions d'ARN à doubles brins (ARNdb) se forment et le duplex formé par l'ARNm et l'ARNsi est dégradé. L'ARNm cible ainsi dégradé n'est plus capable de coder la protéine, ce qui va bloquer le processus de traduction [15].
- Les micros ARN (ARNmi) ont, quant à eux, le rôle de ralentir l'expression génétique [16]. Ces ARNmi s'apparentent partiellement avec des ARNm cibles pour moduler la traduction des protéines. Les ARNm cibles ne sont pas dégradés comme avec les ARNsi, mais ils sont réduits au silence.
- Les petits ARN nucléolaires (ARNsno) jouent un rôle de catalyseur pour modifier les ARN. De plus, dans les cellules eucaryotes, les petits ARN nucléaires (ARNsn) jouent un rôle dans la maturation des ARNr. La maturation des ARN est une étape essentielle afin de rendre les ARN fonctionnels [17,18].

3.3. Les modifications post-transcriptionnelles

De nombreuses modifications très variées sont présentes chez les ARN. Parmi elles, se trouvent les modifications des extrémités (ajout d'un acide aminé à l'extrémité 3' des ARNt) ou encore les changements dans la séquence nucléotidique. Des modifications ont été observées dans tous les types d'ARN. Cependant, c'est chez les ARNt qu'elles sont les plus abondantes.

De nos jours, plus de 160 modifications sont référencées dans la littérature et dans les bases de données telles que Modomics [19]. Ces modifications sont des ajouts de groupements chimiques plus ou moins complexes sur la base et/ou sur le ribose, des réductions/oxydations, mais aussi des isomérisations de la base. Lorsqu'une modification complexe ou que plusieurs modifications sont présentes sur un même nucléotide, on parle alors d'hypermodification. La production des modifications est assurée par de nombreuses enzymes spécifiques [20], et dans le cas des hypermodifications, leur biosynthèse nécessite généralement plusieurs étapes dans un ordre précis.

Pour annoter les nucléosides modifiés, il existe deux nomenclatures. La première est une écriture simplifiée du nom de la modification, basée sur la nature de celle-ci, sa position ainsi que la base canonique du nucléotide modifié (Figure 3.7). Dans cette écriture, la nature de la modification est indiquée selon les abréviations détaillées dans le Tableau 3.1. Une deuxième nomenclature, plus utilisée par les outils bio-informatiques permet d'écrire les modifications en un seul caractère. Par exemple, la modification N6,N6,2'-O-triméthyladenosine est annotée m⁶₂Am ou m^{6,6}Am par la première nomenclature (Figure 3.7) et η par la seconde.

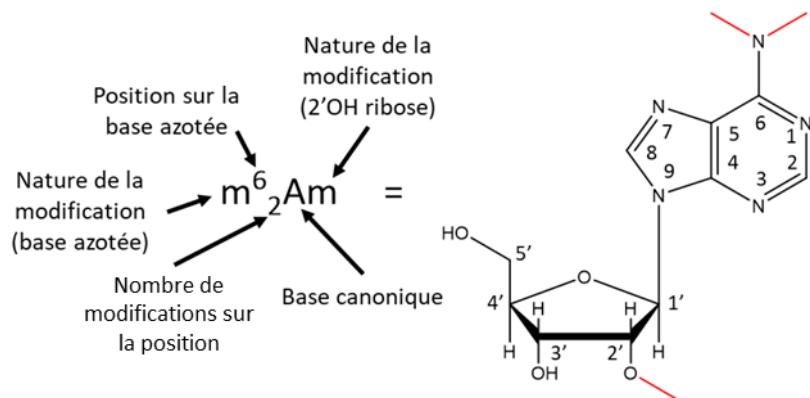


Figure 3.7 : Exemple d'écriture simplifiée d'un nucléotide modifié (N6,N6,2'-O-triméthyladenosine) et correspondance avec la structure chimique. Extrait de [1].

Tableau 3.1 : Principales abréviations utilisées dans l'écriture simplifiée des modifications. Reproduit de [21].

Abréviation	Description complète	Abréviation	Description complète
m	methyl	se	seleno
ac	acetyl	cmo	5-oxyacetic acid
t	threonylcarbamoyl	mcmo	5-oxyacetic acid methyl ester
m	taurinomethyl	mo	methoxy
i	isopentenyl	ho	hydroxy
io	cis-hydroxyisopentenyl	f	formyl
s	thio	gal	galactosyl
ms	methylthio	glu	glutamyl
inm	isopentenylaminomethyl	man	mannosyl
g	glycinylcarbamoyl	r	2'-O-ribosyl
hn	hydroxynorvalylcarbamoyl	I	inosine
hm	hydroxymethyl	ψ	pseudouridine
nm	aminomethyl	Q	queusine
acp	(3-amino-3-carboxypropyl)	yW	wybutosine
cm	carboxymethyl	OHyW	hydroxywybutosine
chm	carboxyhydroxymethyl	C'	agmatidine
cmnm	carboxymethylaminomethyl	k ² C	lysidine
ncm	carbamoylmethyl	D	dihydouridine
mnm	methylaminomethyl	G'	archaeosine

3.3.1. Rôle des modifications naturelles

Les modifications post-transcriptionnelles naturelles sont impliquées dans de nombreuses fonctions biologiques et caractéristiques structurales de l'ARN. En effet, elles peuvent grandement influer sur la rigidité ainsi que la stabilité de la structure de l'ARN [22,23], mais également l'appariement des bases [24] ou encore la traduction des protéines [25]. Dans les prochains paragraphes, quelques exemples de modifications jouant un rôle important dans la stabilité de la structure de l'ARN, mais aussi dans son activité biologique seront détaillées afin de démontrer la nécessité de les caractériser. Il s'agit bien évidemment d'une liste non exhaustive des modifications ayant un rôle important dans les fonctions biologiques des ARN, et notamment la reconnaissance codon-anticodon.

3.3.1.1. Exemples de modifications influençant la stabilité de la structure des ARN

De nombreuses modifications participent au repliement des ARN ainsi qu'à la rigidité et la stabilité de leur structure 3D [22].

Les méthylations font partie des modifications influençant la structure des ARN de différentes façons [26]. Par exemple, les 2'O-méthylation augmentent la rigidité de la structure en double hélice du bras accepteur des ARNt, ce qui facilite le transfert de l'acide aminé spécifique [27]. À la jonction des boucles D et anticodon des ARNt (position 26), une guanosine souvent modifiée en m²G ne peut plus s'apparier avec une cytidine, mais peut alors s'apparier avec une adénosine. Ce changement d'appariement donne à l'ARNt sa structure tertiaire [28].

D'autres modifications telles que les dihydrouridines (D) de la boucle D, les pseudouridines (ψ) ainsi que la thymine (T ou m⁵T) de la boucle T, participent également au maintien et à la stabilité de la structure en L inversé des ARNt [27,29–32].

3.3.1.2. Exemples de modifications influençant le processus de reconnaissance codon-anticodon

Au niveau de la boucle anticodon des ARNt, les positions 34 et 37 sont très souvent modifiées [21].

La première position de l'anticodon est la position 34. Lorsque cette position est modifiée, cela peut avoir une influence sur les appariements de base et donc sur la reconnaissance du codon. Par exemple, lorsqu'une adénosine modifiée en inosine (I, déamination de l'adénosine) est présente à cette position, elle est capable de s'apparier avec trois nucléosides (C, A et U) contrairement à l'unique paire formée entre l'adénosine et l'uridine [33]. De cette manière, un anticodon ayant une inosine en position 34 est capable de s'apparier à 3 codons différents [34]. Des études ont montré que cette modification augmente

la fidélité et l'efficacité du décodage lors de la traduction d'une protéine [35,36]. Cette position 34 a également été détectée avec beaucoup d'autres modifications telles que m⁵C, Q, Gm, mnmm⁵U, ou encore cmo⁵U [37]. En plus d'influencer les appariements de bases codon-anticodon, les modifications en position 34 renforcent également la structure de la boucle anticodon, ce qui augmente l'efficacité du décodage [38].

En ce qui concerne la base à la position 37, il s'agit d'une purine (guanine ou adénosine) très souvent modifiée par un m¹G, yW, t⁶A ou encore i⁶A [39,40]. Ces modifications augmentent la rigidité de la boucle anticodon, ce qui permet une meilleure interaction entre le codon et l'anticodon. Une autre modification de la position 37 ayant une grande influence sur la reconnaissance codon-anticodon est ms²i⁶A. Bien que la position 37 ne fasse pas partie de l'anticodon (positions 34 à 36), cette modification permet de renforcer les interactions codon-anticodon en ajoutant une interaction de van der Walls entre ms²i⁶A de la position 37 et la première position du codon (Figure 3.8) [21].

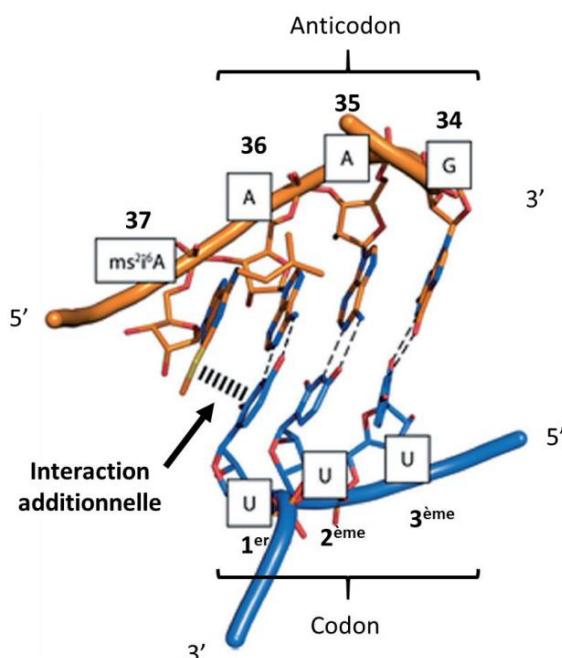


Figure 3.8 : Structure 3D d'une interaction codon-anticodon avec ms²i⁶A en position 37 de l'ARNt. Extrait de [1] et adapté de [21].

3.3.2. Rôle des modifications post-transcriptionnelles dans les maladies

En plus des modifications jouant un rôle dans les fonctions biologiques et les caractéristiques structurales de l'ARN, d'autres sont également impliquées dans le phénomène de résistance aux antibiotiques [41] ou encore dans certaines maladies [42].

En effet, des études ont montré l'implication de modifications spécifiques dans l'obésité [43], les maladies mitochondrielles [44], les cancers [45] ou encore les maladies neurologiques [46]. En ce qui concerne les cancers, certaines modifications peuvent être utilisées comme biomarqueurs [47]. C'est par exemple le cas de la queuosine (Q) dans le

cancer du poumon [48] ou encore de la pseudouridine (Ψ) dans le cancer de la prostate [49]. Grâce à ses études, le suivi des modifications retrouvées dans l'urine permet d'établir un diagnostic précis pour de nombreux cancers [50,51] dont le cancer du côlon [52], le cancer du sein [53], le cancer de la thyroïde [54], ou encore la leucémie [55].

Il est important de noter que des traitements ont également été développés à partir de nucléosides ou oligonucléotides modifiés. À titre d'exemple, on peut retenir le développement de traitements du virus de l'immunodéficience humaine (VIH) basé sur des nucléosides modifiés à partir des années 80. Actuellement, ces traitements sont encore utilisés en combinaison avec d'autres médicaments et permettent de prolonger l'espérance de vie des patients en limitant le développement du virus [56,57]. Un autre exemple marquant est le développement d'oligonucléotides antisens modifiés (ASO), composés d'une à deux dizaines de bases, à la fin des années 1970 [58]. Ces traitements se sont montrés efficaces contre certains cancers [59] ou encore l'atrophie musculaire spinale [60]. Ces ASO peuvent s'apparier à des ARNm cibles, entraîner la dégradation de l'ARN par une enzyme ayant reconnu le duplex ARNm-ASO et ainsi bloquer la traduction de l'information génétique en protéines [61]. Les modifications de ces ASO sont essentielles afin de limiter leur dégradation par les nucléases de la matrice biologique.

Les différents exemples décrits dans les derniers paragraphes mettent en lumière l'importance de pouvoir bien caractériser les modifications post-transcriptionnelles.

3.4. La caractérisation des modifications post-transcriptionnelles des ARN par spectrométrie de masse

Il existe plusieurs techniques et stratégies d'analyse pour caractériser les modifications post-transcriptionnelles. Cette caractérisation englobe à la fois l'identification de la nature de la modification, sa localisation sur la séquence de l'ARN et enfin sa quantification [62]. Comme les modifications ont un impact direct sur l'activité biologique des ARN, il est nécessaire de pouvoir les caractériser afin d'appréhender leurs intérêts biologiques. Afin d'obtenir toutes ces informations, il est nécessaire d'utiliser plusieurs méthodes et plusieurs stratégies d'analyses. Les méthodes de biologie moléculaire, et notamment le séquençage, sont les plus couramment mises en œuvre. La majorité des méthodes de séquençages sont basées sur la réaction en chaîne par polymérase de transcription inverse (RT-PCR) [63]. Cette technique repose sur la duplication de la séquence de l'ARN par une enzyme copiant les séquences par appariements de bases complémentaires. Après duplication, les ARN peuvent ensuite être détectés. Cependant, toutes les modifications n'influencent pas l'appariement des bases et certaines ne peuvent donc pas être détectées avec cette technique. De plus, la RT-PCR permet de localiser les positions des modifications, mais elle ne permet pas d'identifier précisément la nature de celles-ci [64]. Pour pallier ce problème, d'autres méthodes d'analyses ont été développées. Parmi elles, la spectrométrie de masse est notamment devenue une des

méthodes de référence. La stratégie générale d'analyse des modifications par spectrométrie de masse est présentée en Figure 3.9. Chaque étape sera par la suite détaillée.

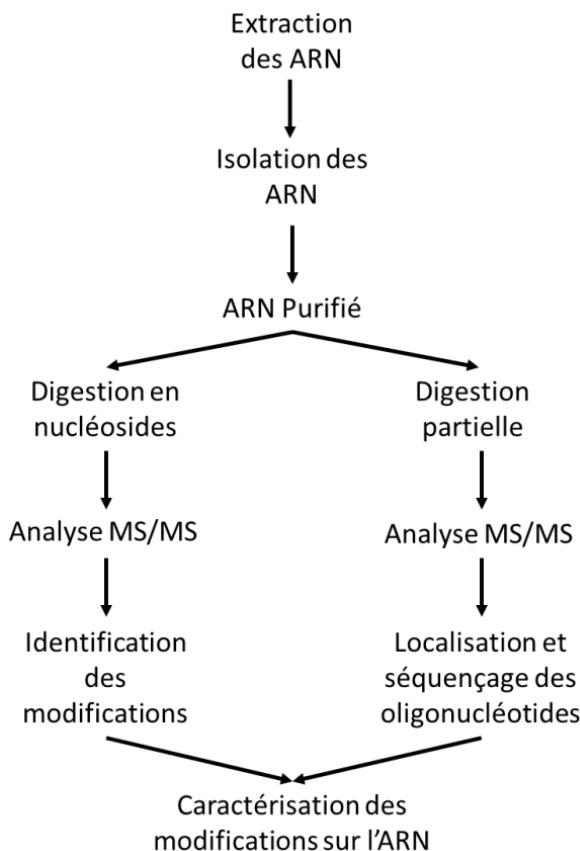


Figure 3.9 : Stratégie générale de caractérisation des modifications post-transcriptionnelles des ARN par spectrométrie de masse. Extrait de [1].

3.4.1. Purification des ARN

La première étape dans la préparation des échantillons d'ARN est leur purification. Dans un premier temps, il est nécessaire de les extraire des cellules. Puis, pour étudier les modifications d'un seul ARN, celui-ci doit être isolé. Dans le cadre de ces travaux, les étapes de purification ont été réalisées par nos collègues biologistes de l'IBMC.

3.4.1.1. Extraction des ARN

L'extraction des ARN présents dans une cellule est divisée en quatre étapes. Tout d'abord, une lyse cellulaire est nécessaire. Celle-ci peut être effectuée chimiquement, à l'aide d'enzymes ou encore mécaniquement [65–67]. Ensuite, les protéines et les complexes nucléoprotéiques sont dénaturés à l'aide de détergents, d'agents réducteurs ou d'agents chaotropiques. Ces derniers sont également utilisés pour inactiver les nucléases [68]. Enfin, les ARN sont séparés de toutes les impuretés issues des précédentes étapes du processus d'extraction. La méthode la plus simple et la plus couramment utilisée est l'extraction liquide-liquide [69]. Elle permet notamment de séparer les ARN des ADN. Ce processus d'extraction permet de récupérer tous les ARN de la cellule dans un extrait total.

Pour ne récupérer que les ARNt, il est également possible de procéder à une extraction phénolique. Le phénol permet à la fois de lyser la cellule et de récupérer les ARNt dans la phase aqueuse. Après une précipitation, la liaison ester aminoacyle qui lie les ARNt à leur acide aminé spécifique est hydrolysée. Plusieurs étapes de lavage permettent ensuite de récupérer un extrait total d'ARNt.

3.4.1.2. Isolement d'un ARN spécifique

De manière à pouvoir étudier les modifications d'un ARN spécifique, il doit ensuite être isolé de l'extrait total. Plusieurs méthodes permettent de séparer les différents ARN et la plus couramment utilisée est l'électrophorèse sur gel de polyacrylamide (PAGE). Sous l'effet d'un champ électrique, les biomolécules déposées à l'extrémité du gel se séparent en fonction de leur taille. Un gel en deux dimensions (2D PAGE) permet de mieux séparer les différents ARN. La première dimension est totalement dénaturante (12,5 % acrylamide/bis-acrylamide (19:1), urée 8 M et tampon 1x Tris Borate EDTA). La bande de gel obtenu après cette première dimension (Figure 3.10.A) est ensuite positionnée sur le gel de la deuxième dimension qui est semi-dénaturante (20 % acrylamide/bis-acrylamide (19:1), urée 4 M et tampon 1x Tris Borate EDTA) [70]. Une révélation aux UV après coloration au bromure d'éthidium permet de visualiser les différents spots obtenus sur le gel 2D PAGE (Figure 3.10.B). Ainsi isolés, les ARNs peuvent ensuite être dessalés avant d'être digérés et étudiés par spectrométrie de masse.

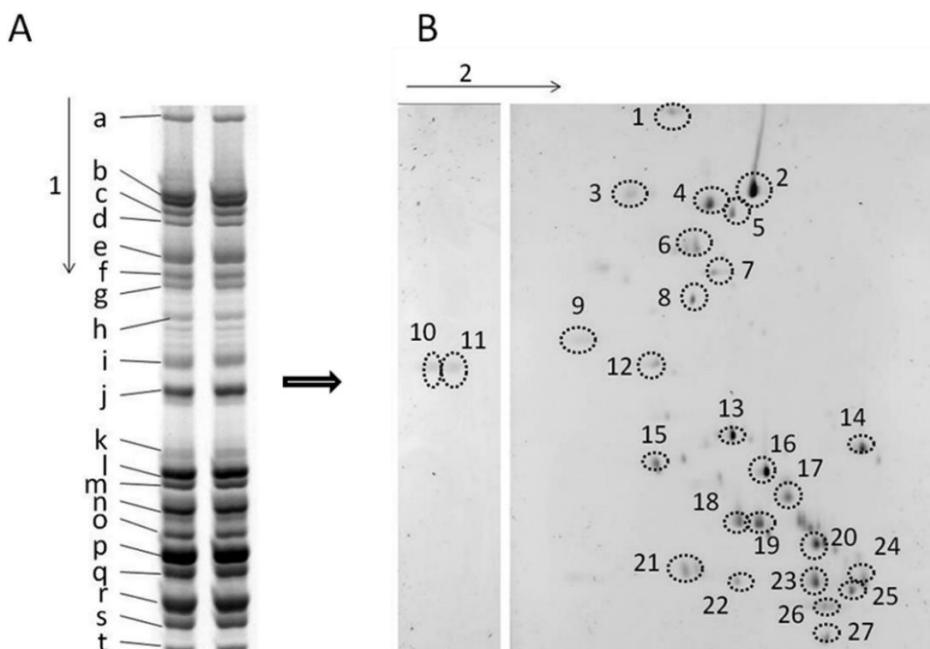


Figure 3.10 : (A) Première dimension dénaturante et (B) seconde dimension semi-dénaturante d'un gel 2D PAGE d'un mélange d'ARNt. Les flèches numérotées 1 et 2 représentent le sens de migration de chacune des deux dimensions. Extrait de [70].

3.4.2. Dessalage des ARN

Les ARN ayant une forte tendance à former des adduits avec les cations potassium et sodium, il est nécessaire de les dessaler avant de les analyser par spectrométrie de masse afin de ne pas perdre en qualité spectrale. Le dessalage des ARN est réalisé généralement soit par

plusieurs cycles de précipitation, soit par extraction en phase solide à l'aide de « ZipTips » (pointes de pipette avec une résine C18).

Notre choix de méthode de dessalage s'est porté sur la précipitation par sa simplicité de mise en œuvre, mais également pour son très faible risque de perte d'échantillon. Les ARN ont été dissous dans une solution d'acétate d'ammonium fortement concentrée, puis un large excès d'éthanol a été ajouté au mélange afin de faire précipiter l'ARN à -20°C. Après une centrifugation de 15 min à 4 °C à au moins 5 000 g, le surnageant peut être retiré et le reste d'éthanol évaporé. Trois cycles de précipitation, d'au moins 2h chacun, ont été réalisés pour dessaler chaque échantillon, avant de resolubiliser l'ARN dans de l'eau ultra-pure.

3.4.3. Digestion des ARN et stratégies d'analyses

Avant de pouvoir analyser les ARN par spectrométrie de masse, il est nécessaire de les digérer. Deux stratégies d'analyses sont disponibles afin d'obtenir le plus d'informations possible sur les modifications post-transcriptionnelles : l'analyse des nucléotides/nucléosides après une hydrolyse totale de l'ARN, et l'analyse des oligonucléotides après une digestion partielle, mais spécifique.

3.4.3.1. 1^{er} niveau : hydrolyse totale

Ce premier niveau d'analyse rend possible l'identification des modifications d'un ARN. Il est aussi possible de quantifier les modifications post-transcriptionnelles avec cette stratégie.

L'ARN est tout d'abord totalement hydrolysé afin d'obtenir un mélange de nucléotides. Plusieurs enzymes, aussi appelées nucléases, peuvent être utilisées afin de rompre les liaisons phosphodiesters entre chaque nucléotide. Elles sont classées selon leur spécificité, leur mode d'attaque (exonucléase ou endonucléase) et leur site de coupure de la liaison phosphodiester (Figure 3.11). Un mélange de différentes nucléases peut être utilisé afin de s'assurer que toutes les liaisons phosphodiesters soient bien hydrolysées [71,72]. En effet, il a été observé que certaines modifications montrent une légère résistance à l'hydrolyse totale par la Nucléase P1 [73].

À la suite de l'hydrolyse totale, les nucléotides sont souvent déphosphorylés à l'aide d'une phosphatase avant analyse par spectrométrie de masse afin de mieux les détecter en mode positif.

Les différentes enzymes principalement utilisées lors de cette hydrolyse totale sont répertoriées dans le Tableau 3.2.

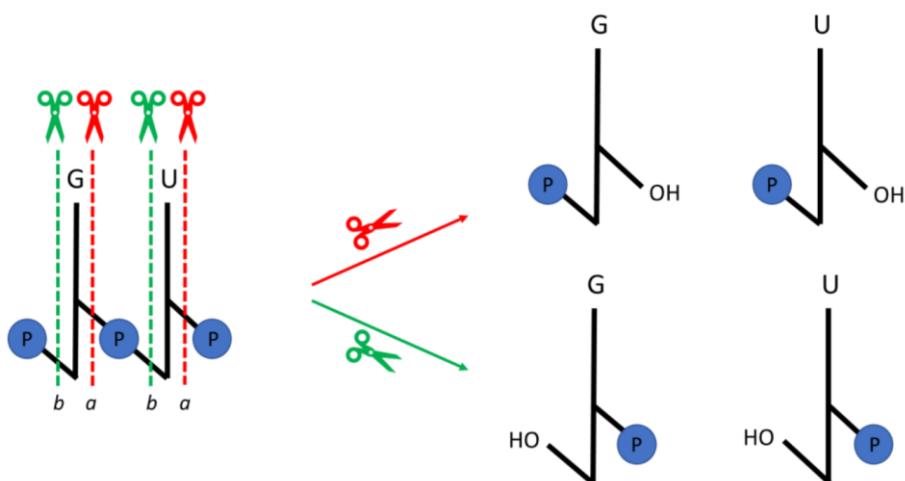


Figure 3.11 : Sites de coupe des liaisons phosphodiesters des ARN et produits formés après coupe. Extrait de [1].

Tableau 3.2 : Principales enzymes utilisées pour l'hydrolyse totale des ARN. NMP : nucléotide monophosphate. Extrait de [1].

Nom de l'enzyme	Résidus spécifiques	Site de coupe	Mode d'action	Produits formés
Nucléase P1	A, C, G, U	a	Endo	5'-NMP
Benzonase	A, C, G, U	a	Endo	5'-NMP
Phosphodiesterase I	A, C, G, U	a	Exo (3'→5')	5'-NMP
Phosphodiesterase II	A, C, G, U	b	Exo (5'→3')	5'-Up-[ARN]-3'-OH
Phosphatase alcaline bactérienne (BAP)	A, C, G, U	a et b	Phosphates terminaux	5'-OH-[ARN]-3'-Cp/C>p

Dans le cadre des travaux présentés dans les prochains chapitres, les ARN ont été hydrolysés selon le protocole suivant [74,75] : à 7 µL d'eau ultra-pure, ont été ajoutés 1 µL de tampon P1 10X (2 mM ZnCl₂, 250 mM NH₄OAc, pH 5,0) puis 1 µL d'ARN à 2 g/L et enfin 1 µL de Nucléase P1 (issue de *Penicillium citrinum* et préalablement préparée à 0,5 U/µL dans 0,02 mM ZnCl₂, 250 mM NH₄OAc). Après une première incubation de 2 h à 37 °C, 10 µL de BAP (issue d'*E. Coli* C90, et préalablement préparée à 1,5 U/µL dans NH₄OAc 100 mM) ont été ajoutés au digestat avant une deuxième incubation de 4 h à 37 °C. Pour finir, les nucléosides ont été évaporés à sec avant d'être repris à la concentration désirée dans de l'acétate d'ammonium à 200 mM et pH 4.

3.4.3.2. 2^{ème} niveau : digestion spécifique

Le second niveau d'analyse permet de localiser les modifications post-transcriptionnelles sur la séquence des ARN [76]. Pour cela, les ARN sont digérés en oligonucléotides, comprenant en général moins d'une dizaine de bases, à l'aide d'enzymes spécifiques, aussi appelées ribonucléases (RNases). Les principales enzymes pouvant être utilisées selon la spécificité désirée sont répertoriées dans le Tableau 3.3.

Tableau 3.3 : Principales enzymes utilisées pour digérer spécifiquement les ARN. Extrait de [1].

Nom de l'enzyme	Résidus spécifiques	Site de coupure	Mode d'action	Oligonucléotides formés
RNase T1	G	b	Endo	5'-OH-[ARN]-3'-Gp/G>p
RNase A	C, U	b	Endo	5'-OH-[ARN]-3'-Cp/C>p, 5'-OH-[ARN]-3'-Up/U>p
RNase U2	G, préférentiellement A	b	Endo	5'-OH-[ARN]-3'-Cp/C>p, 5'-OH-[ARN]-3'-Up/U>p
RNase MC1	Préférentiellement U	b	Endo	5'-Up-[ARN]-3'-OH
Cusativin	Préférentiellement C	b	Endo	5'-OH-[ARN]-3'-Cp/C>p

Du fait de la grande redondance des séquences d'ARN, l'utilisation de plusieurs enzymes spécifiques en parallèle est nécessaire. En effet, si une modification est localisée sur un oligonucléotide redondant, alors la position de la modification sur la séquence totale de l'ARN ne peut pas être connue précisément. En combinant les résultats obtenus pour les différentes digestions, il est alors possible d'obtenir de meilleurs recouvrements des séquences. Ainsi, les RNases T1 et A, enzymes de référence, sont généralement utilisées en parallèle du fait de leur importante spécificité [77,78]. De plus, les digestions obtenues avec ces deux RNases sont très reproductibles puisqu'elles n'induisent pas de « miscleavages ». Il est important de noter qu'elles clivent également les liaisons après certaines modifications telles que m²G pour la RNase T1, ou encore D et ψ pour la RNase A [79].

En plus des enzymes spécifiques, en fonction de la méthode d'analyse et du mode d'ionisation utilisés, il est possible d'ajouter de la BAP lors de la digestion afin de déphosphoryler les extrémités des oligonucléotides.

Dans le cadre des travaux présentés dans les prochains chapitres, les ARN ont été digérés et déphosphorylés avec la RNase T1 associée à la BAP, et en parallèle avec la RNase A associée à la BAP selon le protocole suivant [74] : 10 µL d'un mélange contenant 1000 µL de RNase A ou 100 µL de RNase T1 (préalablement préparées à 1/1000 dans AcONH₄ 100 mM) et 0,5 µL de BAP (150 U/µL) a été ajoutés à 1 µL d'ARN à 0,75 g/L. Après une incubation d'une nuit à 37°C, les oligonucléotides ont été directement analysés.

3.4.4. Analyse par spectrométrie de masse

Au vu de la complexité des échantillons, la spectrométrie de masse peut être couplée à des méthodes séparatives. Pour l'analyse des modifications post-transcriptionnelles des ARN, la chromatographie liquide (LC) et l'électrophorèse capillaire (CE) ont été couplées à la spectrométrie de masse.

3.4.4.1. Les méthodes LC-MS/MS de référence

Les méthodes de référence pour l'analyse des ARN par spectrométrie de masse sont issues du couplage LC-MS/MS.

Au premier niveau d'analyse (le niveau nucléosides), la LC-MS/MS en phase inverse est la méthode de référence pour l'identification et la quantification des modifications post-transcriptionnelles des ARN. Les nucléosides sont obtenus après une hydrolyse totale suivie d'une déphosphorylation, et ils sont détectés en mode positif. Pour identifier les modifications, trois critères sont utilisés : le temps de rétention, la masse de l'ion précurseur et enfin les masses des ions produits. Grâce à la fragmentation des nucléosides par CID, il est possible de déterminer si les modifications sont présentes sur la base ou sur le ribose du nucléoside. En effet, la majorité des nucléosides se fragmentent au niveau de la liaison N-glycosidique. Ainsi, entre une modification sur le ribose et la même modification sur la base, les fragments obtenus par MS/MS ne seront pas les mêmes. Un exemple avec les modifications m⁶A et Am est présenté en Figure 3.12. Dans les méthodes de référence, le mode Multiple Reaction Monitoring (MRM) est utilisé pour suivre les transitions de différentes modifications préalablement sélectionnées [71,72,80–87]. Il s'agit donc d'identifications ciblées.

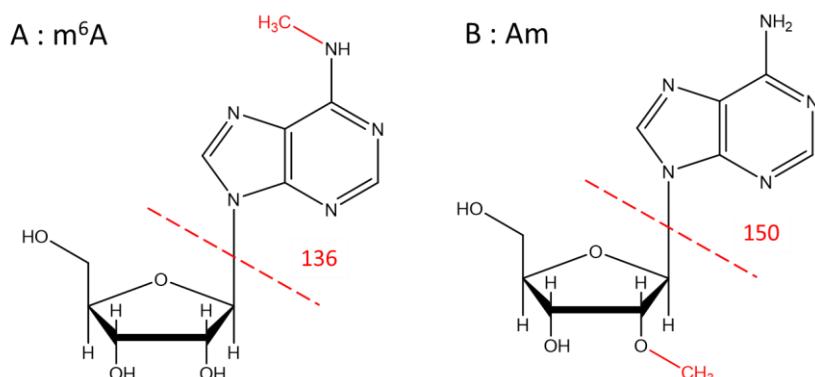


Figure 3.12 : Structures des modifications (A) m⁶A et (B) Am. Les m/z des fragments obtenus par fragmentation CID de la liaison N-glycosidique sont indiqués en rouge.

En ce qui concerne le second niveau d'analyse, la LC-MS/MS est également la méthode de référence pour séquencer les oligonucléotides et ainsi localiser les modifications sur la séquence de l'ARN. Les oligonucléotides sont obtenus après des digestions spécifiques et ils ne sont pas déphosphorylés avant détection en mode négatif. La localisation des modifications sur les oligonucléotides est rendue possible grâce à leur séquençage *De Novo*. Lors de la fragmentation MS/MS des oligonucléotides, les ruptures des liaisons phosphodiesters [88] suivent une nomenclature proposée par McLuckey [89] semblable à celle des peptides (Figure 3.13). Lors de fragmentation par CID, les séries d'ions c et y sont préférentiellement formées [90,91]. Cependant, il est également possible d'observer des ions a-B (perte supplémentaire de la base par rapport à la série a) et des ions w [92]. En faisant correspondre les différences de masses de nucléotides, canoniques ou modifiés, entre les ions d'une même

série d'ions à partir du spectre MS/MS, il est possible de déduire la séquence de l'oligonucléotide fragmenté.

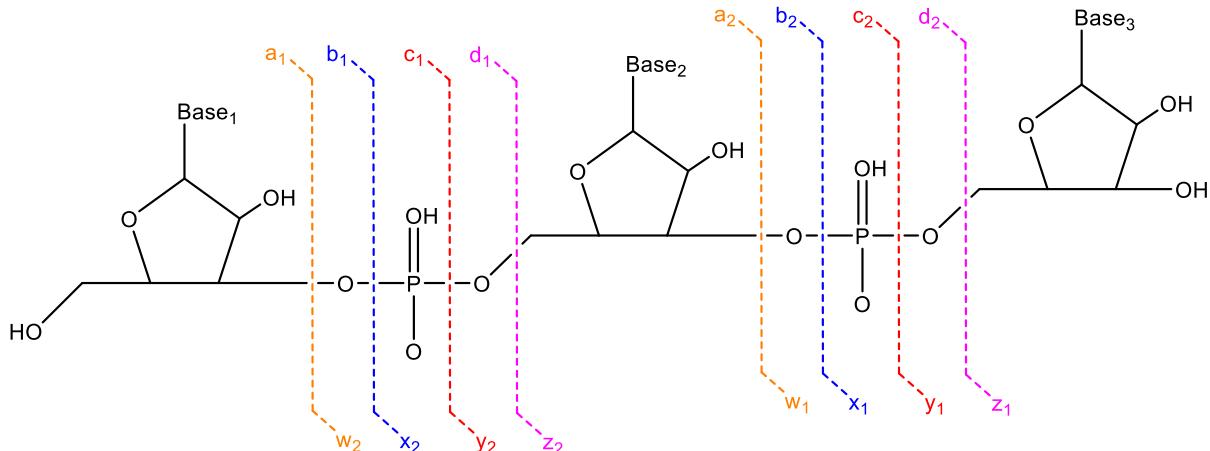


Figure 3.13 : Nomenclature de McLuckey pour la fragmentation des oligonucléotides et le séquençage De Novo.

La méthode LC-MS/MS la plus couramment utilisée est l'IP-RP-LC-MS/MS en mode négatif. Des agents d'appariement d'ions, tel que la triéthylamine, sont nécessaires afin de limiter l'effet de la polarité des groupes phosphodiesters sur la séparation. Cet appariement permet d'augmenter l'hydrophobie des oligonucléotides, et donc d'augmenter les interactions entre les analytes et la phase stationnaire apolaire (C_{18} généralement) afin qu'ils soient mieux retenus. L'utilisation d'un additif, le 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP), est essentielle afin d'assurer une bonne ionisation des oligonucléotides [93]. Cependant la grande concentration de HFIP peut être source de nombreux problèmes techniques. La colonne chromatographique peut facilement se boucher et la constante saturation du détecteur par le HFIP peut réduire sa durée de vie [94].

Au vu de la complexité des échantillons, de la redondance de la séquence des ARN, mais aussi des difficultés techniques amenées par l'usage d'agents d'appariement d'ions et d'additif, il est toujours intéressant d'avoir à disposition un panel de méthodes alternatives et complémentaires.

3.4.4.2. Les méthodes CE-MS/MS

Au laboratoire, une nouvelle méthode CE-MS/MS a été développée sur les deux stratégies d'analyse lors d'une précédente thèse [74]. Il s'agit de la première méthode CZE-MS/MS à avoir été développée pour caractériser les modifications post-transcriptionnelles des ARN.

Le premier niveau d'analyse par CE-MS/MS requiert la même préparation d'échantillon que les méthodes LC-MS/MS de référence : une hydrolyse totale de l'ARN suivie d'une déphosphorylation. La séparation par CE est réalisée dans un capillaire de silice vierge de 100 cm de long et 30 μm de diamètre interne dont l'extrémité devant l'entrée du spectromètre de masse a été rendue poreuse. Une injection hydrodynamique de 5 nL d'échantillon a lieu dans

le capillaire rempli d'électrolyte support (acide acétique 10 %) avant d'appliquer une tension de +30 kV aux extrémités du capillaire pendant 40 min. Au niveau de la détection MS en mode positif, l'acquisition MS/MS n'est pas réalisée en mode MRM comme la méthode de référence, mais en mode « data-dependent » (DDA). Le mode DDA permet une identification non ciblée des modifications post-transcriptionnelles présentes dans l'échantillon.

Pour le second niveau d'analyse, deux analyses d'oligonucléotides sont réalisées après une digestion spécifique, à la RNase T1 pour l'une des analyses et à la RNase A pour l'autre, chacune suivie d'une déphosphorylation à la BAP. La méthode de séparation par CE est identique à celle du premier niveau. C'est un des grands avantages de la méthode CE-MS/MS développée au laboratoire puisqu'elle permet de passer facilement d'un niveau d'analyse à l'autre en routine, contrairement aux méthodes LC-MS/MS de référence. Au niveau de la détection par MS, c'est encore une fois le mode positif qui est utilisé, contrairement au mode négatif des méthodes LC-MS/MS. Le choix du mode positif est basé sur l'utilisation de l'interface « sheathless » pour le couplage CE-MS/MS. En effet, une ionisation en mode négatif demanderait de travailler avec un électrolyte support en conditions basiques. Or de telles conditions engendreraient un fort échauffement du capillaire par effet Joule, et par conséquent, un plus grand risque de casse de la pointe poreuse du capillaire. Enfin, la dernière différence majeure avec la méthode LC-MS/MS est l'étape de déphosphorylation supplémentaire avant analyse par CE-MS/MS. Cette dernière est essentielle pour l'analyse en CE-MS/MS dans nos conditions. En effet, les charges négatives portées par les groupements phosphates participent à l'allongement des temps de migration et à l'élargissement des pics électrophorétiques. Ainsi, en plus de bénéficier d'une meilleure détection en mode positif, la déphosphorylation des phosphates terminaux de chaque oligonucléotide permet d'obtenir une meilleure efficacité de pics ainsi que des temps d'analyse réduits (inférieurs à 1 h 30). La preuve de concept, réalisée sur des ARNt modèles, avait montré qu'il était possible d'obtenir des recouvrements de séquence allant jusque 96% en combinant les résultats des deux analyses d'oligonucléotides [74].

Depuis peu, un kit pour « ZipChip » permettant d'analyser les oligonucléotides est en vente. Le « ZipChip » est un module de CE microfluidique gravé sur une puce, développé par 908devices. Le module se positionne directement au niveau de la source ESI du spectromètre de masse. La méthode pour l'analyse des oligonucléotides est basée sur l'utilisation d'un électrolyte support en conditions basiques, bien que la note d'application disponible pour ce développement soit avec une détection en mode positif [95]. Les avantages apportés par un tel système sont le gain de temps (comme la CE est miniaturisée, les temps d'analyse sont réduits) et la facilité de sa mise en place en routine. Cependant, ce type de « kit » dont on ne connaît pas réellement la composition (d'où le surnom « boîte noire » souvent employé) ne permet pas de faire du développement de méthode. Le module à placer devant le spectromètre de masse, ainsi que la cartouche et la puce en verre où sont gravés les canaux microfluidiques, sont représentés en Figure 3.14.



Figure 3.14 : Système CE-MS microfluidique « ZipChip » développé par 908devices avec (A) un module qui se positionne directement à l'entrée des spectromètres de masse et (B) la cartouche dans laquelle se glisse (C) la puce en verre où sont gravés les canaux microfluidiques (908devices). Extrait et adapté de [96].

D'autres méthodes CE-MS ont été développées par le passé pour les analyses de nucléotides ou d'oligonucléotides [97–101], mais elles n'ont pas pour but de séquencer et de localiser les modifications post-transcriptionnelles. Nous avons donc fait le choix de ne pas les détailler dans ce manuscrit.

3.5. L'interprétation des données

L'envie et la nécessité d'aller toujours plus loin dans la caractérisation des biomolécules, avec des échantillons de plus en plus complexes, et avec un débit de plus en plus élevé, lèvent la question du traitement des données. Bien que les développements de logiciels pour les analyses d'ARN aient commencé depuis les dernières années, il existe à ce jour encore peu de logiciels pour interpréter les données, que ce soit pour l'analyse des nucléosides ou pour celles des oligonucléotides.

3.5.1. Premier niveau d'analyse : identification des nucléosides

Comme expliqué dans le paragraphe 3.4.4.1, la méthode de référence pour l'identification des modifications post-transcriptionnelles est la LC-MS/MS en mode MRM. Le mode MRM permet une interprétation facile et rapide des résultats avec les différents logiciels disponibles.

Dans le cas de la méthode CE-MS/MS, le choix du mode DDA implique la recherche manuelle des différentes modifications. Pour les échantillons dont les modifications sont déjà référencées dans la littérature, les modifications attendues sont recherchées et confirmées. Pour les échantillons dont les modifications n'ont pas encore été identifiées, il est nécessaire de rechercher manuellement toutes les modifications référencées dans la base de données Modomics [19]. Cette base de données est régulièrement mise à jour et compte actuellement plus de 160 modifications. Une telle recherche manuelle non ciblée peut prendre plusieurs heures et être très fastidieuse. Comme certaines modifications sont connues pour être présentes uniquement dans certains des trois domaines des micro-organismes (eubactéries, eucaryotes ou archées), il est possible de concentrer la recherche sur les modifications du

domaine dont fait partie l'échantillon étudié. Cependant, le traitement de données reste malgré tout chronophage.

3.5.2. Second niveau d'analyse : séquençage des oligonucléotides

Une fois encore, la méthode de référence pour l'analyse des oligonucléotides étant la LC-MS/MS en mode négatif, les premiers outils bio-informatiques qui ont à ce jour été développés sont adaptés pour de telles analyses. Le premier outil en libre accès (open-source) pour aider à l'annotation des spectres MS/MS d'oligonucléotides a été publié en 2002, il s'agit de Simple Oligonucleotide Sequencer (SOS) [102]. En 2009, Ariadne, un premier outil en ligne d'interprétation automatique d'analyses LC-MS/MS d'oligonucléotides a été développé [103]. Ces dernières années, il y a eu une accélération du nombre de logiciels open-source développés pour les analyses LC-MS/MS en mode négatif : RoboOligo en 2015 [104], RNAModMapper en 2017 [105] (mis à jour en 2019 [106]), NucleicAcidSearchEngine en 2020 [107] et enfin Pytheas en 2022 [108]. ThermoFisher Scientific a également étendu leur logiciel commercial BioPharma Finder à l'analyse des oligonucléotides.

Cependant, à notre connaissance, uniquement deux outils bio-informatiques ont récemment implémenté le mode positif nécessaire pour être utilisé avec nos données CE-MS/MS. Le premier est l'outil en ligne Ariadne. Le mode positif, ainsi que la RNase T1 associée à la BAP en choix d'enzymes de digestion ont toutes les deux été rajoutées bien après la première publication en 2009. Deux points négatifs peuvent être relevés. Tout d'abord, dans le choix des enzymes de digestion, la BAP n'est associée qu'à la RNase T1, et non aux autres enzymes disponibles telles que la RNase A. De plus, le site ne semble pas totalement opérationnel puisque l'erreur « internal server error » ne cesse d'apparaître malgré de nombreux essais suivant scrupuleusement le manuel d'utilisation. Le deuxième outil prenant en compte le mode positif, Pytheas, a été publié récemment (mai 2022) [108]. Bien qu'une interface graphique soit mentionnée dans la description du logiciel, son utilisation nécessite un appel du code open-source par des lignes de commandes. Il est donc difficile d'utiliser ce logiciel sans avoir de connaissances en informatique. Aucun de ces logiciels ne répondant à nos attentes pour interpréter les résultats d'analyse CE-MS/MS en mode positif, le traitement de données est fait manuellement.

Pour faciliter l'interprétation manuelle, il existe tout de même quelques outils. Tout d'abord, la base de données Modomics recense également les séquences des ARN déjà étudiés [19]. Il est donc possible de se baser sur ces séquences pour interpréter les résultats obtenus pour des échantillons bien décrits dans la littérature. Ensuite, MongoOligo (<http://rna.rega.kuleuven.be/masspec/mongo.htm>), un simulateur en ligne, permet d'effectuer des digestions *in silico* de séquences d'ARN. Cependant ces digestions ne prennent pas en compte la déphosphorylation à la BAP, et ne coupent pas après les modifications. Une vérification manuelle est donc nécessaire avant de réutiliser la liste d'oligonucléotides obtenue. MongoOligo permet également de calculer les différents états de charges des

oligonucléotides ainsi que les fragments attendus lors du séquençage *De Novo*. Ces informations sont d'une grande utilité et facilitent grandement le séquençage manuel, puisqu'il suffit de vérifier que les spectres expérimentaux correspondent bien aux informations théoriques simulées à l'aide de MongoOligo. Cependant, dans le cas d'ARN dont la séquence ou les modifications ne sont pas encore décrites dans la littérature, il n'est pas possible d'utiliser MongoOligo de la même manière.

En fonction de la complexité de l'échantillon, et des connaissances préalables sur ses modifications post-transcriptionnelles, l'interprétation manuelle peut donc s'avérer très chronophage (au minimum plusieurs heures).

3.6. Références

- [1] Lechner, A., "Développement de l'électrophorèse Capillaire Couplée à La Spectrométrie de Masse Pour l'étude Des Modifications Post-Transcriptionnelles Des ARN." Thèse de doctorat, Strasbourg, 2020.
- [2] Watson, J. D. *et al.*, "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid.", *Nature*, **1953**, 171 (4356), 737–738.
- [3] Varani, G. *et al.*, "The G-U Wobble Base Pair.", *EMBO Rep.*, **2000**, 1 (1), 18–23.
- [4] Walsh, E. J., "Biochemistry (Garrett, Reginald H.; Grisham, Charles M.).", *J. Chem. Educ.*, **1997**, 74 (2), 189.
- [5] Sanchez de Groot, N. *et al.*, "RNA Structure Drives Interaction with Proteins.", *Nat. Commun.*, **2019**, 10 (1), 3246.
- [6] Agirrezabala, X. *et al.*, "Structural Insights into tRNA Dynamics on the Ribosome.", *Int. J. Mol. Sci.*, **2015**, 16 (5), 9866–9895.
- [7] Alberts, B. *et al.*, "From RNA to Protein.", *Mol. Biol. Cell 4th Ed.*, **2002**.
- [8] Lilley, D. M. J., "The Origins of RNA Catalysis in Ribozymes.", *Trends Biochem. Sci.*, **2003**, 28 (9), 495–501.
- [9] Statello, L. *et al.*, "Gene Regulation by Long Non-Coding RNAs and Its Biological Functions.", *Nat. Rev. Mol. Cell Biol.*, **2021**, 22 (2), 96–118.
- [10] Krahm, N. *et al.*, "Naturally Occurring tRNAs With Non-Canonical Structures.", *Front. Microbiol.*, **2020**, 11.
- [11] Cech, T. R., "The Ribosome Is a Ribozyme.", *Science*, **2000**, 289 (5481), 878–879.
- [12] Williamson, J. R., "Assembly of the 30S ribosomal subunit.", *Q. Rev. Biophys.*, **2006**, 38 (4), 397–403.
- [13] Morris, K. V. *et al.*, "The Rise of Regulatory RNA.", *Nat. Rev. Genet.*, **2014**, 15 (6), 423–437.
- [14] Wagner, E. G. H. *et al.*, "Chapter Three - Small RNAs in Bacteria and Archaea: Who They Are, What They Do, and How They Do It." in: *Advances in Genetics*; Friedmann, T. *et al.*, Eds.; Academic Press, 2015; Vol. 90, pp 133–208.
- [15] Bernstein, E. *et al.*, "Role for a Bidentate Ribonuclease in the Initiation Step of RNA Interference.", *Nature*, **2001**, 409 (6818), 363–366.
- [16] Bartel, D. P., "MicroRNAs: Genomics, Biogenesis, Mechanism, and Function.", *Cell*, **2004**, 116 (2), 281–297.
- [17] Allmang, C. *et al.*, "Functions of the Exosome in rRNA, SnoRNA and SnRNA Synthesis.", *EMBO J.*, **1999**, 18 (19), 5399–5410.

- [18] Bachellerie, J. P. *et al.*, "The Expanding SnoRNA World.", *Biochimie*, **2002**, 84 (8), 775–790.
- [19] Boccaletto, P. *et al.*, "MODOMICS: A Database of RNA Modification Pathways. 2021 Update.", *Nucleic Acids Res.*, **2022**, 50 (D1), D231–D235.
- [20] El Yacoubi, B. *et al.*, "Biosynthesis and Function of Posttranscriptional Modifications of Transfer RNAs.", *Annu. Rev. Genet.*, **2012**, 46, 69–95.
- [21] Carell, T. *et al.*, "Structure and Function of Noncanonical Nucleobases.", *Angew. Chem. Int. Ed.*, **2012**, 51 (29), 7110–7131.
- [22] Sampson, J. R. *et al.*, "Role of the Tertiary Nucleotides in the Interaction of Yeast Phenylalanine tRNA with Its Cognate Synthetase.", *Biochemistry*, **1990**, 29 (10), 2523–2532.
- [23] Biedenbänder, T. *et al.*, "RNA Modifications Stabilize the Tertiary Structure of tRNAsfMet by Locally Increasing Conformational Dynamics.", *Nucleic Acids Res.*, **2022**, 50 (4), 2334–2349.
- [24] Chawla, M. *et al.*, "An Atlas of RNA Base Pairs Involving Modified Nucleobases with Optimal Geometries and Accurate Energies.", *Nucleic Acids Res.*, **2015**, 43 (14), 6714–6729.
- [25] Agris, P. F. *et al.*, "Chapter One - The Importance of Being Modified: The Role of RNA Modifications in Translational Fidelity." in: *The Enzymes*; Chanfreau, G. F., Ed.; RNA Modification; Academic Press, 2017; Vol. 41, pp 1–50.
- [26] Motorin, Y. *et al.*, "RNA Nucleotide Methylation.", *WIREs RNA*, **2011**, 2 (5), 611–631.
- [27] Wilkinson, M. L. *et al.*, "The 2'-O-Methyltransferase Responsible for Modification of Yeast tRNA at Position 4.", *RNA N. Y. N.*, **2007**, 13 (3), 404–413.
- [28] Steinberg, S. *et al.*, "A Correlation between N2-Dimethylguanosine Presence and Alternate tRNA Conformers.", *RNA N. Y. N.*, **1995**, 1 (9), 886–891.
- [29] Arnez, J. G. *et al.*, "Crystal Structure of Unmodified tRNA(Gln) Complexed with Glutaminyl-tRNA Synthetase and ATP Suggests a Possible Role for Pseudo-Uridines in Stabilization of RNA Structure.", *Biochemistry*, **1994**, 33 (24), 7560–7567.
- [30] Kierzek, E. *et al.*, "The Contribution of Pseudouridine to Stabilities and Structure of RNAs.", *Nucleic Acids Res.*, **2014**, 42 (5), 3492–3501.
- [31] Westhof, E., "Pseudouridines or How to Draw on Weak Energy Differences.", *Biochem. Biophys. Res. Commun.*, **2019**, 520 (4), 702–704.
- [32] Giegé, R. *et al.*, "Structure of Transfer RNAs: Similarity and Variability.", *Wiley Interdiscip. Rev. RNA*, **2012**, 3 (1), 37–61.
- [33] Torres, A. G. *et al.*, "Role of tRNA Modifications in Human Diseases.", *Trends Mol. Med.*, **2014**, 20 (6), 306–314.
- [34] Novoa, E. M. *et al.*, "A Role for tRNA Modifications in Genome Structure and Codon Usage.", *Cell*, **2012**, 149 (1), 202–213.
- [35] Rafels-Ybern, À. *et al.*, "The Expansion of Inosine at the Wobble Position of tRNAs, and Its Role in the Evolution of Proteomes.", *Mol. Biol. Evol.*, **2019**, 36 (4), 650–662.
- [36] Schaub, M. *et al.*, "RNA Editing by Adenosine Deaminases Generates RNA and Protein Diversity.", *Biochimie*, **2002**, 84 (8), 791–803.
- [37] Grosjean, H. *et al.*, "Deciphering Synonymous Codons in the Three Domains of Life: Co-Evolution with Specific tRNA Modification Enzymes.", *FEBS Lett.*, **2010**, 584 (2), 252–264.
- [38] Weixlbaumer, A. *et al.*, "Mechanism for Expanding the Decoding Capacity of Transfer RNAs by Modification of Uridines.", *Nat. Struct. Mol. Biol.*, **2007**, 14 (6), 498–502.

- [39] Agris, P. F. *et al.*, “tRNA’s Wobble Decoding of the Genome: 40 Years of Modification.”, *J. Mol. Biol.*, **2007**, *366* (1), 1–13.
- [40] Thiaville, P. C. *et al.*, “Essentiality of Threonylcarbamoyladenosine (t(6)A), a Universal tRNA Modification, in Bacteria.”, *Mol. Microbiol.*, **2015**, *98* (6), 1199–1221.
- [41] Vester, B. *et al.*, *Antibiotic Resistance in Bacteria Caused by Modified Nucleosides in 23S Ribosomal RNA*; Landes Bioscience, 2009.
- [42] Jonkhout, N. *et al.*, “The RNA Modification Landscape in Human Disease.”, *RNA*, **2017**, *23* (12), 1754–1769.
- [43] Gerken, T. *et al.*, “The Obesity-Associated FTO Gene Encodes a 2-Oxoglutarate-Dependent Nucleic Acid Demethylase.”, *Science*, **2007**, *318* (5855), 1469–1472.
- [44] Kirino, Y. *et al.*, “Codon-Specific Translational Defect Caused by a Wobble Modification Deficiency in Mutant tRNA from a Human Mitochondrial Disease.”, *Proc. Natl. Acad. Sci. U. S. A.*, **2004**, *101* (42), 15070–15075.
- [45] Barbieri, I. *et al.*, “Role of RNA Modifications in Cancer.”, *Nat. Rev. Cancer*, **2020**, *20* (6), 303–322.
- [46] Bednářová, A. *et al.*, “Lost in Translation: Defects in Transfer RNA Modifications and Neurological Disorders.”, *Front. Mol. Neurosci.*, **2017**, *10*.
- [47] Amalric, A. *et al.*, “Quantifying RNA Modifications by Mass Spectrometry: A Novel Source of Biomarkers in Oncology.”, *Crit. Rev. Clin. Lab. Sci.*, **2022**, *59* (1), 1–18.
- [48] Huang, B. S. *et al.*, “Relationship of the Queuine Content of Transfer Ribonucleic Acids to Histopathological Grading and Survival in Human Lung Cancer.”, *Cancer Res.*, **1992**, *52* (17), 4696–4700.
- [49] Stockert, J. A. *et al.*, “Pseudouridine as a Novel Biomarker in Prostate Cancer.”, *Urol. Oncol. Semin. Orig. Investig.*, **2021**, *39* (1), 63–71.
- [50] Liebich, H. M. *et al.*, “Quantitation of Urinary Nucleosides by High-Performance Liquid Chromatography.”, *J. Chromatogr. A*, **1997**, *763* (1–2), 193–197.
- [51] Jiang, Y. *et al.*, “A Fast Capillary Electrophoresis Method for Separation and Quantification of Modified Nucleosides in Urinary Samples.”, *Anal. Chem.*, **2009**, *81* (15), 6474–6480.
- [52] Feng, B. *et al.*, “Normal and Modified Urinary Nucleosides Represent Novel Biomarkers for Colorectal Cancer Diagnosis and Surgery Monitoring.”, *J. Gastroenterol. Hepatol.*, **2005**, *20* (12), 1913–1919.
- [53] Zheng, Y.-F. *et al.*, “Clinical Significance and Prognostic Value of Urinary Nucleosides in Breast Cancer Patients.”, *Clin. Biochem.*, **2005**, *38* (1), 24–30.
- [54] La, S. *et al.*, “Capillary Electrophoretic Profiling and Pattern Recognition Analysis of Urinary Nucleosides from Thyroid Cancer Patients.”, *Anal. Chim. Acta*, **2003**, *486* (2), 171–182.
- [55] Rasmuson, T. *et al.*, “Urinary Excretion of Pseudouridine and Prognosis of Patients with Malignant Lymphoma.”, *Acta Oncol.*, **1995**, *34* (1), 61–67.
- [56] Staszewski, S. *et al.*, “Efavirenz plus Zidovudine and Lamivudine, Efavirenz plus Indinavir, and Indinavir plus Zidovudine and Lamivudine in the Treatment of HIV-1 Infection in Adults.”, *N. Engl. J. Med.*, **1999**, *341* (25), 1865–1873.
- [57] Fletcher, C. V. *et al.*, “Concentration-Controlled Zidovudine Therapy.”, *Clin. Pharmacol. Ther.*, **1998**, *64* (3), 331–338.
- [58] Paterson, B. M. *et al.*, “Structural Gene Identification and Mapping by DNA-MRNA Hybrid-Arrested Cell-Free Translation.”, *Proc. Natl. Acad. Sci. U. S. A.*, **1977**, *74* (10), 4370–4374.

- [59] Sahin, U. *et al.*, "Personalized RNA Mutanome Vaccines Mobilize Poly-Specific Therapeutic Immunity against Cancer.", *Nature*, **2017**, 547 (7662), 222–226.
- [60] Passini, M. A. *et al.*, "Antisense Oligonucleotides Delivered to the Mouse CNS Ameliorate Symptoms of Severe Spinal Muscular Atrophy.", *Sci. Transl. Med.*, **2011**, 3 (72), 72ra18.
- [61] Dias, N. *et al.*, "Antisense Oligonucleotides: Basic Concepts and Mechanisms.", *Mol. Cancer Ther.*, **2002**, 1 (5), 347–355.
- [62] Chen, B. *et al.*, "Analytical Methods for Deciphering RNA Modifications.", *Anal. Chem.*, **2019**, 91 (1), 743–756.
- [63] Motorin, Y. *et al.*, "Identification of Modified Residues in RNAs by Reverse Transcription-Based Methods.", *Methods Enzymol.*, **2007**, 425, 21–53.
- [64] Limbach, P. A., "Indirect Mass Spectrometric Methods for Characterizing and Sequencing Oligonucleotides.", *Mass Spectrom. Rev.*, **1996**, 15 (5), 297–336.
- [65] Kim, J. *et al.*, "Microfluidic Sample Preparation: Cell Lysis and Nucleic Acid Purification.", *Integr. Biol.*, **2009**, 1 (10), 574–586.
- [66] Zhang, J. *et al.*, "Rapid and Label-Free Nanomechanical Detection of Biomarker Transcripts in Human RNA.", *Nat. Nanotechnol.*, **2006**, 1 (3), 214–220.
- [67] Boom, R. *et al.*, "Rapid and Simple Method for Purification of Nucleic Acids.", *J. Clin. Microbiol.*, **1990**, 28 (3), 495–503.
- [68] Akhtar, S. *et al.*, "A Method to Extract Intact and Pure RNA from Mycobacteria.", *Anal. Biochem.*, **2011**, 417 (2), 286–288.
- [69] Chomczynski, P. *et al.*, "The Single-Step Method of RNA Isolation by Acid Guanidinium Thiocyanate-Phenol-Chloroform Extraction: Twenty-Something Years On.", *Nat. Protoc.*, **2006**, 1 (2), 581–585.
- [70] Antoine, L. *et al.*, "Mapping Post-Transcriptional Modifications in *Staphylococcus Aureus* tRNAs by NanoLC/MSMS.", *Biochimie*, **2019**, 164, 60–69.
- [71] Thüring, K. *et al.*, "Analysis of RNA Modifications by Liquid Chromatography–Tandem Mass Spectrometry.", *Methods*, **2016**, 107, 48–56.
- [72] Su, D. *et al.*, "Quantitative Analysis of Ribonucleoside Modifications in tRNA by HPLC-Coupled Mass Spectrometry.", *Nat. Protoc.*, **2014**, 9 (4), 828–841.
- [73] Crain, P. F., "[42] Preparation and Enzymatic Hydrolysis of DNA and RNA for Mass Spectrometry." in: *Methods in Enzymology*; Mass Spectrometry; Academic Press, 1990; Vol. 193, pp 782–790.
- [74] Lechner, A. *et al.*, "Characterization of Post-Transcriptional RNA Modifications by Sheathless Capillary Electrophoresis–High Resolution Mass Spectrometry.", *Anal. Chem.*, **2020**, 92 (10), 7363–7370.
- [75] Gosset-Erard, C. *et al.*, "Optimization of Nucleotides Dephosphorylation for RNA Structural Characterization by Tandem Mass Spectrometry Hyphenated with Separation Methods.", *J. Chromatogr. B*, **2022**, 1208, 123396.
- [76] Kowalak, J. A. *et al.*, "A Novel Method for the Determination of Post-Transcriptional Modification in RNA by Mass Spectrometry.", *Nucleic Acids Res.*, **1993**, 21 (19), 4577–4585.
- [77] Pace, C. N. *et al.*, "Ribonuclease T1: Structure, Function, and Stability.", *Angew. Chem. Int. Ed. Engl.*, **1991**, 30 (4), 343–360.
- [78] Raines, R. T., "Ribonuclease A.", *Chem. Rev.*, **1998**, 98 (3), 1045–1066.

- [79] Ross, R. L. et al., “Mapping Post-Transcriptional Modifications onto Transfer Ribonucleic Acid Sequences by Liquid Chromatography Tandem Mass Spectrometry.”, *Biomolecules*, **2017**, 7 (1), E21.
- [80] Yan, M. et al., “A High-Throughput Quantitative Approach Reveals More Small RNA Modifications in Mouse Liver and Their Correlation with Diabetes.”, *Anal. Chem.*, **2013**, 85 (24), 12173–12181.
- [81] Wolff, P. et al., “Comparative Patterns of Modified Nucleotides in Individual tRNA Species from a Mesophilic and Two Thermophilic Archaea.”, *RNA*, **2020**, rna.077537.120.
- [82] Liu, R. et al., “Determination of Five Nucleosides by LC-MS/MS and the Application of the Method to Quantify N6-Methyladenosine Level in Liver Messenger Ribonucleic Acid of an Acetaminophen-Induced Hepatotoxicity Mouse Model.”, *J. Sep. Sci.*, **2019**, 42 (16), 2668–2678.
- [83] Raćkowska, E. et al., “Development and Validation of a Rapid LC-MS/MS Method for Determination of Methylated Nucleosides and Nucleobases in Urine.”, *J. Chromatogr. B*, **2019**, 1128, 121775.
- [84] Lo, W.-Y. et al., “Urinary Cytidine as an Adjunct Biomarker to Improve the Diagnostic Ratio for Gastric Cancer in Taiwanese Patients.”, *Clin. Chim. Acta*, **2014**, 428, 57–62.
- [85] Feng, Y.-J. et al., “Identification of Inosine and 2'-O-Methylinosine Modifications in Yeast Messenger RNA by Liquid Chromatography-Tandem Mass Spectrometry Analysis.”, *Anal. Chem.*, **2022**, 94 (11), 4747–4755.
- [86] Wang, J. et al., “The Role of Sequence Context, Nucleotide Pool Balance and Stress in 2'-Deoxynucleotide Misincorporation in Viral, Bacterial and Mammalian RNA.”, *Nucleic Acids Res.*, **2016**, 44 (18), 8962–8975.
- [87] Gątarek, P. et al., “LC-MS/MS Determination of Modified Nucleosides in The Urine of Parkinson’s Disease and Parkinsonian Syndromes Patients.”, *Molecules*, **2020**, 25 (21), 4959.
- [88] Phillips, D. R. et al., “A Comprehensive Study of the Low Energy Collision-Induced Dissociation of Dinucleoside Monophosphates.”, *Int. J. Mass Spectrom. Ion Process.*, **1993**, 128 (1), 61–82.
- [89] McLuckey, S. A. et al., “Decompositions of Multiply Charged Oligonucleotide Anions.”, *J. Am. Chem. Soc.*, **1993**, 115 (25), 12085–12095.
- [90] Tromp, J. M. et al., “Gas-Phase Dissociation of Oligoribonucleotides and Their Analogs Studied by Electrospray Ionization Tandem Mass Spectrometry.”, *J. Am. Soc. Mass Spectrom.*, **2005**, 16 (8), 1262–1268.
- [91] Andersen, T. E. et al., “RNA Fragmentation in MALDI Mass Spectrometry Studied by H/D-Exchange: Mechanisms of General Applicability to Nucleic Acids.”, *J. Am. Soc. Mass Spectrom.*, **2006**, 17 (10), 1353–1368.
- [92] Wu, J. et al., “Gas-Phase Fragmentation of Oligonucleotide Ions.”, *Int. J. Mass Spectrom.*, **2004**, 237 (2), 197–241.
- [93] Apffel, A. et al., “Analysis of Oligonucleotides by HPLC-Electrospray Ionization Mass Spectrometry.”, *Anal. Chem.*, **1997**, 69 (7), 1320–1325.
- [94] Lobue, P. A. et al., “Oligonucleotide Analysis by Hydrophilic Interaction Liquid Chromatography-Mass Spectrometry in the Absence of Ion-Pair Reagents.”, *J. Chromatogr. A*, **2019**, 1595, 39–48.
- [95] 908devices, *ZipChip - Rapid characterization of oligonucleotides using microfluidic capillary electrophoresis - mass spectrometry by ZipChip*, Application Note 9.5.

- https://908devices.com/wp-content/uploads/2021/11/ZipChip_APP_NOTE_Oligos_2021_v6.pdf.
- [96] 908devices, *ZipChip -What they are and how they work, Tech Note 1.0.* https://908devices.com/wp-content/uploads/2017/03/ZipChip_TECH-NOTE_1-0_WhatTheyAre-Work_12-19.pdf.
- [97] Khan, N. *et al.*, “Direct Detection of Endogenous MicroRNAs and Their Post-Transcriptional Modifications in Cancer Serum by Capillary Electrophoresis-Mass Spectrometry.”, *Anal. Bioanal. Chem.*, **2016**, 408 (11), 2891–2899.
- [98] Yu, Y. *et al.*, “Ultrasensitive and Simultaneous Determination of RNA Modified Nucleotides by Sheathless Interfaced Capillary Electrophoresis-Tandem Mass Spectrometry.”, *Chem. Commun.*, **2019**, 55 (53), 7595–7598.
- [99] Zhao, Z. (Zack) *et al.*, “On-Line Capillary Electrophoresis-Electrospray Ionization Mass Spectrometry of Nucleotides.”, *ELECTROPHORESIS*, **1995**, 16 (1), 389–395.
- [100] Mateos-Vivas, M. *et al.*, “Capillary Electrophoresis Coupled to Mass Spectrometry Employing Hexafluoro-2-Propanol for the Determination of Nucleosides and Nucleotide Mono-, Di- and Tri-Phosphates in Baby Foods.”, *Food Chem.*, **2017**, 233, 38–44.
- [101] Demelenne, A. *et al.*, “Evaluation of Hydrophilic Interaction Liquid Chromatography, Capillary Zone Electrophoresis and Drift Tube Ion-Mobility Quadrupole Time of Flight Mass Spectrometry for the Characterization of Phosphodiester and Phosphorothioate Oligonucleotides.”, *J. Chromatogr. A*, **2020**, 1614, 460716.
- [102] Rozenski, J. *et al.*, “SOS: A Simple Interactive Program for Ab Initio Oligonucleotide Sequencing by Mass Spectrometry.”, *J. Am. Soc. Mass Spectrom.*, **2002**, 13 (3), 200–203.
- [103] Nakayama, H. *et al.*, “Ariadne: A Database Search Engine for Identification and Chemical Analysis of RNA Using Tandem Mass Spectrometry Data.”, *Nucleic Acids Res.*, **2009**, 37 (6), e47.
- [104] Sample, P. J. *et al.*, “RoboOligo: Software for Mass Spectrometry Data to Support Manual and de Novo Sequencing of Post-Transcriptionally Modified Ribonucleic Acids.”, *Nucleic Acids Res.*, **2015**, 43 (10), e64.
- [105] Yu, N. *et al.*, “RNAModMapper: RNA Modification Mapping Software for Analysis of Liquid Chromatography Tandem Mass Spectrometry Data.”, *Anal. Chem.*, **2017**, 89 (20), 10744–10752.
- [106] Lobue, P. A. *et al.*, “Improved Application of RNAModMapper – An RNA Modification Mapping Software Tool – For Analysis of Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) Data.”, *Methods*, **2019**, 156, 128–138.
- [107] Wein, S. *et al.*, “A Computational Platform for High-Throughput Analysis of RNA Sequences and Modifications by Mass Spectrometry.”, *Nat. Commun.*, **2020**, 11 (1), 926.
- [108] D'Ascenzo, L. *et al.*, “Pytheas: A Software Package for the Automated Analysis of RNA Sequences and Modifications via Tandem Mass Spectrometry.”, *Nat. Commun.*, **2022**, 13 (1), 2424.

PARTIE II : MISE EN PLACE ET ETUDE DU COUPLAGE CE-FTICR-MS

Les progrès en matière d'instrumentation FTICR-MS amènent les laboratoires de recherche spécialisés dans le développement méthodologique sur ce genre d'appareil à acquérir de nouveaux spectromètres permettant d'aller toujours plus loin dans l'exploration du contenu des mélanges complexes. Dans cette optique, le projet RESEX a permis l'acquisition d'un spectromètre FTICR-MS Solarix 7 T 2XR (Bruker Daltonics). Comme l'a montré le chapitre 2 , le passage d'une paire de plaques de détection à deux paires dans la cellule du 2XR permet, soit d'atteindre la performance de mesure d'un aimant de 14 T pour un aimant de 7 T, soit de conserver les performances d'un aimant à 7 T, mais en travaillant deux fois plus vite que les appareils à 7 T d'ancienne génération. Les vitesses d'acquisition deviennent alors plus compatibles avec des techniques séparatives efficaces (à nombre de plateaux théoriques élevé) avec lesquelles les spectromètres FTICR-MS peuvent alors être couplés plus facilement. C'est ainsi que le laboratoire a pu envisager de se lancer dans le couplage CE-FTICR-MS avec l'acquisition d'une électrophorèse capillaire CESI 8000 (Sciex Separations). Le couplage CE-FTICR-MS étant tout nouveau au laboratoire, cette partie du manuscrit est dédiée, dans un premier temps, à la mise en place de ce nouveau couplage, réalisée en grande partie sur un mélange de standards, et dans un second temps, au transfert de méthode sur de vrais échantillons biologiques.

Dans le premier chapitre, la mise en place du couplage sera détaillée, puis l'optimisation de la méthode de détection pour l'analyse de nucléosides réalisée en infusant un digestat de tRNA purifié sera présentée. Il se terminera sur une comparaison entre les performances de ce couplage CE-FTICR-MS, et celles du couplage CE-QTOF-MS, qui est celui le plus couramment mis en place.

Dans le second chapitre, le couplage sera évalué sur divers échantillons biologiques d'ARNt, dont les modifications post-transcriptionnelles sont déjà référencées dans la littérature. Les résultats donnés par les deux couplages seront alors comparés.

CHAPITRE 4 : MISE EN PLACE ET ETUDE DU COUPLAGE CE-FTICR-MS

4.1. Introduction

Afin d'être sûrs des résultats à obtenir, la mise en place des premières expériences en couplage CE-FTICR-MS a été réalisée en s'appuyant sur une application développée au LSMIS et publiée en couplage CE-QTOF-MS : la caractérisation des anticorps monoclonaux (mAbs).

Les mAbs sont des protéines d'environ 150 kDa prenant la forme d'un Y et composées de deux chaînes légères (LC) identiques, et de deux chaînes lourdes (HC) identiques, reliées entre elles par des ponts disulfures. Ces chaînes sont respectivement constituées de 2 domaines (1 constant et 1 variable) et 4 domaines (3 constants et 1 variable). La structure d'un mAb est présentée en Figure 4.1.A. Au cours de la production et de la maturation, des modifications post-traductionnelles (PTMs) peuvent intervenir et sont à l'origine de l'hétérogénéité des mAbs. On y retrouve notamment les N-glycosylations (Figure 4.1.B), les déamidations et les oxydations. Un grand intérêt est porté sur la connaissance de ces modifications, leurs natures et leurs localisations, car elles peuvent avoir une incidence sur l'activité biologique, et donc, également, sur l'efficacité des traitements basés sur l'utilisation de mAbs.

Pour caractériser les mAbs, plusieurs niveaux d'analyses peuvent être utilisés selon le but de l'étude. Ces différentes approches sont représentées en Figure 4.2.

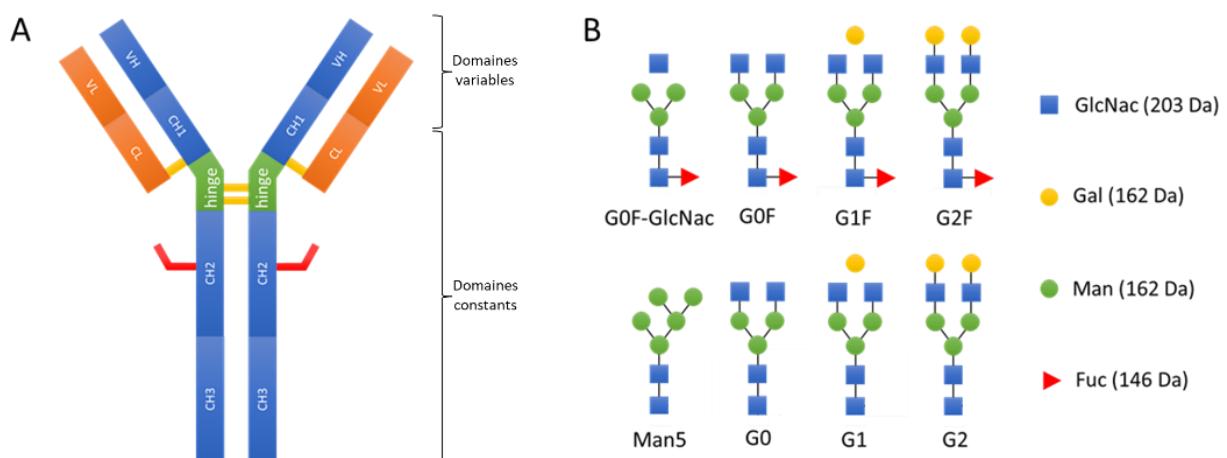


Figure 4.1 : (A) Structure d'un mAb. Les chaînes légères sont représentées en bleu, les chaînes lourdes en orange, les ponts disulfures en jaune et les N-glycosylations en rouge. (B) Structures des principales N-glycosylations présentes sur les mAbs.
Adapté de [1].

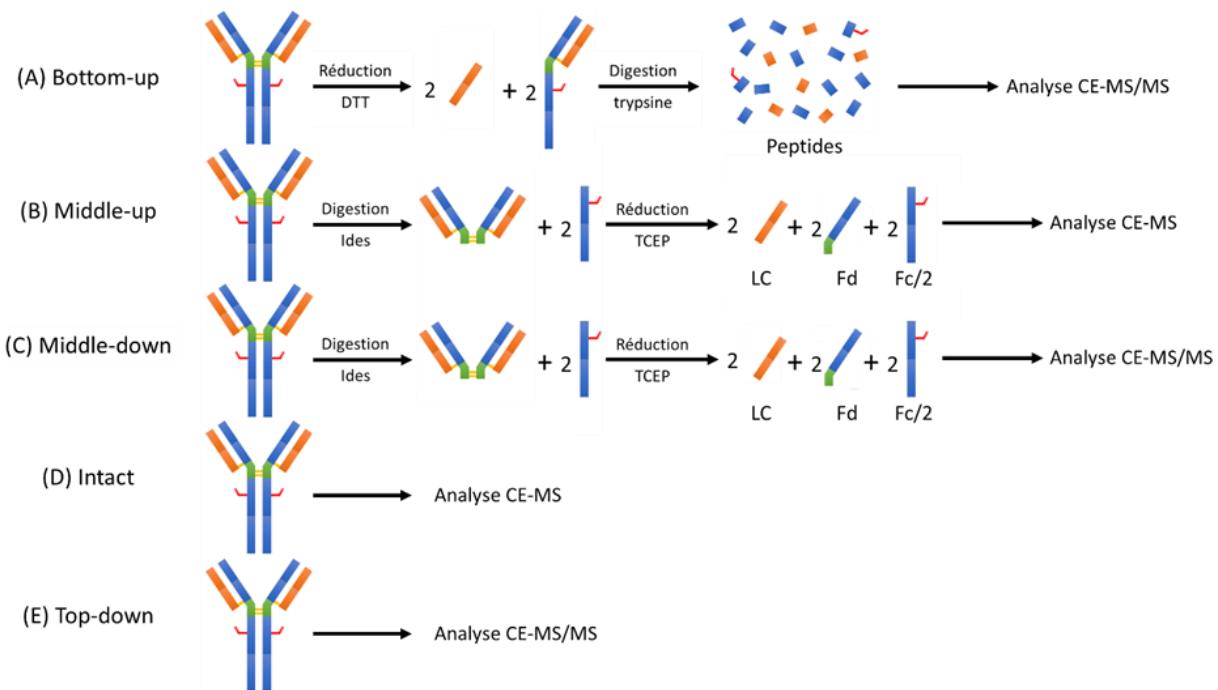


Figure 4.2 : Stratégies d'analyses d'un anticorps monoclonal aux niveaux (A) bottom-up, (B) middle-up et (C) middle-down, (D) intact et (E) top-down. DTT : dithiothréitol ; TCEP : tris(2-carboxyéthyl)phosphine. Adapté de [2].

L'approche bottom-up a été la première à être développée sur le couplage CE-QTOF-MS au laboratoire [3]. Les niveaux middle-up (Figure 4.2.B) et intact (Figure 4.2.D) en CE-MS ont également été développés sur le couplage CE-QTOF-MS lors d'une précédente thèse du laboratoire [2]. Ces travaux aux niveaux intact et middle-up ont prouvé que le couplage CE-MS permettait de séparer partiellement différents variants du mAb et de caractériser certaines de leurs modifications. Au niveau intact, les N-glycosylations présentes sur les différents variants partiellement séparés, ont été identifiées. Un exemple d'analyse de l'adalimumab intact est présenté en Figure 4.3.A. Au niveau middle-up, le couplage CE-QTOF-MS a permis d'identifier les N-glycosylations présentes sur les variants des fragments Fc/2 partiellement séparés, mais aussi de démontrer la présence d'autres modifications sur les fragments Fc/2, LC et Fd telles que des déamidations. Un exemple d'analyse middle-up de l'adalimumab est présenté en Figure 4.3.B. Il est important de noter que les petites PTMs, telles que les déamidations, ne peuvent qu'être supposées. En effet, pour des masses si importantes (150 kDa pour le niveau intact et 25 kDa pour le niveau middle-up), la résolution et la précision de masse du couplage CE-QTOF-MS ne sont pas assez élevées pour identifier, sans ambiguïté, toutes les modifications à l'origine de cette séparation partielle.

Un gain en résolution et en précision de masse, apporté par le couplage CE-FTICR-MS, pourrait permettre d'identifier sans ambiguïté ces modifications. De plus, les différentes techniques de fragmentation disponibles sur le FTICR-MS (CID, ETD et ECD) pourraient être utilisées pour les stratégies middle-down (Figure 4.2.C) et top-down (Figure 4.2.E) afin de tenter de localiser les PTMs sur les différents variants du mAb en combinant les différents résultats. C'est donc avec cet objectif en tête que la mise en place du couplage CE-FTICR-MS a débuté.

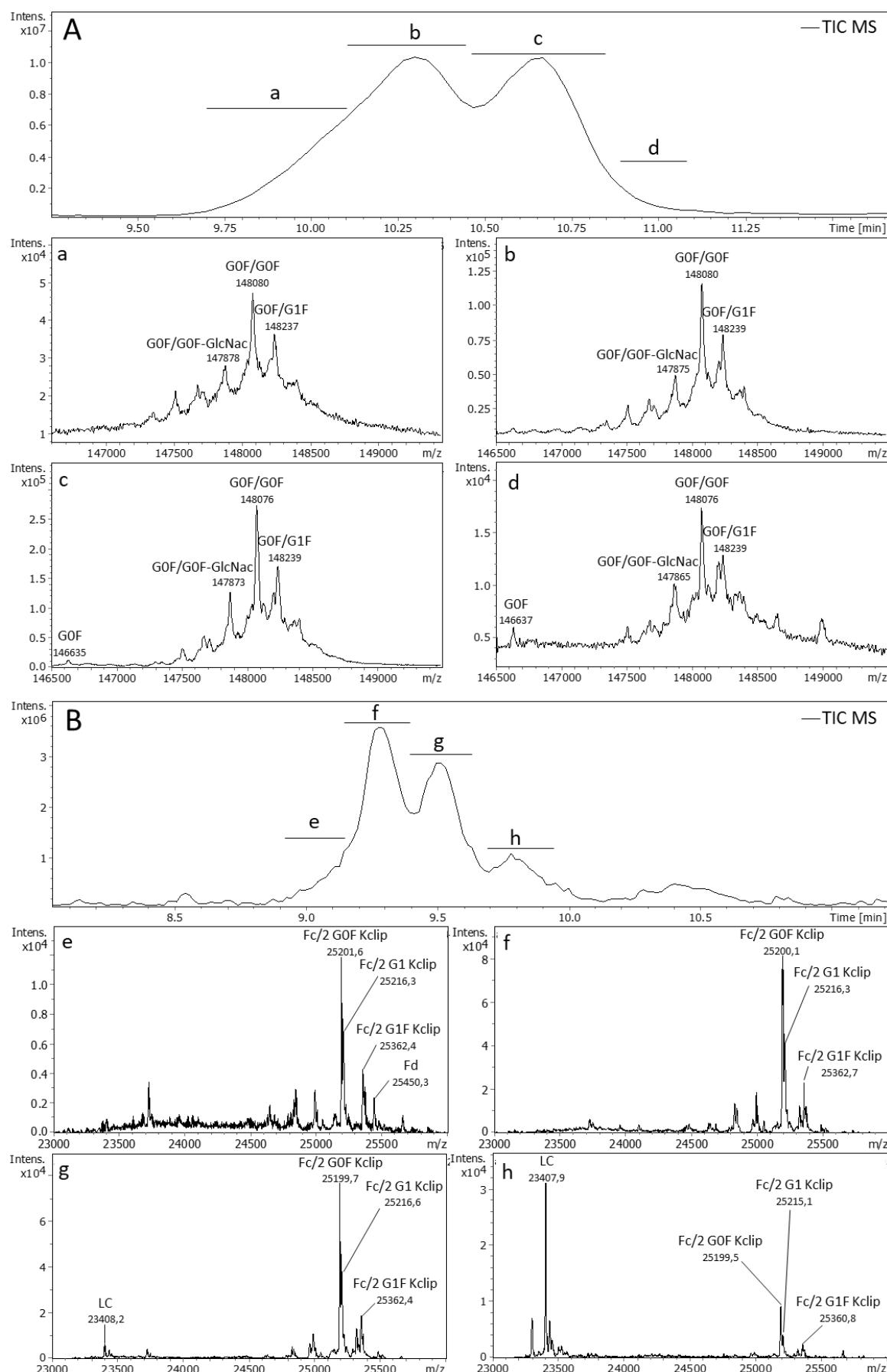


Figure 4.3 : Electrophérogrammes d'analyses de l'adalimumab aux niveaux (A) intact et (B) middle-up obtenus avec le couplage CE-QTOF-MS, et spectres de masse déconvolus associés.

Avant d'atteindre cet objectif, il a tout d'abord fallu parvenir à obtenir un spray stable puisque le FTICR et la CE n'avaient jamais été réellement couplés au laboratoire auparavant. Les différences, entre les deux couplages, au niveau position du capillaire CE devant l'entrée du spectromètre de masse pour obtenir un spray stable seront détaillées ultérieurement (paragraphe 4.6.1). L'étape suivante était de réaliser le tout premier électrophérogramme issu du couplage CE-FTICR-MS au laboratoire. L'approche bottom-up (Figure 4.2.A) a été utilisée pour obtenir le premier électrophérogramme du couplage CE-FTICR-MS puisque c'est une application de référence du laboratoire [3] qui a permis d'acquérir l'expérience du couplage CE-MS. Le premier électrophérogramme obtenu avec le couplage CE-FTICR-MS est présenté en Figure 4.4.A, il s'agit d'une analyse d'infliximab Remicade (EDQM, Council of Europe). Les recouvrements de séquence pour la chaîne lourde et la chaîne légère issus de cette analyse sont également présentés en Figure 4.4.B et Figure 4.4.C.

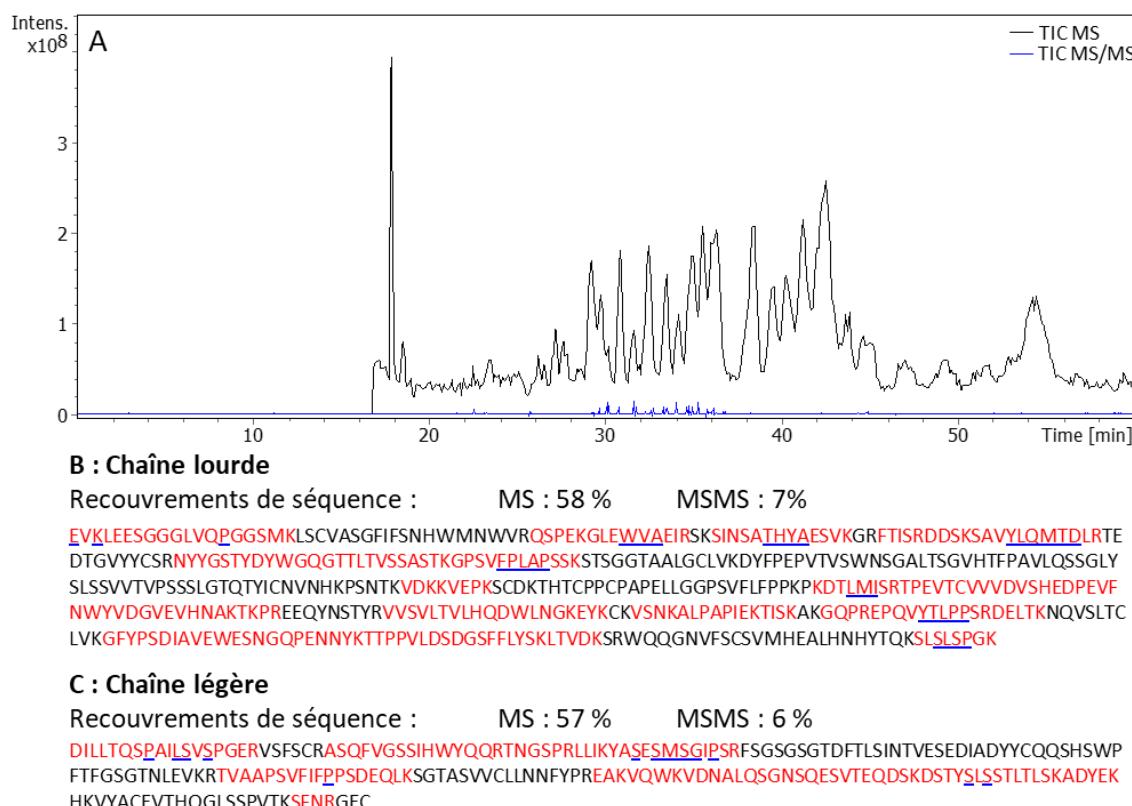


Figure 4.4 : (A) Electrophérogramme d'analyse bottom-up de l'infliximab Remicade obtenu avec le couplage CE-FTICR-MS/MS. et recouvrements de séquence associés à la MS (en rouge) et à la MS/MS (souligné en bleu) (B) de la chaîne lourde et (C) de la chaîne légère.

L'optimisation du couplage sur cette approche n'a pas été poussée beaucoup plus loin, malgré de faibles recouvrements de séquence. En effet, l'approche bottom-up n'est pas la plus adaptée aux besoins en temps de cycle du FTICR, puisque beaucoup de peptides co-migrent sur de faibles fenêtres de temps. La différence entre les temps de cycle accessibles avec le couplage CE-FTICR-MS et ceux du couplage CE-QTOF-MS seront détaillées dans le paragraphe 4.5.1 de ce chapitre. Cette première partie de la mise en place du couplage CE-FTICR-MS a tout de même été bénéfique pour se familiariser avec les difficultés intrinsèques au coulage CE-FTICR-MS, telles que la stabilité du spray, ou le positionnement du capillaire au

niveau de l'interface. Des conseils pratiques concernant ces particularités seront détaillés dans le paragraphe 4.6 de ce chapitre.

La suite des travaux s'est focalisée sur le transfert des stratégies middle-up et intact au couplage CE-FTICR-MS. Dans un premier temps, les méthodes de détection FTICR-MS ont été optimisées en utilisant la CE comme plateforme d'infusion. Les spectres de masse déconvolus, obtenus par les deux couplages, sont représentés en Figure 4.5 pour le niveau intact et en Figure 4.6 pour l'approche middle-up.

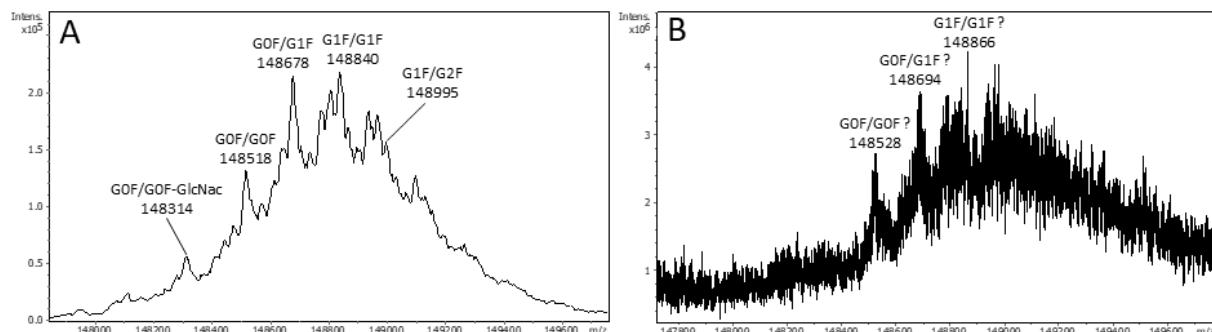


Figure 4.5 : Spectres de masse déconvolus de l'infliximab Remicade au niveau intact obtenus par infusion avec les couplages (A) CE-QTOF-MS et (B) CE-FTICR-MS.

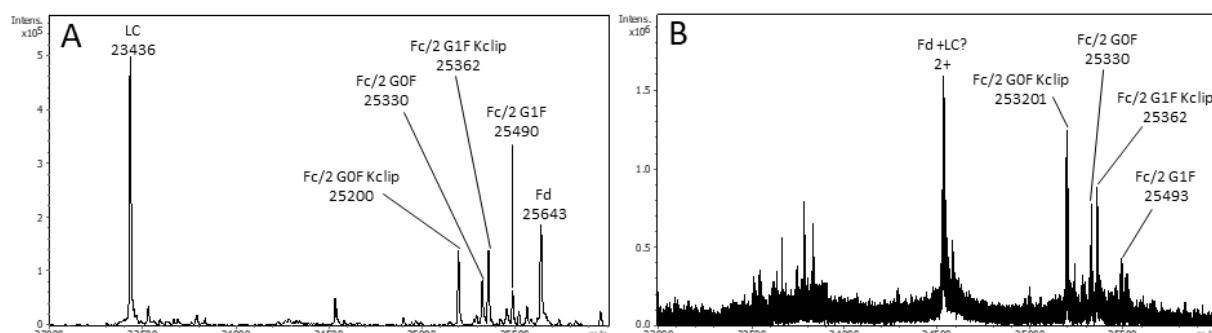


Figure 4.6 : Spectres de masse déconvolus de l'infliximab Remicade au niveau middle-up obtenus par infusion avec les couplages (A) CE-QTOF-MS et (B) CE-FTICR-MS.

Malgré l'optimisation des méthodes de détection, le signal obtenu par le FTICR-MS était beaucoup plus bruité que celui obtenu par le QTOF-MS pour les deux stratégies d'analyse. Au niveau intact, l'identification des différentes formes glycosylées est donc plus délicate sur le spectre de masse obtenu par FTICR-MS (Figure 4.5.B), que sur celui obtenu avec le QTOF (Figure 4.5.A). Au niveau middle-up, il semblerait qu'une agrégation ait eu lieu ou qu'un pont disulfure se soit reformé entre les fragments Fd et LC (Figure 4.6.B) entre l'analyse de l'échantillon sur le QTOF et l'analyse sur le FTICR.

Malgré ce signal très bruité, obtenu lors de l'optimisation des méthodes de détection, et la difficulté rencontrée avec les fragments LC et Fd, un essai de séparation a été réalisé sur le couplage CE-FTICR-MS. La méthode de séparation pour ces deux stratégies d'analyse requiert l'utilisation d'un capillaire greffé positivement au polyéthylèneimine. Malheureusement, aucun spray n'a pu être obtenu avec un tel capillaire sur le couplage CE-FTICR-MS. Plusieurs hypothèses peuvent être émises pour justifier l'absence de spray avec un

capillaire greffé positivement. Tout d'abord, comme cela sera expliqué dans le paragraphe 4.6.1, la hauteur de la CE et le positionnement du capillaire par rapport à l'entrée du spectromètre a pu jouer un rôle dans l'absence du spray. Il est également possible que le capillaire greffé se soit détérioré lors du transport entre Strasbourg et Metz. Il serait intéressant de retenter d'obtenir un spray, une fois le couplage mieux pris en main, avec de nouveaux capillaires greffés.

Ces premiers résultats concernant le couplage CE-FTICR-MS ont montré que l'application initialement choisie était trop ambitieuse (la taille et la complexité des anticorps, la nécessité d'utiliser un capillaire greffé) pour un premier couplage sur un appareillage pour lequel le laboratoire n'avait pas d'expérience et pour un couplage (CE-FTICR-MS et CE-FTICR-MS/MS) relativement peu décrit dans la littérature. Il était donc essentiel de revenir à une application plus simple pour mieux prendre en main ce nouveau couplage, ainsi qu'appréhender et étudier les différents paramètres qui le rendent beaucoup plus délicat à mettre en œuvre qu'en chromatographie liquide.

La caractérisation structurale des ARN, et notamment l'identification de leurs modifications post-transcriptionnelles à partir d'analyses au niveau nucléosides, est aussi un sujet de la plus haute importance tant dans les thématiques du LSMIS que d'un point de vue des développements méthodologiques à réaliser.

4.2. Optimisation des méthodes de détection des nucléosides

La première étape, pour mettre en place le couplage CE-FTICR-MS à partir d'analyses de nucléosides, a été l'optimisation des méthodes MS et MS/MS du FTICR-MS. Pour cela, un échantillon d'ARNt^{Asp GUC} issu de *S. cerevisiae*, digéré à la Nucléase P1 et déphosphorylé à la BAP puis dilué à 0,05 g/L dans de l'acétate d'ammonium à 200 mM pH 4, a été infusé à l'aide de la CE avec une pression de 10 psi, soit environ 100 nL/min. Cela permet de rester dans des conditions compatibles avec la source nanoESI. Les différents paramètres du FTICR-MS optimisés manuellement, sont indiqués en vert sur la Figure 4.7 et leurs valeurs, ainsi que leurs significations respectives, sont détaillées dans le Tableau 4.1. Après optimisation des paramètres de transfert et de détection, le spectre MS présenté en Figure 4.8 a été obtenu.

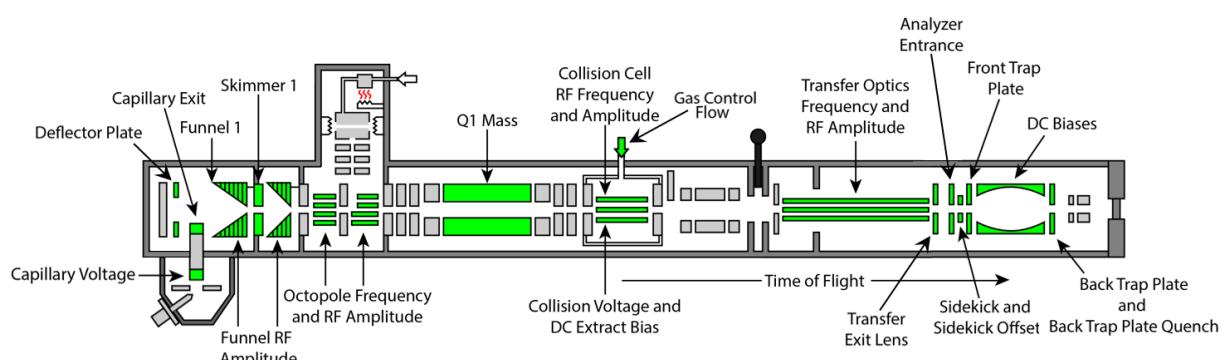


Figure 4.7 : Représentation schématique du spectromètre de masse FTICR SolariX 7T 2XR FTMS. Les paramètres pouvant être changés pour optimiser le transfert et la détection des ions sont représentés en vert. Copyright 2019 Bruker Daltonics

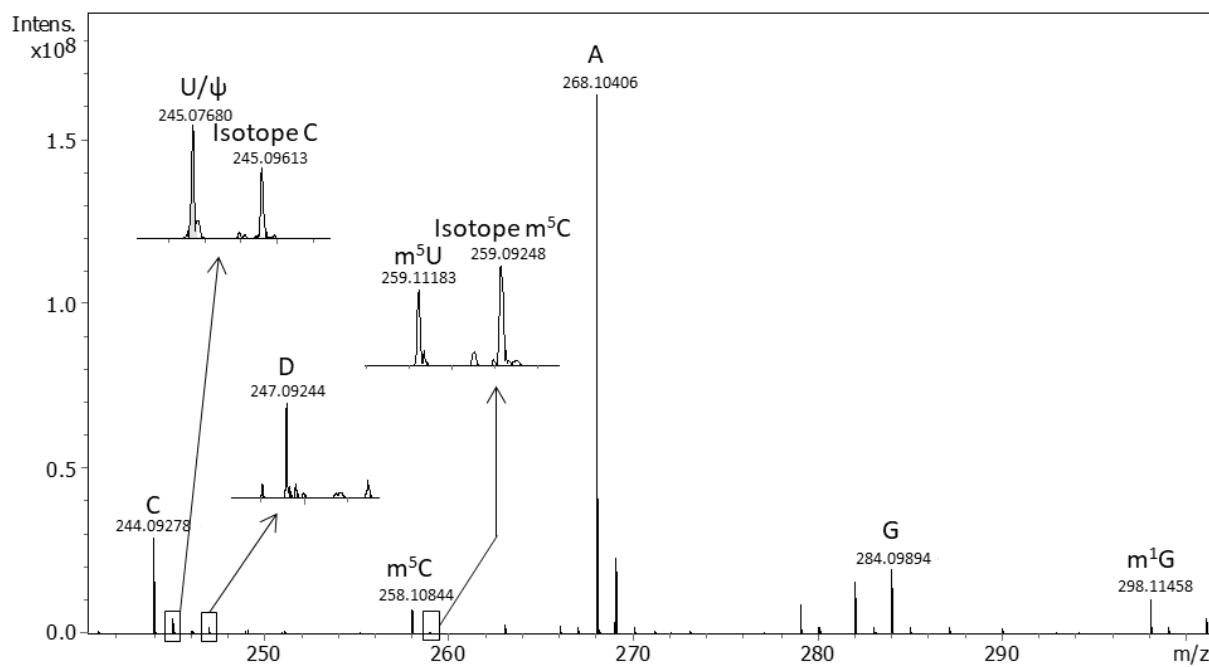


Figure 4.8 : Spectre MS après optimisation de la méthode de détection, issu d'une infusion CE-FTICR-MS d'un échantillon d'ARNt^{Asp GUC} (*S. cerevisiae*) digéré à la P1 et déphosphorylé à la BAP.

Tableau 4.1 : Paramètres du FTICR-MS pouvant être réglés pour optimiser le transfert des ions de la source à la cellule ICR pour la caractérisation des nucléosides.

	Paramètre	Valeur optimisée	Signification et rôle du paramètre
Source	Capillary voltage	1500 V	Tension appliquée à l'entrée du capillaire de transfert, permettant de créer le spray. Cette valeur peut légèrement varier en fonction du capillaire CE utilisé et de son usure, sans toutefois dépasser 1900 V.
	Dry Gas Flow Rate	0 mL/min	Débit de gaz séchant dans la source allant en contre sens des analytes. Lors du couplage CE-FTICR-MS, le spray est stable uniquement en absence de ce gaz.
	Dry Gas Temperature	165 °C	Température du gaz séchant et du capillaire de transfert. Augmenter la température du capillaire de transfert peut améliorer la désolvatation des analytes.
	Capillary Exit	120 V	Tension appliquée à la sortie du capillaire de transfert.
	Deflector Plate	250 V	Tension appliquée à la plaque de déviation permettant de diriger les ions vers la cellule ICR.
	Funnel 1	200 V	Contrôle le gradient de tension entre l'entrée du Funnel 1 et le Skimmer 1. Ce gradient de tension permet de diriger les ions de l'entrée du funnel vers le skimmer.
	Skimmer 1	15 V	Contrôle le gradient de tension entre l'entrée du Funnel 1 et le Skimmer 1. Pour des petites molécules labiles, la tension ne doit pas être trop élevée sous risque de fragmentation. Pour des plus grosses molécules, la tension peut être augmentée pour faciliter la désolvatation.

	Funnel RF Amplitude	150 Vpp	Amplitude de tension entre le funnel 1 et le funnel 2.
Octopole	Frequency	5 MHz	Fréquence appliquée au niveau de l'octopole.
	RF Amplitude	350 Vpp	Amplitude de radiofréquence appliquée au niveau de l'octopole.
Quadrupôle	Q1 Mass	200 m/z	Cette valeur définit le seuil de m/z lors de l'acquisition en mode « broadband ». Elle doit être inférieure aux masses des analytes d'intérêt.
Cellule de collision	Collision Voltage	-3 V	Tension appliquée dans la cellule de collision pour créer un puit de potentiel et accumuler les ions dans la cellule. Augmenter la tension peut permettre d'augmenter le rapport signal-sur-bruit. Cependant, plus la tension est importante, plus il est probable que les ions se fragmentent. Une fois les ions accumulés, la tension change à la valeur du « DC Extract Bias » afin d'expulser les ions.
	DC Extract Bias	0.4 V	Tension appliquée dans la cellule de collision au moment d'expulser les ions et de les envoyer vers la cellule ICR. Une fois les ions expulsés, la tension retourne à la valeur du « Collision Voltage ».
	RF Frequency	2 MHz	Fréquence appliquée dans la cellule de collision.
	Collision RF Amplitude	1600 Vpp	Amplitude de la tension alternative appliquée dans la cellule de collision pour accumuler les ions. Une grande amplitude permet d'accumuler des ions de plus faibles et plus hauts rapports m/z . Pendant la fragmentation par CID, une grande amplitude permet de limiter la diffusion des ions produits lors de la fragmentation des ions monochargés.
	Gas Control Flow	33 %	Débit de gaz dans la cellule de collision.
Transfert de la cellule de collision à l'analyseur	Time Of Flight	0.6 ms	Durée durant laquelle les ions vont voyager entre la cellule de collision et la cellule ICR. Un grand temps de vol privilégiera le transfert des ions de grands m/z , et un petit temps de vol privilégiera le transfert des ions de plus faibles m/z .
	Frequency	6 V	Fréquence appliquée à l'hexapole pour transporter les ions jusqu'à la cellule ICR. Plus les ions sont lourds, plus la fréquence doit être faible, et inversement.
	RF Amplitude	110 Vpp	Amplitude de tension appliquée à l'hexapole.

Analyseur : cellule ICR	Transfer Exit Lens	-40 V	Tension appliquée au niveau de la lentille entre l'hexapole et les lentilles d'entrée dans la cellule ICR.
	Analyzer Entrance	-10 V	Tension appliquée au niveau de la lentille à l'entrée de la cellule ICR lors du transfert des ions de la cellule d'accumulation vers la cellule ICR. En dehors de ce transfert, le voltage appliqué est identique à celui de la « Front Trap Plate ». Cela permet de guider les ions vers l'intérieur de la cellule ICR.
	Sidekick	0 V	Tension appliquée au niveau de la lentille à l'entrée de la cellule ICR lors du transfert des ions de la cellule d'accumulation vers la cellule ICR. En dehors du transfert des ions, le voltage appliqué est identique à celui de la « Front Trap Plate ».
	Sidekick offset	-1.5 V	Différence de tension entre les deux lentilles à l'entrée de la cellule ICR. Augmenter le Sidekick Offset peut augmenter le rapport signal sur bruit, mais également causer une coalescence des pics.
	Front Trap Plate	1.7 %	Tension appliquée au niveau de la plaque de piégeage à l'entrée de la cellule ICR. Cette tension permet de créer un puit de potentiel piégeant les ions dans la cellule.
	Back Trap Plate	1.7 %	Tension appliquée au niveau de la plaque de piégeage au fond de la cellule ICR. Cette tension permet de créer un puit de potentiel piégeant les ions dans la cellule.
	Back Trap Plate Quench	-30 V	Tension appliquée à la plaque de piégeage au fond de la cellule ICR pour expulser les ions présents dans la cellule entre chaque scan.
	Sweep Excitation Power	18 %	Tension appliquée aux plaques d'excitation de la cellule ICR pour exciter les ions. Une tension plus élevée permettra d'exciter des ions de plus grands m/z et inversement. Optimiser cette tension peut améliorer la précision de masse et l'intensité du signal.

L'énergie de collision pour la fragmentation MS/MS a également été optimisée. Une énergie de 15 V correspondait à la meilleure énergie de collision permettant de fragmenter, au niveau de la liaison N-glycosidique, les quatre nucléosides canoniques ainsi que les diverses modifications présentes dans cet échantillon. Les spectres MS/MS correspondants sont présentés en Figure 4.9.

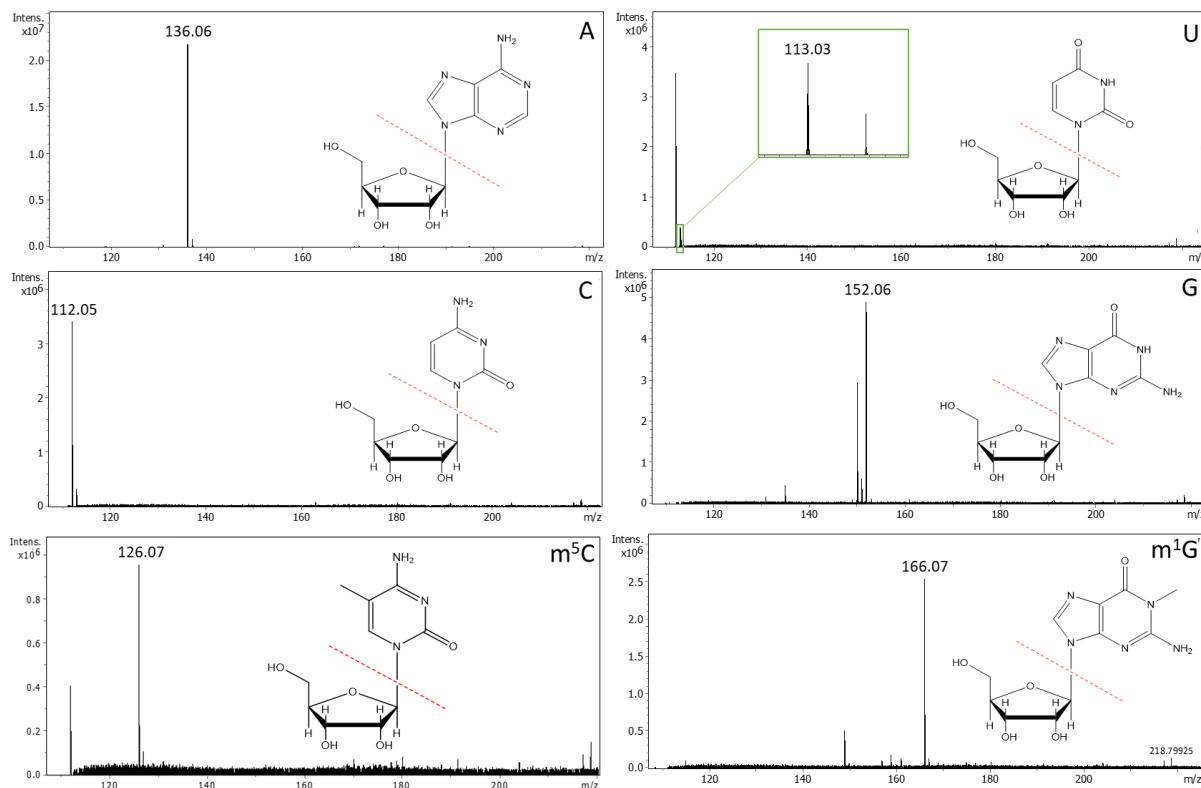


Figure 4.9 : Spectres MS/MS de A, U, C, G, m^5C et m^1G après optimisation de l'énergie de collision à 15eV, issus d'une infusion CE-FTICR-MS/MS d'un échantillon d'ARNT^{Asp} GUC (*S. cerevisiae*) digéré à la P1 et à la BAP. La fragmentation des liaisons N-glycosidiques est représentée en rouge

L'optimisation des autres paramètres de la MS/MS en mode « *data-dependent* », aussi appelée autoMS/MS, sera détaillée dans la partie 4.4 de ce chapitre, puisqu'elle nécessite, dans un premier temps, d'établir les paramètres pour obtenir une détection optimale lors d'une analyse par CE-FTICR-MS.

4.3. Etude des performances du couplage CE-FTICR-MS

Le couplage CE-FTICR-MS étant tout nouveau au laboratoire, nous n'avions que très peu d'expérience et de recul sur celui-ci. Le FTICR-MS est d'utilisation beaucoup plus délicate que le QTOF-MS, suscitant beaucoup plus d'interrogations à la moindre difficulté expérimentale rencontrée. Il a donc été nécessaire de revenir à une application plus simple, qui pourra servir par la suite de référence, afin de contrôler et de s'assurer des performances du couplage, avant de passer à des applications plus complexes et des échantillons précieux au début de chaque campagne d'analyses par CE-FTICR-MS. Pour cela, l'étude du couplage a été effectuée à l'aide de standards des quatre nucléosides canoniques (A, C, U, G), afin de réduire au maximum la complexité du mélange et de la séparation. Travailler avec ces standards permet également de s'affranchir des contraintes liées aux faibles quantités d'échantillons biologiques disponibles.

Au début de l'étude, des difficultés liées à la stabilité et la reproductibilité du spray ont été observées, lorsque le couplage CE-FTICR-MS est utilisé en séparation. Beaucoup de causes peuvent être avancées face à cette instabilité, y compris la hauteur relative de la CE par

rapport à la source. Afin d'y pallier, sans se perdre dans le (ou les) paramètre(s) à l'origine de cette instabilité dans un premier temps, et de s'assurer qu'un flux d'électrolyte support émanait bien du capillaire, une assistance par une pression de 2 psi a été ajoutée dans le capillaire de séparation en plus de la tension de 30 kV.

Une évaluation des performances du couplage a ensuite été réalisée en faisant varier différents paramètres influant sur l'intensité du signal, sur le rapport signal/bruit, sur la résolution, mais également sur le temps de cycle et par conséquent sur le nombre de points par pic électrophorétique. En effet, la CE est une technique séparative efficace donnant des pics très fins, tandis que le FTICR-MS a besoin de temps pour donner des résultats au maximum de ses performances. Il est donc nécessaire de trouver un compromis entre les deux techniques. Trois paramètres peuvent influencer le temps de cycle MS du FTICR : la valeur minimale de la gamme de m/z , le DataPoint Size (taille des points de données) et le temps d'accumulation. L'influence de chaque paramètre sur les performances du couplage CE-FTICR-MS a été étudiée, en ne variant qu'un paramètre à la fois, les autres étant fixés aux valeurs utilisées pour réaliser l'optimisation du transfert des ions vue dans la partie précédente. Dans l'idéal, le couplage devrait fournir le meilleur signal et la meilleure résolution, avec un temps de cycle le plus petit possible afin d'avoir un nombre de points convenable pour chaque pic électrophorétique.

4.3.1. Etude des performances du couplage en fonction du temps d'accumulation des ions dans la cellule de collision

Le temps d'accumulation correspond à la durée durant laquelle les ions vont être accumulés dans la cellule de collision. Théoriquement, plus les ions sont accumulés, plus le signal est intense. Par ailleurs, augmenter le temps d'accumulation augmente également le temps de cycle MS, ce qui peut donc se répercuter sur le nombre de points par pics, lors du couplage du FTICR-MS avec une méthode séparative.

Pour étudier l'influence du temps d'accumulation sur le couplage CE-FTICR-MS, des temps d'accumulation de 10 à 200 ms ont été utilisés. Les valeurs du minimum de la gamme de m/z , du DataPoint Size et du temps d'accumulation utilisées pour cette étude sont présentées dans le Tableau 4.2. Les évolutions de l'aire, de l'intensité, du rapport signal/bruit et du nombre de points des quatre pics électrophorétiques extraits (A, C, U, G) sont présentées dans la Figure 4.10. La résolution, indépendante du temps d'accumulation, n'est pas représentée ici.

Tableau 4.2 : Valeurs du minimum de la gamme de m/z , du DataPoint Size et du temps d'accumulation utilisées pour l'étude de l'influence du temps d'accumulation sur les performances du couplage CE-FTICR-MS.

Paramètres	Valeurs
Low m/z	107
DataPoint Size (M)	2
Accumulation time (ms)	10 ; 50 ; 100 ; 150 ; 200

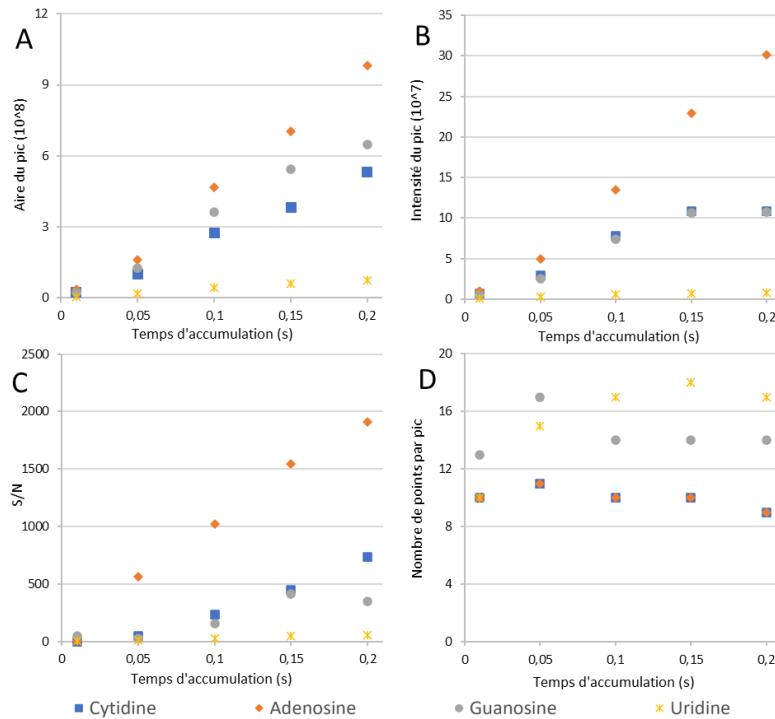


Figure 4.10 : Evolution de (A) l'aire, (B) l'intensité, (C) le rapport signal sur bruit (S/N), (D) le nombre de points pour les quatre pics électrophorétiques extraits, en fonction du temps d'accumulation des ions.

Malgré l'augmentation du temps de cycle liée à l'augmentation du temps d'accumulation, le nombre de points ne diminue pas drastiquement, et le signal, en termes d'intensité, d'aire et de rapport S/N, augmente considérablement. Ainsi, lors du choix de la bonne combinaison de paramètres, il sera utile d'augmenter le temps d'accumulation jusqu'à 100 ms ou 200 ms sans risquer une perte d'information sur les électrophérogrammes.

4.3.2. Etude des performances du couplage en fonction du DataPoint Size

Augmenter le DataPoint Size permet d'augmenter le temps d'acquisition dans la cellule ICR. Pour étudier l'influence du DataPoint Size, et donc du temps d'acquisition dans la cellule ICR, sur les performances du couplage, des valeurs de DataPoint Size allant de 512 k à 4 M ont été utilisées. Les valeurs du minimum de la gamme de m/z , du DataPoint Size et du temps d'accumulation utilisées pour cette étude, sont présentées dans le Tableau 4.3. Les évolutions de l'aire, de l'intensité, du rapport signal/bruit, du nombre de points des quatre pics électrophorétiques extraits et de la résolution sont présentées dans la Figure 4.11. La durée d'acquisition dans la cellule ICR, dépendante du DataPoint Size, est également représentée en Figure 4.11.F.

Tableau 4.3 : Valeurs du minimum de la gamme de m/z , du DataPoint Size et du temps d'accumulation utilisées pour l'étude de l'influence du DataPoint Size sur les performances du couplage CE-FTICR-MS.

Paramètres	Valeurs
Low m/z	107
DataPoint Size (M)	0,512 ; 1 ; 2 ; 4
Accumulation time (ms)	100

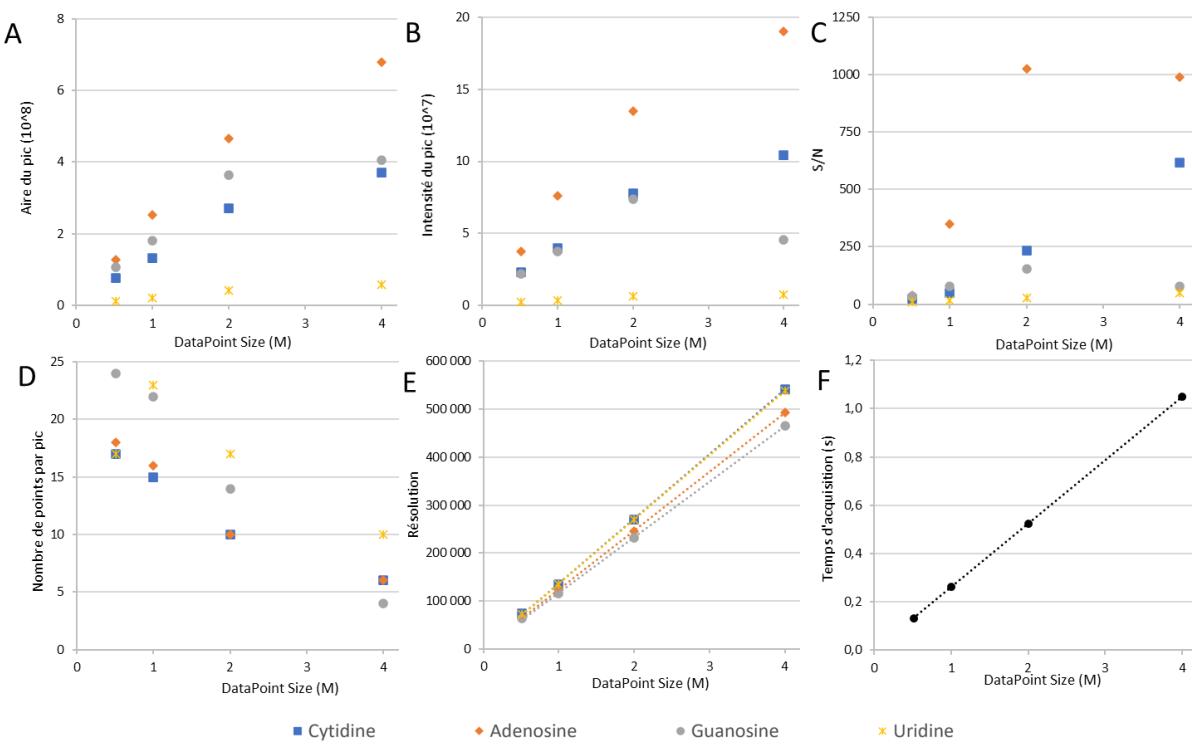


Figure 4.11 : Evolution de (A) l'aire, (B) l'intensité, (C) le rapport signal sur bruit (S/N), (D) le nombre de points, pour les quatre paires électrophorétiques extraits, (E) la résolution MS et (F) le temps d'acquisition dans la cellule ICR en fonction du DataPoint Size.

Le DataPoint Size semble avoir la plus grande influence sur le nombre de points par pic électrophorétique, puisqu'elle influence linéairement le temps d'acquisition dans la cellule ICR. En effet, lorsque le DataPoint Size augmente, le temps d'acquisition augmente également (Figure 4.11.E). Si le temps d'acquisition devient trop grand, des modifications peu abondantes peuvent migrées à un moment où les ions ne sont pas accumulés. Ces modifications peuvent donc ne pas être envoyées dans la cellule ICR et donc ne pas être détectées. Malgré l'augmentation du signal (aire, intensité et rapport S/N) et de la résolution avec l'augmentation du DataPoint Size (Figure 4.11.A, Figure 4.11.B et Figure 4.11.C), il est donc nécessaire de veiller à ne pas trop l'augmenter, et donc ne pas trop augmenter le temps d'acquisition, afin de ne pas risquer une perte d'information sur les modifications peu abondantes. Par ailleurs, il faut garder à l'esprit qu'il sera nécessaire d'ajouter du temps pour la fragmentation MS/MS entre chaque spectre MS, ce qui diminuera encore le nombre de points par pics électrophorétiques.

4.3.3. Etude des performances du couplage en fonction du m/z minimal

Plus la valeur minimale de la gamme de m/z est basse, plus la fréquence cyclotronique des ions correspondants sera élevée, diminuant ainsi le temps d'acquisition dans la cellule ICR pour une même taille de points de données (DataPoint Size). En effet, plus de points sont mesurés pour les petits m/z , et ainsi la taille totale des points de données (DataPoint Size) est plus rapidement atteinte. En théorie, baisser la limite inférieure de la gamme de m/z permettrait donc d'augmenter le nombre de points par pic électrophorétique, mais cela

engendrerait également une diminution de la résolution et de la précision de mesure de masse.

Le choix de la valeur minimale de la gamme de m/z n'est pas infini, puisque celle-ci dépend des fréquences cyclotroniques. Les valeurs du minimum de la gamme de m/z , du Data Point Size et du temps d'accumulation utilisées pour cette étude sont présentées dans le Tableau 4.4. L'évolution de l'aire, de l'intensité, du rapport signal/bruit, du nombre de points des quatre pics électrophorétiques extraits, de la résolution et du temps d'acquisition dans la cellule ICR sont présentés dans la Figure 4.12.

Tableau 4.4 : Valeurs du minimum de la gamme de m/z , du DataPoint Size et du temps d'accumulation utilisées pour l'étude de l'influence du minimum de la gamme de m/z sur les performances du couplage CE-FTICR-MS.

Paramètres	Valeurs
Low m/z	47 ; 86 ; 107
DataPoint Size (M)	2
Accumulation time (ms)	100

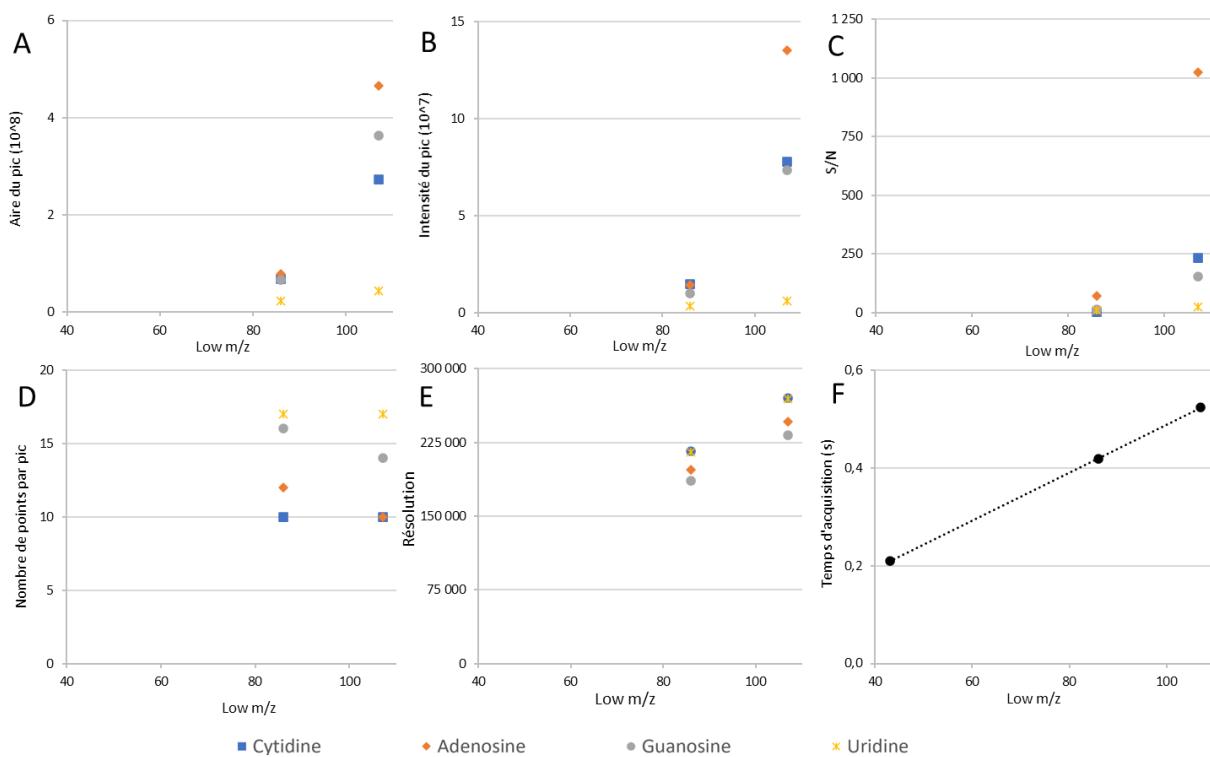


Figure 4.12 : Evolution de (A) l'aire, (B) l'intensité, (C) le rapport signal sur bruit (S/N), (D) le nombre de points, (E) la résolution MS pour les quatre pics électrophorétiques extraits et (F) le temps d'acquisition dans la cellule ICR en fonction du Low m/z du FTICR-MS.

Diminuer la valeur minimale de la gamme de m/z balayée engendre une diminution du signal (aire, intensité, S/N), jusqu'à la perte totale du spray lorsque la valeur basse de la gamme de m/z était de 47. Il n'est donc pas possible de diminuer la valeur minimale de la gamme de m/z pour réduire le temps de cycle du couplage pour cette application.

4.3.4. Combinaison des différents paramètres

Après avoir évalué séparément l'influence des différents paramètres sur les performances du couplage, différentes combinaisons de ces paramètres ont été testées. Les valeurs du minimum de la gamme de m/z , du Data Point Size et du temps d'accumulation utilisées pour cette étude sont présentées dans le Tableau 4.5. L'aire, l'intensité, le rapport signal/bruit, le nombre de points des quatre pics électrophorétiques extraits, la résolution et le temps de cycle MS obtenus pour les différentes combinaisons sont présentés dans la Figure 4.13.

Tableau 4.5 : Valeurs du minimum de la gamme de m/z , du DataPoint Size et le temps d'accumulation utilisées pour étudier l'influence de différentes combinaisons sur les performances du couplage CE-FTICR-MS.

Paramètres	Méthode 1	Méthode 2	Méthode 3	Méthode 4
Low m/z	107	107	107	107
DataPoint Size (M)	1	2	2	1
Accumulation time (ms)	100	100	200	200

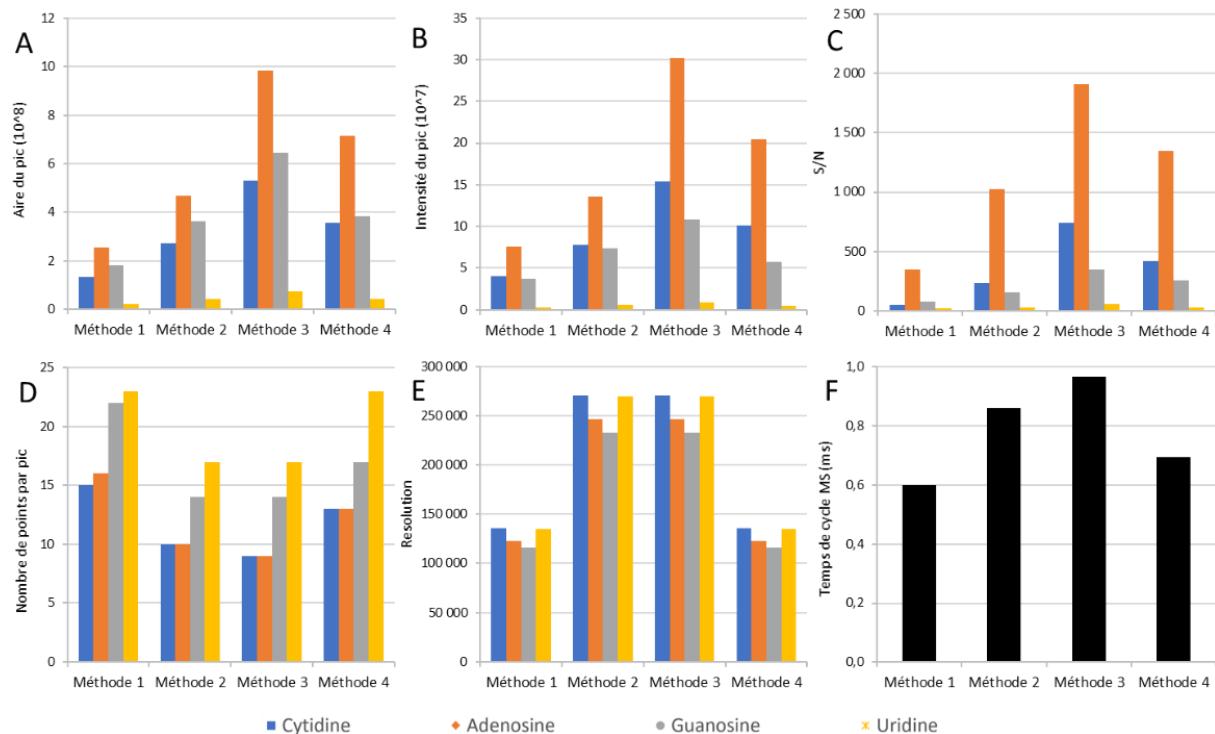


Figure 4.13 : Evolution de (A) l'aire, (B) l'intensité, (C) le rapport signal sur bruit (S/N), (D) le nombre de points, (E) la résolution MS pour les quatre pics électrophorétiques extraits et (F) la durée d'un cycle MS en fonction des différentes combinaisons de temps d'accumulation et de DataPoint Size détaillées dans le Tableau 4.5.

La méthode 4, avec un grand temps d'accumulation et une faible taille de points de données, permet d'augmenter considérablement le signal (en termes d'aire, de rapport S/N et d'intensité) par rapport à la méthode 1 tout en maintenant un nombre de points similaire par pic électrophorétique. La méthode 2, avec un plus faible temps d'accumulation et une taille de points de données plus importante, permet d'obtenir une meilleure résolution que la méthode 4, mais également un signal et un nombre de point par pic plus faibles. La méthode

3, avec un grand temps d'accumulation et une taille de points de données importante, permet d'obtenir un meilleur signal (en termes d'aire, de rapport S/N et d'intensité) et une meilleure résolution par rapport aux autres méthodes, mais un faible nombre de points par pic électrophorétique.

Compte tenu de la nécessité d'obtenir plus d'une dizaine de points par pic électrophorétique pour ajouter, par la suite, la fragmentation MS/MS, le choix de la combinaison optimale pour le couplage CE-FTICR s'est porté sur la méthode 4. Même si la résolution obtenue pour les m/z des quatre nucléosides canoniques n'est pas optimale par rapport aux méthodes 2 et 3, puisqu'il a fallu réduire le temps de cycle et notamment le temps d'acquisition pour augmenter le nombre de points, elle reste deux à trois fois plus élevée que celle obtenue avec le Q-TOF (environ 50 000 pour les quatre nucléosides canoniques). L'électrophérogramme obtenu avec la méthode 4 est présenté en Figure 4.14.

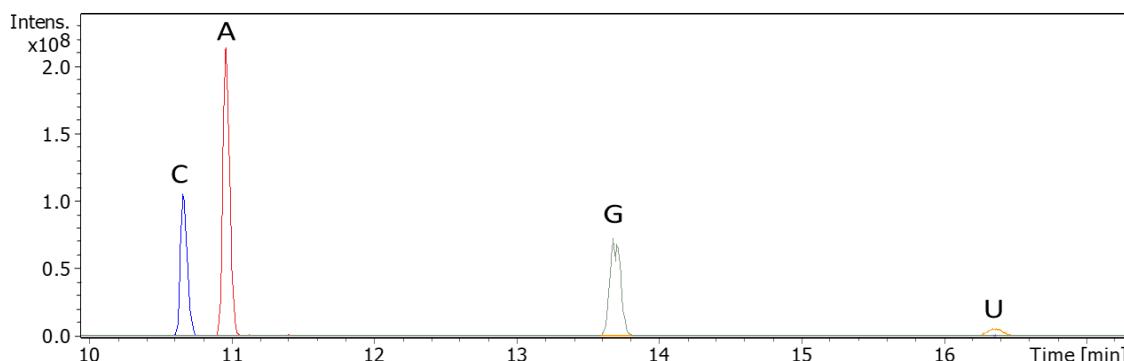


Figure 4.14 : Electrophérogramme obtenu par CZE-MS, issu d'une séparation CE-MS des quatre nucléosides canoniques A, C, U et G à 56,5 μ M. La séparation a été réalisée en appliquant une tension de 30kV et une pression de 2 psi. La détection MS a été réalisée avec une taille de points de données de 1M et un temps d'accumulation de 200 ms (méthode 4). Les m/z des quatre nucléosides canoniques ont été extraits.

Ces premiers résultats de couplage CE-FTICR-MS dans le cadre de l'analyse de nucléosides sont très encourageants et le bon compromis a été trouvé entre la résolution et le nombre de points par pics électrophorétiques. En effet, la résolution d'environ 125 000 pour les quatre nucléosides canoniques est largement suffisante pour cette application et le nombre de points par pics supérieur à 10 permet d'envisager l'ajout de la fragmentation MS/MS. Maintenant qu'une analyse de référence a été mise en place pour le couplage CE-FTICR-MS, il sera nécessaire de compléter ces analyses avec la fragmentation MS/MS par CID, avant de passer à de vrais échantillons d'ARN biologiques contenant des modifications post-transcriptionnelles.

À partir de ce moment, il a été possible de retirer la pression de 2psi ajoutée précédemment pour garantir la stabilité du spray. Cela a été rendu possible grâce à la meilleure maîtrise du positionnement du capillaire à l'entrée du FTICR-MS, qui est délicat et qui sera détaillé dans la partie 4.6.1 de ce chapitre. Tous les résultats suivants seront donc réalisés sans cette pression supplémentaire.

4.4. Optimisation du mode automatique pour la MS/MS

Après l'optimisation de la méthode MS et l'étude des paramètres influant sur les performances du couplage, il a été nécessaire d'optimiser les paramètres MS/MS en mode automatique (autoMS/MS) du FTICR. Le rôle de ces paramètres ainsi que les valeurs optimisées sont détaillés dans le Tableau 4.6.

Tableau 4.6 : Paramètres optimisés de l'AutoMS/MS avec le couplage CE-FTICR-MS pour la caractérisation des nucléosides ainsi que leurs significations respectives.

	Paramètre	Valeur optimisée	Signification et rôle du paramètre
Sélection des précurseurs	Exclude list/Prefer List	Liste d'exclusion contenant les <i>m/z</i> du bruit de fond	La liste d'exclusion permet d'exclure définitivement certaines gammes de <i>m/z</i> de la sélection pour la MS/MS. En général, les zones de « non-intérêt » ainsi que les ions intenses du bruit de fond sont référencés dans cette liste afin d'augmenter le nombre de fragmentations intéressantes. Au contraire, la liste de préférences permet d'indiquer quelles gammes de <i>m/z</i> doivent être préférentiellement choisies lors de la sélection des précurseurs.
	Active Exclusion	Activée	Cette fonction permet d'exclure temporairement les ions déjà sélectionnés pour la MS/MS. Cela permet de sélectionner des ions de plus faible intensité dans le spectre MS, après que les plus intenses aient déjà été fragmentés.
	Strict Active Exclusion Time	Activée	Lorsque cette fonction est activée, s'il n'y a plus de précurseur satisfaisant toutes les conditions pour être sélectionné, aucune MS/MS n'est réalisée.
	Exclude After	1 Spectre	Nombre maximum de spectres MS/MS d'un même ion parent avant que cet ion ne fasse partie de la liste d'exclusion temporaire.
	Release After	0.07 min	Durée durant laquelle les ions sont exclus temporairement lors de l'exclusion active.
	Preferred Charge States	1	Ce paramètre permet de sélectionner préférentiellement les ions aux états de charges prédéfinis.
	No. Precursors	2	Nombre maximum de précurseurs sélectionnés pour la MS/MS à partir du spectre MS. Les ions les plus intenses du spectre MS seront sélectionnés, sauf s'ils font partie de la liste d'exclusion. Pour 2 précurseurs, on parle couramment de Top2.
	Min. threshold	3E5	Seuil d'intensité pour la sélection des ions précurseurs.
	Relative threshold	0.01 %	Seuil d'intensité relative pour la sélection des ions précurseurs.

	Max length group	2	Nombre d'ions du même massif isotopique à ne pas considérer comme des ions précurseurs différents lors du choix des précurseurs de l'autoMS/MS.
	Use strict group lenght	Activée	Lorsque cette fonction est activée, s'il n'y a plus de précurseur satisfaisant toutes les conditions pour être sélectionné, aucune MS/MS n'est réalisée.
Isolement dans le quadrupôle	MS/MS isolation	Activée	Cette fonction permet de sélectionner les ions précurseurs dans le quadrupôle avant fragmentation. Si elle n'est pas activée, la fragmentation sera réalisée sur toute la gamme de masse du spectre MS.
	Isolation Window	2.0 m/z	Largeur de la fenêtre d'isolement pour la sélection de l'ion parent dans le quadrupôle. Usuellement, une fenêtre de 10 m/z permet d'assurer un isolement efficace. Diminuer cette fenêtre d'isolement permet d'isoler des ions possédant des m/z proches, mais une réduction de l'intensité du spectre MS/MS est possible.
Mode de fragmentation	Fragmentation Mode	QCID	Différents modes de fragmentation sont disponibles. Le mode QCID est la fragmentation par CID au niveau du quadrupôle, à différencier du SORI-CID qui est une fragmentation par CID au niveau de la cellule ICR.
	MS/MS Boost	Désactivée	Cette fonction permet d'augmenter le temps d'accumulation uniquement pour la MS/MS, tout en n'augmentant pas le temps d'accumulation pour la MS. L'inverse n'est cependant pas possible.
	Fixed Collision Voltage	15 V	Energie de collision utilisée lors de la fragmentation par CID dans le quadrupôle.

L'objectif de cette optimisation est de permettre de sélectionner un maximum d'ions précurseurs différents afin d'obtenir le maximum d'informations en une seule analyse, tout en ayant des spectres MS/MS de qualité et un temps de cycle compatible avec la séparation apportée par la CE. En effet, il est important d'avoir plus qu'un seul précurseur sélectionné par spectre MS puisqu'au regard de la complexité des échantillons, plusieurs analytes d'intérêt peuvent encore co-migrer. En CZE, c'est notamment le cas des espèces neutres comme l'uridine et une majorité de ses modifications, où une séparation n'est pas possible.

Dans un premier temps, le nombre de précurseurs sélectionnés par cycle a été graduellement augmenté jusqu'à ce qu'il soit impossible d'en ajouter sans risquer de perdre des informations. Un Top2 a donc pu être obtenu. Au-delà, les limites du couplage CE-FTICR-MS commencent à se ressentir, et le nombre de points par pic électrophorétique est bien trop réduit.

Dans un second temps, pour augmenter le nombre et la variété des précurseurs, il existe plusieurs stratégies :

- La 1^{ère} consiste à augmenter la vitesse de scan en diminuant la durée du transient et/ou le temps d'accumulation. Par exemple, si les deux sont réduits de minimum 50%, cela pourrait permettre de doubler le nombre de précurseurs sélectionnés par spectre MS sans augmenter le temps de cycle total.
- La 2^{ème} stratégie consiste à optimiser les paramètres d'exclusion dynamique des ions déjà sélectionnés afin d'alterner la sélection des ions les plus intenses avec notamment ceux qui sont moins intenses.

Usuellement, les deux stratégies sont combinées. Malheureusement, dans notre cas, il n'est pas possible de diminuer la durée du transient, puisque la limite inférieure pour pouvoir activer l'AutoMS/MS a été atteinte. De plus, diminuer uniquement le temps d'accumulation réduisait significativement l'intensité du signal, mais ne permettait pas d'augmenter significativement le nombre de scans par pic électrophorétique comme démontré dans la partie 4.3.1 de ce chapitre. Enfin, il n'est pas possible d'augmenter uniquement la vitesse de scan consacrée à la MS/MS tout en gardant la même pour la MS comme c'est le cas pour un grand nombre de spectromètres de masse, dont le QTOF-MS. Notre choix s'est donc porté sur l'optimisation de l'exclusion lors de la sélection des ions précurseurs. Tous les ions récurrents du bruit de fond ont été ajoutés à la liste d'exclusion pour éviter l'obtention de spectres MS/MS sur du bruit de fond. Les paramètres concernant l'exclusion dynamique ont également été réglés, afin qu'un ion précurseur ne soit au maximum sélectionné qu'un cycle sur deux, ce qui permet d'avoir 4 ions précurseurs différents sur la durée totale de deux cycles (5,6 s pour 2 cycles).

L'évaluation de cette méthode MS/MS sera réalisée sur de vrais échantillons biologiques contenant des modifications post-transcriptionnelles dans le chapitre 5.

4.5. Comparaison des couplages CE-FTICR-MS et CE-QTOF-MS

4.5.1. Temps de cycle

Lorsque l'on s'intéresse aux couplages CE-MS, il est intéressant de noter les différences de temps de cycle entre les deux spectromètres de masses, afin de prendre conscience des limites du couplage CE-FTICR-MS. En effet, le top7 du couplage CE-QTOF-MS est réalisé avec un temps de cycle total de 2,0 s dont la moitié consacrée uniquement à la MS, ce qui permet d'avoir une bonne sensibilité pour les modifications peu abondantes. Le temps de cycle du couplage CE-FTICR-MS est, quant à lui, de 2,8 s avec uniquement un top2 en MS/MS. Le temps de cycle étant plus important, le nombre de points par pic électrophorétique est réduit, tout en balayant moins d'ions précurseurs lors de la MS/MS. Il n'est donc pas improbable que, pour les échantillons biologiques très complexes, certaines modifications ne soient pas détectées en MS ou MS/MS. Si besoin, il est possible de combiner les résultats issus de plusieurs analyses

du même échantillon, en adaptant la liste d'exclusion pour avoir des sélections de précurseurs différents entre les analyses, et ainsi augmenter le nombre de modifications fragmentées.

4.5.2. Résolution

4.5.2.1. Comparaison de la résolution entre les deux couplages

Après optimisation totale du couplage entre la CE et le FTICR SolariX 2XR (Bruker Daltonics), la résolution obtenue pour les quatre nucléosides canoniques est comprise entre 116 000 et 135 000, ce qui correspond à plus du double de la résolution obtenue par le couplage CE-QTOF-MS. En effet, le QTOF Maxis 4G (Bruker Daltonics) permet d'avoir une résolution d'environ 50 000 pour les quatre nucléosides canoniques.

4.5.2.2. Sacrifice nécessaire lors du couplage du FTICR-MS avec la CE

Bien que la résolution apportée par le couplage CE-FTICR-MS reste supérieure à celle du couplage CE-QTOF-MS et que celle-ci suffise amplement pour la caractérisation des modifications post-transcriptionnelles, le sacrifice de résolution lors du couplage entre la CE et le FTICR-MS est significatif. Pour le mesurer, une infusion des quatre nucléosides canoniques a été réalisée en augmentant le DataPoint Size de 1 M à 16 M, ce qui a un effet direct sur le temps d'acquisition au sein de la cellule ICR. Comme indiqué par l'équation (2.7) dans le chapitre 2 et présenté par la Figure 4.15, la résolution est directement proportionnelle au temps d'acquisition. En infusion, la résolution maximale, associée à un temps d'acquisition dans la cellule ICR d'un peu plus de 4 s, est comprise entre 1 860 000 et 2 165 000 pour les quatre nucléosides canoniques. Le couplage du FTICR SolariX 2XR (Bruker) avec la CE pour la caractérisation des modifications post-transcriptionnelles sacrifie donc plus de 90 % de la résolution accessible en infusion.

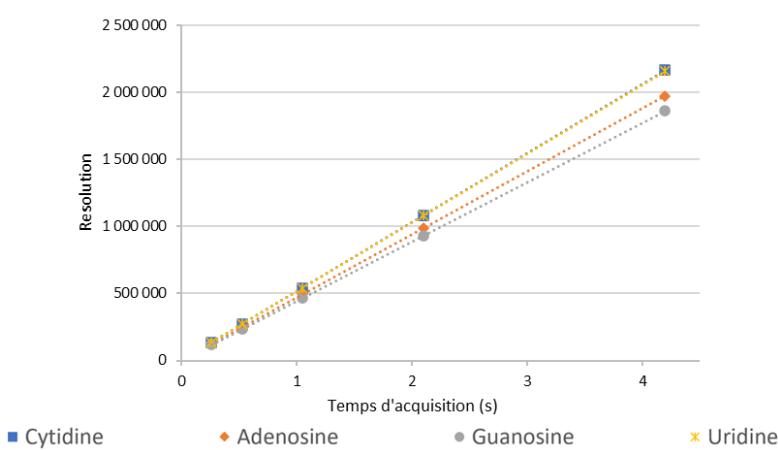


Figure 4.15 : Evolution de la résolution en fonction du temps d'acquisition dans la cellule ICR. Le temps d'acquisition dans la cellule ICR a été défini en changeant le DataPoint Size.

Si la résolution obtenue avec les paramètres optimaux pour le couplage ne suffit pas pour une application particulière, il sera possible de travailler en mode absorption. Celui-ci permet de corriger les défauts de phase des ions dans la cellule ICR en retraitant le signal a

posteriori, menant à une augmentation de la résolution et de la précision de mesure de masse [4]. Cependant, le mode absorption n'est pas simple à mettre en œuvre, puisqu'il nécessite une calibration de toutes les masses unitaires de la gamme de m/z de travail et plusieurs retraitements de signal avant d'obtenir le meilleur résultat possible.

4.5.3. Précision de masse

La précision de masse est un autre avantage du FTICR-MS. Il est donc intéressant de comparer la précision de masse obtenue par le couplage CE-QTOF-MS et celle obtenue par le couplage CE-FTICR-MS. Pour cela, l'erreur sur la mesure de masse a été calculée à partir de 9 analyses (3 analyses par jour sur 3 jours consécutifs), après recalibration interne. En plus des quatre nucléosides canoniques, l'anhydride phtalique a été ajouté à la liste de calibration. Ce dernier, issu des esters phtaliques présents en grande quantité dans le plastique, est un contaminant usuel présent dans le bruit de fond ESI en mode positif. Les différentes masses utilisées pour cette calibration interne sont détaillées dans le Tableau 4.7.

La précision de masse moyenne et la déviation standard ont ensuite été calculées pour chaque nucléoside à partir des valeurs absolues en ppm et sont représentées dans la Figure 4.16.

Tableau 4.7 : Liste des masses références utilisées pour la calibration interne

Nom	Formule chimique	Charge	m/z
Anhydride phtalique	C ₈ H ₅ O ₃	+1	149,023320
Adénosine	C ₁₀ H ₁₄ N ₅ O ₄	+1	268,104030
Cytidine	C ₉ H ₁₄ N ₃ O ₅	+1	244,092797
Guanosine	C ₁₀ H ₁₄ N ₅ O ₅	+1	284,098945
Uridine	C ₉ H ₁₃ N ₂ O ₆	+1	245,076813

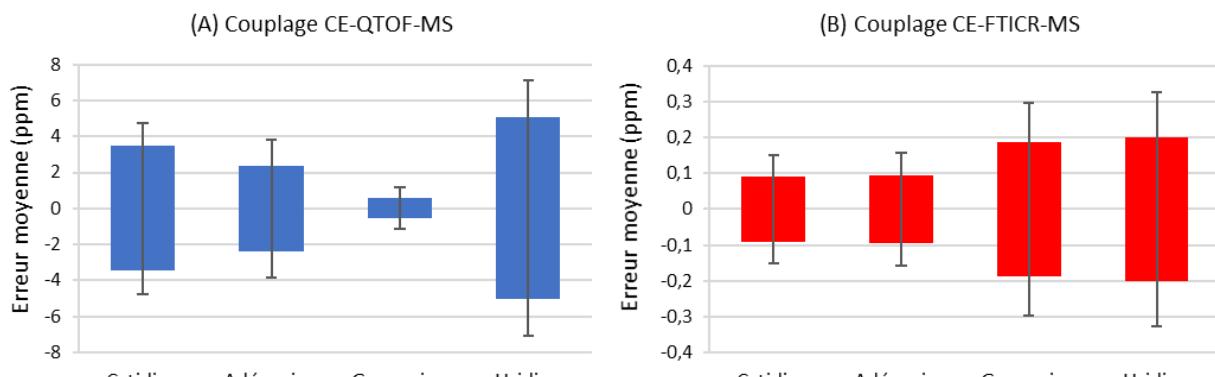


Figure 4.16 : Erreurs moyennes sur la mesure de masse des quatre nucléosides canoniques pour (A) le couplage CE-QTOF-MS et (B) le couplage CE-FTICR-MS ($n=9$).

La précision de masse du FTICR permet de réduire le nombre de formules brutes associées à chaque m/z mesuré. À titre d'exemple, le nombre de formules brutes pouvant être assignées à l'aide de SmartFormula (Bruker Daltonics) aux m/z mesurés des quatre nucléosides canoniques avec des tolérances de 10 ppm pour le couplage CE-QTOF-MS et de

0,5 ppm pour le couplage CE-FTICR-MS, est répertorié dans le Tableau 4.8. Plusieurs formules brutes respectant la tolérance de 10 ppm sont proposées pour chacun des quatre m/z mesurés, alors qu'une unique formule brute est proposée lorsqu'une tolérance maximale de 0,5 ppm est appliquée. Obtenir une formule brute non ambiguë, grâce à la précision de masse du FTICR-MS, pourra être d'une grande utilité pour la caractérisation de modifications inconnues.

Tableau 4.8 : Nombre de formules brutes possibles données par SmartFormula (Bruker Daltonics) pour les m/z mesurés des quatre nucléosides canoniques selon l'erreur tolérée

Nucléoside	m/z mesuré	Nombre de formules brutes assignées	
		Tolérance ≤ 10 ppm	Tolérance $\leq 0,5$ ppm
C	244,092819	2	1
A	268,104020	2	1
G	284,098932	4	1
U	245,076813	3	1

4.5.4. Sensibilité

Une différence de sensibilité entre les deux couplages se fait ressentir. Le couplage CE-FTICR-MS semble en effet au moins deux fois moins sensible que le couplage CE-QTOF-MS. Cette différence de sensibilité ne se ressent pas (ou très peu) sur les standards des nucléosides canoniques. Cependant, elle devrait être plus ressentie sur les échantillons biologiques puisque les quatre nucléosides sont proportionnellement beaucoup plus abondants que les modifications post-transcriptionnelles qui font l'objet de cette étude. Moins la modification est abondante dans l'échantillon, plus la différence de sensibilité se fera ressentir. Si l'on prend l'exemple d'un ARNt constitué de 80 bases, une modification peut être présente une seule fois sur toute la séquence. Celle-ci sera donc environ 20 fois moins abondante qu'un des quatre nucléosides canoniques. Par la suite, la concentration des échantillons biologiques sera donc adaptée en fonction des résultats obtenus.

4.6. Conseils pratiques

A force d'expérience, quelques différences pratiques entre les deux couplages ont été observées et sont détaillées dans cette partie. Divers conseils sont donc donnés afin de faciliter, à chaque début de campagne, la remise en place du couplage CE-FTICR-MS qui reste plus délicat que le couplage CE-QTOF-MS.

4.6.1. Stabilité du spray

La stabilité du spray est un paramètre très important lors du couplage CE-MS. Avant chaque série d'analyses, il est essentiel de s'assurer de la stabilité du spray. Pour cela, le positionnement du capillaire par rapport au spectromètre de masse doit être adapté.

4.6.1.1. Positionnement du capillaire à l'entrée MS

Tout d'abord, la position optimale à l'entrée du spectromètre de masse est différente bien que les deux interfaces nanoESI soient identiques. La pointe du capillaire est placée en position co-axiale devant l'entrée des deux spectromètres de masse (Figure 1.14), mais la position exacte varie entre les deux couplages. Pour le couplage CE-QTOF-MS, il suffit de ne pas être totalement désaxé pour obtenir un spray stable. De plus, la distance entre la pointe du capillaire et la source nanoESI peut être comprise entre 1 et 5 mm. L'intensité du signal peut être légèrement augmentée en rapprochant la pointe du capillaire de l'entrée MS. Au contraire, sur le couplage CE-FTICR-MS, il est nécessaire de positionner précisément la pointe du capillaire dans l'axe de l'entrée MS, tout en étant plus éloigné. La distance minimale pour ce couplage est de 5 mm. Si le capillaire n'est pas exactement dans l'axe ou trop proche de l'entrée de la source, cela engendre une perte de spray ou une apparition de contaminations. Cette dernière pourrait s'expliquer par un effet de succion de la goutte du capillaire de contact, lorsque le capillaire est trop proche de la source.

Le passage d'un capillaire à un autre est donc plus délicat sur le couplage CE-FTICR-MS puisque la position optimale varie plus facilement. Les différentes géométries internes des deux appareils peuvent potentiellement être à l'origine de cette particularité. En effet, le capillaire de transfert est dans l'axe du quadrupôle pour le QTOF, mais perpendiculaire pour le FTICR (Figure 1.6 et Figure 1.7).

4.6.1.2. Réglage de la hauteur de la CE

Sur le couplage CE-QTOF-MS, l'entrée et la sortie du capillaire de séparation sont au même niveau. Ce réglage est idéal pour les couplages CE-MS puisqu'aucun flux supplémentaire n'est engendré du fait de la différence de niveau. Cependant, au début de la mise en place du couplage CE-FTICR-MS, l'obtention d'un spray stable et reproductible s'est montrée très délicate. Afin de pallier ce problème, le niveau de la CE a été légèrement relevé par rapport à l'entrée du FTICR. L'élévation légère de l'entrée du capillaire de séparation par rapport à la pointe dirigée vers l'entrée du FTICR engendre un flux supplémentaire en direction du spectromètre de masse. Cela a facilité la stabilité et la reproductibilité du spray. Cependant, ce flux supplémentaire impacte également la vitesse de migration des analytes et dégrade légèrement la séparation.

4.6.2. « Shimming » du FTICR-MS

Le *shimming* permet de recentrer la trajectoire cyclotronique des ions dans la cellule ICR et de minimiser le mouvement magnetron. Les fréquences harmoniques issues du mouvement magnetron sont à l'origine de pics « fantômes » dans le spectre de masse. Il est donc essentiel de bien « *shimmé* » le FTICR, afin de réduire la complexité des spectres MS/MS et de maximiser le signal. Le *shimming* se décale assez rapidement et peut causer une diminution de signal, ainsi qu'une augmentation de l'intensité des harmoniques qui

complexifient les spectres. Il n'est pas rare que le *shimming* doive être adapté plusieurs fois au cours d'une campagne d'analyses d'une semaine. Cela peut être réalisé sur des ions du bruit de fond sans découpler la CE du FTICR.

4.6.3. Calibration « on-line »

A contrario, il est important de noter que la calibration reste correcte plusieurs jours, ce qui permet de travailler pendant une semaine sans avoir à découpler la CE du FTICR. Une calibration « *on-line* » est réalisable en utilisant une *lock mass* afin de réduire la déviation de la calibration. Cette calibration « en ligne » est surtout nécessaire lorsque la résolution des spectres est très élevée, puisque qu'une petite déviation pourrait induire un dédoublement des pics lorsque plusieurs spectres sont sommés. La *lock mass* peut être soit une contamination présente dans le bruit de fond dont on connaît la formule brute, soit un standard volatile imbibé sur un coton dans un réservoir positionné au-dessus de la source nanoESI (Figure 4.17). Nous avons fait le choix d'utiliser l'anhydride phtalique (Tableau 4.7) comme « *lock mass* », puisque celui-ci est présent dans le bruit de fond tout au long des analyses.

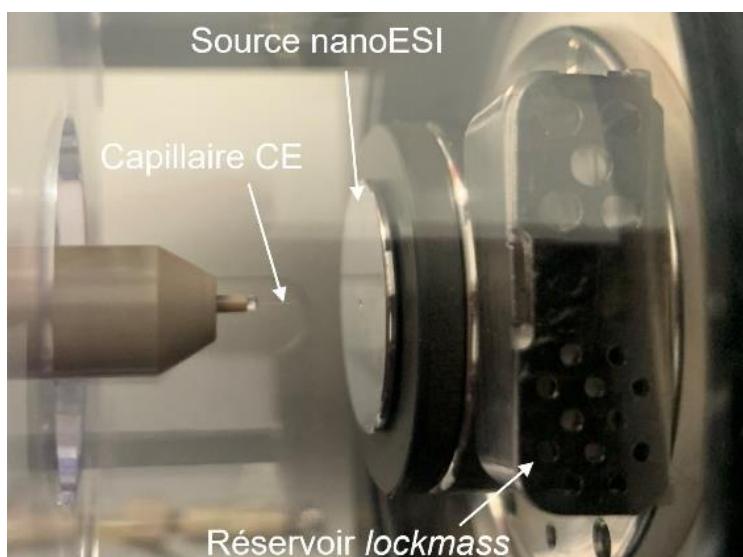


Figure 4.17 : Position de la *lock mass* dans l'interface CESI-MS du FTICR SolariX 7T 2XR (Bruker)

4.6.4. Stockage de données

Les fichiers d'analyses issus du couplage CE-FTICR-MS sont très imposants. Afin de limiter la taille des fichiers, il est essentiel d'appliquer un facteur de réduction d'au moins 95% permettant de supprimer le bruit de fond selon un certain seuil. À titre d'exemple, après réduction de 97%, une analyse CE-FTICR-MS/MS de 35 min a une taille d'environ 10 Go.

4.7. Conclusion

Les travaux présentés dans ce chapitre détaillent la mise en place du couplage CE-FTICR-MS au laboratoire, de l'obtention du premier électrophérogramme basé sur une

analyse d'anticorps par approche bottom-up, au retour sur une application plus simple pour l'étude sur les paramètres influant sur la vitesse d'acquisition réalisée à partir de nucléosides standards. Cette dernière étude a permis de mettre en lumière le compromis à trouver entre performances du FTICR-MS et la vitesse d'acquisition que requière le couplage avec la CE, comme cela avait été annoncé dans le chapitre 2.

Bien que l'étude réalisée dans ce chapitre ait été essentielle pour la mise en place du couplage CE-FTICR-MS au laboratoire et qu'elle ait démontré la nécessité de trouver le bon compromis entre résolution et vitesse d'acquisition, il serait utile de la mener à chaque fois qu'une nouvelle application est développée ou transférée sur ce couplage, afin de s'assurer que les paramètres influant sur la vitesse d'acquisition sont judicieusement choisis. En effet, les besoins en termes de résolution, de séparation, de quantité de données MS/MS, et de temps de cycle maximal, dépendent du couplage avec l'électrophorèse capillaire, mais également de l'application choisie et des objectifs fixés.

Cette étude a également mis en évidence les apports du couplage CE-FTICR-MS par comparaison au couplage CE-QTOF-MS. Notamment, la grande précision de mesure de masse MS apportée par le FTICR pourra être mise à profit lors d'études de modifications inconnues, afin d'assigner, sans ambiguïté, une unique formule brute à chaque modification inconnue.

Enfin, l'ensemble des travaux présentés dans ce chapitre ont permis d'acquérir de l'expérience, de mieux appréhender les particularités et de pallier les difficultés intrinsèques du délicat couplage CE-FTICR-MS. Grâce à cela, la remise en place du couplage est de plus en plus rapide à chaque début de campagne. La méthode développée sur les nucléosides standards est devenue une application de référence pour ce couplage et est utilisée à chaque début de campagne afin de contrôler l'état et les performances de celui-ci.

Dans le chapitre suivant, la méthode développée sera évaluée sur de vrais échantillons d'ARN biologiques dont les modifications sont référencées dans la littérature. Ces échantillons serviront de preuve de concept pour le couplage CE-FTICR-MS.

4.8. Références bibliographiques

- [1] Beck, A. *et al.*, “Biosimilar, Biobetter, and Next Generation Antibody Characterization by Mass Spectrometry.”, *Anal. Chem.*, **2012**, 84 (11), 4637–4646.
- [2] Giorgetti, J. *et al.*, “Combination of Intact, Middle-up and Bottom-up Levels to Characterize 7 Therapeutic Monoclonal Antibodies by Capillary Electrophoresis – Mass Spectrometry.”, *J. Pharm. Biomed. Anal.*, **2020**, 182, 113107.
- [3] Gahoual, R. *et al.*, “Rapid and Multi-Level Characterization of Trastuzumab Using Sheathless Capillary Electrophoresis-Tandem Mass Spectrometry.”, *mAbs*, **2013**, 5 (3), 479–490.
- [4] Da Silva, M. P. *et al.*, “Absorption Mode Spectral Processing Improves Data Quality of Natural Organic Matter Analysis by Fourier-Transform Ion Cyclotron Resonance Mass Spectrometry.”, *J. Am. Soc. Mass Spectrom.*, **2020**, 31 (7), 1615–1618.

CHAPITRE 5 : EVALUATION DU TRANSFERT DE METHODE VERS LE COUPLAGE CE-FTICR-MS

5.1. Introduction

Précédemment, la méthode d'identification des modifications post-transcriptionnelles a été transférée et développée sur le couplage CE-FTICR-MS. Les performances de ce couplage ont été étudiées uniquement à l'aide des standards des quatre nucléosides canoniques, non-modifiés. Cependant, la méthode a été transférée sur le couplage CE-FTICR-MS dans l'idée d'obtenir des informations supplémentaires, par rapport au couplage CE-QTOF-MS, sur de vraies modifications d'ARN. En effet, ce couplage doit pouvoir apporter une plus grande précision de masse et ainsi fournir des formules brutes uniques. Cela est particulièrement intéressant pour les modifications inconnues qui ne sont pas encore répertoriées dans la littérature. Avant de pouvoir s'intéresser aux modifications non référencées dans la littérature, il est essentiel d'évaluer le couplage CE-FTICR-MS sur de vrais échantillons biologiques, plus complexes, mais déjà bien connus et décrits. Cette évaluation fait l'objet de ce chapitre.

Dans un premier temps, quatre ARN de transfert (ARNt) purifiés et issus de *Saccharomyces cerevisiae* ont servi de modèles. En effet, les modifications post-transcriptionnelles de ces ARNt sont bien connues et référencées dans les bases de données telles que Modomics [1]. En plus des quatre nucléosides canoniques, chacun de ces ARNt possède une dizaine de modifications dont une grande majorité de méthylations. Les méthylations font partie des modifications les plus fréquentes pour toutes les classes d'ARN, et sont impliquées dans de nombreuses fonctions biologiques, de la stabilité structurale de l'ARN, au mécanisme de résistance aux antibiotiques de certaines bactéries [2]. Ces échantillons, et leurs modifications, sont donc particulièrement bien adaptés pour évaluer le nouveau couplage sur un cas concret, mais connu. Ce sont également les mêmes ARN que ceux utilisés dans l'étude de Lechner *et al.* [3] pour démontrer la puissance de la méthode CE-MS/MS pour l'étude des modifications d'ARN lors de son développement sur le couplage CE-QTOF-MS.

Puis, le couplage CE-FTICR-MS a été évalué sur un échantillon plus complexe : un extrait total de tous les ARNt issus de *Bos taurus*. Mélange de plusieurs ARNt, cet extrait total est composé de nombreuses méthylations, mais également de modifications plus complexes. Il est un bon modèle pour évaluer le couplage CE-FTICR-MS puisque ces modifications variées sont, encore une fois, bien connues et référencées dans les bases de données telles que Modomics.

Pour chacun de ces échantillons, les analyses issues des deux couplages ont été mises en parallèle et les résultats obtenus ont été comparés afin d'évaluer les performances du couplage CE-FTICR-MS sur de vrais échantillons biologiques.

5.2. Preuve de concept sur des ARNt purifiés

Dans un premier temps, le couplage CE-FTICR-MS a été évalué sur les ARNt purs issus de *Saccharomyces cerevisiae* et utilisés lors de la preuve de concept de la méthode CE-MS/MS sur le couplage avec le QTOF [3]. Les séquences modifiées de ces quatre ARNt sont référencées dans la base de données Modomics [1] et sont présentées en Figure 5.1. À partir de ces séquences, la liste des nucléosides, modifiés et non modifiés, théoriquement présents dans ces échantillons a été recensée dans le Tableau 5.1 avec les différents *m/z* des ions précurseurs et des fragments permettant leurs identifications.

ARNt^{Phe GAA}	G C G G A U U U A [m²G] C U C A G [D] [D] G G G A G A G C [m²G] C C A G A [Cm] U [Gm] A A [yW] A [Ψ] [m⁵C] U G G A G [m⁷G] U C [m⁵C] U G U G [m⁵U] [Ψ] C G [m¹A] U C C A C A G A A U U U C G C A C C A
ARNt^{Trp CCA}	G A A G C G G U [m¹G] [m²G] C U C A A [D] [Gm] G [D] A G A G C [Ψ] [Ψ] [Ψ] C G A [Cm] U [Cm] C A A A [Ψ] C G A A G G [m⁷G] [D] U G C A G G [m⁵U] [Ψ] C A [m¹A] U U C C U G [Ψ] C C G U U U C A C C A
ARNt^{Asp GUC}	U C C G U G A U A G U U [Ψ] A A [D] G G [D] C A G A A U G G G C G C [Ψ] U G U C [m¹G] C G U G C C A G A U [m⁵C] G G G G [m⁵U] [Ψ] C A A U U C C C G U C G C G G A G C C A
ARNt^{Ala AGC}	G G G C G U G U [m¹G] G C G U A G [D] C G G [D] A G C G C [m²G] C U C C C U U [I] G C [m¹I] [Ψ] G G G A G A G G [D] C U C C G G [m⁵U] [Ψ] C G A U U C C G G A C U C G U C C A C C A

Figure 5.1 : Séquences des quatre ARNt étudiés issus de *S. cerevisiae* (**ARNt^{Phe GAA}**, **ARNt^{Trp CCA}**, **ARNt^{Asp GUC}** et **ARNt^{Ala AGC}**) d'après la base de données Modomics [1]. Les modifications post-transcriptionnelles sont indiquées en bleu et entre crochets.

Tableau 5.1 : Récapitulatif des nucléosides présents dans les **ARNt^{Phe GAA}**, **ARNt^{Trp CCA}**, **ARNt^{Asp GUC}** et **ARNt^{Ala AGC}**, issus de *S. cerevisiae*, et des données MS et MS/MS utiles à leur identification

Nucléoside	Symbole	<i>m/z</i> de l'ion précurseur (+1)	<i>m/z</i> des fragments produits (+1)
Cytidine	C	244,092797	112
Adénosine	A	268,104030	136
Guanosine	G	284,098945	152
Uridine	U	245,076813	113
2'-O-methylcytidine	Cm	258,108447	112
5-methylcytidine	m ⁵ C	258,108447	126
1-methyladenosine	m ¹ A	282,119680	150
Inosine	I	269,088046	137
1-methylinosine	m ¹ I	283,103696	151
2'-O-methylguanosine	Gm	298,114595	152
1-methylguanosine	m ¹ G	298,114595	166
N2-methylguanosine	m ² G	298,114595	166
7-methylguanosine	m ⁷ G	298,114595	166
N2,N2,2'-O-trimethylguanosine	m ² ₂ G	312,130245	180
Dihydrouridine	D	247,092463	115 ; 97
Pseudouridine	Ψ	245,076813	209 ; 179 ; 155
5-methyluridine	m ⁵ U	259,092463	127
Wybutosine	yW	509,199053	377

Afin d'éviter des différences issues d'une potentielle dégradation des échantillons entre les deux analyses, chaque échantillon a été réanalysé sur le couplage CE-QTOF-MS juste avant d'entamer la campagne sur le couplage CE-FTICR-MS. De plus, pour pallier les différences de sensibilité observées entre les deux couplages (cf paragraphe 4.5.4), ces quatre échantillons ont été analysés à une concentration de 0,1 g/L au lieu de 0,05 g/L.

Les électrophérogrammes obtenus par le couplage CE-FTICR-MS lors des analyses de nucléosides des ARNt^{Phe GAA}, ARNt^{Asp GUC}, ARNt^{Ala AGC} et ARNt^{Trp CCA} sont présentés en Figure 5.2. Les données, MS et MS/MS, obtenues par les deux couplages pour identifier les nucléosides, contenus dans cet échantillon, sont comparées dans le Tableau 5.2.

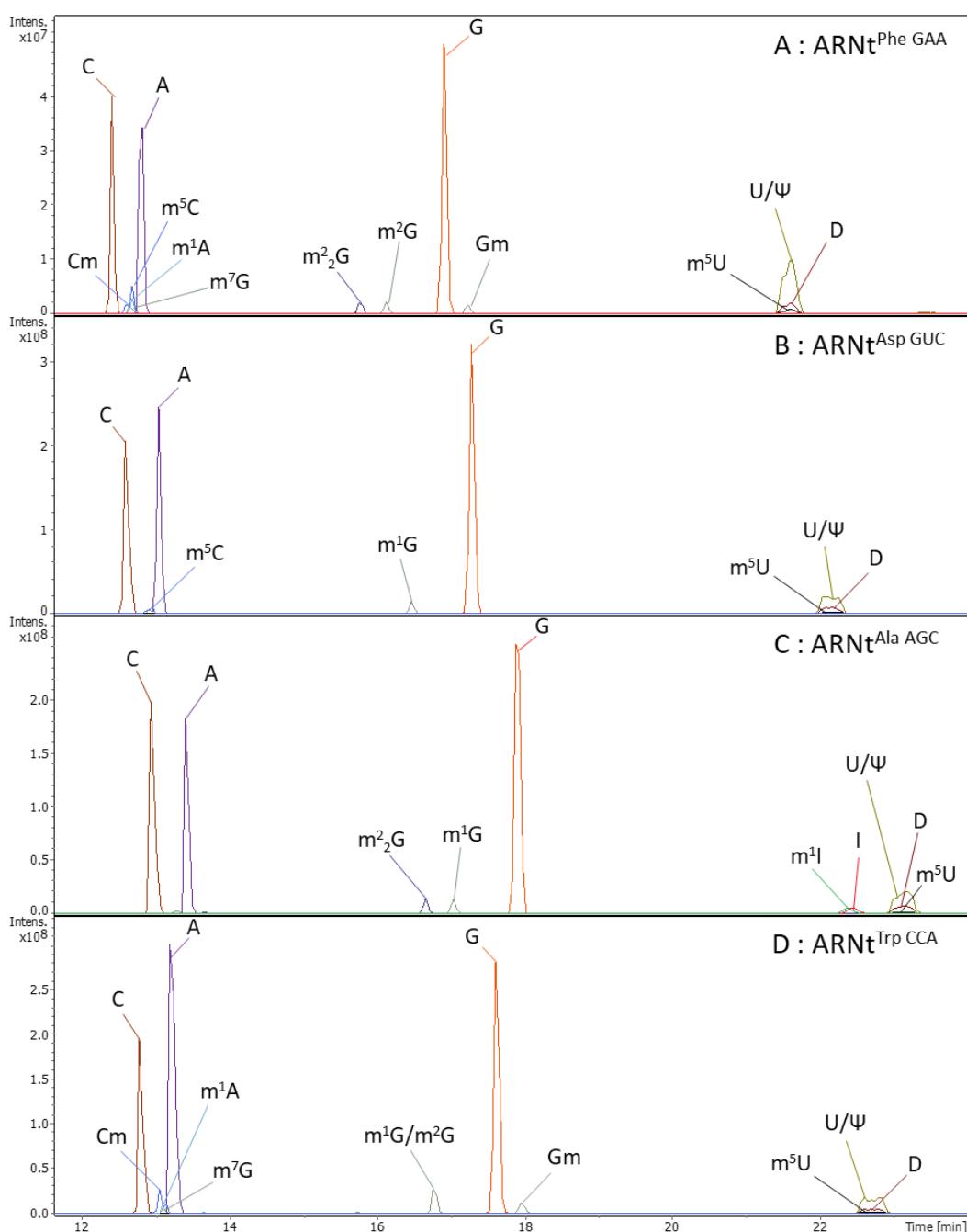


Figure 5.2 : Electrophérogrammes, obtenus par CE-FTICR-MS/MS, des (A) ARNt^{Phe GAA}, (B) ARNt^{Asp GUC}, (C) ARNt^{Ala AGC} et (D) ARNt^{Trp CCA}, issus de *S. cerevisiae*, après digestion à la Nucléase P1 et déphosphorylation à la BAP.

Tableau 5.2 : Comparaison des données MS et MS/MS obtenues pour l'identification des nucléosides des (A) ARNt^{Phe GAA}, (B) ARNt^{Asp GUC}, (C) ARNt^{Ala AGC} et (D) ARNt^{Trp CCA}, issus de *S. cerevisiae*, par les couplages CE-QTOF-MS et CE-FTICR-MS.

A : ARNt ^{Phe GAA}		C	A	G	U	Cm	m ⁵ C	m ¹ A	Gm	m ² G	m ⁷ G	m ² ₂ G	W	m ⁵ U	D	Ψ
CE-QTOF-MS	MS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MS/MS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CE-FTICR-MS	MS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MS/MS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	✗
B : ARNt ^{Asp GUC}		C	A	G	U		m ⁵ C		m ¹ G		m ⁵ U		D		Ψ	
CE-QTOF-MS	MS	✓	✓	✓	✓		✓		✓		✓		✓	✓	✓	✓
	MS/MS	✓	✓	✓	✓		✓		✓		✓		✓	✓	✓	✓
CE-FTICR-MS	MS	✓	✓	✓	✓		✓		✓		✓		✓	✓	✓	✓
	MS/MS	✓	✓	✓	✓		✓		✓		✓		✓	✓	✓	✗
C : ARNt ^{Ala AGC}		C	A	G	U	-	m ¹ I		m ¹ G		m ² ₂ G		m ⁵ U		D	Ψ
CE-QTOF-MS	MS	✓	✓	✓	✓		✓		✓		✓		✓	✓	✓	✓
	MS/MS	✓	✓	✓	✓		✓		✓		✓		✓	✓	✓	✓
CE-FTICR-MS	MS	✓	✓	✓	✓		✓		✓		✓		✓	✓	✓	✓
	MS/MS	✓	✓	✓	✓		✓		✓		✓		✓	✓	✓	✗
D : ARNt ^{Trp CCA}		C	A	G	U	Cm	m ¹ A	Gm	m ¹ G/m ² G	m ⁷ G	m ⁵ U		D		Ψ	
CE-QTOF-MS	MS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MS/MS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CE-FTICR-MS	MS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MS/MS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✗

Les analyses réalisées sur le couplage CE-QTOF-MS, servant de références, montrent que toutes les modifications, attendues pour les quatre ARNt issus de *S. cerevisiae*, ont été identifiées à la fois par MS et MS/MS. Pour ce qui est du couplage CE-FTICR-MS, tous les nucléosides ont été détectés par MS, et la grande majorité des modifications ont pu être confirmées par MS/MS.

Pour les quatre échantillons, la modification Ψ, isomérisation de U appelée pseudouridine, n'a pas été identifiée par MS/MS avec le couplage CE-FTICR-MS/MS. Un spectre MS/MS est disponible, mais aucun de ses fragments caractéristiques n'a été détecté. Cela peut s'expliquer par sa faible abondance et la potentielle faible efficacité d'ionisation de

cette modification neutre. Il s'agit de la seule modification non identifiée par le couplage CE-FTICR-MS/MS pour les ARNt^{Asp GUC} et ARNt^{Ala AGC}.

Deux autres modifications n'ont pas pu être identifiées par MS/MS avec le couplage CE-FTICR-MS/MS pour les ARNt^{Phe GAA} et ARNt^{Trp CCA}. Il s'agit de m⁵U, non identifiée dans les deux échantillons, et yW, non identifiée dans l'ARNt^{Phe GAA}. En effet, du fait de leurs faibles intensités, ces modifications n'ont pas été sélectionnées pour être fragmentées lors du Top2 de l'autoMS/MS.

Cependant, il est également important de noter que, pour l'ARNt^{Phe GAA}, malgré les nombreuses modifications qui co-migrent vers 12,6 min (Cm, m⁵C, m¹A et m⁷G), toutes ces modifications ont pu être identifiées par MS et MS/MS grâce à l'optimisation de la méthode MS/MS du FTICR (cf partie 2 chapitre 4).

Ces premières analyses montrent que le couplage CE-FTICR-MS/MS peut servir pour étudier de vrais échantillons biologiques. Dans les prochains paragraphes, le couplage va donc être évalué sur des échantillons biologiques de plus en plus complexes.

5.3. Preuve de concept sur un extrait total d'ARNt

Pour aller plus loin dans l'évaluation du couplage CE-FTICR-MS, un extrait total d'ARNt de *Bos taurus* a été analysé. Il s'agit d'un échantillon dont les modifications sont également connues et répertoriées dans la littérature, et dans la base de données Modomics [1]. Cependant, les modifications post-transcriptionnelles présentes dans cet extrait sont beaucoup plus nombreuses, puisqu'il est composé d'au moins quinze ARNt, déjà décrits dans la littérature (Figure 5.3), par rapport aux ARNt purifiés étudiés dans la partie précédente. Cet échantillon est ainsi un bon exemple pour évaluer les performances du couplage. La capacité du couplage à identifier de nombreuses modifications, tout en ayant un temps de cycle important (2,8 s) malgré un nombre limité d'ions sélectionnés pour la fragmentation (2), va pouvoir être évaluée. Les séquences des ARNt référencées dans la base de données Modomics [1] pour cet extrait total sont présentées en Figure 5.3.

À partir de ces séquences, la liste de modifications à rechercher dans les analyses a été recensée dans le Tableau 5.3 avec les différents *m/z* des ions précurseurs et des fragments permettant leurs identifications.

Afin de pallier les différences de sensibilité observées entre les deux couplages (cf paragraphe 4.5.4), cet échantillon a été analysé à une concentration de 0,1 g/L sur le couplage CE-QTOF-MS et à une concentration de 0,8 g/L sur le couplage CE-FTICR-MS.

L'électrophérogramme obtenu sur le couplage CE-FTICR-MS lors de l'analyse des nucléosides de l'extrait total d'ARNt de *B. taurus* est présenté en Figure 5.4 , et la comparaison

des données MS et MS/MS obtenues par les deux couplages pour identifier les différentes modifications est présentée dans le Tableau 5.4.

ARNt^{Arg CGG}	G A C C C A G U [m ¹ G] [m ² G] C C U A A [D] [Gm] G A [D] A A G G C A [ψ] C A G C [Cm] U C C G [m ¹ G] A G C U G G G A [D] U G [ψ] G G G [m ⁶ U] [ψ] C G [m ¹ A] G U C C C A U C U G G G U C G C C A
ARNt^{Arg CCU}	G C C C C A G U [m ¹ G] [m ² G] C C U A A [D] G G A [D] A A G G C A [ψ] U G G C [m ³ C] U C C U [t ⁶ A] A G C C A G G G A [D] U G [ψ] G G G [m ⁶ U] [ψ] C G [m ¹ A] G U C C C A U C U G G G G [ψ] C C C A
ARNt^{Arg [m³C]C[ψ]}	G G C U C C G U [m ¹ G] G C G C A A [D] G G A [D] A G G C A [ψ] [ψ] G G A [m ³ C] U [mcm ⁵ s ² U] C [ψ] [t ⁶ A] A U [ψ] C A A A G [m ⁷ G] [D] U [m ⁵ C] C G G G [m ⁵ U] [ψ] C G [m ¹ A] G U C C C G G G A G U C G C C A
ARNt^{Asn [Q]UU}	G U C U C U G U [m ¹ G] [m ² G] C G C A A [D] C G G [D] [acp ³ U] A G C G C [m ² ,G] [ψ] [ψ] C G G C U [Q] U U [t ⁶ A] A C C G A A A G [m ⁷ G] [D] U G G U G G [ψ] [ψ] C G [m ¹ A] G C C C A C C C A G G G A C G C C A
ARNt^{Asp [manQ]UC}	G G U G C C G U A [m ² G] C G [ψ] A G [D] [Gm] G C A [...] C G [...] G A C U C U [Cm] U [manQ] U C A A A G A G U G G [m ⁷ G] A [m ⁵ C] [...] U G A G [m ⁵ U] [ψ] C G [m ¹ A] U A C U C A C G G A C C G C C A
ARNt^{Gln CUG}	G G U U C C A U [m ¹ G] G U G [ψ] A A [D] [Gm] G [D] [D] A G C A C U C [ψ] G G A [Cm] U C U G A A [ψ] C C A G C G A U [m ⁵ C] C G A G [ψ] [ψ] C A [m ¹ A] A U C U C G G U G G A A C C U C C A
ARNt^{Leu [I]AG}	G G U A G C G U G [m ² G] C [ac ⁴ C] G A G C G G [D] C [ψ] A A G G C [m ² ,G] C U G G A [ψm] U [I] A G [m ¹ G] C [ψ] C C A G U C [ψ] C [ψ] U C G G G G G [m ⁵ C] G U G G G [m ⁵ U] [ψ] C G [m ¹ A] A U C C C A C C G C U G C C A C C A
ARNt^{Leu [f⁵Cm]AA}	G U C A G [m ² G] A U G [m ² G] C [ac ⁴ C] G A G U G G [D] C [ψ] A A G G C [m ² G] C C A G A C U [f ⁵ Cm] A A [m ¹ G] [ψ] C U G G [Um] C [ψ] C C G U A U G G A G G [m ⁵ C] G U G G G [m ⁵ U] [ψ] C G [m ¹ A] A U C C C A C U U C U G A C A C C A
ARNt^{Leu CAG}	G U C A G [m ² G] A U G [m ² G] C [ac ⁴ C] G A G C G G [D] [D] C A A G G C [m ² ,G] C U G C G [Um] U C A G [m ¹ G] [...] C G C A G [Um] C [ψ] C C C U G G A G G [m ⁵ C] G U G G G [m ⁵ U] [ψ] C G [m ¹ A] A U C C C A C U U C U G A C A C C A
ARNt^{Phe [Gm]AA}	G C C G A A A U A [m ² G] C U C [m ¹ A] G [D] [D] G G G A G A G C [m ² ,G] [ψ] [ψ] A G A [Cm] U [Gm] A A [o ₂ yW] A [ψ] C U A A A G [m ⁷ G] [D] C [m ⁵ C] C U G G [m ⁵ U] [ψ] C G [m ¹ A] U C C C G G U U U C G G C A C C A
ARNt^{Sec [mcm⁵U]CA}	G C C C G G A U G A C C C U C A G U G G U C U G G G G U G C A G G C U [mcm ⁵ U] C A [i ⁶ A] A C C U G U A G C U G U C U A G C G A C A G A G U G G U [ψ] C A [m ¹ A] U U C C A C C U U U C G G G C G C C A
ARNt^{Thr [I]GU}	G G C C C U G U [m ¹ G] [m ² G] C U A G C [D] G G [D] C A A A G C [m ² ,G] C C U G U [m ³ C] U [I] G U [t ⁶ A] A A C A G G A G A [D] [m ⁵ C] C U G G G U [ψ] C G [m ¹ A] A U C C C A G C G G G G [m ⁵ C] U C C A
ARNt^{Trp [Cm]CA}	G A C C U C [m ² G] U [m ¹ G] [m ² G] C G C A A [D] [Gm] G [D] A G C G C [m ² ,G] [ψ] C U G A [Cm] U [Cm] C A [m ¹ G] A [ψm] C A G A A G [m ⁷ G] [D] U G C G U G [ψ] [ψ] C G [m ¹ A] A U C A C G U C G G G G U C A C C A
ARNt^{Tyr [galQ][ψ]A}	C C U U C [m ² G] A U A [m ² G] C U C A G [D] [D] G G [acp ³ U] U A G A G C [m ² ,G] [m ² ,G] A G G A C U [galQ] [ψ] A [m ¹ G] A [m] C C U U A G [m ⁷ G] [D] [m ⁵ C] G C U G G [m ⁵ U] [ψ] C G [m ¹ A] U U C C G G C U C G A A G G A C C A
ARNt^{Lys [mcm⁵s²U]UU}	G C C C G [m ² G] A U A [m ² G] C U C A G [D] C G G [D] A G A G C A [ψ] C A G A C U [mcm ⁵ s ² U] U U A [ψ] C U G A G G [m ⁷ G] [D] [m ⁵ C] [m ⁵ C] A G G G [m ⁵ Um] [ψ] C A G [m ¹ A] G [m ¹ A] G U C C C U G U U U C G G G C G C C A

Figure 5.3 : Séquences des ARNt de l'extrait total de *B. taurus*, d'après la base de données Modomics [1]. Les modifications post-transcriptionnelles sont indiquées en bleu et entre crochets. Les modifications [...] et [...] sont décrites comme inconnues.

Tableau 5.3 : Récapitulatif des nucléosides présents dans l'extrait total d'ARNt de *B. taurus*, et des données MS et MS/MS utiles à leur identification.

Nucléoside	Symbole	<i>m/z</i> de l'ion précurseur (+1)	<i>m/z</i> des fragments produits (+1)
Cytidine	C	244,092797	112
Adénosine	A	268,104030	136
Guanosine	G	284,098945	152
Uridine	U	245,076813	113
1-methyladenosine	m^1A	282,119680	150
3-methylcytidine	m^3C	258,108447	126
5-methylcytidine	m^5C	258,108447	126
2'-O-methylcytidine	Cm	258,108447	112
1-methylguanosine	m^1G	298,114595	166
N2-methylguanosine	m^2G	298,114595	166
7-methylguanosine	m^7G	298,114595	166
2'-O-methylguanosine	Gm	298,114595	152
N2,N2,2'-O-trimethylguanosine	m^2_2G	312,130245	180
N6-isopentenyladenosine	i^6A	336,166631	204 ; 136 ; 148
3-(3-amino-3-carboxypropyl)uridine	acp ³ U	346,124491	214 ; 197 ; 168
Queuosine	Q	410,167025	295 ; 163
Galactosyl-queuosine	galQ	572,219848	295 ; 163 ; 440
Mannosyl-queuosine	manQ	572,219848	295 ; 163 ; 440
Pseudouridine	ψ	245,076813	209 ; 179 ; 155
Dihydouridine	D	247,092463	115 ; 97
5-methyluridine	m^5U	259,092463	127
2'-O-methyluridine	Um	259,092463	113
2'-O-methylpseudouridine	ψm	259,092463	223
5,2'-O-dimethyluridine	m^5Um	273,108113	127
N4-acetylcytidine	ac ⁴ C	286,103362	154 ; 112
5-methoxycarbonylmethyl-2-thiouridine	mcm ⁵ s ² U	333,075098	172 ; 255 ; 273 ; 141
Inosine	I	269,088046	137
N6-threonylcarbamoyladenosine	t ⁶ A	413,141538	281
Methoxycarbonylmethyluridine	mcm ⁵ U	317,097942	185 ; 153 ; 125
Peroxywybutosine	o ₂ yW	541,188882	409
5-formyl-2'-O-methylcytidine	f ⁵ Cm	286,103362	140

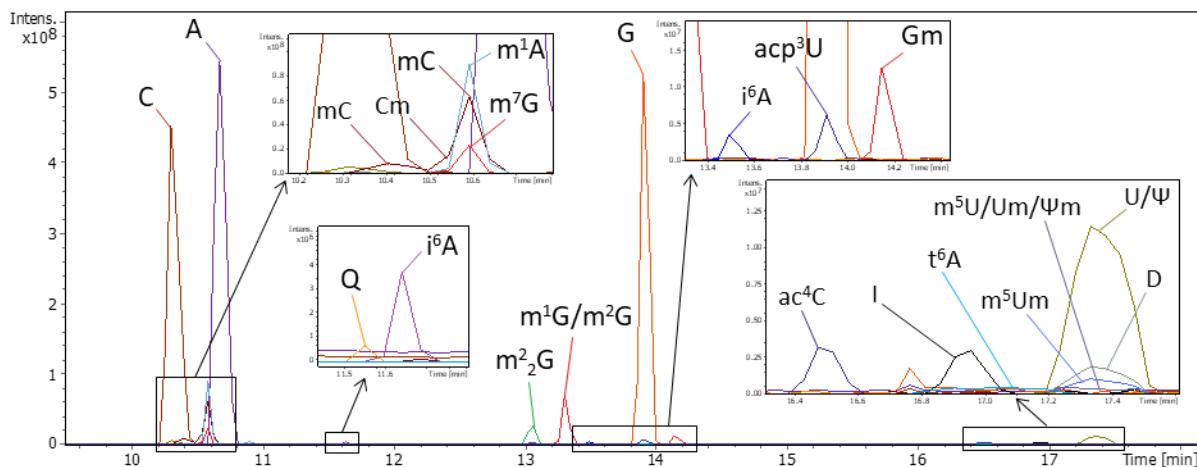


Figure 5.4 : Electrophérogramme, obtenu par CE-FTICR-MS/MS, de l'extrait total d'ARNt de *B. taurus*, après digestion à la Nucléase P1 et déphosphorylation à la BAP.

Tableau 5.4 : Comparaison des données MS et MS/MS obtenues pour l'identification des nucléosides de l'extrait total d'ARNt de *B. taurus*, par les couplages CE-QTOF-MS et CE-FTICR-MS.

	C	A	G	U	Cm	m^3C	m^5C	ac^4C	f^5Cm	m^1A	i^6A	t^6A	-	Gm	m^1G/m^2G
CE-QTOF-MS	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓
	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓
CE-FTICR-MS	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓
	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓

	m^7G	m^2G	Q	$galQ/manQ$	$\sigma_2\gamma W$	Um	m^5U	m^5Um	acp^3U	mcm^5U	mcm^5s^2U	D	Ψ	Ψm
CE-QTOF-MS	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓
	✓	✓	✓	✓	✗	✓	✓	✓	✓	✗	✗	✓	✓	✓
CE-FTICR-MS	✓	✓	✓	✓	✓	✗	✓	✓	✓	✗	✗	✓	✓	✓
	✓	✓	✓	✓	✗	✓	✓	✗	✓	✗	✗	✓	✗	✗

L'analyse, de cet extrait total d'ARNt issu de *B. taurus*, réalisée sur le couplage CE-QTOF-MS, a permis d'identifier, à la fois par MS et MS/MS, 27 nucléosides (dont 23 modifications post-transcriptionnelles) sur les 31 référencés dans le Tableau 5.3. Deux modifications supplémentaires (mcm^5U et mcm^5s^2U) n'ont pas été confirmées par MS/MS par manque de fragmentation dû à leurs faibles intensités. Les deux dernières modifications n'ont pas du tout été détectées avec le couplage CE-QTOF-MS et ne sont donc pas attendues sur le couplage CE-FTICR-MS.

Mis à part les deux modifications mcm^5U et mcm^5s^2U , tous les nucléosides ont été détectés par MS, et la grande majorité des modifications ont pu être confirmées par MS/MS

avec le couplage CE-FTICR-MS. Seules trois modifications (m^5Um , ψ et ψm) n'ont pas pu être identifiées par MS/MS avec ce couplage. En effet, du fait de la faible intensité de m^5Um , ce nucléoside n'a pas été sélectionné pour être fragmenté. Quant à ψ et ψm , isomérisations respectives de U et Um, un spectre MS/MS est disponible, mais aucun de ses fragments caractéristiques n'a été détecté comme cela a été le cas pour ψ .

Dans tous ces échantillons, ψ est la seule modification à poser un problème récurrent d'identification par le couplage CE-FTICR-MS. En effet, cette modification, neutre dans les conditions d'analyses (pH 2.2), co-migre avec tous les autres analytes neutres, notamment avec son isomère U. Les données MS combinées aux mobilités électrophorétiques, ne permettent pas de discriminer les deux nucléosides. Par conséquent, ils ne peuvent être différenciés que par la MS/MS. Ainsi l'absence des fragments caractéristiques de ψ impacte directement son identification. Cette absence de signal pourrait être causée par un effet de suppression plus important que lors du couplage CE-QTOF-MS, ou par les différences de sensibilité observée lors de l'étude du couplage (cf paragraphe 4.5.4). La pseudouridine, ψ , est une des modifications post-transcriptionnelles les plus courantes chez les ARN codants et non-codants [4] permettant d'augmenter les interactions ARN-ARN et ARN-protéines [5]. Comme elle est présente chez la majorité des ARN, il serait intéressant de trouver une alternative pour améliorer sa détection par FTICR-MS. Une des stratégies serait, dans un premier temps, d'analyser un standard de ψ pour s'affranchir de la potentielle compétition d'ionisation avec U, et diagnostiquer la source du problème.

L'absence de données MS/MS concernant les nucléosides non sélectionnés pour la fragmentation s'explique par le nombre d'ions précurseurs fragmentés par cycle. En effet, comme cela est expliqué dans le chapitre 4, afin de ne pas avoir un temps de cycle trop long, seulement 2 ions précurseurs sont sélectionnés pour chaque cycle de 2,8 s lors du couplage CE-FTICR-MS/MS, contre 7 pour chaque cycle de 2,0 s lors du couplage CE-QTOF-MS/MS. Nécessairement, moins d'ions précurseurs différents sont fragmentés avec le couplage CE-FTICR-MS. Cela est d'autant plus marqué lorsque de nombreuses modifications de faibles intensités co-migrent, comme c'est le cas pour m^5Um . Pour pallier cette différence, il est envisageable d'analyser plusieurs fois le même échantillon, en excluant les modifications déjà confirmées par MS/MS, et en combinant les résultats obtenus.

5.4. Conclusion

Dans le chapitre précédent, le couplage CE-FTICR-MS a été mis en place à partir de nucléosides standards. L'objectif de ce chapitre était d'évaluer, sur des échantillons biologiques connus, les performances du nouveau couplage pour la caractérisation des modifications post-transcriptionnelles.

Le couplage CE-FTICR-MS a d'abord été évalué sur quatre ARNt issus de *S. cerevisiae* dont les modifications, en grande majorité des méthylations, sont déjà décrites dans la littérature. Puis il a été évalué sur un extrait total d'ARNt issu de *B. taurus*, possédant plus de

modifications variées, elles aussi préalablement décrites dans la littérature. Les résultats issus des analyses avec les couplages CE-QTOF-MS et CE-FTICR-MS ont démontré que le couplage CE-FTICR-MS peut être utilisé pour identifier les modifications post-transcriptionnelles. Cependant, il faut être conscient que les compromis, qui ont été nécessaires lors de l'optimisation de la méthode, peuvent poser des limites pour les analyses d'échantillons biologiques. Ces limites ont été constatées sur le nombre d'ions précurseurs différents sélectionnés pour la fragmentation, notamment lors de la co-migration de plusieurs analytes de faible intensité. Hormis pour ψ et ses dérivés (tel que ψm), des résultats similaires ont été obtenus sur les deux couplages en termes de nombre de modifications identifiées. De plus, il est important de rappeler que le couplage CE-FTICR-MS offre une meilleure précision de masse.

Ainsi, les résultats présentés, sur des échantillons biologiques connus, démontrent que l'utilisation du couplage CE-FTICR-MS est possible pour la caractérisation des modifications post-transcriptionnelles au niveau nucléosides. En outre, il s'agit ici des premières analyses de nucléosides réalisées avec un couplage CE-FTICR-MS/MS pour la caractérisation des modifications post-transcriptionnelles. À partir de ces résultats, il est donc envisageable d'utiliser ce couplage sur des ARN dont les modifications ne sont pas décrites dans la littérature.

Après que la méthode développée par Lechner *et al.* [3] ait été transférée avec succès, pour des échantillons biologiques, sur le couplage CE-FTICR-MS, de nouveaux développements autour de cette méthode ainsi que l'analyse d'échantillons encore plus complexes, ont pu être envisagés et sont détaillés dans les prochains chapitres.

5.5. Références

- [1] Boccaletto, P. *et al.*, “MODOMICS: A Database of RNA Modification Pathways. 2021 Update.”, *Nucleic Acids Res.*, **2022**, 50 (D1), D231–D235.
- [2] Motorin, Y. *et al.*, “RNA Nucleotide Methylation.”, *WIREs RNA*, **2011**, 2 (5), 611–631.
- [3] Lechner, A. *et al.*, “Characterization of Post-Transcriptional RNA Modifications by Sheathless Capillary Electrophoresis–High Resolution Mass Spectrometry.”, *Anal. Chem.*, **2020**, 92 (10), 7363–7370.
- [4] Charette, M. *et al.*, “Pseudouridine in RNA: What, Where, How, and Why.”, *IUBMB Life*, **2000**, 49 (5), 341–351.
- [5] Lin, T.-Y. *et al.*, “Pseudouridines in RNAs: Switching Atoms Means Shifting Paradigms.”, *FEBS Lett.*, **2021**, 595 (18), 2310–2322.

PARTIE III : DEVELOPPEMENT METHODOLOGIQUE POUR L'ETUDE DES MODIFICATIONS POST-TRANSCRIPTIONNELLES DES ARN

Dans la partie précédente, le couplage CE-FTICR-MS a été mis en place sur des standards et des échantillons biologiques bien référencés dans la littérature. Afin de pousser plus loin l'étude des modifications post-transcriptionnelles des ARN par CE-MS/MS, une partie de ces travaux de thèse est basée sur de nouveaux développements méthodologiques. Cette troisième partie, divisée en trois chapitres, est consacrée aux développements méthodologiques qui ont été réalisés de la préparation des échantillons à l'interprétation des données. Ces développements ont été réalisés sur le couplage CE-QTOF-MS dans l'objectif de les transposer sur le couplage CE-FTICR-MS dans la quatrième et dernière partie de ces travaux.

Le premier chapitre présente l'optimisation de la déphosphorylation des nucléotides pour la caractérisation structurale des ARN. Plusieurs protocoles sont proposés dans la littérature pour cette étape de déphosphorylation. Cependant, aucun ne ressort comme une méthode de référence garantissant une déphosphorylation totale. L'objectif de cette étude était donc de proposer à la communauté scientifique un protocole fixé et vérifié, garantissant la déphosphorylation totale des échantillons.

Le second chapitre porte sur l'analyse d'un échantillon plus complexe, dont les modifications post-transcriptionnelles ne sont pas encore référencées dans la littérature. Il s'agit d'un ARN ribosomique 16S pur issu de *P. aeruginosa*. Les analyses par CE-QTOF-MS ont été réalisées aux niveaux nucléosides et oligonucléotides. L'utilisation du couplage CE-QTOF-MS, essentielle avant de pouvoir envisager le transfert de cette application au couplage CE-FTICR-MS, permet aussi de tester la méthode CE-MS/MS à deux niveaux du laboratoire, sur un cas réel et plus complexe.

Enfin, pour faciliter l'interprétation des deux niveaux d'analyses (nucléosides et oligonucléotides), le dernier chapitre détaille le développement de deux outils bio-informatiques permettant de caractériser les modifications post-transcriptionnelles des ARN. Le développement de ces deux logiciels a découlé directement des besoins identifiés lors de l'analyse de l'ARN ribosomique.

CHAPITRE 6 : OPTIMISATION DE LA DEPHOSPHORYLATION DES NUCLEOTIDES POUR LA CARACTERISATION STRUCTURALE DES ARN

6.1. Introduction

Comme cela a été expliqué dans le chapitre 3 : (paragraphe 3.4.4), l'identification des modifications post-transcriptionnelles des ARN s'effectue à l'aide d'analyses par CE-MS/MS, ou LC-MS/MS, au niveau nucléosides. Pour obtenir les nucléosides à partir des ARN entiers, la préparation de l'échantillon s'effectue en deux étapes. Tout d'abord, l'ARN est hydrolysé totalement afin d'obtenir des nucléotides. Cette digestion peut être réalisée avec plusieurs enzymes, et dans notre cas, il s'agit de la Nucléase P1. Les nucléotides sont ensuite déphosphorylés à l'aide d'une phosphatase alcaline bactérienne (BAP).

Plusieurs protocoles sont décrits dans la littérature [1–3] et utilisés dans de nombreuses études [4–6], mais aucun ne semble être considéré comme le protocole de référence. De plus, des études d'efficacité n'ont pas toujours été réalisées pour chaque nouveau protocole, laissant la possibilité d'une déphosphorylation partielle. Cependant, une digestion et une déphosphorylation efficaces sont essentielles pour éviter des complications supplémentaires dans l'analyse et l'interprétation des données.

Dans ce sixième chapitre, nous proposons donc d'optimiser l'étape de déphosphorylation sur un extrait d'ARNt modèle pour fixer deux protocoles reposant sur deux impératifs classiques : un protocole économique avec une consommation réduite de BAP ou un plus rapide avec un temps de déphosphorylation court.

L'optimisation de l'étape de déphosphorylation a fait l'objet d'une publication, intitulée « Optimization of nucleotides dephosphorylation for RNA structural characterization by tandem mass spectrometry hyphenated with separation methods » et parue dans *Journal of Chromatography B* (doi:10.1016/j.jchromb.2022.123396) [7].

6.2. Publication : Optimization of nucleotides dephosphorylation for RNA structural characterization by tandem mass spectrometry hyphenated with separation methods

Clarisso Gosset-Erard, Antony Lechner, Philippe Wolff, Frédéric Aubriet, Emmanuelle Leize-Wagner, Patrick Chaimbault, and Yannis-Nicolas François

[DOI:10.1016/j.jchromb.2022.123396](https://doi.org/10.1016/j.jchromb.2022.123396)

6.2.1. Abstract

As part of RNA characterization, the identification of post-transcriptional modifications can be performed using hyphenation of separation methods with mass spectrometry. To identify RNA modifications with those methods, a first total digestion followed by a dephosphorylation step are usually required to reduce RNA to nucleosides. Even though effective digestion and dephosphorylation are essential to avoid further complications in analysis and data interpretation, to our knowledge, no fixed protocol is yet referenced in the literature. Therefore, the aim of this work is to optimize the dephosphorylation step using a total extract of transfer RNA (tRNA) from *B. taurus* as a model and to determine and fix two protocols, leading to complete dephosphorylation, based on time and bacterial alkaline phosphatase (BAP) consumptions. Capillary electrophoresis-tandem mass spectrometry (CE-MS/MS) was used to estimate the dephosphorylation efficiency of both protocols on many canonical and modified nucleotides. For a timesaving protocol, we established that full dephosphorylation was obtained after a 4-hour incubation at 37 °C with 7.5 U of BAP for 1 µg of tRNA. And for a BAP-saving protocol, we established that full dephosphorylation was obtained 3.0 U of BAP after an overnight incubation at 37 °C. Both protocols are suitable for quantitative analyses as no loss of analytes is expected. Moreover, they can be widely used for all other RNA classes, including messenger RNA or ribosomal RNA.

6.2.2. Introduction

Over the past few years, RNA studies have shown a growing interest due to the huge potential of RNA in therapeutic strategies such as diagnostic [8], treatment [9] or vaccination [10]. Structural information, including the identification and location of post-transcriptional modifications, is required to characterize RNA. Nowadays, more than 150 modified nucleosides are described in the literature and referenced in databases like Modomics [11]. These modifications are found in all types of RNA and are of great interest due to their involvement in the RNA structural stability [12] and many biological functions [13] but also in several diseases [14] as diabetes [15] or cancers [16].

Many techniques have been developed to analyze RNA post-transcriptional modifications, including separative methods such as high-performance liquid chromatography (LC) [2,3,5,15,17–22] or capillary electrophoresis (CE) [4] hyphenated to tandem mass spectrometry (MS/MS). Indeed, on one hand, these separative methods, which have complementary selectivity, enable the characterization of several modified nucleosides including isomers [4]. On the other hand, MS has become a powerful tool for the identification and structural characterization of biomolecules due to its sensitivity and specificity. Using those couplings, the identification of RNA modifications is commonly performed at the nucleosides level, while their localization is determined at the oligonucleotides level. Both, CE and LC, could have been used in this study. However, CE offers the possibility to study RNA at

both levels, nucleosides and oligonucleotides, with the same method [4], contrary to LC-MS, while maintaining the advantages of isomeric separation.

In order to digest RNA into nucleosides, the protocols usually described in the literature are based on two steps. The first one is a total digestion with nuclease P1 [4,17] or with a combination of nuclease P1 and phosphodiesterase I [1,5], leading to nucleotides. The second step is a dephosphorylation with bacterial alkaline phosphatase (BAP) [1,4–6,17]. Effective digestion and dephosphorylation are essential to avoid further complications in analysis and data interpretation. In-solution, nuclease P1 seems to be quite efficient with a reported complete digestion of DNA in less than 2 hours [23]. In 1990, Crain *et al* proposed one protocol for both steps leading to full digestion and dephosphorylation [1]. In 2014, Su *et al*, proposed an entire protocol, from the sample preparation to the data interpretation, for the quantitative analysis of RNA modifications by LC-MS. The first digestion step was performed with benzonase and phosphodiesterase I, and the dephosphorylation step with alkaline phosphatase [2]. In 2016, Thüring *et al* published a similar protocol using nuclease P1 instead of benzonase [3]. However, none of them are describing the optimization process of the dephosphorylation step to obtain fully dephosphorylated nucleosides. Moreover, since then, several studies used various protocols [4–6]. A kit is also sold by New England BioLabs enabling one-step digestion and dephosphorylation of RNA and DNA, however no information on its enzyme composition is available. Therefore, we consider that there is no fixed protocol for the dephosphorylation of nucleotides. Moreover, to the best of our knowledge, analyte recovery rates were never discussed as no step may implies a loss of analytes, and performance studies have not always been carried out for each new protocol, leaving the possibility of partial dephosphorylation. Indeed, most of the studies are using Multiple Reaction Monitoring (MRM) detection [2,3,5,6,15,18–20,22,24] as it is really useful to increase sensitivity. However, it is not appropriate for untargeted studies in which nucleotides are remaining in the sample due to poor dephosphorylation or when nucleoside modifications are still unknown.

In this work, we propose to optimize the dephosphorylation step using BAP to fix two different protocols based on two classical imperatives: (i) economical with reduced BAP consumption or (ii) fast with a short dephosphorylation time.

6.2.3. Materials and methods

6.2.3.1. Chemicals

The chemicals used were of analytical grade or high purity. Water was obtained using a Direct-Q3 UV-R water purification system (Merck Millipore, Amsterdam, The Netherlands). Nuclease P1 from *Penicillium citrinum*, ammonium acetate and zinc chloride ($ZnCl_2$) were purchased from Sigma-Aldrich (Saint Louis, MO, U.S.A.). Glacial acetic acid and *Bacterial Alkaline Phosphatase* (BAP) (150 U/ μ L) from *Escherichia coli* C90 was purchased from ThermoFisher Scientific (Marietta, OH, U.S.A.).

6.2.3.2. tRNA sample

A total extract of transfer RNA (tRNA) from eukaryotic cytosol of *B. taurus* was used to perform this study. The modifications from *B. taurus* tRNA are numerous and well known, making it a good model to track dephosphorylation performances. All modifications reported in Modomics database [11] for *B. taurus* tRNA total extract are listed in Table 6.1.

Table 6.1: List of modifications reported in Modomics database [11] for *B. taurus* tRNA total extract.

Nucleoside full name	Nucleosides abbreviations	Nucleotides abbreviations
Adenosine	A	pA
Cytidine	C	pC
Guanosine	G	pG
Uridine	U	pU
1-methyladenosine	m ¹ A	pm ¹ A
3-methylcytidine	m ³ C	pm ³ C
5-methylcytidine	m ⁵ C	pm ⁵ C
2'-O-methylcytidine	Cm	pCm
1-methylguanosine	m ¹ G	pm ¹ G
N2-methylguanosine	m ² G	pm ² G
7-methylguanosine	m ⁷ G	pm ⁷ G
2'-O-methylguanosine	Gm	pGm
N2,N2,2'-O-trimethylguanosine	m ² ₂ G	pm ² ₂ G
N6-isopentenyladenosine	i ⁶ A	pi ⁶ A
3-(3-amino-3-carboxypropyl)uridine	acp ³ U	pacp ³ U
Queuosine	Q	pQ
Galactosyl-queuosine	galQ	pgalQ
Mannosyl-queuosine	manQ	pmanQ
Pseudouridine	Y	pY
Dihydouridine	D	pD
5-methyluridine	m ⁵ U	pm ⁵ U
2'-O-methyluridine	Um	pUm
2'-O-methylpseudouridine	Ym	pYm
5,2'-O-dimethyluridine	m ⁵ Um	pm ⁵ Um
N4-acetylcystidine	ac ⁴ C	pac ⁴ C
5-methoxycarbonylmethyl-2-thiouridine	mcm ⁵ s ² U	pmcm ⁵ s ² U
Inosine	I	pl
N6-threonylcarbamoyladenosine	t ⁶ A	pt ⁶ A
Methoxycarbonylmethyluridine	mcm ⁵ U	pmcm ⁵ U
Peroxywybutosine	o ₂ yW	po ₂ yW
5-formyl-2'-O-methylcytidine	f ⁵ Cm	pf ⁵ Cm

6.2.3.3. Sample preparation

B. taurus tRNA was desalted by three successive precipitations with ammonium acetate and ethanol. A single batch of total tRNA extract was digested with nuclease P1 following a previously described protocol [4]. Briefly, P1 10x buffer (2 mM ZnCl₂, 250 mM ammonium acetate, pH 5.0) and tRNA extract were successively added into deionized water before adding 0.5 U of P1 (previously prepared at 0.5 U/μL in 100 mM ammonium acetate) per μg of tRNA. This single batch of tRNA was incubated at 37 °C for 2 hours, and then used for different dephosphorylation conditions. Nucleotides were dephosphorylated using seven different ratios from 0.15 to 7.5 U of BAP (previously prepared at 1.5 U/μL in 100 mM ammonium acetate) for 1 μg of tRNA and incubated at 37 °C for two periods of time: 4 hours and overnight. The final concentration of tRNA in each sample was 0.1 μg/μL. For each BAP quantity and incubation time combinations, samples were prepared in triplicate and directly analyzed. Analyses of sample triplicates were carried out on different days.

6.2.3.4. CE-MS/MS analysis

CE-MS/MS analyses were performed following a previously described method [4]. Briefly, CE was performed on a bare-fused silica capillary (100 cm total length, 30 μm i.d.) filled with 10% acetic acid as background electrolyte. Analyses were performed with 5 nL injections and +30kV separation voltage and 2 psi pressure. A porous tip sheathless interface was used for the CE-MS hyphenation and MS detection was performed in positive mode. MS/MS detection was performed using data-dependent mode (theoretical precursor and product ions used for the identification are available in Supplemental Information in Table S6.1). To ensure the full dephosphorylation of a maximum of various nucleotides, dephosphorylation efficiencies of as many modifications from *B. taurus* tRNA as possible were monitored. However, two modifications reported in Table 6.1 (o₂yW and f⁵Cm) were withdrawn for the dephosphorylation study due to the lack of signal of corresponding nucleotides or nucleosides in control samples without any dephosphorylation step or with full dephosphorylation. Therefore, the dephosphorylation of 4 canonical and only 25 modified nucleotides was monitored (Table 6.1, except o₂yW and f⁵Cm).

6.2.3.5. Verification of P1 digestion step

As some miscleavages of modified nucleosides were previously reported when using only nuclease P1 [1], it was essential to check the performance of the P1 digestion step in our protocols (as P1 is not retrieved before the second incubation, total P1 incubation times are 2 h + 4 h and 2 h + overnight) before making any conclusion. Indeed, it is worth noting that any difference in the digestion efficiency could have an impact on the results obtained for the dephosphorylation step. To ensure that P1 digestion was complete, the ion currents of dinucleotides combining each modification with all four canonical nucleotides were extracted with a mass tolerance of ± 0.02 Da. Since none of those dinucleotides were detected, we considered that the digestion with nuclease P1 was complete for all studied nucleosides and

nucleotides in all samples. Indeed, only dinucleotides have been reported in the literature after incomplete digestion with P1 [1]. Consequently, the following study can be focused only on the dephosphorylation step. The ion current of all nucleosides and their corresponding nucleotide species were also extracted with a mass tolerance of ± 0.02 Da.

6.2.3.6. Estimation of the dephosphorylation efficiency

The dephosphorylation efficiency was then estimated for each canonical and modified nucleotides based on the following equation:

$$\%_{\text{dephosphorylation}} \approx \frac{\text{Area}_{\text{nucleoside}}}{\text{Area}_{\text{nucleoside}} + \text{Area}_{\text{nucleotide}}} \times 100 \quad (6-6.1)$$

$\text{Area}_{\text{nucleoside}}$ and $\text{Area}_{\text{nucleotide}}$ are peak areas obtained from extracted ions electropherograms. As nucleosides and nucleotides may have various ionization efficiencies, it is important to note that this dephosphorylation efficiency is only an estimation. However, confidence in full dephosphorylation is obtained when its efficiency is estimated to 100%, which is the aim of this study.

For post-transcriptional modifications leading to some isomers partially or not separated by CE, the sum of their areas was used and a common dephosphorylation efficiency was calculated. This is the case of m^3C , m^5C and Cm, as well as galQ and manQ, m^1G and m^2G , or U and Y (Table S6.2 in supplemental information).

6.2.4. Results and discussion

An example of incomplete dephosphorylation CE-MS/MS analysis is presented in Figure 6.1, and the influence of the dephosphorylation efficiency on the analysis is illustrated by the electropherograms in Figure 6.2. Full overall dephosphorylation was obtained for 7.5 U of BAP for 1 μg of RNA incubated at 37 °C for 4 hours (Figure 6.2.B) and for 3.0 U of BAP for 1 μg of RNA incubated overnight at 37 °C (Figure 6.2.C). Among these protocols enabling full dephosphorylation of nucleotides, the first one with a 4-hour incubation fits with the protocol developed in our laboratory and described by Lechner *et al* [4], and the overnight protocol is a new combination of incubation time and amount of BAP compared to what is described in the literature.

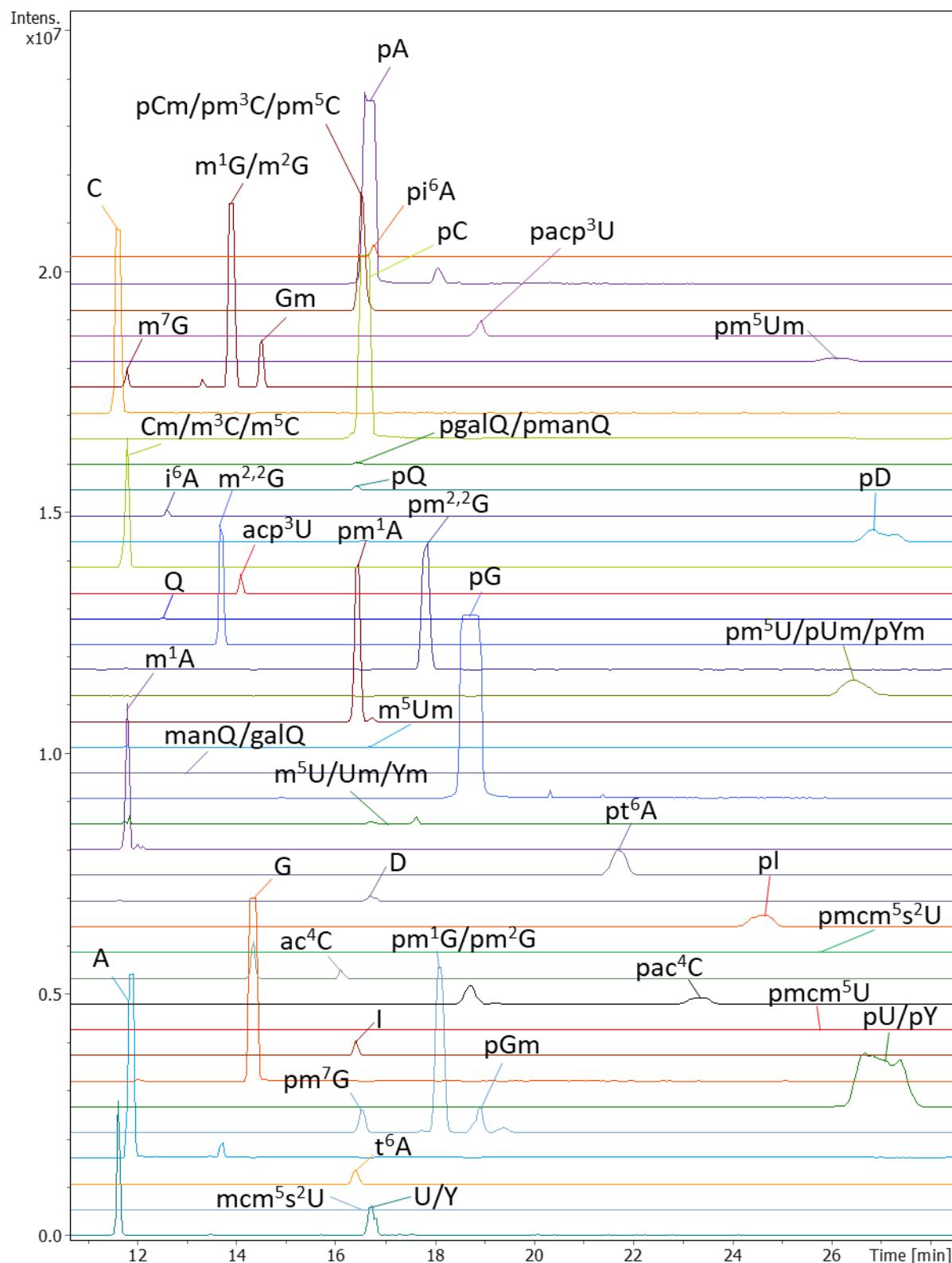


Figure 6.1: Multiple extracted ion electropherograms for the analysis of total tRNA extract from *B. taurus* after digestion with nuclease P1 and dephosphorylation after a 4-hour incubation at 37°C with 0.75 U of BAP per µg of RNA. Migration times of all nucleosides and nucleotides are presented in Table S6.3 in Supplemental Information.

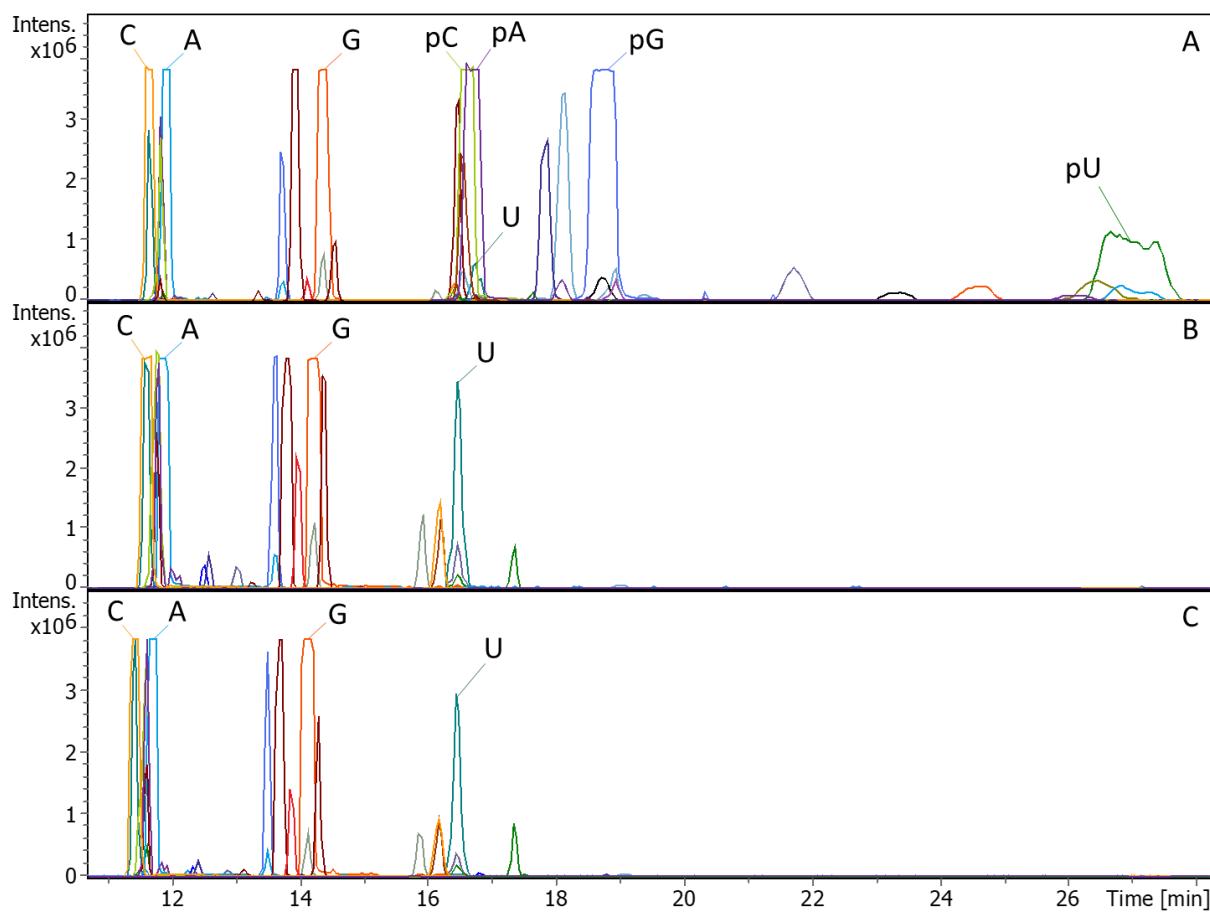


Figure 6.2: Multiple extracted ion electropherograms for the analysis of total tRNA extract from *B. taurus* after digestion with nuclease P1 and dephosphorylation with (A) 0.75 U of BAP for 1 µg of RNA incubated for 4 hours at 37 °C, (B) 7.5 U of BAP for 1 µg of RNA incubated for 4 hours at 37 °C and (C) 3.0 U of BAP for 1 µg of RNA incubated for 4 hours at 37 °C. All 4 canonical and 25 modified nucleosides and nucleotides were extracted. Full annotation of nucleoside and nucleotides of Figure 6.2.A is available in Figure 6.1.

The full dephosphorylation enabled the reduction of CE analysis time (Figure 6.2.B and Figure 6.2.C) due to late migration of many nucleotides negatively charged. For instance, the migration time of pU was around 27 min (Figure 6.2.A), thus increasing the analysis time that is around 17 min obtained with full dephosphorylation (Figure 6.2.B and Figure 6.2.C). Furthermore, it prevents the potential ion suppression caused by the comigration of some nucleotides with nucleosides. To illustrate this point, all nucleosides and nucleotides have been annotated in Figure 6.1, and canonical nucleosides and nucleotides have been annotated in Figure 6.2, particularly showing the peak overlapping of U, D, $m^5U/Um/Ym$, m^5Um , $pm^3C/pm^5C/pCm$, pA pC, and other species around 17 min (Figure 6.2.A). As U and other modified nucleosides are neutral, ionization competition with co-migrating nucleotides can have a huge impact on their detection. Indeed, ion suppression phenomenon can lead to a loss of sensitivity, particularly for low concentrated compounds, which is often the case of modified nucleosides. In turn, low peak intensities can lead to a failure of the automatic MS/MS fragmentation process in data-dependent mode resulting in a loss of characterization for some nucleosides. MRM is a way to face this limitation in MS/MS fragmentation, but it does not prevent potential ion suppression during ionization. Finally, with uncompleted dephosphorylation, the occurrence of each modification is split between both nucleoside and

nucleotide species (Figure 6.1 and Figure 6.2.A). Full dephosphorylation is therefore essential for quantification studies avoiding the underestimation or overestimation of RNA modifications. Our fixed protocols should be well suited for relative quantitation analysis as well as absolute quantitation studies with the use of internal standards. Indeed, no step that could cause analyte loss, such as post-digestion desalting, is required.

Figure 6.3 represents the dephosphorylation efficiency for 29 nucleosides for 4-hour and overnight incubations carried out with seven increasing quantities of BAP (0.15 to 7.5 U) for 1 µg of tRNA. The results for each condition correspond to average dephosphorylation efficiencies obtained from triplicates. Full dephosphorylation was considered for efficiencies equal or higher than 98 %. All standard errors were below ± 19% for uncompleted dephosphorylations and below ± 3% for full dephosphorylations (Table S6.4 and Table S6.5 in supplemental information). As analyses of sample triplicates were carried out on different days, this shows that the estimated day-to-day variability is relatively low.

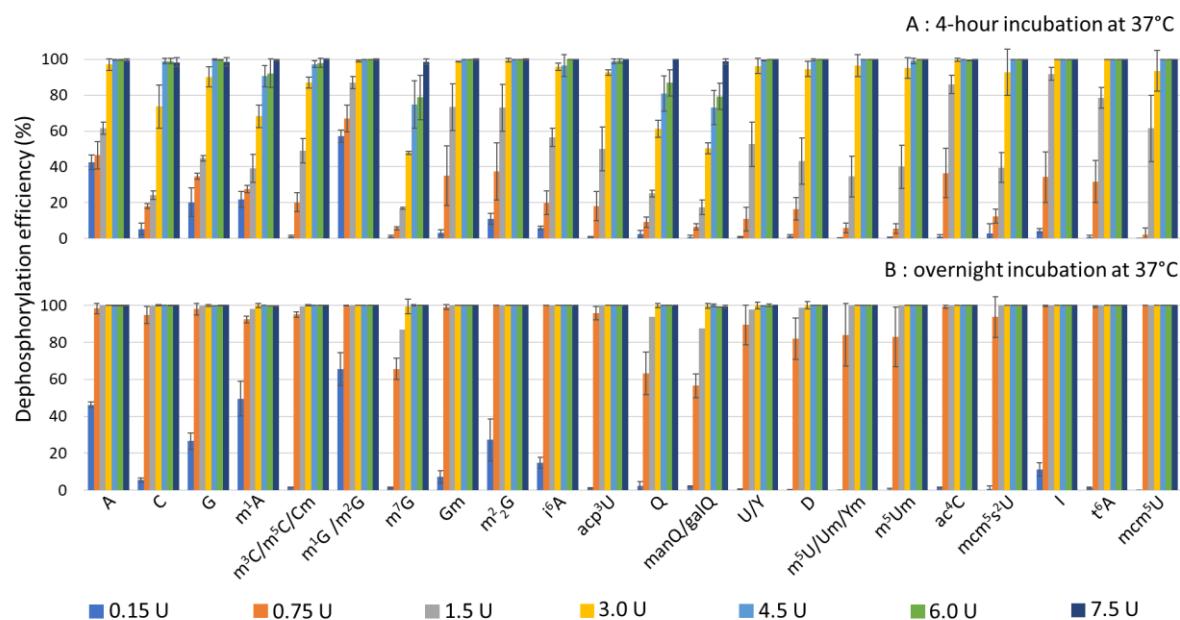


Figure 6.3: Estimated dephosphorylation efficiencies for 29 nucleotides after (A) 4-hour incubation and (B) overnight incubation at 37°C with seven quantities from 0.15 to 7.5 U of BAP for 1 µg of RNA. Standard errors were obtained from sample triplicates. More precise data are available in Suplemental Information (Table S6.4 and Table S6.5)

For a given incubation time, the minimum quantity of BAP needed to obtain full dephosphorylation is not the same for all modifications. For instance, many modifications including m¹G, m²G, Gm, acp³U and ac⁴C are fully dephosphorylated with 4.5 U of BAP after a 4-hour incubation at 37°C, whereas dephosphorylation efficiencies of m⁷G, Q, manQ and galQ do not exceed 81 % for the same BAP quantity and incubation time (Figure 6.3.A).

Even if small differences in the dephosphorylation efficiencies could have been expected between large and distinctive modifications, the huge gap in the dephosphorylation efficiencies is more surprising between small-sized isomeric modifications with close structures. For instance, no conclusion can be made regarding a potential difference in

dephosphorylation efficiencies between m^1G and m^2G as they are not separated by the CE. However, the sum of m^1G and m^2G (annotated m^1G/m^2G) is already more than half dephosphorylated with the softest condition (0.15 U of BAP for 1 μ g of tRNA incubated for 4 hours at 37 °C) while m^7G and Gm are barely dephosphorylated (Figure 6.3.A). The proportion shift of pm^7G , pm^1G/pm^2G and pGm between 0.15 U and 1.5 U of BAP incubated for 4 hours at 37 °C, as well as their chemical structures, are illustrated in Figure 6.4. Even if pm^1G/pm^2G were after a 4-hour incubation at 37°C with 0.15 U of BAP per μ g of RNA (Figure 6.4.A), pm^7G was the only isomer not fully dephosphorylated with 4.5 U of BAP per μ g of RNA (Figure 6.4.E). This proportion shift indicates that m^7G has the lowest dephosphorylation rate of the four methylated guanosine isomers.

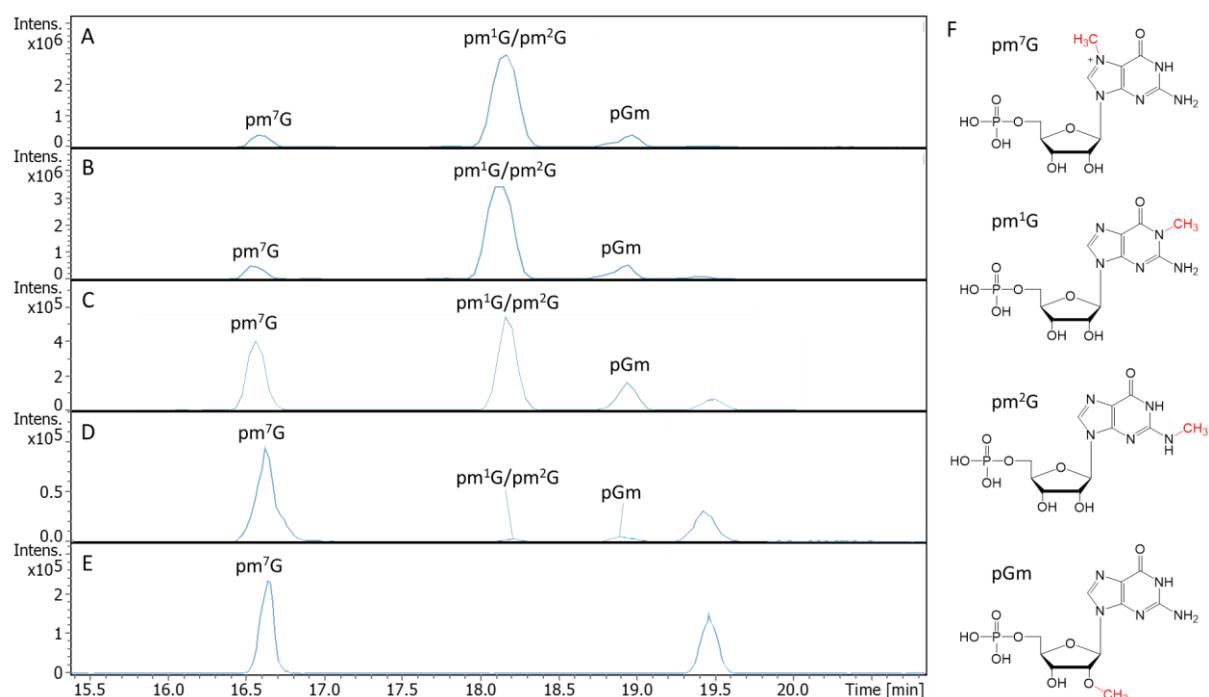


Figure 6.4: Extracted ion electropherogram (m/z 378.08) for the analysis total tRNA extract from *B. taurus* after digestion with nuclease P1 and dephosphorylation after a 4-hour incubation at 37°C with (A) 0.15 U of BAP per μ g of RNA, (B) 0.75 U of BAP per μ g of RNA, (C) 1.5 U of BAP per μ g of RNA, (D) 3.0 U of BAP per μ g of RNA and (E) 4.5 U of BAP per μ g of RNA. (F) Structures of all four isomers of methylated monophosphate guanosine.

After digestion with nuclease P1, nucleotides with 3'-OH and 5'-NMP ends are obtained [25]. Therefore, the bad dephosphorylation of m^7G compared to m^1G/m^2G and Gm is quite surprising as the residual phosphate group is linked to the ribose and not to the modified base. Due to the positions of methylation in Gm and m^7G , based on the reaction mechanism of BAP described in Yang and coworker research [26], we rather expected Gm than the other isomers to be the most difficult to dephosphorylate or at least to have similar dephosphorylation efficiencies because of the light steric hindrance of the methylations. Besides these inhomogeneous dephosphorylation rates, it is also worth noting that the first nucleoside at the 5'-end of each tRNA sequence does not possess a phosphate group and moreover is usually not modified. Thus, there is already a small amount of unmodified nucleosides in the sample even without any dephosphorylation step. This could correlate to

the higher dephosphorylation efficiency of G compared to Gm and m⁷G displayed for the softest condition (0.15 U of BAP for 1 µg of tRNA incubated for 4 hours at 37 °C) in Figure 6.3.A.

However, regarding each modification, dephosphorylation efficiencies higher than 98 % were obtained for 7.5 U of BAP for 1 µg of RNA incubated at 37 °C for 4 hours and for 3.0 U of BAP for 1 µg of RNA incubated overnight at 37 °C.

6.2.5. Conclusion

The most remarkable feature of this study is that two equivalent fixed protocols can be established for the nucleotides dephosphorylation in routine for CE-MS/MS but also LC-MS/MS analysis purposes. According to individual needs:

- (i) the first one is quicker but the amount of enzyme matters (7.5 U of BAP for 1 µg of RNA incubated at 37 °C for 4 hours),
- (ii) the second one is longer but more economical in enzyme quantity (3.0 U of BAP for 1 µg of RNA incubated overnight at 37 °C).

6.2.6. Supplemental information

Table S6.1: List of modifications reported in Modomics database [11] for *B. taurus* tRNA total extract and theoretical m/z used for their MS and MS/MS identification.

	Nucleoside		Nucleotide	
	m/z of [M+H] ⁺ precursor ions	m/z of [M+H] ⁺ product ions	m/z of [M+H] ⁺ precursor ions	m/z of [M+H] ⁺ product ions
A	268.10	136	348.08	136
C	244.09	112	324.07	112
G	284.12	152	364.08	152
U	245.08	113	325.06	113
m ¹ A	282.12	150	362.10	150
m ³ C	258.11	126	338.09	126
m ⁵ C	258.11	126	338.09	126
Cm	258.11	112	338.09	112
m ¹ G	298.11	166	378.09	166
m ² G	298.11	166	378.09	166
m ⁷ G	298.11	166	378.09	166
Gm	298.11	152	378.09	152
m ² ₂ G	312.13	180	392.11	180
i ⁶ A	336.17	204 / 136 / 148	416.15	204 / 136 / 148
acp ³ U	346.13	214 / 197 / 168	426.11	214 / 197 / 168
Q	410.17	295 / 163	490.15	375 / 163
galQ	572.22	132 / 163 / 295 / 495	652.20	132 / 163 / 395
manQ	572.22	132 / 163 / 295 / 495	652.20	132 / 163 / 395
Y	245.08	209	325.06	209
D	247.09	115	327.07	115
m ⁵ U	259.09	127	339.07	127
Um	259.09	113	339.07	113
Ym	259.09	209	339.07	209
m ⁵ Um	273.11	127	353.09	127
ac ⁴ C	286.10	154 / 112	366.08	154 / 112
mcm ⁵ s ² U	333.08	201 / 169 / 141	413.06	201 / 169 / 141
I	269.09	137	349.07	137
t ⁶ A	413.14	281	493.12	281
mcm ⁵ U	317.10	185 / 153 / 125	397.08	185 / 153 / 125
o ₂ yW	541.19	409	621.17	409
f ⁵ Cm	286.10	140	366.08	140

Table S6.2: CE separation status of isomeric nucleosides and nucleotides.

Isomeric nucleosides and nucleotides	CE separation
$m^3C/m^5C/Cm$	partially separated
m^1G/m^2G	not separated
m^7G	fully separated
Gm	fully separated
manQ/galQ	not separated
U/Y	not separated
$m^5U/Um/Ym$	not separated
$pm^3C/pm^5C/pCm$	partially separated
pm^1G/pm^2G	not separated
pm^7G	fully separated
pGm	fully separated
pmanQ/pgalQ	not separated
pU/pY	not separated
$pm^5U/pUm/pYm$	not separated

Table S6.3: Migration times of nucleosides and nucleotides annotated in Figure 6.1.

Nucleosides	Migration time (min)	Nucleotides	Migration time (min)
A	11.9	pA	16.7
C	11.7	pC	16.6
G	14.4	pG	18.7
m ¹ A	11.8	m ¹ A	16.5
m ³ C/m ⁵ C/Cm	11.8	pm ³ C/pm ⁵ C/pCm	16.5
m ¹ G/m ² G	13.9	pm ¹ G/pm ² G	18.1
m ⁷ G	11.8	pm ⁷ G	16.6
Gm	14.6	pGm	19.0
m ² G	13.7	pm ² G	17.9
i ⁶ A	12.6	pi ⁶ A	16.8
acp ³ U	14.1	pacp ³ U	19
Q	12.6	pQ	16.4
manQ/galQ	13.1	pmanQ/pgalQ	16.5
U/Y	16.8	pU/pY	26.7
D	16.7	pD	26.9
Um/Ym	16.8	pm ⁵ U/pUm/pYm	26.5
m ⁵ Um	16.7	pm ⁵ Um	26.0
ac ⁴ C	16.1	pac ⁴ C	23.4
mcm ⁵ s ² U	16.6	pmcm ⁵ s ² U	25.7
I	16.4	pI	24.6
t ⁶ A	16.4	pt ⁶ A	21.7
mcm ⁵ U	not detected	pmcm ⁵ U	25.7

Chapitre 6 : Optimisation de la déphosphorylation des nucléotides

Table S6.4: Estimated dephosphorylation efficiencies for 29 nucleotides after 4-hour incubation at 37°C with seven quantities from 0.15 to 7.5 U of BAP per µg of RNA. Averages and standard errors were obtained from sample triplicates.

BAP quantity (U/µg of tRNA)	Averaged dephosphorylation efficiencies (%)							Standard errors (%)						
	0.15	0.75	1.5	3.0	4.5	6.0	7.5	0.15	0.75	1.5	3.0	4.5	6.0	7.5
A	43	47	62	97	100	100	100	4	8	3	3	0	0	1
C	5	18	24	74	99	99	98	3	1	2	12	2	2	3
G	20	35	45	90	100	100	99	8	2	2	6	0	0	2
m ¹ A	22	28	39	68	91	92	100	5	2	8	6	6	8	0
m ³ C/m ⁵ C/Cm	1	20	49	87	97	98	100	1	5	7	3	2	2	0
m ¹ G/m ² G	57	67	87	99	100	100	100	4	8	3	1	0	0	0
m ⁷ G	1	6	17	48	75	79	99	1	1	1	1	13	12	2
Gm	3	35	73	99	100	100	100	2	17	13	0	0	0	0
m ² ₂ G	11	37	73	99	100	100	100	3	16	13	1	0	0	0
i ⁶ A	6	20	56	96	97	100	100	1	7	5	2	6	0	0
acp ³ U	1	18	50	93	99	99	100	0	8	12	1	1	1	0
Q	2	9	25	61	81	87	100	2	3	2	5	10	7	0
manQ/galQ	1	7	18	50	73	79	99	1	2	4	3	10	7	2
U/Y	1	11	53	96	99	100	100	0	7	12	4	0	0	0
D	1	17	43	95	100	100	100	1	6	13	4	1	0	0
m ⁵ U/Um/Ym	0	6	35	97	100	100	100	0	3	11	6	0	0	0
m ⁵ Um	0	6	40	95	99	100	100	1	3	12	6	2	0	0
ac ⁴ C	1	37	86	100	100	100	100	1	14	5	1	0	0	0
mcm ⁵ s ² U	3	12	40	93	100	100	100	5	4	8	13	0	0	0
I	4	34	92	100	100	100	100	1	14	4	0	0	0	0
t ⁶ A	1	32	79	100	100	100	100	1	12	6	0	0	0	0
mcm ⁵ U	0	2	61	94	100	100	100	0	4	19	11	0	0	0

Chapitre 6 : Optimisation de la déphosphorylation des nucléotides

Table S6.5: Estimated dephosphorylation efficiencies for 29 nucleotides after overnight incubation at 37°C with seven quantities from 0.15 to 7.5 U of BAP per µg of RNA. Averages and standard errors were obtained from sample triplicates.

BAP quantity (U/µg of RNA)	Averaged dephosphorylation efficiencies (%)							Standard errors (%)						
	0.15	0.75	1.5	3.0	4.5	6.0	7.5	0.15	0.75	1.5	3.0	4.5	6.0	7.5
A	46	98	100	100	100	100	100	1	3	0	0	0	0	0
C	6	95	100	100	100	100	100	1	5	0	0	0	0	0
G	27	98	100	100	100	100	100	5	3	1	0	0	0	0
m ¹ A	50	92	98	100	100	100	100	9	2	1	0	0	0	0
m ³ C/m ⁵ C/Cm	2	95	99	100	100	100	100	0	2	0	0	0	0	0
m ¹ G/m ² G	66	100	100	100	100	100	100	9	0	0	0	0	0	0
m ⁷ G	1	66	87	100	100	100	100	0	6	4	0	0	0	0
Gm	7	99	100	100	100	100	100	3	1	0	0	0	0	0
m ² ₂ G	27	100	100	100	100	100	100	11	0	0	0	0	0	0
i ⁶ A	15	100	100	100	100	100	100	3	0	0	0	0	0	0
acp ³ U	1	96	100	100	100	100	100	0	4	0	0	0	0	0
Q	3	63	94	100	100	100	100	2	12	1	0	0	0	0
manQ/galQ	2	57	88	100	100	99	99	0	6	1	1	0	1	1
U/Y	0	89	98	100	100	100	100	0	11	2	0	1	0	0
D	0	82	99	100	100	100	100	0	11	2	0	0	0	0
m ⁵ U/Um/Ym	0	84	100	100	100	100	100	0	17	0	0	0	0	0
m ⁵ Um	0	83	100	100	100	100	100	1	16	0	0	0	0	0
ac ⁴ C	2	99	100	100	100	100	100	0	1	0	0	0	0	0
mcm ⁵ s ² U	1	94	100	100	100	100	100	2	11	0	0	0	0	0
I	11	100	100	100	100	100	100	4	0	0	0	0	0	0
t ⁶ A	1	99	100	100	100	100	100	0	1	0	0	0	0	0
mcm ⁵ U	0	100	100	100	100	100	100	0	0	0	0	0	0	0

6.3. Conclusion

Dans ce chapitre, l'étape de déphosphorylation des nucléotides pour l'identification des modifications au niveau nucléosides a été optimisée. Pour cela, un extrait total d'ARNt issu de *B. taurus* a été digéré à la Nucléase P1 puis déphosphorylé avec différentes quantités de BAP lors d'incubations à 37 °C, soit pendant 4 h, soit pendant une nuit.

Afin de s'assurer que la déphosphorylation soit complète, l'aire de chacune des espèces phosphophorylées a été comparée à l'aire des espèces déphosphorylées correspondantes. Il est important de noter que l'échantillon a été considéré comme complètement déphosphorylé lorsque chacun des quatre nucléosides canoniques et chacune des 25 modifications référencées et détectées dans cet extrait étaient complètement déphosphorylés. En effet, comme toutes les modifications n'ont pas montré les mêmes vitesses de déphosphorylation, il est essentiel de s'assurer qu'elles sont bien toutes déphosphorylées avant de les analyser.

Selon les besoins de chacun, deux protocoles, permettant d'obtenir une déphosphorylation complète, ont été proposés. Ils sont 100% compatibles avec les analyses par CE-MS/MS, mais aussi celles par LC-MS/MS, puisqu'aucun sel non volatile n'est utilisé.

Le premier protocole est le plus rapide, mais il requiert une plus grande quantité d'enzyme : 7,5 U de BAP par µg d'ARN et une incubation à 37 °C pendant 4 h.

Le second protocole est plus long, mais il requiert une plus faible quantité d'enzyme : 3,0 U de BAP par µg d'ARN et une incubation à 37 °C pendant une nuit (protocole « overnight »).

Par choix d'économie d'enzyme, le protocole « overnight » sera utilisé pour toutes les analyses présentées dans les prochains chapitres.

6.4. Références

- [1] Crain, P. F., “[42] Preparation and Enzymatic Hydrolysis of DNA and RNA for Mass Spectrometry.” in: *Methods in Enzymology*; Mass Spectrometry; Academic Press, 1990; Vol. 193, pp 782–790.
- [2] Su, D. *et al.*, “Quantitative Analysis of Ribonucleoside Modifications in TRNA by HPLC-Coupled Mass Spectrometry.”, *Nat. Protoc.*, **2014**, 9 (4), 828–841.
- [3] Thüring, K. *et al.*, “Analysis of RNA Modifications by Liquid Chromatography–Tandem Mass Spectrometry.”, *Methods*, **2016**, 107, 48–56.
- [4] Wolff, P. *et al.*, “Comparative Patterns of Modified Nucleotides in Individual TRNA Species from a Mesophilic and Two Thermophilic Archaea.”, *RNA*, **2020**, rna.077537.120.
- [5] Lechner, A. *et al.*, “Characterization of Post-Transcriptional RNA Modifications by Sheathless Capillary Electrophoresis–High Resolution Mass Spectrometry.”, *Anal. Chem.*, **2020**, 92 (10), 7363–7370.

- [6] Wang, J. et al., "The Role of Sequence Context, Nucleotide Pool Balance and Stress in 2'-Deoxynucleotide Misincorporation in Viral, Bacterial and Mammalian RNA.", *Nucleic Acids Res.*, **2016**, *44* (18), 8962–8975.
- [7] Gosset-Erard, C. et al., "Optimization of Nucleotides Dephosphorylation for RNA Structural Characterization by Tandem Mass Spectrometry Hyphenated with Separation Methods.", *J. Chromatogr. B*, **2022**, *1208*, 123396.
- [8] Lu, D. et al., "RNA-Based Diagnostic and Therapeutic Strategies for Cardiovascular Disease.", *Nat. Rev. Cardiol.*, **2019**, *16* (11), 661–674.
- [9] Adams, D. et al., "Patisiran, an RNAi Therapeutic, for Hereditary Transthyretin Amyloidosis.", *N. Engl. J. Med.*, **2018**, *379* (1), 11–21.
- [10] Pardi, N. et al., "mRNA Vaccines — a New Era in Vaccinology.", *Nat. Rev. Drug Discov.*, **2018**, *17* (4), 261–279.
- [11] Boccaletto, P. et al., "MODOMICS: A Database of RNA Modification Pathways. 2021 Update.", *Nucleic Acids Res.*, **2022**, *50* (D1), D231–D235.
- [12] Boo, S. H. et al., "The Emerging Role of RNA Modifications in the Regulation of mRNA Stability.", *Exp. Mol. Med.*, **2020**, *52* (3), 400–408.
- [13] Agris, P. F., "The Importance of Being Modified: Roles of Modified Nucleosides and Mg²⁺ in RNA Structure and Function." in: *Progress in Nucleic Acid Research and Molecular Biology*; Cohn, W. E. et al., Eds.; Academic Press, 1996; Vol. 53, pp 79–129.
- [14] Bohnsack, M. T. et al., "The Mitochondrial Epitranscriptome: The Roles of RNA Modifications in Mitochondrial Translation and Human Disease.", *Cell. Mol. Life Sci.*, **2018**, *75* (2), 241–260.
- [15] Yan, M. et al., "A High-Throughput Quantitative Approach Reveals More Small RNA Modifications in Mouse Liver and Their Correlation with Diabetes.", *Anal. Chem.*, **2013**, *85* (24), 12173–12181.
- [16] Amalric, A. et al., "Quantifying RNA Modifications by Mass Spectrometry: A Novel Source of Biomarkers in Oncology.", *Crit. Rev. Clin. Lab. Sci.*, **2022**, *59* (1), 1–18.
- [17] Sakaguchi, Y. et al., "Chapter Two - Nucleoside Analysis by Hydrophilic Interaction Liquid Chromatography Coupled with Mass Spectrometry." in: *Methods in Enzymology*; He, C., Ed.; RNA Modification; Academic Press, 2015; Vol. 560, pp 19–28.
- [18] Liu, R. et al., "Determination of Five Nucleosides by LC-MS/MS and the Application of the Method to Quantify N6-Methyladenosine Level in Liver Messenger Ribonucleic Acid of an Acetaminophen-Induced Hepatotoxicity Mouse Model.", *J. Sep. Sci.*, **2019**, *42* (16), 2668–2678.
- [19] Raćkowska, E. et al., "Development and Validation of a Rapid LC-MS/MS Method for Determination of Methylated Nucleosides and Nucleobases in Urine.", *J. Chromatogr. B*, **2019**, *1128*, 121775.
- [20] Lo, W.-Y. et al., "Urinary Cytidine as an Adjunct Biomarker to Improve the Diagnostic Ratio for Gastric Cancer in Taiwanese Patients.", *Clin. Chim. Acta*, **2014**, *428*, 57–62.
- [21] Chen, F. et al., "Simultaneous Determination of 10 Nucleosides and Nucleobases in *Antrodia Camphorata* Using QTRAP LC-MS/MS.", *J. Chromatogr. Sci.*, **2014**, *52* (8), 852–861.
- [22] Feng, Y.-J. et al., "Identification of Inosine and 2'-O-Methylinosine Modifications in Yeast Messenger RNA by Liquid Chromatography-Tandem Mass Spectrometry Analysis.", *Anal. Chem.*, **2022**, *94* (11), 4747–4755.
- [23] Ma, Y. et al., "Quantitation of DNA by Nuclease P1 Digestion and UPLC-MS/MS to Assess Binding Efficiency of Pyrrolobenzodiazepine.", *J. Pharm. Anal.*, **2020**, *10* (3), 247–252.

- [24] Gątarek, P. *et al.*, “LC-MS/MS Determination of Modified Nucleosides in The Urine of Parkinson’s Disease and Parkinsonian Syndromes Patients.”, *Molecules*, **2020**, 25 (21), 4959.
- [25] Yang, W., “Nucleases: Diversity of Structure, Function and Mechanism.”, *Q. Rev. Biophys.*, **2011**, 44 (1), 1–93.
- [26] Yang, K. *et al.*, “A New Activity for an Old Enzyme: Escherichia Coli Bacterial Alkaline Phosphatase Is a Phosphite-Dependent Hydrogenase.”, *Proc. Natl. Acad. Sci.*, **2004**, 101 (21), 7919–7924.

CHAPITRE 7 : CARACTERISATION DES MODIFICATIONS POST-TRANSCRIPTIONNELLES D'UN ARN RIBOSOMIQUE

7.1. Introduction

Les chapitres précédents se sont concentrés sur la caractérisation des modifications post-transcriptionnelles, au niveau nucléosides, et sur l'analyse d'échantillons dont les modifications sont déjà référencées dans la littérature. La complexité de l'échantillon a été graduellement augmentée, en travaillant, dans un premier temps, avec les standards des quatre nucléosides canoniques, pour mettre en place le couplage CE-FTICR-MS, puis sur les ARNr purs de *S. cerevisiae* utilisés lors de la preuve de concept de la méthode CZE-MS/MS [1], précédemment développée au laboratoire, et enfin sur un extrait total d'ARNr de *B. taurus*. Bien que ces échantillons soient de plus en plus complexes, les séquences de ces ARN et leurs modifications sont déjà référencées dans la littérature [2,3].

Dans ce chapitre, un ARN ribosomique (ARNr) 16S pur issu de *P. aeruginosa* a été analysé, aux niveaux nucléosides et oligonucléotides, dans l'objectif d'augmenter encore la complexité de l'échantillon. L'utilisation du couplage CE-QTOF-MS, essentielle avant de pouvoir envisager le transfert de cette application au couplage CE-FTICR-MS, permet aussi de tester cette méthode CZE-MS/MS à deux niveaux [1], sur un cas réel et plus complexe.

Cet ARN a été choisi, car il présente deux caractéristiques sur lesquelles la méthode va pouvoir être évaluée :

- La séquence de l'ARN ribosomique choisi est vingt fois plus longue que celles des ARNr (1537 bases contre 75 pour les ARNr), ce qui rend les modifications potentiellement moins abondantes.
- Seule la séquence canonique génomique est, pour le moment, référencée dans la littérature [4]. Ces modifications ne sont donc pas connues.

Dans un premier temps, la méthode a été appliquée telle quelle sur le couplage CE-QTOF-MS, sans que la concentration de l'échantillon ne soit adaptée, afin d'évaluer son application directe à des échantillons plus complexes et de cibler les adaptations nécessaires.

L'analyse au niveau nucléosides a été réalisée en premier afin d'identifier les modifications présentes sur cet ARNr. L'analyse au niveau oligonucléotides a ensuite été réalisée pour localiser ces modifications. Afin d'évaluer la faisabilité de l'étude, avant d'aller plus loin, les oligonucléotides ont été générés par une simple digestion spécifique (RNase T1 puis BAP), au lieu de deux digestions en parallèle (RNase T1 puis BAP, et RNase A puis BAP) décrites dans la méthode. En ce qui concerne l'interprétation des données, il ne suffit plus de

vérifier la présence des modifications connues et leurs localisations sur la séquence. C'est une recherche totalement à l'aveugle, sans aucun à priori, qui doit être effectuée.

7.2. Premier niveau d'analyse : identification des nucléosides modifiés

La recherche d'informations sur les modifications post-transcriptionnelles de l'ARNr 16S de *P. aeruginosa* commence par l'analyse des nucléosides afin d'identifier ses modifications post-transcriptionnelles. Comme aucune information n'est connue, la recherche manuelle des modifications a été effectuée sur la liste complète des modifications décrites dans la littérature et répertoriées sur la base de données Modomics [2]. Cette liste comprend plus de 150 modifications.

Après analyse par CZE-MS/MS, il est nécessaire d'extraire les masses de toutes les modifications. L'électrophérogramme présenté en Figure 7.1 est alors obtenu.

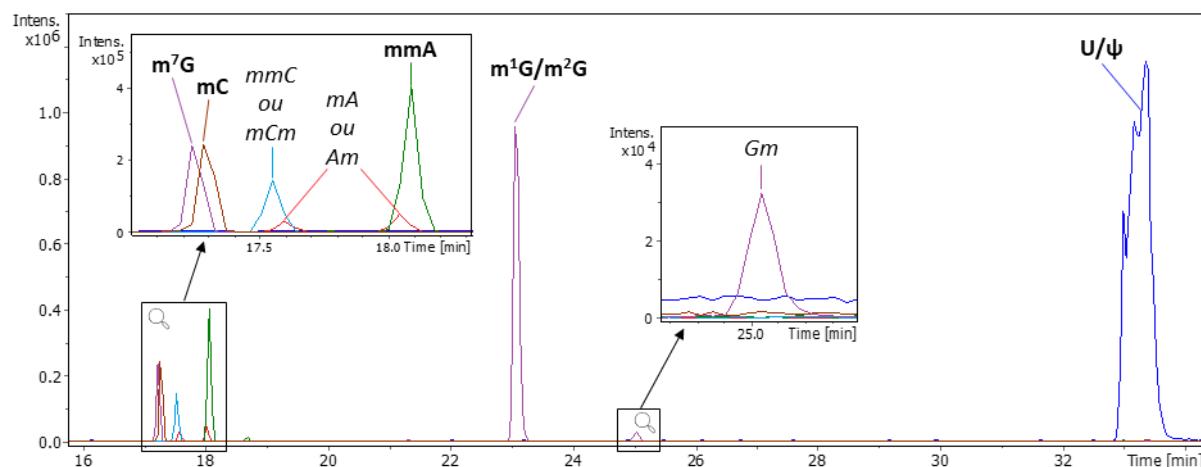


Figure 7.1 : Electrophérogramme des modifications de l'ARNr 16S issu de *P. aeruginosa*, obtenu par CZE-MS/MS après digestion à la Nucléase P1 et déphosphorylation à la BAP. Les nucléosides annotés en gras ont été identifiés par MS et MS/MS. Les nucléosides annotés en italique ont été uniquement identifiés par MS.

Cette analyse nous a permis d'identifier, par MS et MS/MS, les quatre nucléosides canoniques ainsi que les modifications suivantes : une pseudouridine (notée « ψ »), minimum deux isomères de guanine méthylée, notés « m^1G/m^2G » et « m^7G », une cytosine méthylée, notée « mC », et une adénine diméthylée, notée « mmA ». D'autres modifications ont pu être détectées par MS, mais leurs positions sur les nucléosides (sur la base ou sur le ribose) n'ont pas pu être déterminées par MS/MS par manque de signal. C'est le cas d'un troisième isomère de guanosine méthylée, notée « Gm », de deux isomères d'adénosines méthylées, notées « mA ou Am », et d'une cytosine diméthylée, notée « mCm ou mmC ». Pour confirmer ces modifications à l'aide de la MS/MS, il serait intéressant de travailler à plus grande concentration.

Grâce à cette première identification, au niveau nucléosides, le nombre de modifications potentielles à localiser est grandement réduit. De plus de 150 modifications, il n'en reste plus que 11. Il est important de noter que seules les modifications déjà décrites

dans la littérature ont été cherchées. Il n'est donc pas impossible que des modifications inconnues soient présentes, ou que d'autres soient inférieures à la limite de détection.

7.3. Second niveau d'analyse : oligonucléotides

Après avoir identifié 11 modifications post-transcriptionnelles, l'analyse au niveau oligonucléotides a été réalisée dans le but de localiser ces modifications. Dans un premier temps, uniquement la digestion à la RNase T1 suivie de la déphosphorylation à la BAP a été effectuée afin d'évaluer l'utilisation de la méthode sur un ARNr.

L'interprétation manuelle des données étant déjà longue et fastidieuse pour les ARNr étudiés dans la preuve de concept, la tâche est d'autant plus complexe pour cet ARNr au vu de sa taille, mais aussi de la méconnaissance des positions modifiées.

Pour tenter de réduire au maximum ce traitement de données, une première étape consiste à rechercher uniquement les oligonucléotides non modifiés. L'objectif était de réduire le nombre de positions, potentiellement modifiées, à rechercher sur la séquence. La seconde étape se focalise sur la localisation des modifications sur le reste des oligonucléotides.

7.3.1. Séquençage des oligonucléotides non modifiés

La séquence utilisée pour la digestion *in silico* a été établie à partir de la séquence génomique de cet ARN décrite par Toschka *et al.* en 1988 [4]. Pour cela, toutes les thymines (T) ont été remplacées par des uraciles (U).

L'électrophérogramme présenté en Figure 7.2 est obtenu après extractions des différents *m/z* correspondant aux oligonucléotides non modifiés. Les séquences des oligonucléotides détectés sont vérifiées par séquençage *De Novo* à partir des spectres MS/MS. Un exemple de séquençage *De Novo* est présenté en Figure 7.3 pour l'oligonucléotide CAAG.

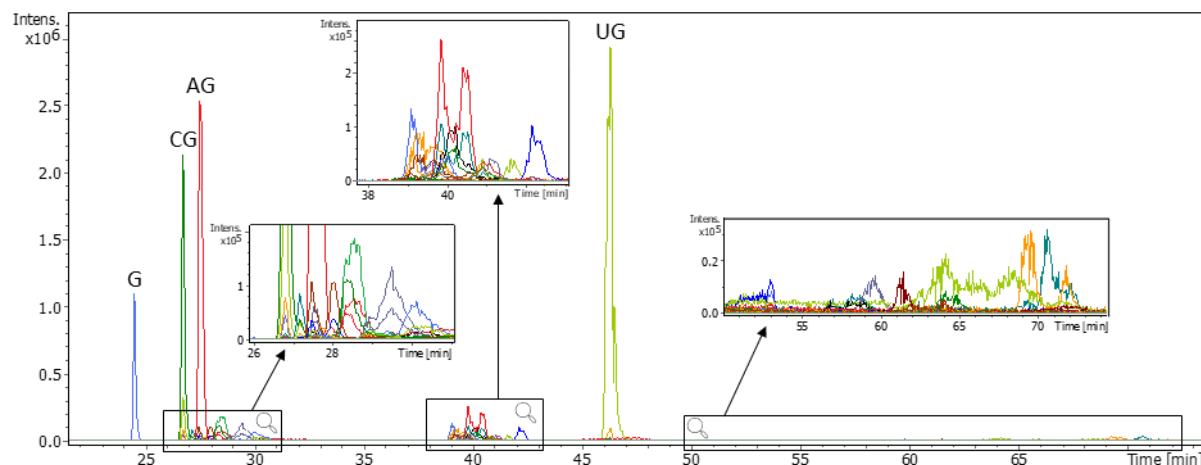


Figure 7.2 : Electrophérogramme des oligonucléotides de l'ARNr 16S issu de *P. aeruginosa*, obtenu par CZE-MS/MS après digestion à la RNase T1 et déphosphorylation à la BAP.

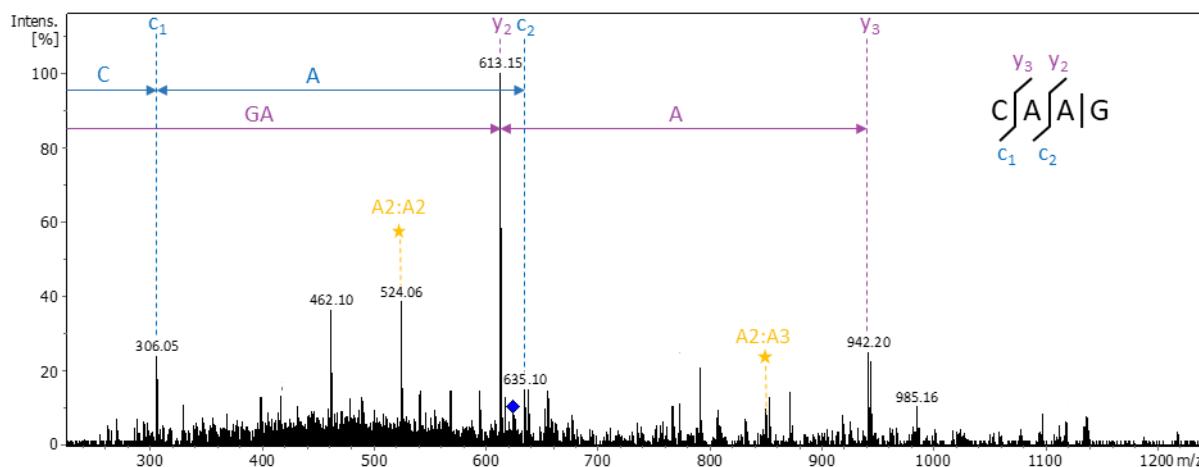


Figure 7.3 : Spectre MS/MS de CAAG issu de la fragmentation par CID (15,8 eV) de l'ion m/z 624,13 ($+2$). Deux fragments internes sont indiqués par les étoiles jaunes.

Sur cet électrophérogramme (Figure 7.2), deux groupes assez caractéristiques se dégagent. Tout d'abord, les pics obtenus avant 35 min correspondent à des oligonucléotides ne comprenant aucun uracile. Le deuxième groupe, après 35 min, correspond quant à lui aux oligonucléotides comprenant au minimum un uracile. Cette différence de migration peut s'expliquer par l'absence de charge sur les uraciles et son impact sur la charge globale des oligonucléotides. En effet, à pH 2.2 (acide acétique 10%), les trois autres bases possèdent une charge positive, et les phosphates possèdent, quant à eux, une charge négative. Comme il y a un phosphate de moins par rapport au nombre de bases, il suffit qu'un uracile (U) soit présent dans la séquence pour rendre la charge globale nulle, ou même négative si plus d'un uracile est présent. Plus il y a d'uraciles, plus la charge globale est négative. Les grands oligonucléotides possédant de nombreux uraciles ont, probablement, une mobilité apparente négative trop importante par rapport au flux électroosmotique, ce qui les empêche de migrer vers le spectromètre de masse. Une autre explication est liée à la très faible quantité d'ARNr injectée (0,7 fmol). En effet la preuve de concept [1] a montré que les pics des oligonucléotides contenant plusieurs uraciles sont peu efficaces, réduisant ainsi l'intensité de leur signal. Comme les oligonucléotides détectés sont déjà de faible intensité (à l'exception de G, CG, AG et UG qui sont très redondants), il est possible que certains oligonucléotides, dont les pics sont moins efficaces, soient en dessous de la limite de détection.

Le recouvrement de la séquence canonique de l'ARN ribosomique est présenté en Figure 7.4. La digestion par la RNase T1 est représentée par les espaces après les guanosines (G). Le recouvrement de séquence issu de cette analyse est de 1,4 % lorsque seulement les oligonucléotides identifiés et uniques sont pris en compte. Les oligonucléotides identifiés, mais redondants représentent 40,3 % de la séquence. Malheureusement, l'identification des oligonucléotides redondants ne permet pas d'obtenir des informations sur les positions non modifiées. C'est pourquoi ils ne sont pas compris dans le calcul du recouvrement de séquence. 29,3% de la séquence n'ont pas pu être confirmés par le séquençage *De Novo* par manque de signal. Enfin, les oligonucléotides qui n'ont pas été détectés lors de cette analyse représentent 29,0% de la séquence.

G AACUG AAG AG UUUG AUCAUG G CUCAG AUUG AACG CUG G CAG CAG G G G CCUUCAACACAUG
 CAAG UCG AG CUUAUG AAG G G AG CUUG CCUUG G AUUCAG CG G CG G ACG G G UG AG UAAUG
 CCUAG G AAUCUG CCUG G UAG UAG G G G AUAACG UCCG G AAACG G CCG CUAAUACCG CAUACG
 UCCUG AG G G AG AAAG UCG G G G AUCUUCG G ACCUCACG CUAUCAG AUG AG CCUAG G UCG G
 AUUAG CUAG UUG G UG G G G UAAAG G CCUACCAAG G CG ACG AUCCG UAACUG G UCUG AG AG G
 AUG AUCAG UCACACUG G AACUG AG ACACG G UCCAG ACUCCUACG G G AG G CAG CAG UG G G G
 AAUAUUG G ACAAUG G G CG CAAG CCUG AUCCAG CCAUG CCG CG UG UG UG AAG AAG G UCUUCG
 G AUUG UAAAG CACUUUAAG UUG G G AG G AAG G G CAG UAAG UUAAUACCUUG CUG UUUG ACG
 UUACCAACAG AAUAAG CACCG G CUACUUCG UG CCAG CAG CCG CG G UAAUACG AAG G G UG
 CAAG CG UUAAUCG G AAUUAUCUG G G CG UAAAG CG CG UUAG UG G UUCAG CAAG CUUG AUG
 UG AAAUCCCCG G G CUCAACCUG G G AACUG CAUCCAAAAG CUACUG AG CUAG AG UACG G UAG
 AG G UG G UG AAUUUCCUG UG UAG CG G UG AAAUG CG UAG AUAUAG G AAG G AACACCAG UG G
 CG AAG G CG ACCACCUG G ACUG UACUG ACACUG AG G UG CG AAAG CG UG G G G AG CAAACAG G
 AUUAG AUACCCUG G UAG UCCACG CCG UAAACG AUG UCG ACUAG CCG UUG G G AUCCUUG AG
 AUCUUAG UG G CG CACG UAACG CG AUAAG UCG ACCG CCUG G G G AG UACG G CCG CAAG G
 UUAAAACUAAAUG AAUUG ACG G G G CCCG CACAAG CG G UG G AG CAUG UG G UUUAAUUCG
 AAG CAACG CG AAG AACCUUACCUUG G CCUUG ACAUG CUG AG AACUUUCCAG AG AUG G AUUG G
 UG CCUUCG G G AACAG AG ACACAG G UG CUG CAUG G CUG UCG UCAG CUCG UG UCG UG AG AUG
 UUG G G UUAAG UCCCG UAACG AG CG CAACCCUUG UCCUUAG UUACCG CACCUCG G G UG G G
 CACUCUAAG G AG ACUG CCG G UG ACAAAACCG G AG G AAG G UG G G G AUG ACG UCAAG
 UCAUCAUG G CCCUUACG G CCAG G G CUACACACG UG CUACAAUG G UCG G UACAAAG G G UUG
 CCAAG CCG CG AG UG G G AG CUAAUCCCAAAAACCG AUCG UAG UCCG G AUCG CAG UCUG
 CAACUCG ACUG CG UG AAG UCG G AAUCG CUAG UAAUCG UG AAUCAG AAUG UCACG G UG
 AAUACG UCCCCG G G CCUUG UACACACCG CCCG UCACACCAUG G G AG UG G G UUG CUCCAG AAG
 UAG CUAG UCUAACCG CAAG G G G G ACG G UUACCACG G AG UG AAUCAUG ACUG G G G UG AAG
 UCG UAACAG G UAG CCG UAG G G G AACCUUG CG G CUG G AUCACCUUUA

Figure 7.4 : Recouvrement de la séquence canonique de l'ARNr 16S issu de *P. aeruginosa* digéré à la RNase T1 et à la BAP. Les oligonucléotides, générés par cette digestion, sont séparés par un espace. Les oligonucléotides verts ont été identifiés par MS et MSMS et sont uniques (non redondants). Les oligonucléotides oranges ont été identifiés par MS et MSMS, mais sont redondants. Les oligonucléotides bleus sont à confirmer (identifiés uniquement par MS, et potentiellement une seule fois pour plusieurs isomères), et les oligonucléotides rouges n'ont pas été détectés.

À l'exception d'un unique oligonucléotide (AACACCAG), tous ceux qui n'ont pas été détectés contiennent au moins un uracile. Cette tendance peut déjà être remarquée, dans une moindre mesure, à partir des résultats présentés par Lechner *et al.* [1]. Cependant, la double digestion spécifique était suffisante pour compenser ses effets de redondance, passant ainsi d'un recouvrement de séquence de l'ARNt^{Phe GAA} issu de *S. cerevisiae* de 67%, après une première digestion à la RNase T1, à 96%, en combinant les deux digestions à la RNase T1 et à la RNase A [1]. En effet, plus les séquences d'ARN sont longues, plus le risque de redondance dans les séquences d'oligonucléotides, issus d'une seule digestion spécifique, est élevé. Sur l'ARNr 16S issu de *P. aeruginosa*, digéré uniquement à la RNase T1 et déphosphorylé à la BAP, 54% de la séquence canonique totale correspond à des oligonucléotides redondants. Une double digestion spécifique, comme celle proposée dans la méthode initiale, est d'autant plus intéressante et nécessaire lorsque les séquences d'ARN sont plus longues. Par comparaison, l'ARNt^{Phe GAA} issu de *S. cerevisiae*, utilisé lors de la preuve de concept et comptant au total 76 nucléotides, présente seulement 14% de redondance après une digestion à la RNAse T1 [1]. Au vu de la taille de l'ARN ribosomique, il est probable qu'une troisième digestion spécifique, par exemple, par la RNase U2 qui coupe après les purines (G et A) non modifiées [5], soit

nécessaire pour réduire au maximum l'impact de la redondance, et ainsi obtenir des recouvrements de séquence aussi importants que ceux obtenus lors de la preuve de concept sur les ARNr.

Seuls les cinq oligonucléotides uniques qui ont été identifiés sont nécessairement non modifiés. À cause de la grande redondance de la séquence, ce premier séquençage ne permet pas réellement de réduire significativement le nombre de potentielles positions modifiées. La recherche concernant la position des modifications doit donc être effectuée sur la quasi-totalité de la séquence.

7.3.2. Localisation des modifications post-transcriptionnelles

Pour connaître les rapports *m/z* à extraire pour localiser les modifications, toutes les combinaisons possibles doivent être testées. Néanmoins, en se basant sur les modifications identifiées au préalable par l'analyse de nucléosides, il existe seulement 15 formes modifiées potentielles pour un petit oligonucléotide tel que AG, mais plus de 7 millions pour un oligonucléotide tel que CUAAUCCCAUAAAACCG (qui est le plus long de la séquence). L'identification des modifications par l'analyse des nucléosides permet de réduire significativement le nombre de formes possibles, mais pas suffisamment pour être capable de les localiser manuellement. Par ailleurs, certaines modifications sur les guanosines (notamment m^7G) empêchent l'hydrolyse par la RNase T1, ce qui induit un « miscleavage » à prendre en compte dans la simulation de la digestion. Ces « miscleavages » induisent une augmentation du nombre de possibilités à rechercher.

Avant de poursuivre sur les ARN ribosomiques, il sera nécessaire de mettre en place un nouveau logiciel permettant le séquençage de l'ARN en mode positif.

7.4. Conclusion

Les résultats préliminaires présentés dans ce chapitre ont démontré que la méthode CZE-MS/MS à deux niveaux, précédemment développée au laboratoire [1], peut être appliquée à des échantillons plus complexes dont les modifications ne sont pas connues. Cette méthode demande, cependant, à être légèrement adaptée avant de transférer ces applications plus complexes sur le couplage CE-FTICR-MS.

L'analyse au niveau nucléosides a permis d'identifier 11 modifications post-transcriptionnelles. Bien que cette recherche à l'aveugle de nucléosides modifiés soit réalisable manuellement, il est important de noter qu'elle est également fastidieuse et chronophage. De plus, la faible concentration en échantillon, lors de cette analyse, et la faible abondance en modifications due à la taille de l'ARNr 16S, ont contribué à la faible intensité du signal. Certaines modifications n'ont probablement pas été détectées, car trop peu concentrées, bien qu'elles soient présentes dans l'échantillon. Il serait donc intéressant, par

la suite, de refaire cette analyse en concentrant plus l'échantillon, afin de vérifier s'il n'est pas possible d'identifier d'autres modifications.

Au second niveau d'analyse, le séquençage de certains oligonucléotides uniques et non modifiés a pu être réalisé. Cependant, la localisation des modifications, préalablement identifiées par le premier niveau d'analyse, n'a pas pu être réalisée manuellement et requiert le développement d'un logiciel. De plus, l'utilisation d'une deuxième, voire d'une troisième RNase spécifique, sera nécessaire pour réduire au maximum l'impact de la redondance de la séquence, et d'augmenter au maximum le recouvrement de la séquence. Une augmentation de la concentration en échantillon pourra également être bénéfique pour la détection des oligonucléotides uniques, modifiés ou non-modifiés.

Dans le chapitre suivant, le développement de deux logiciels, qui permettent de pallier aux besoins d'outils bio-informatiques d'interprétation des données pour les deux niveaux d'analyse, sera détaillé.

7.5. Références

- [1] Lechner, A. *et al.*, “Characterization of Post-Transcriptional RNA Modifications by Sheathless Capillary Electrophoresis–High Resolution Mass Spectrometry.”, *Anal. Chem.*, **2020**, 92 (10), 7363–7370.
- [2] Boccaletto, P. *et al.*, “MODOMICS: A Database of RNA Modification Pathways. 2021 Update.”, *Nucleic Acids Res.*, **2022**, 50 (D1), D231–D235.
- [3] Kellner, S. *et al.*, “Profiling of RNA Modifications by Multiplexed Stable Isotope Labelling.”, *Chem. Commun.*, **2014**, 50 (26), 3516–3518.
- [4] Toschka, H. Y. *et al.*, “Complete Nucleotide Sequence of a 16S Ribosomal RNA Gene from *Pseudomonas Aeruginosa*.”, *Nucleic Acids Res.*, **1988**, 16 (5 Pt B), 2348.
- [5] Houser, W. M. *et al.*, “Combining Recombinant Ribonuclease U2 and Protein Phosphatase for RNA Modification Mapping by Liquid Chromatography–Mass Spectrometry.”, *Anal. Biochem.*, **2015**, 478, 52–58.

CHAPITRE 8 : DEVELOPPEMENT DE DEUX LOGICIELS POUR LA CARACTERISATION DES ARN A PARTIR D'ANALYSES EN MODE POSITIF

8.1. Introduction

Dans le chapitre précédent, l'analyse par CE-MS/MS de l'ARNr 16S issu de *P. aeruginosa* aux niveaux nucléosides et oligonucléotides a mis en lumière la nécessité de développer des outils bio-informatiques de manière à faciliter, et même rendre possible, l'interprétation de telles analyses. En effet, comme la nature des modifications post-transcriptionnelles de cet ARN ne sont pas décrites dans la littérature, l'interprétation manuelle de l'analyse des nucléosides a été possible, mais chronophage. De la même manière, l'interprétation de l'analyse au niveau oligonucléotides a également été très chronophage, même en ne considérant dans un premier temps uniquement la séquence canonique de l'ARN. La localisation de ces modifications sur la séquence a, quant à elle, été impossible tant le nombre de combinaisons possibles pour chaque oligonucléotide modifié était gigantesque.

Actuellement, la LC-MS/MS reste la méthode de référence pour la caractérisation des ARN, que ce soit pour l'analyse des nucléosides ou celle des oligonucléotides. Au niveau nucléosides, la méthode LC-MS/MS de référence utilise le mode « Multiple Reaction Monitoring » (MRM) pour suivre différentes transitions spécifiques aux modifications post-transcriptionnelles [1–11]. Bien que la MRM permette une interprétation facile et rapide des données, elle ne permet pas de réaliser une identification non ciblée puisqu'il est nécessaire d'établir au préalable une liste de modifications d'intérêt à suivre. Au laboratoire, nous avons décidé de développer la méthode CE-MS/MS en la basant sur une acquisition de la MS/MS en mode « data-dependent » (DDA) afin d'identifier les modifications de manière non ciblée [12]. Or, aucun logiciel n'est disponible pour interpréter de telles données et identifier les modifications post-transcriptionnelles. Au niveau oligonucléotides, la méthode LC-MS/MS est en mode négatif et les oligonucléotides ne sont pas déphosphorylés [1,13–15]. A contrario, notre méthode CE-MS/MS est en mode positif et requiert la déphosphorylation des oligonucléotides avant analyse [12]. Bien que quelques logiciels aient été développés pour les analyses en mode négatif (Ariadne [16], RoboOligo [17], RNAModMapper [18,19], NucleicAcidSearchEngine [20], et Pytheas [21]), et que certains aient ajouté par la suite un mode positif (Ariadne et Pytheas [22]), aucun logiciel ne nous permettait d'interpréter facilement nos analyses CE-MS/MS autrement que manuellement.

Ce huitième chapitre porte donc sur le développement de deux logiciels pour faciliter l'interprétation des données pour la caractérisation des modifications post-transcriptionnelles des ARN. Le premier logiciel, Nucleos'ID, permettant l'identification non ciblée des modifications post-transcriptionnelles à partir d'analyses MS/MS de nucléosides, a fait l'objet d'une publication, intitulée « Optimization of nucleotides dephosphorylation for RNA

structural characterization by tandem mass spectrometry hyphenated with separation methods » et parue dans le journal scientifique à comité de lecture *Analytical Chemistry* (doi.org/10.1021/acs.analchem.2c04722).

Le second logiciel, OligoMod+, ayant pour objectif de localiser les modifications post-transcriptionnelles à partir d'analyses MS/MS d'oligonucléotides en mode positif, est en cours de développement. Actuellement, il permet de retrouver les oligonucléotides modifiés d'un échantillon d'ARN dont les modifications sont bien connues et localisées. Ainsi, son fonctionnement actuel sera brièvement détaillé pour clôturer ce chapitre.

8.2. Publication : “Nucleos’ID: new search engine enabling the untargeted identification of RNA post-transcriptional modifications from tandem mass spectrometry analyses of nucleosides”

Clarisse Gosset-Erard[‡], Mévie Didierjean[‡], Jérôme Pansanel[‡], Antony Lechner, Philippe Wolff, Lauriane Kuhn, Frédéric Aubriet, Emmanuelle Leize-Wagner, Patrick Chaimbault and Yannis-Nicolas François.

[‡]These authors contributed equally.

[DOI:10.1021/acs.analchem.2c04722](https://doi.org/10.1021/acs.analchem.2c04722)

8.2.1. Abstract

As RNA post-transcriptional modifications have shown a growing interest, several methods were developed for their characterization. One of them established for their identification, at the nucleosidic level, is the hyphenation of separation methods, such as liquid chromatography or capillary electrophoresis, to tandem mass spectrometry. However, to our knowledge, no software is yet available for the untargeted identification of RNA post-transcriptional modifications from MS/MS data-dependent acquisitions. Thus, very long and tedious manual data interpretations are required. To meet the need of easier and faster data interpretation, a new friendly-user search engine, called Nucleos’ID, was developed for CE-MS/MS and LC-MS/MS users. Performances of this new software were evaluated on CE-MS/MS data from nucleoside analyses of already well-described *S. cerevisiae* transfer RNA and *B. taurus* total tRNA extract. All samples showed great true positive, true negative and false discovery rates considering the database size containing all modified and unmodified nucleosides referenced in the literature. The true positive and true negative rates obtained were above 0.94 while the false discovery rates were between 0.09 and 0.17. To increase the level of sample complexity, untargeted identification of several RNA modifications from *P. aeruginosa* 70S ribosome was achieved by the Nucleos’ID search following CE-MS/MS analysis.

8.2.2. Introduction

Due to their implication in many biological and structural functions, RNA post-transcriptional modifications are of great interest. Indeed, RNA modifications were reported to play an important role in RNA structural folding and stability [23], base pairing [24], translation fidelity [25], regulation of gene expression [26], but also cell development [27]. Some modifications are also involved in antibiotic resistance mechanisms [28] or several diseases [29]. Over the years, therapeutic strategies based on post-transcriptional RNA modifications have received growing interest. For instance, modified messenger RNA were reported as potential treatment for infectious disease [30] and some RNA modifications were reported as cancer biomarkers for diagnostic purposes [31].

Nowadays, more than 150 post-transcriptional modifications are described in the literature and referenced in the Modomics database [32] for all types of RNA. The variety of modifications goes from simple chemical groups such as methylations, which are the most naturally abundant, to more complex modifications such as total rearrangements of the nucleic bases.

One of the methods enabling the characterization of those modifications is mass spectrometry (MS) and especially tandem mass spectrometry (MS/MS). Over the years, MS has become a benchmark method for the characterization of biomolecules thanks to its high sensitivity, high specificity, and ability to obtain structural information. For RNA characterization, MS is usually hyphenated with a separation method, like liquid chromatography (LC) [1–11,13–15,33–36] or capillary electrophoresis (CE) [12]. Such hyphenations enable lower sample consumption, isomers separation and higher sensitivity, making them suitable, and essential, techniques for the characterization of RNA post-transcriptional modifications.

Characterization of RNA modifications with CE-MS/MS or LC-MS/MS is based on a two-level strategy. Briefly, the first level consists in identifying all the different post-transcriptional modifications present in the sample after a total hydrolysis of RNA followed by a dephosphorylation to obtain nucleosides [1–5,10–12,14,15,33]. In a second level, the identified modifications are then precisely mapped on the sequence after specific RNase digestions to obtain oligonucleotides [1,12–15].

To get information on the modified positions on the nucleosides, collision-induced dissociation (CID) is classically used for MS/MS fragmentation. Indeed, CID leads to the fragmentation of the N-glycosidic bond and thus, to the neutral loss of the modified or unmodified ribose, enabling the discrimination between a modified base or a modified ribose [12,37].

Two modes have been reported for MS/MS analyses of nucleotides. The most common is the Multiple Reaction Monitoring (MRM) mode [1–11]. MRM enables fast and easy data

interpretation, including for quantification studies [2,3,11]. However, MRM is a targeted strategy which requires a pre-established list of modified and non-modified nucleosides of interest to be monitored. If one modification is not well predicted and monitored, it will not be detected. Thus, MRM is not really suited for untargeted analyses. The second one, much less common, is the classical MS/MS using data-dependent acquisition mode [12]. This mode enables identification of RNA modification without any preconceptions. However, to our knowledge, no software is yet available for the untargeted identification of RNA post-transcriptional modifications from classical data dependent tandem MS data. Therefore, a long and tedious manual interpretation is required.

To meet the need in bioinformatic tools to fasten and ease data interpretation of nucleosides analyses, a new friendly-user search engine, called Nucleos'ID, was developed. After entering different evaluation parameters, like MS and MS/MS mass tolerances, MS/MS absolute intensity threshold, MS/MS score threshold, and active exclusion time, the software will compare experimental MS/MS data with a database to identify modifications. Each identified modification will then be scored, and results will be saved in different formats (CSV, XLSX) after being sorted to ease the user's reading. All modifications described in the literature and referenced in the Modomics database [32] are considered, thus enabling untargeted searches.

8.2.3. Materials and methods

8.2.3.1. Chemicals

The chemicals used were of analytical grade or high purity. Water was obtained using a Direct-Q3 UV-R water purification system (Merck Millipore, Amsterdam, The Netherlands). Nuclease P1 from *Penicillium citrinum*, ammonium acetate and zinc chloride ($ZnCl_2$) were purchased from Sigma-Aldrich (Saint Louis, MO, U.S.A.). Glacial acetic acid and Bacterial Alkaline Phosphatase (BAP) (150 U/ μ L) from *Escherichia coli* C90 were purchased from ThermoFisher Scientific (Marietta, OH, U.S.A.).

8.2.3.2. RNA samples

Two RNA samples were used to evaluate the results obtained with Nucleos'ID. Those samples are a purified tRNA^{Phe GAA} from *Saccharomyces cerevisiae* and a total tRNA extract from *Bos taurus*. The purity of both samples was controlled by polyacrylamide gel electrophoresis. In addition to the four canonical nucleosides (A, C, G, U), the *S. cerevisiae* tRNA^{Phe GAA} is composed of 11 post-transcriptional modifications, and the *B. taurus* total tRNA extract is composed of 27 modifications. Those modifications are well known and already described in the literature (Table 8.1) [32]. Thus, both samples are good models of increasing complexity to evaluate the performance of the software. A third sample, whose modifications are not described in the literature yet, was analyzed to apply Nucleos'ID on a more complex case. This sample is a purified 70S ribosome containing 5S, 16S and 23S ribosomal RNA (rRNA)

from *Pseudomonas aeruginosa*. The purity of this sample was controlled by agarose gel electrophoresis. All nucleosides described in this study annotated with their short name. All full names are available in Supporting Information (Table S8.1).

Table 8.1 : List of modified and non-modified nucleosides reported in Modomics database [32] for *S. cerevisiae* tRNA^{Phe GAA} and *B. taurus* total tRNA extract. Modified RNA sequences for *S. cerevisiae* tRNA^{Phe GAA} and *B. taurus* total tRNA extract are available in Supplemental information (Figure S8.1 and Figure S8.2).

Nucleoside	<i>S. cerevisiae</i> tRNA ^{Phe GAA}	<i>B. taurus</i> total tRNA extract
A	✓	✓
C	✓	✓
G	✓	✓
U	✓	✓
m ¹ A	✓	✓
t ⁶ A	-	✓
i ⁶ A	-	✓
I	-	✓
m ³ C	-	✓
m ⁵ C	✓	✓
Cm	✓	✓
ac ⁴ C	-	✓
f ⁵ Cm	-	✓
m ¹ G	-	✓
m ² G	✓	✓
m ⁷ G	✓	✓
Gm	✓	✓
m ² ₂ G	✓	✓
Q	-	✓
galQ	-	✓
manQ	-	✓
yW	✓	-
o ₂ yW	-	✓
Y	✓	✓
D	✓	✓
Um	-	✓
m ⁵ U	✓	✓
Ym	-	✓
m ⁵ Um	-	✓
mcm ⁵ U	-	✓
mcm ⁵ s ² U	-	✓
acp ³ U	-	✓

8.2.3.3. Sample preparation

First, three successive precipitations with ammonium acetate and ethanol were performed to desalt RNA samples. Then, nuclease P1 and BAP were used to perform total hydrolysis of RNA, following a previously described protocol [12]. Briefly, P1 10x buffer (2 mM ZnCl₂, 250 mM ammonium acetate, pH 5.0) and RNA samples were successively diluted into deionized water before adding 0.5 U of P1 (previously prepared at 0.5 U/µL in 100 mM ammonium acetate) per µg of tRNA. A first incubation was carried out at 37 °C for 2 hours. After adding 7.5 U of BAP (previously prepared at 1.5 U/µL in 100 mM ammonium acetate) for each µg of RNA, a second incubation was carried out at 37 °C for 4 hours.

Finally, samples were fully evaporated and dissolved in ammonium acetate 200 mM pH 4 to a final concentration of 0.05 µg/µL for the tRNA^{Phe GAA} from *S. cerevisiae*, 0.1 µg/µL for the total tRNA extract from *B. taurus*, and 1.0 µg/µL for the 70S ribosome of *P. aeruginosa*.

8.2.3.4. Capillary electrophoresis-tandem mass spectrometry (CE-MS/MS)

A CESI8000 capillary electrophoresis system (Sciex Separation, Brea, CA, USA) was hyphenated to a maXis 4G mass spectrometer (Bruker Daltonics, Bremen, Germany) thanks to a porous tip sheathless interface. CE-MS/MS analyses were performed following a previously described method [12]. Briefly, 5nL of sample were injected in the CE capillary and CE separation was carried out by applying +30 kV voltage on a bare-fused silica capillary (100 cm total length, 30 µm i.d.) filled with 10 % acetic acid as background electrolyte. MS detection was performed in ion positive mode over the range 80-800 *m/z*, with an accumulation time of 1 s and a resolution of 49 426 at 410 *m/z*. Data dependent MS/MS was performed using an accumulation time from 0.14 s to 0.56 s depending on the number of precursor ions, CID fragmentation mode and a collision energy of 20 eV for all precursor ions. From 3 to 7 most intense precursor ions were selected in the quadrupole with a window of 1.5 *m/z*. Dynamic exclusion was activated for 0.7 min after each spectrum. Finally, MS/MS data were converted into MGF (Mascot Generic Format) files for input in Nucleos'ID. The conversion was performed using DataAnalysis software (Bruker Daltonics, Bremen, Germany). However, MSconvert can also be used.

8.2.3.5. Nucleos'ID workflow

Nucleos'ID is a software developed in Python 3 and providing an efficient and sober user interface. It is available for Linux, Mac OS X, and Windows operating systems. Both executables and source code can be downloaded from the Github repository (<https://github.com/MSARN/NucleosID>). The code is distributed under an open-source license (Apache 2) and is registered on Zenodo [38] .

The global workflow for data analysis and data interpretation using Nucleos'ID is represented in Figure 8.1. For the identification of nucleosides, RNA is reduced into

nucleosides before CE-MS/MS or LC-MS/MS analysis. MS/MS data are then converted to a MGF file for interpretation with the software. After suppressing the background noise, Nucleos'ID compares MS/MS data with a database to identify modifications. Each identified modification is then scored, and results are displayed after being sorted to ease the user's reading. Nucleos'ID can be used with low- or high-resolution MS/MS data, however the best results would be obtained while using with MS resolution higher than 25 000 and mass errors lower than 30 ppm.

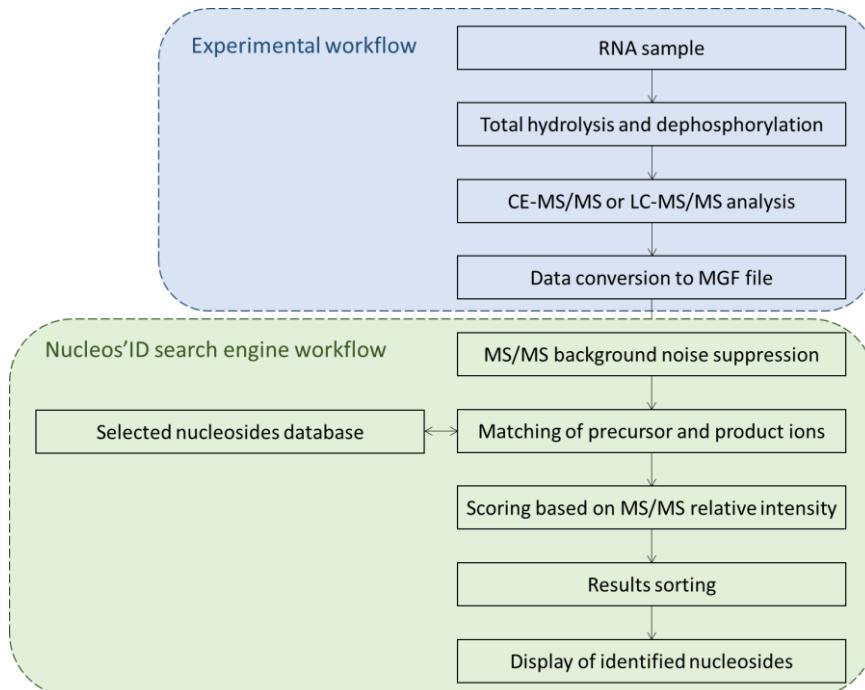


Figure 8.1: Workflow of identification of RNA modifications by CE-MS/MS or LC-MS/MS.

8.2.3.6. Databases

Three databases, recalculated from the Modomics database [32], are available depending on the primary phylogenetic domain of the sample (Eubacteria, Archaea and Eukarya) [39,40]. They can be reused, as long as the attribution to Modomics is cited. The databases will be kept up to date in the future thanks to the release of new versions. Those databases contain *m/z* values of precursor ions (+1 charge state) and product ions (+1 charge state) generated by CID for all non-modified and modified nucleosides (thus, after dephosphorylation). Most nucleosides only have one product ions in the database because the N-glycosidic linkage is the one that is preferentially fragmented by CID. Moreover, several modifications, such as m^1G , m^2G and m^7G , are grouped into sets as they share the same precursor and product ions. Therefore, it is important to know that the discrimination of isomers grouped into a set is not possible with Nucleos'ID. The user can select one of those databases, or a combination of two or three of them, for the search. By default, a combination of the three databases is considered.

8.2.3.7. MS/MS background noise suppression

The first step of Nucleos'ID workflow is a cleaning procedure of the MS/MS background noise. This process removes, in the MGF file, all product ions below an absolute intensity threshold defined by the user. By default, the absolute intensity threshold is set to 0, such as no cleaning is performed. Indeed, noise is instrument dependent, and a wrong threshold setting can lead to many false negatives. To estimate this threshold, the intensity of MS/MS spectra with only low intensity background noise must be manually checked. For instance, with our set-up a threshold around 500 AU could have been set (one MS/MS spectrum used for this estimation is available in Figure S8.3 in supplemental information).

When used, this step also enables the reduction of the number of comparisons. Therefore, the execution time and the needed computer resources are also reduced.

8.2.3.8. Modifications matching

To identify post-transcriptional modifications, MS/MS data are compared to a database containing the *m/z* of precursor ions and products ions. Once a precursor ion is matching experimentally and theoretically within MS mass tolerance, all the product ions from this precursor are searched in the MS/MS spectra within MS/MS mass tolerance. A modification is identified when the precursor ion and at least one of the product ions are matching. MS and MS/MS mass tolerances used for the comparison are defined by the user, either in Da or in ppm [41]. As the result file is well detailed and includes accurate measured *m/z* of all matching precursor and product ions, it is possible to reprocess the results afterwards with lower mass tolerances.

8.2.3.9. Scoring system

As with any data interpretation software, it is essential to associate a confidence score to each result. This score is used to differentiate which result can be a false positive among all the true positives. To our knowledge, scoring systems in proteomics or oligonucleotides sequencing, are always probability based. Indeed, the number of matched fragments among all theoretical fragments is used to calculate a probability that the match is only due to chance [16,18,42,43]. However, in the case of nucleosides, a similar scoring function would not be representative since most of them have only one specific product ion after CID fragmentation. Moreover, for some modifications, all MS/MS fragments in the database are not always detected. This are the reasons why our scoring system is solely based on the MS/MS relative intensity of the most intensive matched product ion. For instance, if a matched product ion is the most intense peak of its MS/MS spectrum, the score associated with this modification will be 100 %.

8.2.3.10. Result sorting

Before saving the output, results are filtered according to two user-defined criteria. The first sort is based on the score. If a result has a score below the user-defined threshold, it will not be displayed in order to reduce the number of false positives. The second sorting is done to reduce the redundancy of the modifications identified in a short period of time. Indeed, if the same modification is selected several times to be fragmented in MS/MS, it will be identified as many times by the software. For this, the user can define an exclusion time. If the same modification is identified several times during this exclusion time, then only the result with the best MS/MS intensity is selected. Both filters make it convenient to select the most significant matches.

8.2.3.11. Results display

Final results are then saved using one of the two available file formats (CSV or XLSX). By default, CSV format is proposed, as it enables easier additional post-processing of the data and can also be open by any spreadsheet software.

8.2.3.12. Data analysis

Data interpretation was performed with Nucleos'ID software on a Dell Inc. laptop using Windows 10 operating system with an Intel Core i5 processor and 8 GB RAM. Nucleos'ID search was performed with default parameters for the purified *S. cerevisiae* tRNA^{Phe GAA} and *B. taurus* total tRNA extract. The default settings are presented on the friendly-user interface in Figure 8.2.

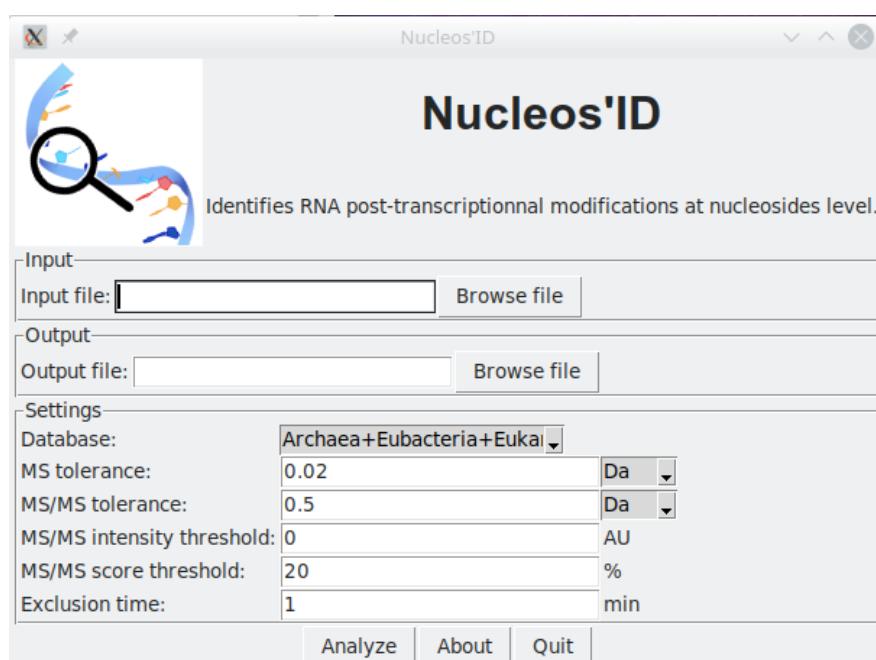


Figure 8.2: Friendly-user interface of Nucleos'ID with its default settings.

For the *P. aeruginosa* 70S ribosome, the score threshold was set to 0 % to ensure that no modification was missed as none are described in the literature for this sample, and only the Eubacteria database was used. Manual search of all nucleosides in the database was also performed, and results were compared.

8.2.3.13. Evaluation of True Positive, True Negative and False Discovery Rates

To assess the performances of the new search engine, three common rates based on the confusion matrix containing the numbers of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) results (Table 8.2) were chosen. Those rates are the True Positive Rate, the True Negative Rate and the False Discovery Rate [44]. They are commonly used as statistically metrics to assess software performances [16,45,46].

Table 8.2: Confusion matrix for the classification of true positives, true negatives, false positives and false negatives depending on the Nucleos'ID results and the actual nucleosides content of the sample.

		Actual nucleoside content of the sample	
		Present nucleoside	Non-present nucleoside
Nucleos'ID search	Identified nucleoside	True positive (TP)	False positive (FP)
	Non identified nucleoside	False negative (FN)	True negative (TN)

To identify the actual nucleoside content of the sample, manual interpretation was performed using the same databases (Eubacteria, Archaea and Eukarya) as the Nucleos'ID search. Each set of nucleosides from the Nucleos'ID search database was counted as 1 true positive, true negative, false positive or false negative. When using all three databases for the search, a total of 119 sets were considered. When using only the Eubacteria database for the search, a total of 69 sets were considered. As only the manual interpretation enables the discrimination of some isomers with similar fragmentation pattern (based on electrophoretic mobilities comparison), only the identifications of the sets are compared with the Nucleos'ID search.

The True Positive Rate (TPR), also called sensitivity, indicates the probability that an actual nucleoside present in the sample will be correctly identified during the Nucleoside search. To avoid missing a positive result, TPR must be as high as possible. TPR was calculated with the Equation 8.1 :

$$TPR = \frac{TP}{TP + FN} \quad (8.1)$$

The True Negative Rate (TNR), also called specificity, indicates the probability that a nucleoside which is not present in the sample will correctly be undetected with the Nucleos'ID search. TNR was calculated with the Equation 8.2:

$$TNR = \frac{TN}{TN + FP} \quad (8.2)$$

Both TPR and TNR take on values between 0 and 1. A TPR equal to 1 means that all the nucleosides present in the sample have been identified by the software. Similarly, a TNR equal to 1 means that all the nucleosides that are not present in the sample have not been identified by the Nucleos'ID search.

The False Discovery Rate (FDR) indicates the proportion of false positives among all significant results identified by Nucleos'ID. Usually, in proteomics and transcriptomics, the FDR is calculated using a decoy sequence database [46]. However, for the identification at the nucleoside level, no decoy database can be constructed. Therefore, we decided, as mentioned before, to use only the numbers of false positives and true positives from the selected database. The FDR was calculated with the Equation 8.3:

$$FDR = \frac{FP}{FP + TP} \quad (8.3)$$

The FDR also takes on values between 0 and 1. A FDR equal to 0 means that all the nucleosides identified by the software are present in the sample. And if, for instance, a FDR equal to 0.10, 10 % of the nucleosides identified by Nucleos'ID are false positives and thus, not present in the sample.

To summarize, high TPR and TNR, and low FDR, indicate a high level of confidence in the results obtained with the Nucleos'ID search.

8.2.4. Results and discussion

8.2.4.1. Evaluation on a noncomplex sample: identification of post-transcriptional modifications from *S. cerevisiae* tRNA^{Phe GAA}.

The output file of the Nucleos'ID search is presented in Table 8.3. Table 8.3: Output CSV file open in excel for the Nucleos'ID search of *S. cerevisiae* tRNAPhe^{GAA}. To evaluate the results obtained from Nucleos'ID search, all expected and unexpected nucleosides were checked with manual interpretation. The electropherogram obtained with manual interpretation is represented in Figure 8.3.

Table 8.3: Output CSV file open in excel for the Nucleos'ID search of *S. cerevisiae* tRNA^{Phe GAA}.

Modification	Observed MS	Theoretical MS	Observed MS/MS	Theoretical MS/MS	Relative intensity (%)	Detection time (min)
A	268.1	268.1	136	136	100.0	16.5
C	244.09	244.09	112	112	100.0	16.0
Cm	258.11	258.11	112	112	22.6	16.3
cnm5U	284.1	284.09	152	152	39.4	22.0
cnm5U	284.1	284.09	152	152	100.0	22.1
G	284.1	284.1	152	152	100.0	22.1
Gm	298.11	298.11	152	152	100.0	22.6
m1A/m2A/m6A/m8A	282.12	282.12	150	150	100.0	16.4
m1G/m2G/m7G	298.11	298.11	166	166	100.0	16.3
m1G/m2G/m7G	298.11	298.11	166	166	100.0	21.1
m2,2G/m2,7G/preQ1	312.13	312.13	180	180	100.0	20.5
m3C/m4C/m5C	258.11	258.11	126	126	33.6	16.3
yW	509.19	509.2	377	377	100.0	20.6

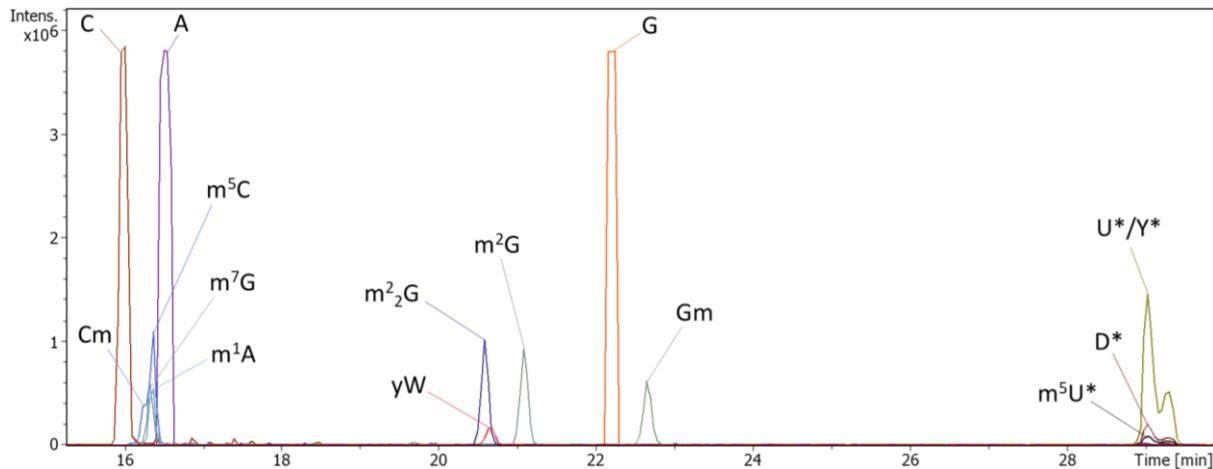


Figure 8.3 : Multiple extracted ion electropherograms for the nucleosides analysis of tRNA^{Phe GAA} from *S. cerevisiae* obtained after manual interpretation. Nucleosides annotated with * are expected nucleosides that were only detected by MS as they were not selected for fragmentation.

First, both interpretation methods, Nucleos'ID search and manual interpretation, identified the 4 canonical nucleosides and 7 expected sets of post-transcriptional modifications (Table 8.4). Those results are considered as true positives. Second, the expected modifications identified manually but not found by the software are modifications that were not selected for fragmentation due to low MS relative intensity. As no MS/MS spectrum is available in the input MGF file, it is therefore normal that they do not appear in the Nucleos'ID search results. Thus, those 4 modifications are not false negatives and can be considered as true negatives.

Table 8.4: Identification after manual interpretation and Nucleos'ID search of expected nucleosides *S. cerevisiae* tRNA^{Phe GAA}.

The symbol ✓ shows that the identification was based on MS and MS/MS data. The symbol ≈ shows that the manual identification was only based on MS due to no selection of those nucleosides for MS/MS fragmentation. Nucleosides identified solely on MS are not considered as valid as they cannot be differentiated from their isomers. The symbol ✗ shows that the nucleoside was not identified.

	A	C	G	U	m ¹ A	m ⁵ C	Cm	m ² G / m ⁷ G	Gm	m ^{2,2} G	yW	Y	D	m ⁵ U
Manual	✓	✓	✓	≈	✓	✓	✓	✓	✓	✓	✓	≈	≈	≈
Nucleos'ID	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗

Finally, only one additional modification, cnm^5U , was found by Nucleos'ID when using 0.02 Da of mass tolerance (Table 8.3) but was not identified during the manual interpretation (Figure 8.3). cnm^5U was not expected as it is not annotated on the sequence of *S. cerevisiae* tRNA^{Phe} GAA (Figure S8.1) referenced in the Modomics database [32]. The modification cnm^5U was identified at a detection time of 22.1 min (1327.72 s) which is the same detection time of G. The mass differences between the theoretical precursor ions, and between the theoretical product ions, of cnm^5U and G is respectively below 0.02 Da and 0.5 Da (0.011233 Da for both to be exact). Thus, with MS and MS/MS mass tolerances of respectively 0.02 Da and 0.5 Da, Nucleos'ID software does not differentiate the two nucleosides. With an MS and MS/MS mass tolerances set at 0.02 Da and 0.5 Da, cnm^5U is then considered as a false positive. For instance, setting the MS mass tolerance below 0.01 Da could have been useful to enable differentiation of both nucleosides.

With a total of 10 true positives, 108 true negatives, 1 false positive and 0 false negative, the TPR of this Nucleos'ID search is 1.00, the TNR is 0.99 and the FDR is 0.09 with the default parameters. These excellent rates indicate great confidence in the results obtained by the software and show that a high level of confidence can be given to these results. To reduce the number of false positives, the search parameters can be adjusted depending on the performances of the mass spectrometer. For instance, lowering the mass tolerance to less than 0.01 Da could withdraw the false positive cnm^5U and thus a true negative rate of 1.00 could be achieved.

*8.2.4.2. Increase in sample complexity: identification of post-transcriptional modifications from *B. taurus* total tRNA extract*

As in the previous example, to evaluate the results obtained from Nucleos'ID search (output file from Nucleos'ID search available in supporting information in Table S8.2), all expected and unexpected nucleosides were checked with manual interpretation (the electropherogram obtained with manual interpretation is available in supporting information in Figure S8.4). The comparison from both interpretation methods is summarized in Table 8.5 for expected modifications, and in Table 8.6 for unexpected modifications found by Nucleos'ID.

Among the expected unmodified and modified nucleosides for *B. taurus* total tRNA extract, 19 sets of nucleosides have been identified manually and with the Nucleos'ID search. Those results are considered as true positives. One of those modifications (I) was detected at two different times (Table S8.2) by Nucleos'ID despite having no expected isomer. The second one is a true positive and the first one is due to the first isotope of A. The mass differences between the theoretical precursor ions, and between the theoretical product ions, of I and the first isotope of A are respectively below 0.02 Da and 0.5 Da. As I was truly detected manually at the second detection time, this modification is not considered as a false positive.

Chapitre 8 : Développement de deux logiciels pour la caractérisation des ARN

Table 8.5: Identification after manual interpretation and Nucleos'ID search of expected nucleosides in B. taurus total tRNA extract. The symbol ✓ shows that the identification was based on MS and MS/MS data. The symbol ≈ shows that the manual identification was only based on MS due to no selection for MS/MS fragmentation. Nucleosides identified solely on MS are not considered as valid as they can not be differentiated from their isomers. The symbol ✗ shows that the nucleoside was not identified.

	A	C	G	U	m^1A	t^6A	i^6A	-	m^3C / m^5C	Cm	ac^4C	f^5Cm	$m^1G / m^2G / m^7G$	Gm	$m^{2,2}G$	Q	galQ/manQ	O_2yW	Y	D	m^5U	Um	Ym	m^5Um	mcm^5U	mcm^5s^2U	acp^3U
Manual	✓	✓	≈	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	≈	✓	✓	✓	≈	≈	≈	✓
Nucleos'ID	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓	✓	✗	✗	✗	✓

*Table 8.6: Identification after manual interpretation and Nucleos'ID search of unexpected nucleosides in B. taurus total tRNA extract. The symbol ✓ shows that the identification was based on MS and MS/MS data. The symbol ✗ shows that the nucleoside was not identified manually. The symbol ? indicates an ambiguous case that is considered as not valid. *The number of detected isomers is based on the manual interpretation.*

Nucleoside full name	Abbreviation	Nucleos'ID search	Manual search	Number of detected isomers*
5-hydroxymethylcytidine / 5-aminomethyluridine	hm^5C / nm^5U	✓	✗	✗
2'-O-methyl-5-hydroxymethylcytidine	hm^5Cm	✓	✓	1
2,8-dimethyladenosine/ N6,N6-dimethyladenosine	$m^{2,8}A / m^{6,6}A$	✓	✓	1
2-methylthio-N6-methyladenosine	ms^2m^6A	✓	✓	1
N6-acetyladenosine	ac^6A	✓	✓	1
N6-hydroxynorvalylcarbamoyladenosine / N6-methyl-N6-threonylcarbamoyladenosine	hn^6A / m^6t^6A	✓	✓	1
N6-hydroxymethyladenosine	hm^6A	✓	✓	1
5-carboxymethylaminomethyl-2-thiouridine	$cmnm^5s^2U$	✓	✗	✗
1-methylpseudouridine / 3-methylpseudouridine	m^1Y / m^3Y	✓	✓	1
3-(3-amino-3-carboxypropyl)pseudouridine	acp^3Y	✓	?	1
3-(3-amino-3-carboxypropyl)-5,6-dihydrouridine	acp^3D	✓	✓	1

In addition, 2 expected modifications (f^5Cm and o_2yW) were not identified either manually or by the software. This can be explained either by a degradation of RNA during storage, or by their low concentration in the sample which could be below the detection limit. Indeed, on the one hand, f^5Cm is only present in tRNA^{Leu CAA} which is already low abundant, and this modification can be partial meaning that a portion of this specific position could also be modified with Cm. On the other hand, the 37th position of tRNA^{Phe GAA} must be modified with yW or one of its derivatives, such as o_2yW . During manual interpretation, yW was found with very low MS signal intensity and thus, was not selected for fragmentation. As those modifications were not detected at all with Nucleos'ID and manual searches, they are considered as true negatives.

Only 1 expected modification (Ym) was identified manually but not identified with the Nucleos'ID search. This is due to the low relative intensity of its MS/MS fragment due to the co-migration with many nucleosides, including isomeric ones. Indeed, when several isomeric nucleosides are co-migrating, all isomers are selected for fragmentation at the same time and combined MS/MS spectra are obtained. Matching of several isomers to one combined MS/MS spectrum is possible, however it is potentially leading to low MS/MS relative intensity, thus low score, for some of the isomers. In the case of Ym, as its MS/MS relative intensity (3.2 %) is lower than the default score threshold (20 %), this result was withdrawn from the output file during the results sorting. This modification is considered as a false negative.

The remaining 5 expected nucleosides were not identified by the software since they were not selected for MS/MS fragmentation. As mentioned in the previous example, these modifications are also considered as true negatives.

The Table 8.6 shows that 11 sets of unexpected modifications have been identified by the Nucleos'ID search. Among those unexpected modifications, 8 sets were also found manually and are considered as true positives. The identification of unexpected modifications can be explained, either by the low amount of rRNA in the tRNA extract despite purification, or by potential missing tRNA sequences for *B. taurus* in the Modomics database [32]. Two other sets of unexpected modifications (hm^5C/nm^5U and $cmnm^5s^2U$) have not been identified manually as their MS/MS intensities were too low to be considered as a real signal, compared to the unusual intensity of the background noise in those MS/MS spectra.

Finally, 1 modification (acp^3Y) was identified by the Nucleos'ID search. However, manual interpretation indicated that this is an ambiguous case. Indeed, according to the Modomics database [32], acp^3Y and acp^3U share the same *m/z* for their precursor ions, as well as for the theoretical product ion. The modification acp^3Y has only one theoretical product ion, the one in common, while acp^3U has additional theoretical product ions. Those additional product ions have been detected during the CE-MS/MS analysis. Therefore, there is no confident information leading to an unambiguous identification of acp^3Y . Due to this ambiguity, acp^3Y was considered as false positive.

With a total of 27 true positives, 88 true negatives, 3 false positive and 1 false negative, the TPR of this Nucleos'ID search is 0.96, the TNR is 0.97 and the FDR is 0.10. Although the sample is more complex than the previous *S. cerevisiae* tRNA^{Phe} GAA (more numerous and diverse modifications), the TPR, TNR and FDR are still very good. Once again, search parameters can be optimized to minimize the probability of obtaining false positives. For instance, in this case, the absolute intensity threshold, that is set to 0 by default, can be a powerful tool to eliminate some false positives. However, if the absolute intensity threshold is too high, it could lead to more false negatives.

To assess the suitability of the default score threshold (set at 20 %), the same research was performed using thresholds from 0 to 100 %. The number of true positives, false positives, true negatives and false negatives, as well as the TPR, TNR and FDR obtained with these different thresholds, are represented in Figure 8.4.

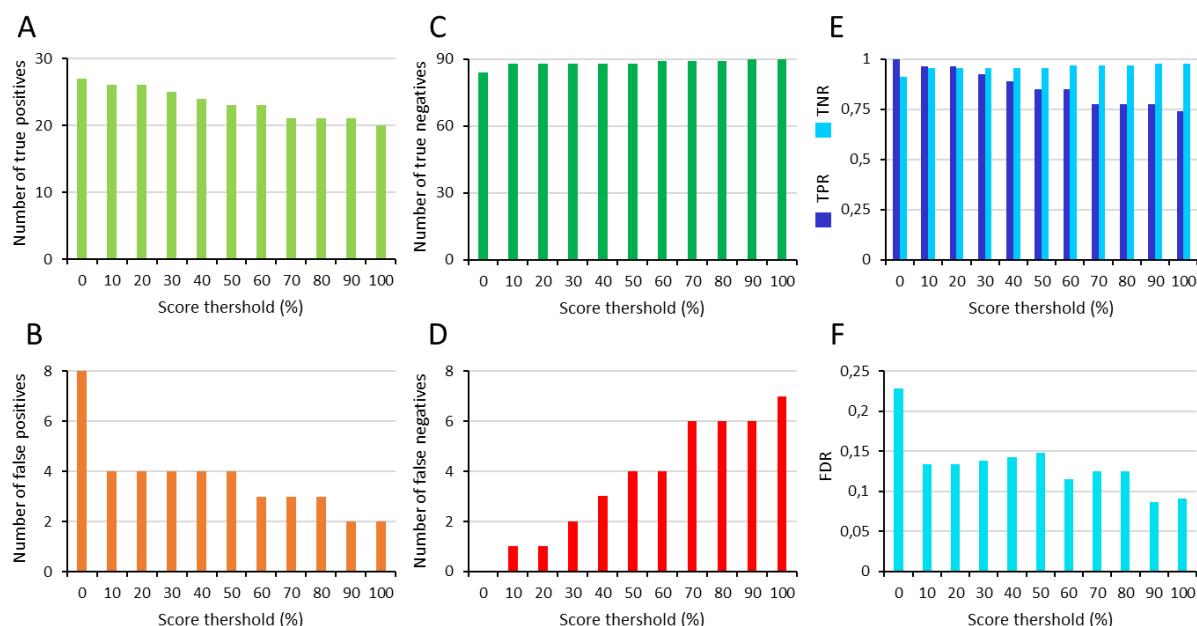


Figure 8.4: Evolution of the numbers of (A) true positives, (B) false positives, (C) true negatives, (D) false negatives, as well as (E) the TPR, the TNR and (F) the FDR, depending on the score threshold used for the Nucleos'ID search for *B. taurus* total tRNA extract.

With a score threshold at 0 %, there is 0 false negative (Figure 8.4D) and the number of true positives is the highest (Figure 8.4A), meaning that all nucleosides that should be found are identified. However, the number of false positives, and thus the FDR, are quite high (Figure 8.4B and Figure 8.4F). On the contrary, with a score threshold at 100 %, the number of true negatives is the highest (Figure 4C), there are fewer false positives (2 instead of 7) and the FDR is way lower (0.09 instead of 0.20). However, the number of true positives is much lower (20 instead of 28), and there are many more false negatives (8 instead of 0). This means that results found during the Nucleos'ID search are unlikely to be wrong, but some nucleosides will be missed by the software. Thus, finding the right score threshold is all about compromises.

To identify the best compromise, the score threshold should provide the best TPR and TNR. Score thresholds of 10 % and 20 % gave the highest TPR and TNR (Figure 8.4E). Thus, thresholds of 10 % or 20 % are the best compromise. As it appeared to be a good choice for many samples, the default score threshold was set to 20 %.

8.2.4.3. Untargeted identification of RNA post-transcriptional modifications of *P. aeruginosa* 70S ribosome

To increase sample complexity after the first results evaluation, Nucleos'ID search engine was used to identify RNA post-transcriptional modifications of *P. aeruginosa* 70S ribosome. Indeed, only well described samples were previously used to evaluate the software, whereas RNA modifications from *P. aeruginosa* 70S ribosome are not described in the literature yet.

As *P. aeruginosa* is part of the Eubacteria domain, an untargeted search was performed on the Eubacteria database. In order to avoid missing any modification, the score threshold was set to 0 %, and all results obtained in the output file (available in Supporting Information in Table S8.3), were compared to manual search (the electropherogram obtained with manual interpretation is available in supporting information in Figure S8.5). No additional nucleosides were found manually, thus no false negative was obtained. All modified and non-modified nucleosides found by the Nucleos'ID search are summarized in Table 8.7.

In addition to the 4 canonical nucleosides, 11 sets of modifications detected by the software were confirmed by manual interpretation. All of them were considered as true positives. The last three modifications (m^3Y , cmm^5Um and cmo^5U) were considered as false positives as their product ions absolute intensities were too low to be considered as a real signal and not as background noise.

Even though, RNA modifications in *P. aeruginosa* 70S ribosome have not been described yet, those of others bacterial species, such as *E. coli*, have already been reported in the literature [47,48]. As species from the same domain, here the Eubacteria, usually show similarities in their RNA post-transcriptional modifications, it is interesting to compare our results to those from *E. coli* ribosome.

All the identified sets of modifications except Q, ncm^5U and acp^3U are similar to those described in *E. coli* 16S or 23S rRNA [48]. Only three additional modifications (D, Um and Cm) are reported in *E. coli*. D was not identified with MS/MS due the very intense signal of U. Um and Cm were not detected at all in our analysis of *P. aeruginosa* 70S ribosome.

Finally, it is important to acknowledge that Q, ncm^5U and acp^3U were detected with very low signal intensities in *P. aeruginosa* 70S ribosome, and are usually found in bacterial tRNA, but not in bacterial rRNA. As ribosome is where the translation of proteins takes place,

some tRNA can be trapped between both 30S and 50S ribosome subunits during the purification process, leading to a very small proportion of tRNA in the sample.

Table 8.7: Manual verification of all nucleosides found by Nucleos'ID search in *P. aeruginosa* 70S ribosome. The symbol ✓ shows that the identification was based on MS and MS/MS data. The symbol ✗ shows that the nucleoside was not identified manually. *The number of detected isomers is based on the manual interpretation.

Nucleosides full names	Abbreviations	Nucleos'ID score (%)	Manual search	Number of detected isomers*
Adenosine	A	100	✓	1
Cytidine	C	100	✓	1
Guanosine	G	100	✓	1
Uridine	U	100	✓	1
1-methyladenosine / 2-methyladenosine / N6-methyladenosine / 8-methyladenosine	m ¹ A / m ² A / m ⁶ A / m ⁸ A	100	✓	2
2,8-dimethyladenosine / N6,N6-dimethyladenosine	m ^{2,8} A / m ^{6,6} A	100	✓	1
3-methylcytidine / N4-methylcytidine / 5-methylcytidine	m ³ C / m ⁴ C / m ⁵ C	100	✓	1
N4,2'-O-dimethylcytidine	m ⁴ Cm	100	✓	1
1-methylguanosine / N2-methylguanosine / 7-methylguanosine	m ¹ G / m ² G / m ⁷ G	100	✓	2
2'-O-methylguanosine	Gm	100	✓	1
Queuosine	Q	100	✓	1
3-methyluridine / 5-methyluridine	m ³ U / m ⁵ U	100	✓	1
Pseudouridine	Y	43	✓	1
Methoxycarbonylmethyluridine	ncm ⁵ U	90	✓	1
3-(3-amino-3-carboxypropyl)uridine	acp ³ U	100	✓	1
5-carboxymethylaminomethyl-2'-O-methyluridine	cmm ⁵ Um	55	✗	✗
Uridine 5-oxyacetic acid	cmo ⁵ U	6	✗	✗

With a total of 15 true positives, 51 true negatives, 3 false positives and 0 false negative, the TPR of this Nucleos'ID search is 1.00, the TNR is 0.94 and the FDR is 0.17. Once again, the TPR and TNR are very good, and the FDR is also quite good considering the relatively small size of the database.

As mentioned before, search parameters could be optimized to reduce the number of false positives or to increase the number of true negatives. Nevertheless, it is important to remember that, while parameter optimization can be useful in reducing the number of false positives, it can also lead to a loss of information and an increase of false negative. For instance, for the *P. aeruginosa* 70S ribosome, as none of the true positives has a score below

20%, lower FDR could have been achieved by using the default score threshold. However, using a very high score threshold of 75 % would lead to three less false positives (m^3Y , $cmmn^5Um$ and cmo^5U), but one true positive (Y) would be lost. The right compromise must be found by the user depending on the performance of the mass spectrometer, on nature and knowledge of the sample, but also on the aim of the Nucleos'ID search. For this sample, the aim of the Nucleos'ID search was to save time in the data interpretation before manual checking the results to avoid any missed modifications.

The great results obtained for all three samples demonstrate the strong potential of the Nucleos'ID software for the untargeted identification of post-transcriptional modifications after MS/MS analyses. Indeed, from several hours for manual interpretation if all modifications from the three databases are searched, to less than 1 s for the Nucleos'ID search engine, the gain in time is impressive, as for instance, results showed in this article are based on MGF files containing from around 500 spectra to more than 15 600 spectra. In case the user wants to manually check the Nucleos'ID results, only a few more minutes are necessary, as the number of potential modifications was significantly reduced compared to the full database.

8.2.5. Conclusion

To conclude, a new software enabling untargeted identification of RNA modifications from MS/MS data was developed and is freely available for CE-MS/MS or LC-MS/MS users. Nucleos'ID was evaluated on two well-known samples and good TPR, TNR and FDR were achieved. The evaluation was performed using default settings, however, search settings can be tuned, with common sense, to match the mass spectrometer performances and the aim of the study. Nucleos'ID was then used to characterize *P. aeruginosa* 70S ribosome modifications that were not described in the literature yet. In addition to the 4 canonical nucleosides, 11 sets of modifications have been successfully identified. As the results are displayed instantly when the Nucleos'ID search is performed, it is a timesaving tool compared to several hours of untargeted manual interpretation. In addition to providing structural information on RNA, the use of this new software is undeniably a time saver for the identification of RNA modifications, especially when the sample is not yet described in the literature, even with a manual verification following the Nucleos'ID search.

8.2.6. Code availability

Source code for Nucleos'ID is available on Github (<https://github.com/MSARN/NucleosID>) under Apache 2 license.

8.2.7. Supplemental information

Table S8.1: Corresponding short names and full names of all nucleosides cited in this study.

Nucleoside short name	Nucleoside full name
A	Adenosine
C	Cytidine
G	Guanosine
U	Uridine
m ¹ A	1-methyladenosine
m ² A	2-methyladenosine
m ⁶ A	N6-methyladenosine
m ^{2,8} A	2,8-dimethyladenosine
m ^{6,6} A	N6,N6-dimethyladenosine
hm ⁶ A	N6-hydroxymethyladenosine
hn ⁶ A	N6-hydroxynorvalylcarbamoyladenosine
ms ² m ⁶ A	2-methylthio-N6-methyladenosine
t ⁶ A	N6-threonylcarbamoyl-adnosine
m ⁶ t ⁶ A	N6-methyl-N6-threonylcarbamoyladenosine
i ⁶ A	N6-isopentenyladenosine
I	Inosine
ac ⁶ A	N6-acetyladenosine
Cm	2'-O-methylcytidine
m ³ C	3-methylcytidine
m ⁴ C	N4-methylcytidine
m ⁵ C	5-methylcytidine
m ⁴ Cm	N4,2'-O-dimethylcytidine
ac ⁴ C	N4-acetylcytidine
f ⁵ Cm	5-formyl-2'-O-methylcytidine
hm ⁵ C	5-hydroxymethylcytidine
hm ⁵ Cm	2'-O-methyl-5-hydroxymethylcytidine
Gm	2'-O-methylguanosine
m ¹ G	1-methylguanosine
m ² G	N2-methylguanosine
m ⁷ G	7-methylguanosine
m ^{2,2} G	N2,N2-dimethylguanosine
Q	Queuosine
galQ	Galactosyl-queuosine
manQ	Mannosyl-queuosine
yW	Wybutosine
o ₂ yW	Peroxywybutosine
Y	Pseudouridine
Um	2'-O-methyluridine
m ³ U	3-methyluridine
m ⁵ U	5-methyluridine
m ⁵ Um	5,2'-O-dimethyluridine
mcm ⁵ U	Methoxycarbonylmethyluridine
mcm ⁵ s ² U	5-methoxycarbonylmethyl-2-thiouridine
acp ³ U	3-(3-amino-3-carboxypropyl)uridine
ncm ⁵ U	Methoxycarbonylmethyluridine
nm ⁵ U	5-aminomethyluridine
cnm ⁵ U	cyanomethyluridine
cmnm ⁵ s ² U	5-carboxymethylaminomethyl-2-thiouridine
cmnm ⁵ Um	5-carboxymethylaminomethyl-2'-O-methyluridine
cmo ⁵ U	Uridine 5-oxyacetic acid
Y	Pseudouridine
m ¹ Y	1-methylpseudouridine
m ³ Y	1-methylpseudouridine
acp ³ Y	3-(3-amino-3-carboxypropyl)pseudouridine
D	Dihydrouridine
acp ³ D	3-(3-amino-3-carboxypropyl)-5,6-dihydrouridine

ARNt^{Phe GAA} G C G G A U U U A [m²G] C U C A G [D] [D] G G G A G A G C [m²G] C C A G A [Cm] U [Gm] A A [yW] A [Y] [m⁵C] U G G A G [m⁷G] U C [m⁵C] U G U G [m⁵U] [Y] C G [m¹A] U C C A C A G A A U U C G C A C C A

Figure S8.1: *S. cerevisiae* tRNA^{Phe GAA} sequence referenced in the Modomics database [32].

ARNt^{Arg CGG}	G A C C C A G U [m ¹ G] [m ² G] C C U A A [D] [Gm] G A [D] A A G G C A [Y] C A G C [Cm] U C C G [m ¹ G] A G C U G G G G A [D] U G [Y] G G G [m ⁵ U] [Y] C G [m ¹ A] G U C C C A U C U G G G U C G C C A
ARNt^{Arg CCU}	G C C C C A G U [m ¹ G] [m ² G] C C U A A [D] G G A [D] A A G G C A [Y] U G G C [m ³ C] U C C U [t ⁶ A] A G C C A G G G A [D] U G [Y] G G G [m ⁵ U] [Y] C G [m ¹ A] G U C C C A U C U G G G G [Y] C C C A
ARNt^{Arg [mcm⁵s²U]C[Ψ]}	G G C U C C G U [m ¹ G] G C G C A A [D] G G A [D] A G C G C A [Y] [Y] G G A [m ³ C] U [mcm ⁵ s ² U] C [Y] [t ⁶ A] A U [Y] C A A A G [m ⁷ G] [D] U [m ⁵ C] C G G G [m ⁵ U] [Y] C G [m ¹ A] G U C C C G G C G G A G U C G C C A
ARNt^{Asn [Q]UU}	G U C U C U G U [m ¹ G] [m ² G] C G C A A [D] C G G [D] X A G C G C [m ² , ² G] [Y] [Y] C G G C U [Q] U U [t ⁶ A] A C C G A A A G [m ⁷ G] [D] U G G U G G [Q] [Y] C G [m ¹ A] G C C C A C C C A G G G A C G C C A
ARNt^{Asp [manQ]UC}	G G U G C C G U A [m ² G] C G [Y] A G [D] [Gm] G C A [.] C G [.] G A C U C U [Cm] U [manQ] U C A A A G A G U G G [m ⁷ G] A [m ⁵ C] [.] U G A G [m ⁵ U] [Y] C G [m ¹ A] U A C U C A C G G A C C G C C A
ARNt^{Gln CUG}	G G U U C C A U [m ¹ G] G U G [Y] A A [D] [Gm] G [D] [D] A G C A C U C [Y] G G A [Cm] U C U G A A [Y] C C A G C G A U [m ⁵ C] C G A G [Y] [Y] C A [m ¹ A] A U C U C G G U G G A A C C U C C A
ARNt^{Leu [I]AG}	G G U A G C G U G [m ² G] C [ac ⁴ C] G A G C G G [D] C [Y] A A G G C [m ² , ² G] C U G G A [Ym] U [I] A G [m ¹ G] C [Y] C C A G U C [Y] C [Y] U C G G G G G [m ⁵ C] G U G G G [m ⁵ U] [Y] C G [m ¹ A] A U C C C A C C G C U G G C C A C C A
ARNt^{Leu [f⁵Cm]AA}	G U C A G [m ² G] A U G [m ² G] C [ac ⁴ C] G A G U G G [D] C [Y] A A G G C [m ² G] C C A G A C U [f ⁵ Cm] A A [m ¹ G] [Y] [Y] C U G G [Um] C [Y] C C G U A U G G A G G [m ⁵ C] G U G G G [m ⁵ U] [Y] C G [m ¹ A] A U C C C A C C A C U U C U C U G A C A C C A
ARNt^{Leu CAG}	G U C A G [m ² G] A U G [m ² G] C [ac ⁴ C] G A G C G G [D] [D] C A A G G C [m ² , ² G] C U G C G [Um] U C A G [m ¹ G] [.] C G C A G [Um] C [Y] C C C U G G A G G [m ⁵ C] G U G G G [m ⁵ U] [Y] C G [m ¹ A] A U C C C A C U U C U G A C A C C A
ARNt^{Phe [Gm]AA}	G C C G A A A U A [m ² G] C U C [m ¹ A] G [D] [D] G G G A G A G C [m ² , ² G] [Y] [Y] A G A [Cm] U [Gm] A A [o ₂ yW] A [Y] C U A A A G [m ⁷ G] [D] C [m ⁵ C] C U G G [m ⁵ U] [Y] C G [m ¹ A] U C C C G G U U U C G G C A C C A
ARNt^{Sec [mcm⁵U]CA}	G C C C G G A U G A C C C U C A G U G G U C U G G G G U G C A G G C U [mcm ⁵ U] C A [i ⁶ A] A C C U G U A G C U G U C U A G C G A C A G A G U G G U [Y] C A [m ¹ A] U U C C A C C U U U C G G G C G C C A
ARNt^{Thr [I]GU}	G G C C C U G U [m ¹ G] [m ² G] C U A G C [D] G G [D] C A A A G C [m ² , ² G] C C U G U [m ³ C] U [I] G U [t ⁶ A] A A C A G G A G A [D] [m ⁵ C] C U G G G U [Y] C G [m ¹ A] A U C C C A G C G G G G [m ⁵ C] U C C A
ARNt^{Trp [Cm]CA}	G A C C U C [m ² G] U [m ¹ G] [m ² G] C G C A A [D] [Gm] G [D] A G C G C [m ² , ² G] [Y] C U G A [Cm] U [Cm] C A [m ¹ G] A [Ym] C A G A A G [m ⁷ G] [D] U G C G U G [Y] [Y] C G [m ¹ A] A U C A C G U C G G G G U C A C C A
ARNt^{Tyr [galQ][Y]A}	C C U U C [m ² G] A U A [m ² G] C U C A G [D] [D] G G [acp ³ U] U A G A G C [m ² , ² G] [m ² , ² G] A G G A C U [galQ] [Y] A [m ¹ G] A [Ym] C C U U A G [m ⁷ G] [D] [m ⁵ C] G C U G G [m ⁵ U] [Y] C G [m ¹ A] U U C C G G C U C G A A G G A C C A
ARNt^{Lys [mcm⁵s²U]UU}	G C C C G [m ² G] A U A [m ² G] C U C A G [D] C G G [D] A G A G C A [Y] C A G A C U [mcm ⁵ s ² U] U U A [Y] C U G A G G [m ⁷ G] [D] [m ⁵ C] [m ⁵ C] A G G G [m ⁵ U] [Y] C A [m ¹ A] [m ¹ A] G U C C C U G U U C G G G G C G C C A

Figure S8.2: *B. taurus* tRNA sequences referenced in the Modomics database [32].

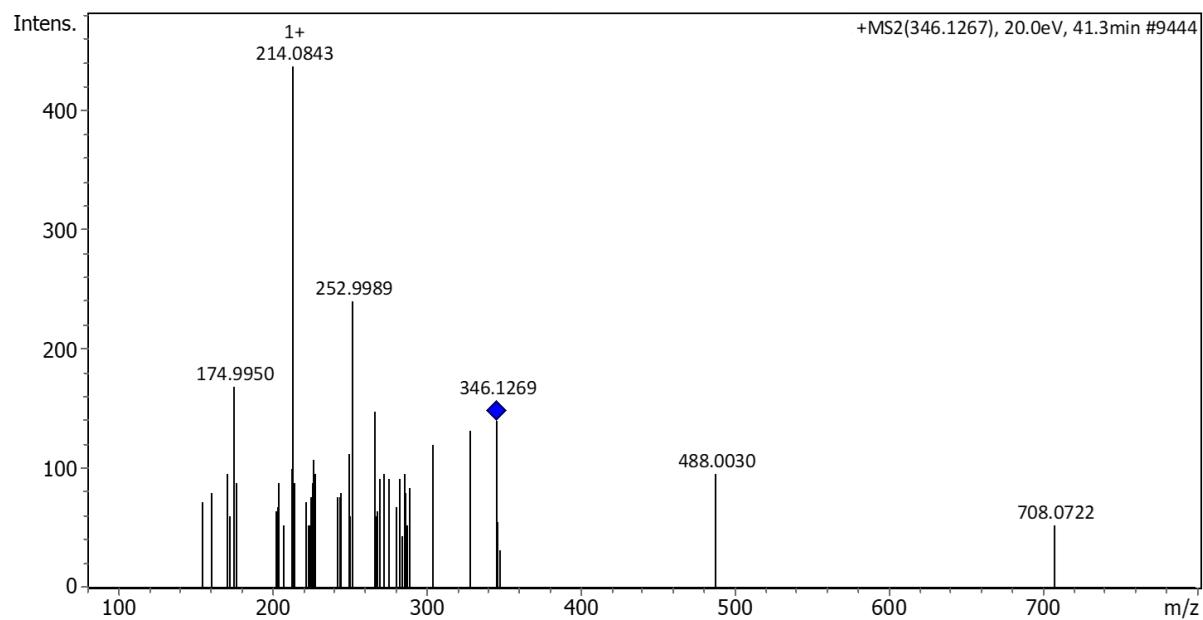


Figure S8.3: MS/MS spectrum with only background noise that can be used to estimate the MS/MS absolute intensity threshold.

Table S8.2: Output CSV file open in excel for the Nucleos'ID search of *B. taurus* total tRNA extract.

Modification	Observed MS (Da)	Theoretical MS (Da)	Observed MS/MS (Da)	Theoretical MS/MS (Da)	Score (%)	Detection time (s)
C	244.0956	244.0928	112.0524	112.050538	100	681.694
m3C/m4C/m5C	258.1068	258.1084	126.06667	126.066188	100	687.069
Cm	258.1069	258.1084	112.0512	112.050538	35.83	689.754
m1A/m2A/m6A/m8A	282.1118	282.1197	150.07841	150.077422	100	692.439
m1G/m2G/m7G	298.1121	298.1146	166.07251	166.072336	100	692.718
I	269.1057	269.088	137.06518	137.045787	77.85	697.812
A	268.1055	268.104	136.06356	136.061772	100	700.498
ac6A	310.1115	310.1146	178.07206	178.072336	100	703.742
hm5Cm	288.1165	288.119	142.06141	142.061103	25.55	708.835
m2,8A/m6,6A	296.1325	296.1353	164.09298	164.093072	81.72	713.928
hm5C/nm5U	274.1009	274.1034	142.06128	142.061103	59.61	733.781
Q	410.1607	410.167	163.06136;295.10073	163.061437;295.103696	100	736.188
i6A	336.1628	336.1666	136.0619;204.12366	204.124372;136.061772	100	738.874
galQ/manQ	572.2074	572.2198	163.06142;295.10063	440.177589;163.061437;295.103696	100	765.741
hm6A	298.1116	298.1146	136.06187;268.10169	268.10403;136.061772	100	779.169
m2,2G/m2,7G/preQ1	312.1279	312.1302	180.08844	180.087986	100	800.66
m1G/m2G/m7G	298.1133	298.1146	166.07362	166.072336	100	811.404
acp3D	348.1353	348.1401	216.09657	216.097882	65.8	817.055
acp3U	346.1204	346.1245	168.07645;197.055;214.08136	214.082232;197.055683;168.076753	100	822.149
acp3Y	346.1204	346.1245	214.08136	214.082232	100	822.149
ms2m6A	328.1101	328.1074	196.07094	196.065143	93.72	827.521
Gm	298.1121	298.1146	152.057	152.056686	100	846.323
cmm5s2U	348.0814	348.086	216.04134;254.99316	255.043404;273.053969;141.01171;216.043739;237.03284;225.03284	100	932.847
ac4C	286.1009	286.1034	112.05112;154.06103	154.061103;112.050538	100	934.976
t6A	413.1353	413.1415	281.09673	281.099279	100	951.091
I	269.086	269.088	137.04597	137.045787	100	953.776
U	245.0754	245.0768	113.03529	113.034554	100	964.521
Y	245.0754	245.0768	155.04527;179.04479;209.05485	209.055683;179.045119;155.045119	34.39	964.521
Um	259.0907	259.0925	113.03507	113.034554	100	967.207
m3U/m5U	259.0907	259.0925	127.05057	127.050204	98.94	967.207
m1Y/m3Y	259.0907	259.0925	169.06065;179.04432;209.05482;227.06469	169.060769;179.045119;227.066248;209.055683	42.98	967.207
hn6A/m6t6A	427.1504	427.1572	295.11197	295.114929	100	978.232

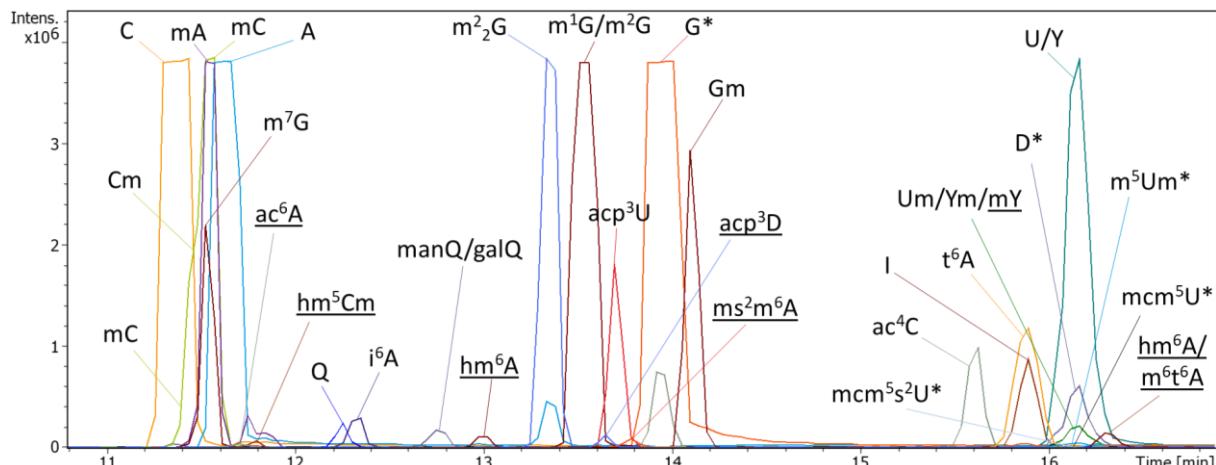


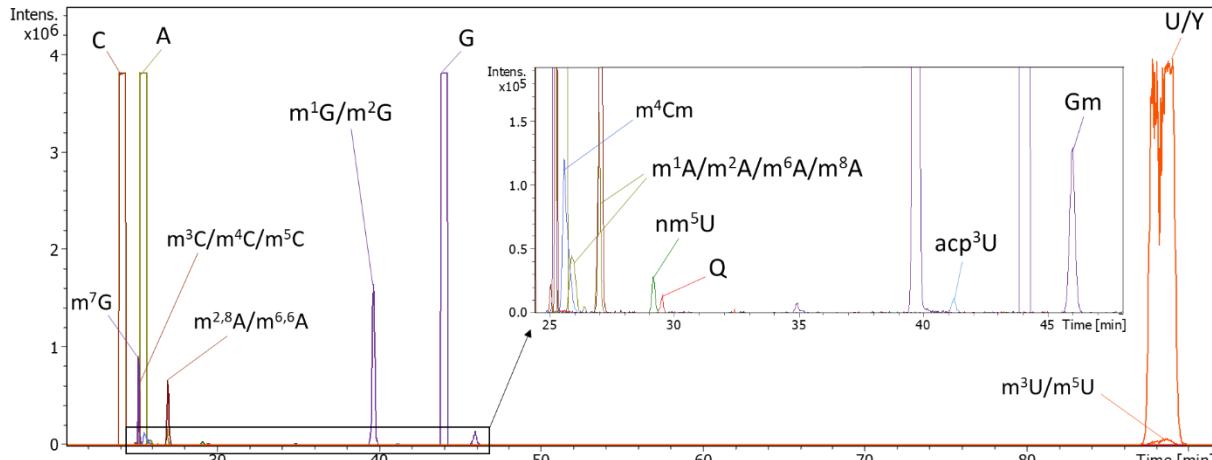
Figure S8.4: Multiple extracted ion electropherograms for the analysis of *B. taurus* total tRNA extract obtained after manual interpretation. Nucleosides annotated with * are expected nucleosides that were only detected by MS as they were not selected for fragmentation. Underlined nucleosides were not expected.

Migration times obtained using capillary electrophoresis are known not to be the most reproducible as they are easily subject to temperature change and capillary degradation. Electrophoretic mobilities were calculated based on the calculation of the electroosmotic flow of uridine which is a neutral marker. They were compared between the different analyses, and to some separately analyzed standards, to ensure correct identification of the analytes despite the variation in migration times.

Chapitre 8 : Développement de deux logiciels pour la caractérisation des ARN

*Table S8.3: Output CSV file open in excel for the Nucleos'ID search of *P. aeruginosa* 70S ribosome.*

Modification	Observed MS (Da)	Theoretical MS (Da)	Observed MS/MS (Da)	Theoretical MS/MS (Da)	Score (%)	Detection time (s)
C	244.0941	244.0928	112.04988	112.050538	100	1443.21
m1G/m2G/m7G	298.1164	298.1146	166.07267	166.072336	100	1518.69
m3C/m4C/m5C	258.1098	258.1084	126.06614	126.066188	100	1522.86
m4Cm	272.1251	272.1241	126.06508	126.066188	100	1533.1
A	268.1066	268.104	136.06188	136.061772	100	1553.58
m1A/m2A/m6A/m8A	282.1214	282.1197	150.07795	150.077422	100	1561.7
A	268.1055	268.104	136.06166	136.061772	100	1586.05
m1A/m2A/m6A/m8A	282.1213	282.1197	150.07758	150.077422	100	1622.25
m2,8A/m6,6A	296.1373	296.1353	164.0936	164.093072	100	1626.35
A	268.1057	268.104	136.06181	136.061772	100	1676.29
nm5U	274.105	274.1034	142.06102	142.061103	89.5	1755.36
A	268.1057	268.104	136.062	136.061772	100	1765.52
Q	410.1696	410.167	163.06169;295.10578	163.061437;295.103696	100	1771.74
Q	410.1758	410.167	295.11217	163.061437;295.103696	74.63	1848.83
A	268.1058	268.104	136.06204	136.061772	37.48	2275.19
m1G/m2G/m7G	298.1168	298.1146	166.07302	166.072336	100	2386
cmmn5Um	346.1267	346.1245	252.99893	271.092463;221.031874;253.058088;200.066582	55.05	2475.16
acp ³ U	346.1267	346.1245	214.08428	214.082232;197.055683;168.076753	100	2475.16
G	284.1032	284.0989	152.05852	152.056686	100	2634.31
G	284.1063	284.0989	152.0602	152.056686	100	2655.02
Gm	298.1163	298.1146	152.05687	152.056686	100	2759.02
A	268.1064	268.104	136.06093	136.061772	100	2783.01
G	284.101	284.0989	152.05729	152.056686	52.75	2881.95
G	284.1005	284.0989	152.05651	152.056686	32.75	2986.45
G	284.1006	284.0989	152.05687	152.056686	68.91	3158.77
A	268.1058	268.104	136.06235	136.061772	73.95	3420.95
G	284.1007	284.0989	152.05707	152.056686	70.94	4193.74
G	284.1006	284.0989	152.05685	152.056686	100	4287.2
A	268.1062	268.104	136.06184	136.061772	100	4422.87
G	284.1008	284.0989	152.05643	152.056686	53.32	4652.33
cmoSU	319.0725	319.0772	187.07018	187.034948;169.024383;141.029468	5.85	5239.41
G	284.1008	284.0989	152.05705	152.056686	100	5249.21
Y	245.0786	245.0768	179.04546;209.06085	209.055683;179.045119;155.045119	43.3	5249.35
U	245.0785	245.0768	113.03421	113.034554	100	5292.67
m3U/m5U	259.0939	259.0925	127.04975	127.050204	100	5310.65
m3Y	259.094	259.0925	169.06112	169.060769;179.045119;227.066248;209.055683	13.61	5314.75
U	245.0796	245.0768	113.03421	113.034554	100	5314.9
Y	245.0796	245.0768	155.0464;179.0447;209.05656	209.055683;179.045119;155.045119	2.23	5314.9
G	284.1009	284.0989	152.05627	152.056686	68.46	5415.1



*Figure S8.5: Multiple extracted ion electropherograms for the analysis of *P. aeruginosa* 70S ribosome obtained after manual interpretation.*

Migration times obtained using capillary electrophoresis are known not to be the most reproducible as they are easily subject to temperature change and capillary degradation. Electrophoretic mobilities were calculated based on the calculation of the electroosmotic flow of uridine which is a neutral marker. They were compared between the different analyses, and to some separately analyzed standards, to ensure correct identification of the analytes despite the variation in migration times.

8.3. OligoMod+ : nouveau logiciel pour localiser les modifications post-transcriptionnelles à partir d'analyses MS/MS d'oligonucléotides en mode positif

Comme cela a été montré dans le chapitre 7 portant sur l'analyse des modifications d'un ARN ribosomique, l'interprétation des données au niveau oligonucléotides est un réel défi lorsque les modifications post-transcriptionnelles, et notamment leurs positions, ne sont pas déjà décrites dans la littérature. Certains logiciels, tels que RNAModMapper [18], Ariadne [16], RoboOligo [17] ou NucleicAcidSearchEngine [20], ont été développés pour les analyses MS/MS ou LC-MS/MS en mode négatif, mais aucun logiciel ne permettait de facilement interpréter les analyses CE-MS/MS en mode positif. C'est avec l'objectif final de localiser les modifications non décrites dans la littérature qu'un second logiciel a été développé en collaboration avec des bio-informaticiens. Ce nouveau logiciel, appelé OligoMod+ permet, pour le moment, de vérifier que les modifications connues (nature de la modification et positions sur la séquence de l'ARN) sont bien retrouvées. L'interface de cette première version est présentée en Figure 8.5.

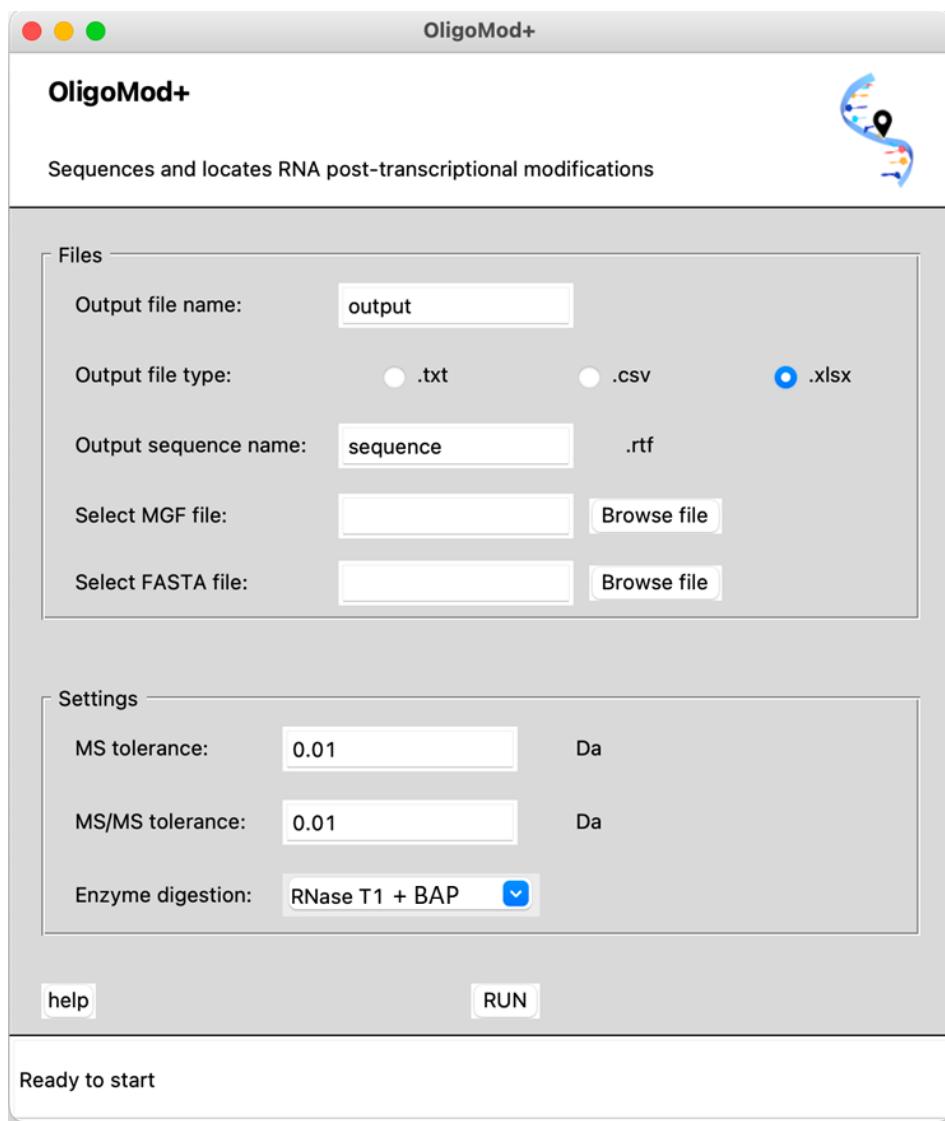


Figure 8.5 : Interface de la première version du logiciel OligoMod+.

Le fonctionnement de ce second logiciel, schématisé en Figure 8.6, est brièvement expliqué dans les prochains paragraphes.

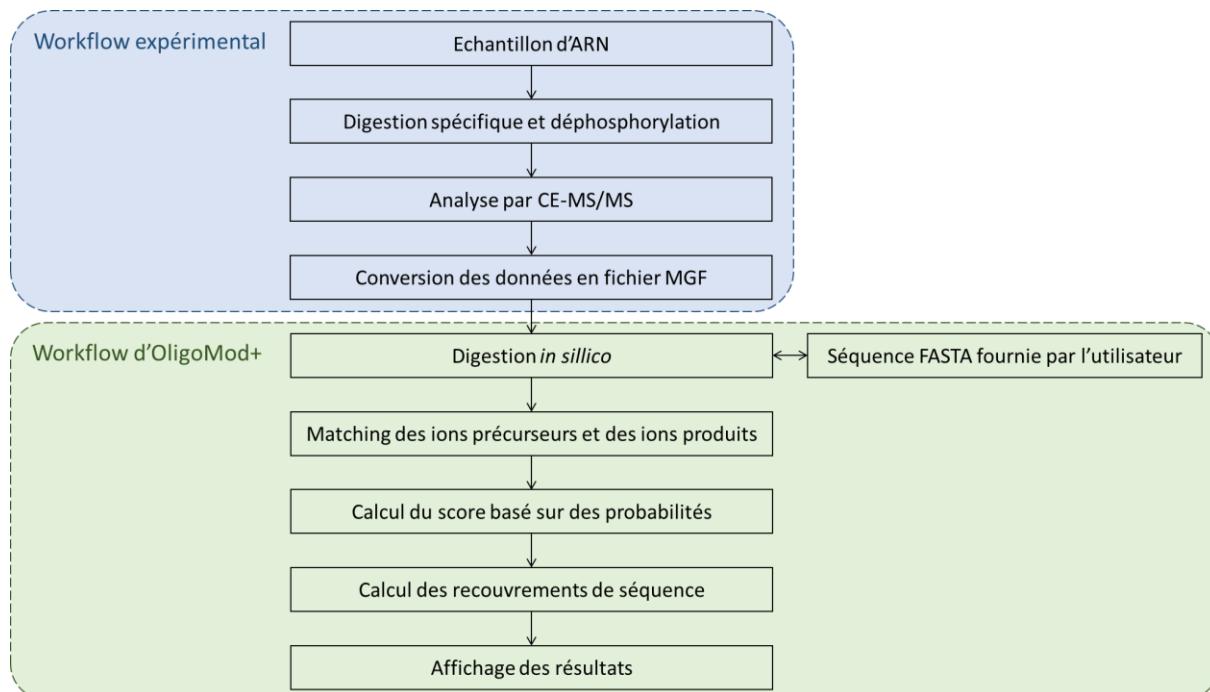


Figure 8.6 : Principe expérimental et fonctionnement du logiciel OligoMod+ pour la vérification des positions des modifications post-transcriptionnelles à partir d'analyse CE-MS/MS en mode positif.

À partir de la séquence FASTA, contenant les modifications connues, fournie par l'utilisateur, la digestion de l'ARN est simulée par OligoMod+ selon l'enzyme choisie. Actuellement, deux duos d'enzyme sont disponibles : la RNase T1 associée à la BAP, et la RNase A associée à la BAP. À partir de la digestion *in silico*, une base de données est créée pour l'échantillon étudié. Celle-ci contient les *m/z* théoriques des oligonucléotides simulés, ainsi que tous les fragments issus des séries d'ions c, y, w et a-B, qui sont favorisées lors de la fragmentation par CID. Les pertes de base anioniques et neutres (annotées M-B et M-BH) sont également référencées dans cette base de données. Les oligonucléotides sont identifiés par OligoMod+ après comparaison entre les spectres expérimentaux, MS et MS/MS, et cette base de données.

De la même manière que pour Nucleos'ID, chaque identification est associée à un score permettant de donner un niveau de confiance au résultat. Cependant, dans le cas d'OligoMod+, le système de score est basé sur ce qui a déjà été réalisé en protéomique (notamment dans le logiciel Peppy [42]), mais aussi dans les quelques logiciels en transcriptomiques [18,20] qui prennent en compte uniquement le mode négatif. Tout d'abord, la probabilité que le « *match* » entre l'oligonucléotide théorique et le spectre MS/MS ne soit pas dû uniquement au hasard est calculée. Ensuite, un deuxième calcul de probabilités est réalisé afin de prendre en compte le fait que les fragments, qui ont « *matchés* » pour l'identification de l'oligonucléotide, ne sont pas uniquement par hasard parmi les pics les plus intenses du spectre MS/MS. Les deux calculs de probabilités sont combinés pour obtenir un

unique score pour chaque identification. Le score est ensuite normalisé afin de ne pas dépendre de la longueur de l'oligonucléotide identifié.

Les résultats obtenus par OligoMod+ sont ensuite affichés sous deux formats. Le premier affichage des résultats est semblable à celui de Nucleos'ID. Il s'agit d'un tableau indiquant la séquence des oligonucléotides et leur emplacement dans la séquence de l'ARN, les différents *m/z* théoriques et expérimentaux permettant les identifications MS et MS/MS, le temps de détection et le score associé à chaque identification. Un deuxième affichage des résultats permet de mieux visualiser le recouvrement de séquence obtenu par la recherche. La séquence de l'ARN apparaît selon un code couleur (Figure 8.7).

```
>tRNA LE Phe GAA sequence
G CG G AUUUAL CUCAG DDG G G AG AG CRCCAG ABU#AAY AP?UG G AG 7UC?UG UG TPCG "UCCACAG AAUUCG CACCA
```

Figure 8.7 : Affichage du recouvrement de séquence d'une analyse CE-MS/MS de l'ARN^{tRNA Phe GAA} issu de *S. cerevisiae* et digéré à la RNase T1 et à la BAP. Les oligonucléotides identifiés par OligoMod+ et uniques dans la séquence de l'ARN sont indiqués en vert. Les oligonucléotides identifiés par OligoMod+, mais redondants sont indiqués en orange. Les oligonucléotides non identifiés par OligoMod+ sont indiqués en rouge.

De très bons résultats préliminaires ont été obtenus, puisqu'en combinant les deux digestions spécifiques (RNase T1+BAP et RNase A+BAP) pour l'ARN^{tRNA Phe GAA} issu de *S. cerevisiae* (dont les modifications sont bien décrites dans la littérature), un recouvrement total de séquence de 90% a été obtenu avec OligoMod+ en ne prenant en compte uniquement les oligonucléotides uniques. Ce recouvrement de séquence est à comparer aux 96% obtenus manuellement lors de la preuve de concept réalisée précédemment au laboratoire [12]. La différence observée est notamment due aux « *miscleavages* » au niveau de certaines hypermodifications, aux nucléosides et dinucléotides dont les fragments peuvent être difficiles à détecter et qui sont par conséquent peu identifiés par OligoMod+ contrairement à l'interprétation manuelle.

OligoMod+ n'étant pas finalisé, les résultats obtenus ne seront pas plus détaillés dans ce manuscrit. Cependant, il est important de noter le gain de temps considérable déjà observé lorsque les modifications sont connues. Lorsque la version 2 sera finalisée et prendra en compte la localisation des modifications non décrites dans la littérature, le gain de temps et l'apport d'OligoMod+ à la communauté scientifique sera d'autant plus appréciable, puisqu'à ce jour, il est impossible de localiser les modifications manuellement si les positions de celle-ci ne sont pas connues.

8.4. Conclusion

Afin de pallier le manque d'outils bio-informatiques pour caractériser les modifications post-transcriptionnelles des ARN à partir d'analyses MS/MS en mode positif, deux logiciels ont été développés, un pour chacun des deux niveaux d'analyses (nucléosides et oligonucléotides).

Le premier, Nucleos'ID, permet d'identifier les modifications post-transcriptionnelles de manière non-ciblée, à partir d'analyses par CE-MS/MS ou LC-MS/MS. Ce logiciel a été évalué sur deux échantillons dont les modifications sont bien décrites dans la littérature : un ARN^{tPhe} GAA issu de *S. cerevisiae* et un extrait total d'ARNt issu de *B. taurus*. D'excellents taux de vrais positifs (TPR, respectivement 1,00 et 0,96), vrais négatifs (TNR, respectivement 0,99 et 0,97) et faux positifs (FDR, respectivement 0,09 et 0,10) ont été obtenus sur ces deux recherches. Ensuite, Nucleos'ID a permis d'identifier 11 modifications dans un ribosome 70S de *P. aeruginosa*, composé de plusieurs ARNr longs de 150 à presque 3000 nucléotides. Malgré la grande complexité de l'échantillon, de très bons TPR (1,00), TNR (0,96) et FDR (0,12) ont de nouveau été obtenus. D'une interprétation manuelle non ciblée de plusieurs heures à une recherche à l'aide de Nucleos'ID qui affiche les résultats instantanément, le gain de temps apporté par ce nouvel outil bio-informatique est indéniable. Cela démontre le champ des possibles dans l'identification non ciblée des modifications post-transcriptionnelles des ARN, d'autant plus lorsque l'échantillon n'est pas encore décrit dans la littérature.

Le second logiciel, OligoMod+, permet actuellement de retrouver la position des modifications post-transcriptionnelles sur la séquence de l'ARN lorsque celles-ci sont connues et indiquées directement dans la séquence FASTA fournie par l'utilisateur. Il s'agit du premier outil informatique open-source entièrement pensé pour interpréter les analyses CE-MS/MS en mode positif au niveau oligonucléotides. De premiers résultats prometteurs ont été obtenus et laissent présager un réel apport pour la localisation des modifications post-transcriptionnelles. Une seconde version est encore en développement, afin de localiser les modifications post-transcriptionnelles des ARN lorsque leurs positions ne sont pas connues. Lorsque cette version sera finalisée, il sera alors possible de localiser les modifications des ARN ribosomiques de *P. aeruginosa*. Tous les ARN dont les modifications ne sont pas décrites dans la littérature pourront donc être étudiés plus facilement par CE-MS/MS.

Dans le prochain chapitre, le logiciel Nucleos'ID sera utilisé, à la fois sur le couplage CE-QTOF-MS/MS et sur le couplage CE-FTICR-MS/MS, pour identifier les modifications dans des échantillons d'ARN qui ne sont pas totalement décrits dans la littérature.

8.5. Références

- [1] Wolff, P. *et al.*, “Comparative Patterns of Modified Nucleotides in Individual tRNA Species from a Mesophilic and Two Thermophilic Archaea”, *RNA*, **2020**, rna.077537.120.
- [2] Yan, M. *et al.*, “A High-Throughput Quantitative Approach Reveals More Small RNA Modifications in Mouse Liver and Their Correlation with Diabetes”, *Anal. Chem.*, **2013**, 85 (24), 12173–12181.
- [3] Zhu, Y. *et al.*, “LC-MS-MS Quantitative Analysis Reveals the Association between FTO and DNA Methylation”, *PLoS ONE*, **2017**, 12 (4), e0175849.
- [4] Sakaguchi, Y. *et al.*, Chapter Two - Nucleoside Analysis by Hydrophilic Interaction Liquid Chromatography Coupled with Mass Spectrometry. In *Methods in Enzymology*; He, C., Ed.; RNA Modification; Academic Press, 2015; Vol. 560, pp 19–28.

- [5] Liu, R. *et al.*, “Determination of Five Nucleosides by LC-MS/MS and the Application of the Method to Quantify N6-Methyladenosine Level in Liver Messenger Ribonucleic Acid of an Acetaminophen-Induced Hepatotoxicity Mouse Model”, *J. Sep. Sci.*, **2019**, 42 (16), 2668–2678.
- [6] Raćkowska, E. *et al.*, “Development and Validation of a Rapid LC-MS/MS Method for Determination of Methylated Nucleosides and Nucleobases in Urine”, *J. Chromatogr. B*, **2019**, 1128, 121775.
- [7] Lo, W.-Y. *et al.*, “Urinary Cytidine as an Adjunct Biomarker to Improve the Diagnostic Ratio for Gastric Cancer in Taiwanese Patients”, *Clin. Chim. Acta*, **2014**, 428, 57–62.
- [8] Chen, F. *et al.*, “Simultaneous Determination of 10 Nucleosides and Nucleobases in *Antrodia Camphorata* Using QTRAP LC-MS/MS”, *J. Chromatogr. Sci.*, **2014**, 52 (8), 852–861.
- [9] Feng, Y.-J. *et al.*, “Identification of Inosine and 2'-O-Methylinosine Modifications in Yeast Messenger RNA by Liquid Chromatography–Tandem Mass Spectrometry Analysis”, *Anal. Chem.*, **2022**, 94 (11), 4747–4755.
- [10] Thüring, K. *et al.*, “Analysis of RNA Modifications by Liquid Chromatography–Tandem Mass Spectrometry”, *Methods*, **2016**, 107, 48–56.
- [11] Su, D. *et al.*, “Quantitative Analysis of Ribonucleoside Modifications in tRNA by HPLC-Coupled Mass Spectrometry”, *Nat. Protoc.*, **2014**, 9 (4), 828–841.
- [12] Lechner, A. *et al.*, “Characterization of Post-Transcriptional RNA Modifications by Sheathless Capillary Electrophoresis–High Resolution Mass Spectrometry”, *Anal. Chem.*, **2020**, 92 (10), 7363–7370.
- [13] Paulines, M. J. *et al.*, “Using Spectral Matching to Interpret LC-MS/MS Data during RNA Modification Mapping”, *J. Mass Spectrom.*, **2019**, 54 (11), 906–914.
- [14] Cao, X. *et al.*, “Enhanced Detection of Post-Transcriptional Modifications Using a Mass-Exclusion List Strategy for RNA Modification Mapping by LC-MS/MS”, *Anal. Chem.*, **2015**, 87 (16), 8433–8440.
- [15] Ross, R. *et al.*, “Sequence Mapping of Transfer RNA Chemical Modifications by Liquid Chromatography Tandem Mass Spectrometry”, *Methods*, **2016**, 107, 73–78.
- [16] Nakayama, H. *et al.*, “Ariadne: A Database Search Engine for Identification and Chemical Analysis of RNA Using Tandem Mass Spectrometry Data”, *Nucleic Acids Res.*, **2009**, 37 (6), e47.
- [17] Sample, P. J. *et al.*, “RoboOligo: Software for Mass Spectrometry Data to Support Manual and de Novo Sequencing of Post-Transcriptionally Modified Ribonucleic Acids”, *Nucleic Acids Res.*, **2015**, 43 (10), e64.
- [18] Yu, N. *et al.*, “RNAModMapper: RNA Modification Mapping Software for Analysis of Liquid Chromatography Tandem Mass Spectrometry Data”, *Anal. Chem.*, **2017**, 89 (20), 10744–10752.
- [19] Lobue, P. A. *et al.*, “Improved Application of RNAModMapper – An RNA Modification Mapping Software Tool – For Analysis of Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) Data”, *Methods*, **2019**, 156, 128–138.
- [20] Wein, S. *et al.*, “A Computational Platform for High-Throughput Analysis of RNA Sequences and Modifications by Mass Spectrometry”, *Nat. Commun.*, **2020**, 11 (1), 926.
- [21] D'Ascenzo, L. *et al.*, “Pytheas: A Software to Map RNA Modifications via Tandem Mass Spectrometry”, *Biophys. J.*, **2019**, 116 (3), 287a.

- [22] D'Ascenzo, L. et al., "Pytheas: A Software Package for the Automated Analysis of RNA Sequences and Modifications via Tandem Mass Spectrometry", *Nat. Commun.*, **2022**, 13 (1), 2424.
- [23] Helm, M., "Post-Transcriptional Nucleotide Modification and Alternative Folding of RNA", *Nucleic Acids Res.*, **2006**, 34 (2), 721–733.
- [24] Chawla, M. et al., "An Atlas of RNA Base Pairs Involving Modified Nucleobases with Optimal Geometries and Accurate Energies", *Nucleic Acids Res.*, **2015**, 43 (14), 6714–6729.
- [25] Agris, P. F., The Importance of Being Modified: Roles of Modified Nucleosides and Mg²⁺ in RNA Structure and Function. In *Progress in Nucleic Acid Research and Molecular Biology*; Cohn, W. E. et al., Eds.; Academic Press, 1996; Vol. 53, pp 79–129.
- [26] Roundtree, I. A. et al., "Dynamic RNA Modifications in Gene Expression Regulation", *Cell*, **2017**, 169 (7), 1187–1200.
- [27] Wilkinson, E. et al., "Roles of RNA Modifications in Diverse Cellular Functions", *Front. Cell Dev. Biol.*, **2022**, 10 (828683).
- [28] Vester, B. et al., *Antibiotic Resistance in Bacteria Caused by Modified Nucleosides in 23S Ribosomal RNA*; Landes Bioscience, 2009.
- [29] Jonkhout, N. et al., "The RNA Modification Landscape in Human Disease", *RNA*, **2017**, 23 (12), 1754–1769.
- [30] Gao, M. et al., "Synthetic Modified Messenger RNA for Therapeutic Applications", *Acta Biomater.*, **2021**, 131, 1–15.
- [31] Amalric, A. et al., "Quantifying RNA Modifications by Mass Spectrometry: A Novel Source of Biomarkers in Oncology", *Crit. Rev. Clin. Lab. Sci.*, **2022**, 59 (1), 1–18.
- [32] Boccaletto, P. et al., "MODOMICS: A Database of RNA Modification Pathways. 2021 Update", *Nucleic Acids Res.*, **2022**, 50 (D1), D231–D235.
- [33] Grobe, S. et al., "Identification and Quantification of (t)RNA Modifications in *Pseudomonas Aeruginosa* by Liquid Chromatography-Tandem Mass Spectrometry", *ChemBioChem*, **2019**, 20 (11), 1430–1437.
- [34] Dai, Y. et al., "Sensitive and Simultaneous Determination of Uridine Thiolation and Hydroxylation Modifications in Eukaryotic RNA by Derivatization Coupled with Mass Spectrometry Analysis", *Anal. Chem.*, **2021**, 93 (18), 6938–6946.
- [35] Chen, M.-Y. et al., "Comprehensive Profiling and Evaluation of the Alteration of RNA Modifications in Thyroid Carcinoma by Liquid Chromatography-Tandem Mass Spectrometry", *Chin. Chem. Lett.*, **2022**, 33 (8), 3772–3776.
- [36] Chen, M.-Y. et al., "Adolescent Alcohol Exposure Alters DNA and RNA Modifications in Peripheral Blood by Liquid Chromatography-Tandem Mass Spectrometry Analysis", *Chin. Chem. Lett.*, **2022**, 33 (4), 2086–2090.
- [37] Jensen, S. S. et al., "Collision-Induced Dissociation of Cytidine and Its Derivatives", *J. Mass Spectrom.*, **2007**, 42 (1), 49–57.
- [38] Didierjean, M. et al., MSARN/NucleosID: NucleosID v0.8.0, 2022.
- [39] McCown, P. J. et al., "Naturally Occurring Modified Ribonucleosides", *WIREs RNA*, **2020**, 11 (5), e1595.
- [40] Grosjean, H., *Nucleic Acids Are Not Boring Long Polymers of Only Four Types of Nucleotides: A Guided Tour*; Landes Bioscience, 2013.
- [41] Brenton, A. G. et al., "Accurate Mass Measurement: Terminology and Treatment of Data", *J. Am. Soc. Mass Spectrom.*, **2010**, 21 (11), 1821–1835.

- [42] Risk, B. A. *et al.*, “Peppy: Proteogenomic Search Software”, *J. Proteome Res.*, **2013**, *12* (6), 3019–3025.
- [43] Perkins, D. N. *et al.*, “Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data”, *Electrophoresis*, **1999**, *20* (18), 3551–3567.
- [44] Brereton, R. G., “False Discovery Rates, Power and Related Concepts”, *J. Chemom.*, **2021**, *35* (6), e3241.
- [45] Kapp, E. A. *et al.*, “An Evaluation, Comparison, and Accurate Benchmarking of Several Publicly Available MS/MS Search Algorithms: Sensitivity and Specificity Analysis”, *PROTEOMICS*, **2005**, *5* (13), 3475–3490.
- [46] Aggarwal, S. *et al.*, False Discovery Rate Estimation in Proteomics. In *Statistical Analysis in Proteomics*; Jung, K., Ed.; Methods in Molecular Biology; Springer: New York, NY, 2016; pp 119–128.
- [47] Kowalak, J. A. *et al.*, “Posttranscriptional Modification of the Central Loop of Domain V in Escherichia Coli 23 S Ribosomal RNA (*)”, *J. Biol. Chem.*, **1995**, *270* (30), 17758–17764.
- [48] Wilson, D. N. *et al.*, “The Weird and Wonderful World of Bacterial Ribosome Regulation”, *Crit. Rev. Biochem. Mol. Biol.*, **2007**, *42* (3), 187–219.
- [49] Risk, B. A. *et al.*, “Peppy: Proteogenomic Search Software”, *J. Proteome Res.*, **2013**, *12* (6), 3019–3025.

PARTIE IV : MISE EN APPLICATION DU COUPLAGE CE-FTICR-MS

CHAPITRE 9 : MISE EN APPLICATION DU COUPLAGE CE-FTICR-MS

9.1. Introduction

Précédemment, la méthode pour caractériser les modifications post-transcriptionnelles par analyse des nucléosides a été transférée du couplage CE-QTOF-MS au couplage CE-FTICR-MS sur des échantillons bien connus et décrits dans la littérature. Différents développements méthodologiques ont ensuite été présentés, tels que l'optimisation de l'étape de déphosphorylation, ainsi que le développement de logiciels permettant de faciliter l'interprétation des données, notamment pour l'identification non ciblée des modifications post-transcriptionnelles. Ces différents travaux permettent maintenant d'envisager la mise en application du couplage CE-FTICR-MS sur des échantillons dont les modifications n'ont pas encore été décrites dans la littérature.

Dans un premier temps, l'analyse par CE-FTICR-MS de l'ARNr 16S issu de *P. aeruginosa*, qui a fait l'objet du chapitre 7, sera détaillée. Cet ARNr est composé de 1537 bases et seule sa séquence canonique génomique est référencée dans la littérature [1]. De par la longueur de sa séquence et le peu de modifications généralement présentes sur les ARNr, cet échantillon est un premier cas d'application complexe. L'analyse de cet échantillon représente donc un grand défi pour le couplage CE-FTICR-MS.

Dans un second temps, l'analyse par CE-FTICR-MS d'un extrait total d'ARNt issu de *Thermus thermophilus* sera détaillé. *T. thermophilus* est connue pour être une bactérie extrémophile puisqu'elle est capable de croître à des températures allant jusqu'à 82°C selon les souches [2]. Les modifications post-transcriptionnelles de cette bactérie ont très peu été décrites. Seuls 6 ARNt ont été étudiés avec des méthodes telles que chromatographie 2D sur couche mince [3–6]. Ces 6 séquences d'ARNt sont répertoriées dans Modomics [7]. A notre connaissance, aucune étude en spectrométrie de masse n'a été réalisée, bien qu'elle soit particulièrement intéressante puisque les modifications ont une grande influence sur la stabilité de la structure des ARN. L'étude des modifications de cet extrait total s'effectue en collaboration avec des biologistes de l'IBMC qui ont, en parallèle de nos analyses de nucléosides, réalisé des analyses au niveau oligonucléotides par LC-MS/MS à partir de spots d'un gel 2D d'électrophorèse. D'après ces analyses LC-MS/MS, une adénosine modifiée inconnue (non décrite dans Modomics et annotée xA) a été détectée. L'identification de cette modification ne pouvant se faire qu'au niveau nucléosides, elle est devenue notre principal objectif avec le couplage CE-FTICR-MS/MS dont la précision de masse est une grande alliée pour établir une formule brute unique.

Pour chacun de ces échantillons, les analyses ont d'abord été interprétées à l'aide du logiciel Nucleos'ID avant d'être vérifiées manuellement. Les analyses issues du couplage CE-FTICR-MS ont également été mises en parallèle avec celles issues du couplage CE-QTOF-MS

afin d'évaluer les performances du couplage CE-FTICR-MS sur des échantillons encore non décrits dans la littérature.

9.2. ARNr 16S de *Pseudomonas aeruginosa*

La première mise en application du couplage CE-FTICR est l'analyse de l'ARNr 16S de *P. aeruginosa*. Cet ARN a été digéré par la Nucléase P1 puis déphosphorylé à la BAP selon le protocole « overnight » établi au laboratoire et décrit dans la section 3.4.3 du chapitre 3 et dans le chapitre 6 [8,9]. Un premier test a été conduit dans le chapitre 7. Dans le but de potentiellement identifier plus de modifications, et de les confirmer autant que possible par MS/MS, la concentration de l'échantillon pour analyse par CE-QTOF-MS/MS a été augmentée de 0,05 g/L à 1,0 g/L. Afin d'éviter des différences issues d'une potentielle dégradation des échantillons entre les deux analyses, chaque échantillon a été analysé sur le couplage CE-QTOF-MS juste avant d'entamer la campagne sur le couplage CE-FTICR-MS. De plus, pour pallier les différences de sensibilité observées entre les deux couplages (cf paragraphe 4.5.4 du chapitre 4), cet échantillon a été analysé sur le FTICR-MS à des concentrations de 5,5 g/L, 2,75 g/L et 2,0 g/L au lieu de 1,0 g/L sur le QTOF.

9.2.1. Recherche Nucleos'ID

L'interprétation des données a été réalisée à l'aide du logiciel Nucleos'ID paramétré de la manière suivante pour toutes les analyses CE-QTOF-MS/MS et CE-FTICR-MS/MS :

- Base de données : Eubactéries
- Tolérance MS : 0,02 Da
- Tolérance MS/MS : 0,5 Da
- Seuil d'intensité MS/MS absolue : 0 UA
- Seuil pour le score : 0 %
- Temps d'exclusion dynamique : 1 min

Le seuil du score a été fixé à 0% afin de ne manquer aucune modification, car elles ne sont pas décrites dans la littérature pour cet échantillon. Les nucléosides identifiés par Nucleos'ID ont ensuite été vérifiés manuellement (Tableau 9.1). L'électrophérogramme obtenu après la vérification manuelle pour le couplage CE-QTOF-MS est présenté en Figure 9.1 et celui obtenu après la vérification manuelle de l'analyse à 2,0 g/L pour le couplage CE-FTICR-MS est présenté en Figure 9.2.

Avec un total de 20 vrais positifs et 3 faux positifs, un FDR de 0,17 a été obtenu pour le couplage CE-QTOF-MS. Parmi les 3 modifications induisant des faux positifs, deux sont dues à du bruit de fond (pour cmnm⁵s²U et cmnm⁵Um). La troisième modification induisant un faux positif est m¹I. Ce nucléoside a été identifié par Nucleos'ID au même temps de migration que mA, avec une différence en MS de 0,02 Da avec la valeur théorique de m¹I. Il s'agit en réalité du premier isotope de mA (*m/z* 282,12) qui a été fragmenté et le spectre MS/MS associé n'est

qu'un spectre de bruit de fond. Quant au couplage CE-FTICR-MS, avec entre 6 et 9 vrais positifs et 0 faux positif, des FDR de 0,00 ont été obtenus.

Tableau 9.1 : Nucléosides identifiés, par la recherche Nucleos'ID et après vérification manuelle, dans l'ARNr 16S issu de *Pseudomonas aeruginosa*, analysé par CE-QTOF-MS/MS et CE-FTICR-MS/MS. Le symbole \checkmark indique que l'identification est basée sur les données MS et MS/MS. Le symbole \approx indique que l'identification manuelle est basée uniquement sur la MS à cause de l'absence de fragmentation par CID. Le symbole \times indique que le nucléoside n'a pas été identifié.

	CE-QTOF-MS		CE-FTICR-MS					
	1,0 g/L		2,0 g/L		2,75 g/L		5,5 g/L	
	Nucleos' ID	Vérification manuelle	Nucleos' ID	Vérification manuelle	Nucleos' ID	Vérification manuelle	Nucleos' ID	Vérification manuelle
A	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
C	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
G	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
U	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\times	\times
$m^1A / m^2A / m^6A / m^8A$	\checkmark	\checkmark (2 isomères)	\checkmark	\checkmark (1 isomère) \approx (1 isomère)	\checkmark	\checkmark (1 isomère) \approx (1 isomère)	\checkmark	\checkmark (1 isomère) \approx (1 isomère)
Am	\checkmark	\checkmark	\times	\approx	\times	\approx	\times	\approx
$m^{2,8}A / m^{6,6}A$	\checkmark	\checkmark (1 isomère)	\checkmark	\checkmark (1 isomère)	\checkmark	\checkmark (1 isomère)	\checkmark	\checkmark (1 isomère)
t^6A	\checkmark	\checkmark	\times	\times	\times	\times	\times	\times
I	\checkmark	\checkmark	\times	\approx	\times	\approx	\times	\times
m^1I	\checkmark	\times	\times	\times	\times	\times	\times	\times
$m^3C / m^4C / m^5C$	\checkmark	\checkmark (1 isomère)	\checkmark	\checkmark (1 isomère)	\times	\approx	\times	\times
Cm	\checkmark	\checkmark	\times	\times	\times	\approx	\times	\times
m^4Cm	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\times	\times
nm^5U	\checkmark	\checkmark	\times	\times	\times	\times	\times	\times
$m^1G / m^2G / m^7G$	\checkmark	\checkmark (m^1G/m^2G et m^7G)	\checkmark	\checkmark (m^1G/m^2G et m^7G)	\checkmark	\checkmark (m^1G/m^2G) \approx (m^7G)	\checkmark	\checkmark (m^1G/m^2G)
Gm	\checkmark	\checkmark	\times	\approx	\checkmark	\checkmark	\times	\times
$m^{2,2}G / preQ_1$	\checkmark	\checkmark	\times	\times	\times	\times	\times	\times
Q	\checkmark	\checkmark	\times	\times	\times	\times	\times	\times
ψ	\checkmark	\checkmark	\times	\approx	\times	\approx	\times	\times
m^3U / m^5U	\checkmark	\checkmark	\times	\approx	\times	\approx	\times	\times
acp ³ U	\checkmark	\checkmark	\times	\times	\times	\times	\times	\times
cmmn ⁵ s ² U	\checkmark	\times	\times	\times	\times	\times	\times	\times
cmmn ⁵ Um	\checkmark	\times	\times	\times	\times	\times	\times	\times

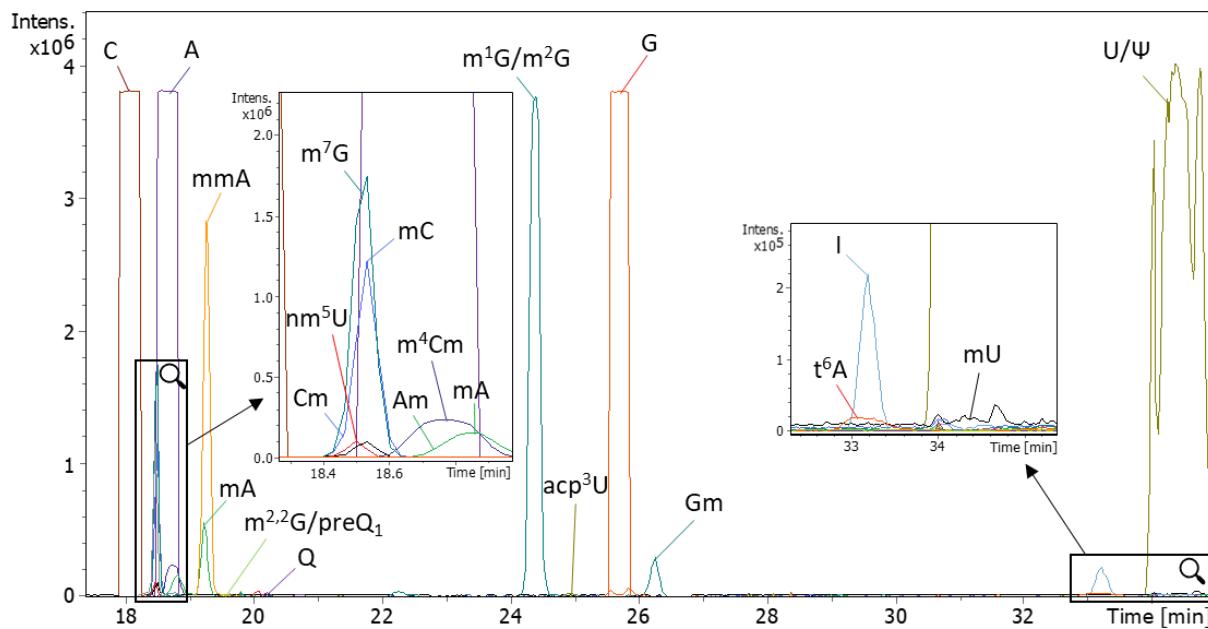


Figure 9.1 : Electrophérogramme des nucléosides de l'ARNr 16S issu de *P. aeruginosa*, obtenu par CE-QTOF-MS/MS après digestion à la Nucléase P1 et déphosphorylation à la BAP. Tous les nucléosides ont été identifiés par MS et MS/MS.

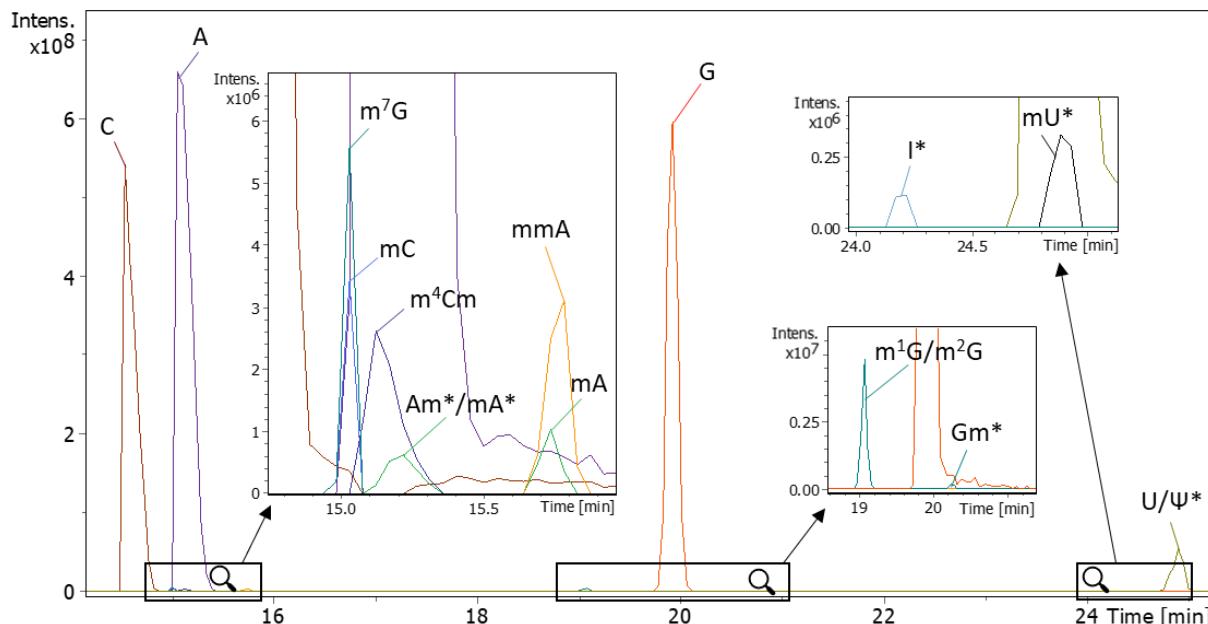


Figure 9.2 : Electrophérogramme des modifications de l'ARNr 16S issu de *P. aeruginosa*, obtenu par CE-FTICR-MS/MS à une concentration de 2,0 g/L après digestion à la Nucléase P1 et déphosphorylation à la BAP. Les nucléosides annotés avec * ont été uniquement identifiés par MS.

Comme cela a déjà été expliqué dans la section 8.2.4.3 du chapitre 8, les modifications de l'ARNr 16S de *P. aeruginosa* n'ont pas encore été décrites, mais celles d'autres espèces bactériennes, comme *E. coli*, ont déjà été rapportées dans la littérature et les espèces d'un même domaine présentent généralement des modifications similaires. Certaines modifications identifiées avec l'un ou l'autre couplage (mA, mmA, mC, mU, m⁴Cm, mG et ψ) sont similaires à celles décrites pour l'ARNr 16S de *E. coli* [10]. D'autres, telles que t⁶A, I, nm⁵U, m^{2,2}G, preQ₁, Q ou encore acp³U, sont généralement présentes dans les ARNt de bactéries, mais pas dans les ARNr [11]. Leur présence peut être due à une très faible proportion d'ARNt résiduelle dans l'échantillon après purification.

9.2.2. Comparaison des résultats avec le couplage CE-QTOF-MS

Après vérification manuelle des recherches Nucleos'ID, les résultats obtenus par les deux couplages ont été comparés afin d'évaluer les résultats obtenus par le couplage CE-FTICR-MS.

Alors que l'on pourrait s'attendre à détecter de plus en plus de modifications avec le couplage CE-FTICR-MS en augmentant la concentration de l'échantillon, étant donné les différences de sensibilité entre les deux systèmes et la très faible intensité de certaines modifications post-transcriptionnelles détectées, l'inverse a été observé (Figure 9.3). En augmentant la concentration de 2,0 g/L à 2,75 g/L, le nombre total de nucléosides n'a pas diminué, mais une modification identifiée par MS et MS/MS à 2,0 g/L n'a pas été fragmentée à 2,75 g/L. En augmentant à 5,5 g/L, le nombre total de nucléosides détectés a été divisé par deux (passant de 16 à 8), avec moins d'identifications par MS/MS, mais également moins de nucléosides identifiés uniquement par MS.

Cette observation peut être expliquée par une saturation de la cellule ICR entraînant une perte d'informations sur les modifications très peu abondantes. En effet, dans le cas de l'ARNr 16S, long de 1537 bases, le nombre de positions non modifiées est bien supérieur au nombre de positions modifiées. Ainsi, les quatre nucléosides canoniques sont bien plus abondants que les modifications post-transcriptionnelles. Ce phénomène est d'autant plus amplifié pour les modifications issues des ARNt résiduels détectées avec le couplage CE-QTOF-MS. La gamme dynamique nécessaire pour bien détecter toutes les modifications est donc importante. En augmentant la concentration de l'échantillon, la cellule ICR est remplie de plus en plus d'ions, jusqu'à saturation. Cela induit alors une réduction de la gamme dynamique accessible et une perte d'information lors de l'analyse.

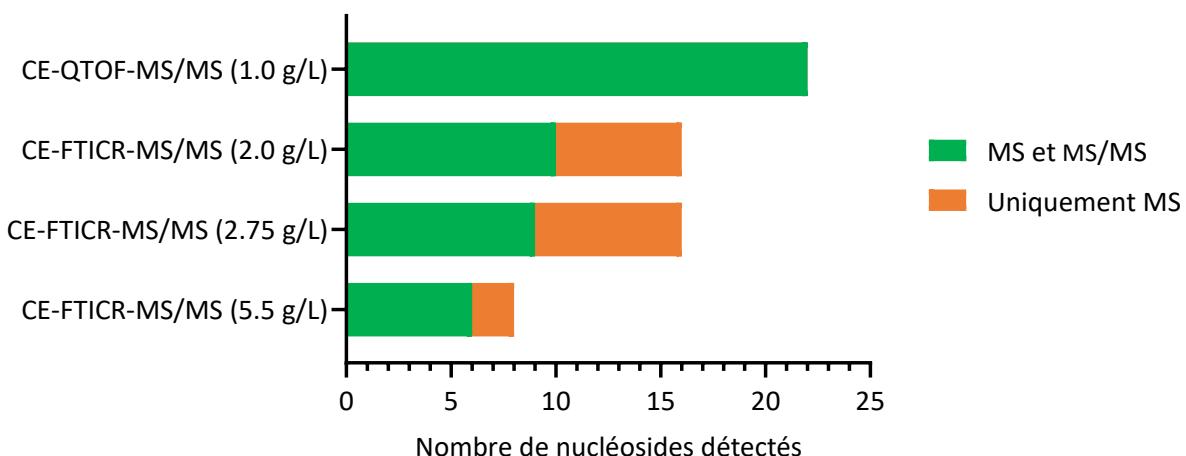


Figure 9.3 : Nombre de nucléosides détectés pour les 4 analyses en CE-QTOF-MS/MS et CE-FTICR-MS/MS de l'ARNr 16S issu de *Pseudomonas aeruginosa*, après vérification manuelle des résultats de Nucleos'ID.

Faute de temps, des concentrations plus faibles et intermédiaires n'ont pas pu être testées à ce jour. Il serait également intéressant de descendre en concentration petit à petit afin de trouver la concentration optimale pour minimiser la surcharge de la cellule ICR tout en

compensant les différences de sensibilité entre les deux couplages pour permettre un maximum d'identification. Les meilleurs résultats, obtenus lors de l'analyse à 2,0 g/L pour le couplage CE-FTICR-MS, seront comparés en détail à l'analyse à 1,0 g/L sur le couplage CE-QTOF-MS dans le paragraphe suivant.

L'analyse, de cet ARNr 16S issu de *P. aeruginosa*, réalisée sur le couplage CE-QTOF-MS, a permis d'identifier, à la fois par MS et MS/MS, 22 nucléosides (dont 18 modifications post-transcriptionnelles). Sur le couplage CE-FTICR-MS, parmi ces 22 nucléosides, 10 (A, C, G, U, mA, mmA, mC, m⁴Cm, m¹G/m²G et m⁷G) ont également été identifiés par MS et MS/MS. Six autres modifications (1 isomère de mA, Am, I, Gm, mU et ψ) ont été détectées uniquement par MS. En effet, du fait de la faible intensité des modifications post-transcriptionnelles par rapport aux nucléosides canoniques, les 5 premiers nucléosides n'ont pas été sélectionnés pour être fragmentés. Quant à ψ, isomérisation de U, un spectre MS/MS est disponible, mais aucun de ses fragments caractéristiques n'a été détecté comme cela était déjà le cas dans le chapitre 5. En effet, ses fragments semblent, encore une fois, difficiles à détecter avec le couplage CE-FTICR-MS. Les six dernières modifications (t⁶A, Cm, nm⁵U, m^{2,2}G/preQ₁, Q et acp³U) n'ont pas du tout été détectées avec le couplage CE-FTICR-MS. Ces modifications étaient déjà très peu intenses dans l'analyse CE-QTOF-MS. Il n'est donc pas étonnant de ne pas les avoir détectées avec le couplage CE-FTICR-MS pour les raisons mentionnées précédemment (différence de sensibilité et saturation de la cellule ICR induisant une réduction de la gamme dynamique).

9.3. Extrait total d'ARNr de *Thermus thermophilus*

La seconde mise en application du couplage CE-FTICR est un extrait total d'ARNr issu de la bactérie thermophile *Thermus thermophilus*. De la même manière, cet extrait total d'ARNr a été digéré par la Nucléase P1 puis déphosphorylé à la BAP selon le protocole « overnight » établi au laboratoire et décrit dans le chapitre 6 [8,9]. Encore une fois, afin d'éviter des différences issues d'une potentielle dégradation des échantillons entre les deux analyses, chaque échantillon a été analysé en couplage CE-QTOF-MS juste avant d'entamer la campagne sur le couplage CE-FTICR-MS. De plus, pour pallier les différences de sensibilité observées entre les deux couplages (cf paragraphe 4.5.4 du chapitre 4), cet échantillon a été analysé sur le FTICR à une concentration de 2,0 g/L au lieu de 0,1 g/L sur le QTOF. Par manque de temps lors de la dernière campagne sur le couplage CE-FTICR-MS, des concentrations plus faibles n'ont pas pu être testées.

9.3.1. Recherche Nucleos'ID

L'interprétation des données a été réalisée à l'aide du logiciel Nucleos'ID pour les analyses CE-QTOF-MS/MS et CE-FTICR-MS/MS avec les mêmes paramètres que l'ARNr 16S de *P. aeruginosa*. De la même manière, le seuil du score a été fixé à 0% afin de ne manquer aucune modification, car cet échantillon n'est pas décrit dans la littérature. Les nucléosides identifiés par Nucleos'ID ont ensuite été vérifiés manuellement (Tableau 9.2).

L'électrophérogramme obtenu après la vérification manuelle pour le couplage CE-QTOF-MS est présenté en Figure 9.4 et l'électrophérogramme obtenu après la vérification manuelle pour le couplage CE-FTICR-MS est présenté en Figure 9.5.

Tableau 9.2 : Nucléosides identifiés, par la recherche Nucleos'ID et après vérification manuelle, dans l'extrait total d'ARNt issu de *Thermus thermophilus*, analysé par CE-QTOF-MS/MS et CE-FTICR-MS/MS. Le symbole ✓ indique que l'identification est basée sur les données MS et MS/MS. Le symbole ≈ indique que l'identification manuelle est basée uniquement sur la MS à cause de l'absence de fragmentation par CID. Le symbole ✗ indique que le nucléoside n'a pas été identifié.

	CE-QTOF-MS/MS (0.1 g/L)	CE-FTICR-MS/MS (2.0 g/L)		
	Recherche par Nucleos'ID	Vérification manuelle	Recherche par Nucleos'ID	Vérification manuelle
A	✓	✓	✓	✓
C	✓	✓	✓	✓
G	✓	✓	✓	✓
U	✓	✓	✓	✓
$m^1A / m^2A / m^6A / m^8A$	✓	✓ (1 isomère)	✓	✓ (1 isomère)
$m^{2,8}A / m^{6,6}A$	✓	✓ (1 isomère)	✗	≈ (1 isomère)
t^6A	✓	✓	✓	✓
ms^2t^6A	✓	✓	✓	✓
ms^2i^6A	✓	✓	✗	≈
ms^2io^6A	✓	✓	✓	✓
I	✓	✓	✓	✓
$m^3C / m^4C / m^5C$	✓	✓ (1 isomère)	✗	✗
Cm	✓	✓	✓	✓
$m^{4,4}C$	✓	✓	✗	≈
ho^5C	✓	✓	✗	≈
ac ⁴ C	✓	✓	✗	✗
$m^1G / m^2G / m^7G$	✓	✓ (m^1G / m^2G et m^7G)	✓	✓ (m^1G / m^2G et m^7G)
Gm	✓	✓	✓	✓
$m^{2,2}G / preQ_1$	✓	✓ (2 isomères)	✗	✗
Q	✓	✓	✗	✗
ψ	✓	✓	✗	≈
D	✓	✓	✓	✓
m^3U / m^5U	✓	✓	✗	≈
Um	✓	✓	✗	≈
m^5s^2U	✓	✓	✓	✓
mnm^5s^2U	✓	✓	✗	≈
$cmmnm^5U$	✓	✓	✗	✗
mo^5U	✓	✗	✓	✗
s^2U / s^4U	✓	✗	✗	✗

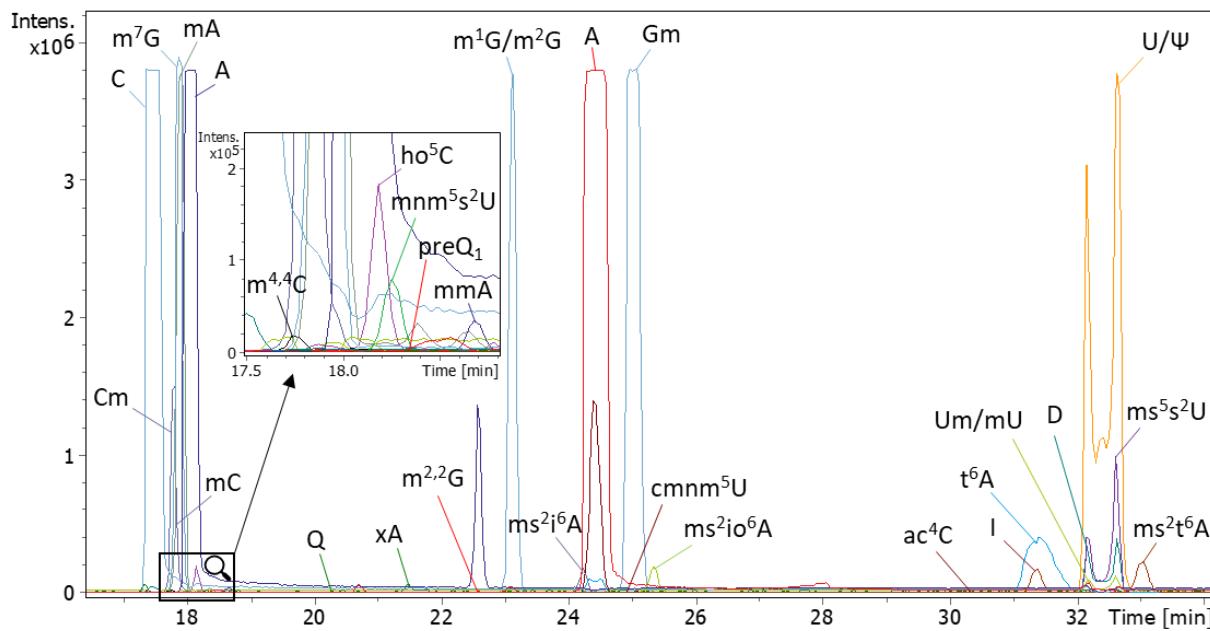


Figure 9.4 : Electrophérogramme des nucléosides de l'extrait total d'ARNt issu de *Thermus thermophilus*, obtenu par CE-QTOF-MS/MS après digestion à la Nucléase P1 et déphosphorylation à la BAP. Tous les nucléosides ont été identifiés par MS et MS/MS. xA (m/z 314) est une adénosine modifiée non répertoriée dans Modomics.

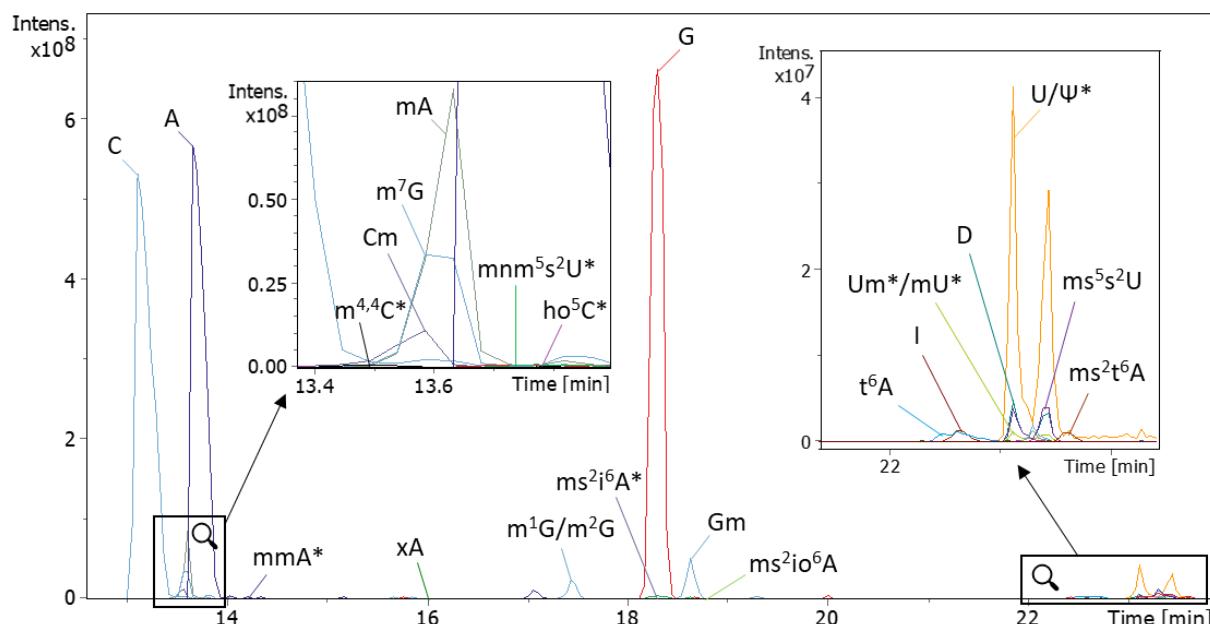


Figure 9.5 : Electrophérogramme des modifications de l'extrait total d'ARNt issu de *Thermus thermophilus*, obtenu par CE-FTICR-MS/MS après digestion à la Nucléase P1 et déphosphorylation à la BAP. Les nucléosides annotés avec * ont été uniquement identifiés par MS. xA (m/z 314) est une adénosine modifiée non répertoriée dans Modomics.

Avec 29 vrais positifs et 2 faux positifs, un FDR de 0,07 a été obtenu pour le couplage CE-QTOF-MS. Quant au couplage CE-FTICR-MS, avec 15 vrais positifs et 1 faux positif, un FDR de 0,07 a été obtenu.

Parmi les 2 modifications induisant des faux positifs, s^2U / s^4U est dûe au bruit de fond et mo^5U a été confondue avec m^5s^2U . En effet, la différence de masse entre mo^5U et m^5s^2U est de moins de 0,02 Da, et les deux modifications possèdent le même m/z comme ion produit dans la base de données.

9.3.2. Comparaison des résultats avec le couplage CE-QTOF-MS

Après vérification manuelle des recherches Nucleos'ID, les résultats obtenus par les deux couplages ont été comparés afin d'évaluer le couplage CE-FTICR-MS. L'analyse par CE-QTOF-MS a permis d'identifier dans cet extrait total d'ARNt issu de *Thermus thermophilus*, à la fois par MS et MS/MS, 29 nucléosides dont 25 modifications post-transcriptionnelles. L'analyse par CE-FTICR-MS a, quant à elle, permis d'identifier par MS et MS/MS, 15 nucléosides parmi les 29 de l'analyse par CE-QTOF-MS. Parmi ces nucléosides, m^5s^2U a fait l'objet de nombreuses études qui ont relié sa présence à la stabilité des ARN de bactéries thermophiles, et notamment de *T. thermophilus* [12]. Huit modifications ont également été détectées uniquement en MS, et 6 n'ont pas été détectées par le couplage CE-FTICR-MS.

Parmi les 16 nucléosides identifiés par MS/MS avec le couplage CE-FTICR-MS, 7 modifications (m^2G , D, Gm, Cm, $m7G$, m^5s^2U , m^1A) sont communes avec les 6 ARNt déjà étudiés par 2D-TLC. Parmi celles identifiées uniquement avec la MS, ψ est commune. Deux autres modifications (s^4U et i^6A) issues des analyses par 2D-TLC n'ont pas été détectées par nos analyses. Cependant, l'analyse de l'extrait total d'ARNt par le couplage CE-FTICR-MS a permis de détecter 11 modifications supplémentaires, dont 4 identifiées par MS/MS.

Comme pour l'ARNr, il serait par la suite intéressant de tester plusieurs concentrations (notamment plus faibles que 2,0 g/L) afin de trouver la concentration optimale pour identifier le maximum de modifications.

Lors d'analyses LC-MS de cet extrait total d'ARNt après digestion en oligonucléotides, nos collègues biologistes ont détecté une adénosine modifiée inconnue (non référencée dans la base de données Modomics), dont la masse a pu être déduite lors du séquençage *De Novo*. Cependant, l'analyse d'oligonucléotides ne leur permet pas d'identifier la nature de la modification et sa position sur l'adénosine. Cette modification n'est pas référencée dans la base de données Modomics [7] et est annotée, dans un premier temps, xA. À partir du $\Delta M=375$ Da obtenu par nos collègues lors du séquençage *De Novo*, nous avons donc cherché manuellement un nucléoside de m/z 314 dans nos analyses CE-MS/MS. Un pic électrophorétique a été détecté (annoté xA dans les Figure 9.4 et Figure 9.5) avec un m/z 314.09, et une perte de -132 Da lors de la MS/MS. Cela montre que ce pic peut en effet correspondre à un nucléoside possédant un ribose non modifié (-132 Da). Cette modification a été détectée par les deux couplages CE-QTOF-MS/MS et CE-FTICR-MS/MS.

9.3.3. Identification d'une modification inconnue (non référencée dans Modomics)

Bien que la modification xA ait été détectée et fragmentée par les deux couplages, son identification a été déterminée à partir de l'analyse par CE-FTICR-MS. En effet, grâce à la grande précision de masse apportée par le FTICR, une unique formule brute peut être assignée dans une tolérance maximale de 0,5 ppm (cf paragraphe 4.5.3 du chapitre 4) aux m/z mesurés

expérimentalement en MS et MS/MS. Les spectres MS et MS/MS obtenus par CE-FTICR-MS/MS sont présentés en Figure 9.6.

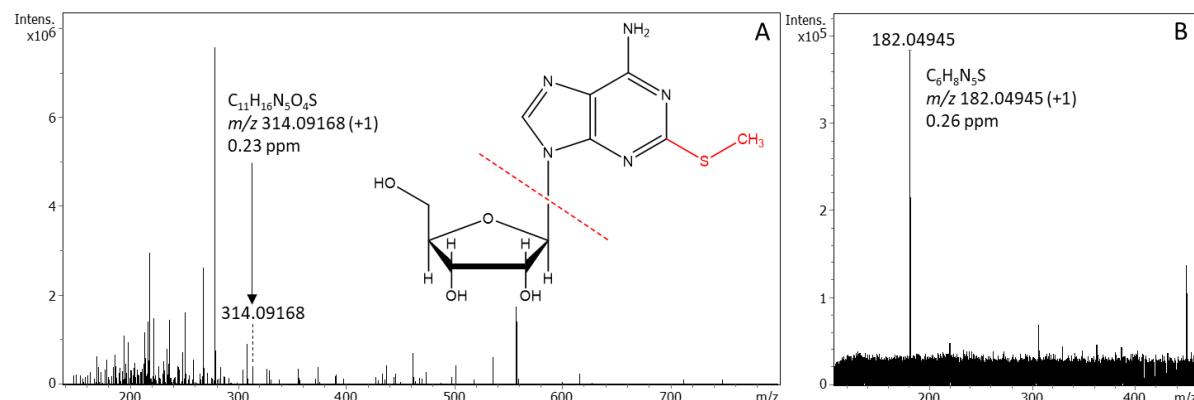


Figure 9.6 : (A) Spectres MS et (B) spectre MS/MS (fragmentation CID de m/z 314,09168) obtenus lors de l'analyse par CE-FTICR-MS/MS de l'extrait total d'ARNT issu de *Thermus thermophilus*. La structure potentielle de la modification xA déterminée grâce aux spectres MS et M/SMS est présentée.

À partir du spectre MS (Figure 9.6.A) et du m/z 314,09168 mesuré, la formule brute $C_{11}H_{16}N_5O_4S$ (0,23 ppm) a pu être associée à la modification xA. À partir du spectre MS/MS (Figure 9.6.B) et du m/z 182,04945 mesuré, la formule brute $C_6H_8N_5S$ (0,26 ppm) a pu être associée au fragment de la modification xA après fragmentation par CID. Ces informations confirment bien qu'il s'agit d'un nucléoside avec la perte du ribose et indiquent que la modification est un groupe thiométhyle positionné sur la base de l'adénosine. Sans l'obtention de fragments supplémentaires au niveau de la base, il n'est pas possible d'affirmer avec certitude la position de la modification sur celle-ci. Cependant, la potentielle structure de la modification, présentée en Figure 9.6, peut être déduite à partir des modifications référencées dans Modomics [7] portant un groupe thiométhyle sur la base (Figure 9.7). Cette potentielle structure peut ainsi être annotée ms^2A (2-methylthio-adenosine).

Malgré le fait qu'elle ne soit pas référencée dans Modomics, la modification ms^2A a déjà été identifiée dans des extraits d'ARNT issus de *Bacillus subtilis* [13] et *Bacillus stearothermophilus* [14], deux bactéries à Gram positif également thermophiles. De plus, dans le cadre de la collaboration avec les biologistes de l'IBMC, d'autres extraits totaux d'ARNT de bactéries extrémophiles ont été analysés par CE-MS/MS, et ms^2A a été identifiée dans *Thermotoga maritima*, une bactérie à Gram négatif hyperthermophile.

Bien que ms^2t^6A , ms^2i^6A et ms^2io^6A aient été détectées dans l'extrait total de *Thermotoga maritima*, ms^2A n'est pas connue pour être un précurseur de ces modifications. En effet, d'après la littérature, c'est le groupe thiométhyle qui se rajoute à t^6A et i^6A grâce à l'enzyme MiaB, et quant à ms^2io^6A , son précurseur est ms^2i^6A et non io^6A . Deux hypothèses ont été formulées par Vold *et al.* [13] pour expliquer la présence de ms^2A . Soit il s'agit d'une dégradation des modifications ms^2t^6A , ms^2i^6A et ms^2io^6A , soit il existe une ou plusieurs enzymes, encore non décrites, permettant d'obtenir ms^2t^6A , ms^2i^6A et ms^2io^6A à partir de ms^2A .

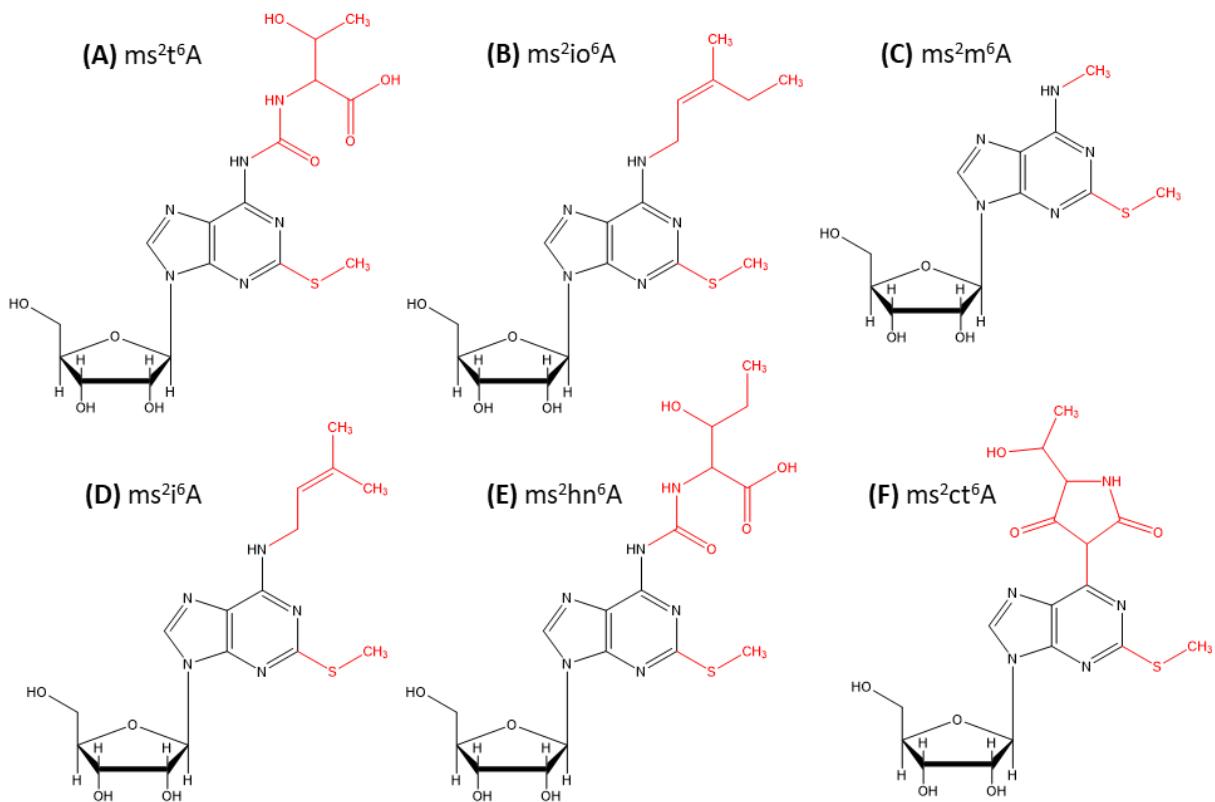


Figure 9.7 : Structures des modifications (A) 2-methylthio-N6-threonylcarbamoyladenosine (ms^2t^6A), (B) 2-methylthio-N6-(*cis*-hydroxyisopentenyl) adenosine (ms^2io^6A), (C) 2-methylthio-N6-methyladenosine (ms^2m^6A), (D) 2-methylthio-N6-isopentenyladenosine (ms^2i^6A), (E) 2-methylthio-N6-hydroxynorvalylcarbamoyladenosine (ms^2hn^6A) et (F) 2-methylthio cyclic N6-threonylcarbamoyladenosine (ms^2ct^6A) issues de Modomics [7].

9.4. Conclusion

Dans les chapitres précédents, le couplage CE-FTICR-MS a été mis en place puis testé sur des échantillons connus bien décrits dans la littérature. Divers développements méthodologiques ont ensuite été présentés, dont l'optimisation de l'étape de déphosphorylation et le développement de logiciels facilitant l'interprétation des données, notamment pour l'identification non ciblée des modifications post-transcriptionnelles. Tout cela a été mis à contribution afin de mettre en application le couplage CE-FTICR-MS sur deux échantillons dont les modifications post-transcriptionnelles n'étaient pas, ou pas entièrement, référencées dans la littérature.

Tout d'abord, l'ARNr 16S de *P. aeruginosa* a constitué un réel défi pour le couplage CE-FTICR-MS. En effet, cet ARN long de 1537 bases ayant peu de positions modifiées, la gamme dynamique nécessaire pour analyser les modifications de cet ARN est importante puisqu'il n'est pas possible de retirer les bases canoniques présentes en très grande majorité dans la séquence. Bien que le couplage du FTICR avec une méthode séparative telle que la CE permette d'augmenter la gamme dynamique, la différence de sensibilité entre les deux couplages et la complexité de l'échantillon n'ont pas permis d'obtenir d'aussi bons résultats avec les deux couplages. Cependant, de nombreuses modifications ont pu, tout de même, être identifiées par le couplage CE-FTICR-MS.

Ensuite, l'extrait total d'ARNt de *T. thermophilus* a permis de mettre en lumière la capacité du couplage CE-FTICR-MS pour identifier des modifications inconnues (non référencée dans Modomics). En effet, en combinant la gamme dynamique apportée par le couplage et la grande précision de masse du FTICR, il a été possible d'assigner une unique formule brute à une modification inconnue. Cette modification s'est révélée être ms²A.

Les deux applications détaillées dans ce chapitre montrent qu'il est nécessaire de choisir les bonnes applications pour chacun des couplages CE-QTOF-MS et CE-FTICR-MS afin d'obtenir les meilleurs résultats possibles. Alors que l'identification de modifications référencées dans les bases de données est meilleure avec le couplage CE-QTOF-MS, l'assignement d'une formule brute unique pour la modification xA n'a été possible qu'avec le couplage CE-FTICR-MS.

9.5. Références

- [1] Toschka, H. Y. *et al.*, “Complete Nucleotide Sequence of a 16S Ribosomal RNA Gene from Pseudomonas Aeruginosa.”, *Nucleic Acids Res.*, **1988**, 16 (5 Pt B), 2348.
- [2] Ohtani, N. *et al.*, “An Extreme Thermophile, *Thermus Thermophilus*, Is a Polyploid Bacterium.”, *J. Bacteriol.*, **2010**, 192 (20), 5499–5505.
- [3] WATANABE, K. *et al.*, “Nucleotide Sequence of Formylmethionine TRNA from an Extreme Thermophile, *Thermus Thermophilus HB8*.”, *J. Biochem. (Tokyo)*, **1979**, 86 (4), 893–905.
- [4] Horie, N. *et al.*, “Two TRNA^{Alle} Species from an Extreme Thermophile, *Thermus Thermophilus HB8*: Effect of 2-Thiolation of Ribothymidine on the Thermostability of TRNA.”, *Biochemistry*, **1985**, 24 (21), 5711–5715.
- [5] Watanabe, K. *et al.*, “Replacement of Ribothymidine by 5-Methyl-2-Thiouridine in Sequence GTΨC in TRNA of an Extreme Thermophile.”, *FEBS Lett.*, **1974**, 43 (1), 59–63.
- [6] Keith, G. *et al.*, “Sequence of TRNA Asp Form *Thermus Thermophilus HB8*.”, *Nucleic Acids Res.*, **1993**, 21 (18), 4399.
- [7] Boccaletto, P. *et al.*, “MODOMICS: A Database of RNA Modification Pathways. 2021 Update.”, *Nucleic Acids Res.*, **2022**, 50 (D1), D231–D235.
- [8] Lechner, A. *et al.*, “Characterization of Post-Transcriptional RNA Modifications by Sheathless Capillary Electrophoresis–High Resolution Mass Spectrometry.”, *Anal. Chem.*, **2020**, 92 (10), 7363–7370.
- [9] Gosset-Erard, C. *et al.*, “Optimization of Nucleotides Dephosphorylation for RNA Structural Characterization by Tandem Mass Spectrometry Hyphenated with Separation Methods.”, *J. Chromatogr. B*, **2022**, 1208, 123396.
- [10] Wilson, D. N. *et al.*, “The Weird and Wonderful World of Bacterial Ribosome Regulation.”, *Crit. Rev. Biochem. Mol. Biol.*, **2007**, 42 (3), 187–219.
- [11] McCown, P. J. *et al.*, “Naturally Occurring Modified Ribonucleosides.”, *WIREs RNA*, **2020**, 11 (5), e1595.
- [12] Davanloo, P. *et al.*, “Role of Ribothymidine in the Thermal Stability of Transfer RNA as Monitored by Proton Magnetic Resonance.”, *Nucleic Acids Res.*, **1979**, 6 (4), 1571–1581.
- [13] Vold, B. S. *et al.*, “Thiolation and 2-Methylthio- Modification of *Bacillus Subtilis* Transfer Ribonucleic Acids.”, *J. Bacteriol.*, **1981**, 148 (3), 869–876.

- [14] Agris, P. F. *et al.*, “The Effect of Growth Temperatures on the in Vivo Ribose Methylation of *Bacillus Stearothermophilus* Transfer RNA.”, *Arch. Biochem. Biophys.*, **1973**, 154 (1), 277–282.

CONCLUSION GENERALE ET PERSPECTIVES

L'objectif principal de ces travaux était de mettre en place et développer le couplage CE-FTICR-MS au laboratoire en s'appuyant sur des analyses de biomolécules à visée thérapeutique.

La première partie de ce manuscrit était consacrée à la revue bibliographique du couplage de l'électrophorèse capillaire avec la spectrométrie de masse à transformées de Fourier et résonance cyclotronique des ions (FTICR-MS), ainsi qu'à un état de l'art sur les acides ribonucléiques et la caractérisation de leurs modifications post-transcriptionnelles par MS. Dans un premier temps, la présentation des couplages existants entre les méthodes séparatives et la FTICR-MS a permis de comprendre l'utilité du couplage CE-FTICR-MS, mais aussi de prendre connaissance des défis qui lui sont propres. En effet, les deux techniques peuvent sembler incompatibles de prime abord, avec d'un côté une grande efficacité de séparation et de l'autre côté des temps d'acquisition relativement longs. Cependant, en trouvant de bons compromis entre séparation et performances de détection (résolution, précision de mesure de masse), de nombreuses applications ont montré qu'un tel couplage a tout de même des avantages non négligeables et que le champ des possibles est très large, notamment en ce qui concerne les analyses de biomolécules. Dans un second temps, l'état de l'art consacré à l'application finalement choisie pour réaliser ces travaux, les acides ribonucléiques (ARN), a permis de mettre en lumière la nécessité de caractériser leurs modifications post-transcriptionnelles. Cette application nous a semblé être un bon choix pour la mise en place du couplage CE-FTICR-MS. En effet, elle est connue au laboratoire puisqu'une précédente thèse avait développé une méthode CE-MS/MS à deux niveaux d'analyses pour caractériser les modifications des ARN. Le premier niveau consiste à identifier les modifications de l'ARN après analyse des nucléosides qui le compose. Le second niveau est, quant à lui, basé sur une analyse des oligonucléotides obtenus après digestion spécifique des ARN. Il permet de séquencer les oligonucléotides et ainsi localiser les modifications post-transcriptionnelles. Les travaux présentés dans ce manuscrit sont en très grande partie focalisés sur le premier niveau d'analyse. En effet cette application permet de travailler dans un premier temps avec des échantillons simples et de monter petit à petit en complexité.

Le couplage CE-FTICR-MS étant délicat et tout nouveau au laboratoire, il a d'abord fallu le mettre en place et l'"apprivoiser". Ce travail est décrit dans la deuxième partie de ce manuscrit. Après quelques essais de transposition de méthodes avec l'analyse d'anticorps monoclonaux (par approches bottom-up, intact, et middle-up), le besoin de revenir à une application plus simple pour appréhender les difficultés de ce couplage s'est fait ressentir.

Ainsi, les débuts du couplage CE-FTICR-MS pour la caractérisation des modifications post-transcriptionnelles des ARN ont été menés sur un mélange de standards des quatre nucléosides canoniques. Après optimisation de la méthode de détection pour l'analyse des nucléosides, une étude des performances du couplage a été menée afin de trouver le bon compromis entre, d'une part la résolution et la précision de mesure de masse du FTICR, et d'autre part la vitesse d'acquisition. Afin d'obtenir des temps de cycle compatibles avec la séparation par CE, il a été nécessaire de "sacrifier" une partie de la résolution et de la précision de masse du FTICR. Cependant, malgré ce sacrifice, la résolution et la précision de mesure de masse atteignables avec le couplage CE-FTICR-MS restent plus importantes que celles du couplage CE-QTOF-MS plus couramment utilisé. Cette étude a également permis d'acquérir de l'expérience et de mieux en appréhender les particularités et de pallier les difficultés intrinsèques de ce délicat couplage au comportement parfois capricieux. Ce dernier a ensuite été évalué sur de vrais échantillons biologiques bien connus et référencés dans la littérature. Deux niveaux de complexité ont été évalués : tout d'abord, des ARNt purs issus de *S. cerevisiae* contenant chacun une dizaine de modifications, puis un extrait total d'ARNt issu de *B. taurus* contenant une trentaine de modifications variées. Hormis pour la pseudouridine (ψ) et ses dérivés (tel que ψm), des résultats similaires ont été obtenus sur les deux couplages CE-QTOF-MS et CE-FTICR-MS en termes de nombre de modifications identifiées. Pour pallier l'absence de sélection pour la fragmentation de certains nucléosides lorsque beaucoup d'entre eux co-migrent, il est envisageable de combiner plusieurs analyses en adaptant les paramètres d'exclusion dynamique de la MS/MS. Ces premiers résultats biologiques ont permis de démontrer que l'utilisation du couplage CE-FTICR-MS est possible pour la caractérisation des modifications post-transcriptionnelles au niveau nucléosides.

Après le succès du transfert de méthode de la CE-QTOFMS à la CE-FTICR-MS pour des échantillons biologiques, de nouveaux développements autour de cette méthode ainsi que l'analyse d'échantillons encore plus complexes, ont pu être envisagés. Ces développements ont fait l'objet de la troisième partie de ce manuscrit. Le premier développement s'est focalisé sur l'optimisation de l'étape de déphosphorylation lors de la préparation des échantillons d'ARN pour l'analyse des nucléosides. Aucun consensus ne semblait se dégager dans la littérature pour départager les différentes méthodes mises en place pour cette préparation d'échantillon. Cette optimisation a débouché sur la proposition de deux protocoles permettant d'obtenir une déphosphorylation complète des échantillons. Le premier est rapide, mais demande une grande quantité d'enzyme, et le second est plus long, mais plus économique en enzyme. Les deux niveaux de la méthode CE-MS ont ensuite été testés sur un échantillon bien plus complexe dont les modifications ne sont pas décrites dans la littérature : l'ARNr 16 S pur de *P. aeruginosa*. Il a été essentiel de réaliser ces analyses en couplage CE-QTOF-MS avant de pouvoir envisager le transfert de cette application en couplage CE-FTICR-MS. Grâce à l'analyse des nucléosides, une dizaine de modifications post-transcriptionnelles a pu être identifiée dans cet échantillon. L'analyse des oligonucléotides a permis de séquencer quelques oligonucléotides non modifiés. Cependant, la localisation des modifications,

préalablement identifiées par l'analyse des nucléosides, s'est révélée impossible sans l'aide d'outils bio-informatiques. C'est ainsi que les travaux décrits dans cette partie se sont ensuite dirigés vers le développement d'outils bio-informatiques pour caractériser les modifications post-transcriptionnelles des ARN à partir d'analyses MS/MS en mode positif. Deux logiciels, un pour chaque niveau d'analyse, ont été développés. Le premier, Nucleos'ID, permet d'identifier les modifications post-transcriptionnelles de manière non-ciblée, à partir d'analyses par CE-MS/MS ou LC-MS/MS. D'une interprétation manuelle non ciblée de plusieurs heures à une recherche dont les résultats sont instantanés, le gain de temps apporté par ce nouvel outil bio-informatique open-source est indéniable. Le second, OligoMod+, est le premier outil bio-informatique open-source entièrement pensé pour interpréter les analyses CE-MS/MS d'oligonucléotides en mode positif. Il permet actuellement de retrouver la position des modifications post-transcriptionnelles sur la séquence de l'ARN lorsque celles-ci sont connues. À long terme, il permettra de localiser les modifications post-transcriptionnelles des ARN lorsque leurs positions ne sont pas connues. Il sera alors possible de reprendre l'analyse complexe des oligonucléotides de l'ARNr de *P. aeruginosa*.

Enfin, tous les travaux précédents ont permis de mettre en application le couplage CE-FTICR-MS sur des échantillons biologiques dont les modifications post-transcriptionnelles n'étaient pas, ou pas entièrement, référencées dans la littérature. Dans un premier temps, les nucléosides de l'ARNr étudiés en couplage CE-QTOF-MS dans la partie précédente ont été analysés en couplage CE-FTICR-MS. Bien que le couplage du FTICR avec une méthode séparative telle que la CE permette d'augmenter la gamme dynamique, la différence de sensibilité entre les deux couplages et la complexité de l'échantillon n'ont pas permis d'obtenir d'aussi bons résultats que ceux du couplage CE-QTOF-MS. Cependant, de nombreuses modifications ont pu, tout de même, être identifiées par le couplage CE-FTICR-MS. La deuxième mise en application est l'analyse d'un extrait total d'ARNt de *T. thermophilus*, une bactérie extrémophile. Cette fois-ci, en plus d'identifier les modifications de cet extrait parmi toutes celles décrites dans la littérature et référencées dans les bases de données pour les eubactéries, une modification inconnue a été identifiée. Grâce à la combinaison de la séparation électrophorétique, de l'augmentation de la gamme dynamique et de la grande précision de mesure de masse unique au couplage CE-FTICR-MS, des formules brutes uniques ont pu être assignées à la modification et à son produit issu de la fragmentation MS/MS. À titre de comparaison, alors que pour les échantillons complexes l'identification de modifications référencées dans les bases de données est meilleure avec le couplage CE-QTOF-MS, ce dernier n'avait pas permis d'identifier sans ambiguïté cette modification.

Par la suite, plusieurs pistes de développement de méthode pour la caractérisation des modifications post-transcriptionnelles pourront être envisagées. Tout d'abord, il sera essentiel de finaliser le logiciel OligoMod+ pour être capable de localiser de manière non ciblée les modifications sur des ARN lorsque les positions de celles-ci ne sont pas connues. Ensuite, des développements au niveau de la CE sont envisageables. Parmi eux, l'utilisation de capillaires greffés pour la séparation pourra être testée pour améliorer la détection des gros

oligonucléotides. En effet, la méthode CE-MS actuelle ne permet pas de bien détecter les gros oligonucléotides puisque ceux-ci sont souvent chargés négativement. D'autre part, à ce jour, la fragmentation MS/MS par CID permet de discriminer la position de la modification entre le ribose et la base du nucléoside, mais pas de localiser précisément la modification sur la base. Il serait donc intéressant de développer une méthode de fragmentation MS/MS permettant de discriminer la position des modifications sur la base nucléotidique. Associée à la séparation des isomères par CE, une telle méthode permettrait d'approfondir la caractérisation des modifications post-transcriptionnelles. Pour finir, d'un point de vue du couplage CE-FTICR-MS, maintenant qu'il est mis en place et que ses capacités ont été démontrées, il va être nécessaire de choisir précautionneusement les prochaines applications. En effet, ce couplage étant délicat à mettre en œuvre, et parfois en difficulté lorsqu'il s'agit de fragmenter trop de composés simultanément (co-migrations), il faut savoir ne le réserver qu'aux applications requérant absolument sa précision en masse inégalée et son ultra haute résolution. Les différents travaux détaillés dans ce manuscrit montrent l'importance de choisir les bonnes applications pour le couplage CE-FTICR-MS, telles que l'identification de modifications inconnues. Cela permet de profiter au maximum de ses avantages comme la séparation électrophorétique d'isomères, l'ultra-haute résolution et la grande précision de mesure de masse.

LISTE DES COMMUNICATIONS SCIENTIFIQUES ET AUTRES RESPONSABILITES

PUBLICATIONS À COMITÉ DE LECTURE

- **Clarisse Gosset-Erard**, Zhao M., Lordel-Madeleine S., Ennahar S., “Identification of punicalagin as the bioactive compound behind the antimicrobial activity of pomegranate (*Punica granatum* L.) peels.”, 2021, *Food Chemistry*, 352, 129396, [doi:10.1016/j.foodchem.2021.129396](https://doi.org/10.1016/j.foodchem.2021.129396).
- **Clarisse Gosset-Erard**, Antony Lechner, Philippe Wolff, Frédéric Aubriet, Emmanuelle Leize-Wagner, Chaimbault P. and Yannis-Nicolas François, “Optimization of nucleotides dephosphorylation for RNA structural characterization by tandem mass spectrometry hyphenated with separation methods”, 2022, *Journal of Chromatography B*, 1208, 123396, [doi:10.1016/j.jchromb.2022.123396](https://doi.org/10.1016/j.jchromb.2022.123396).
- **Clarisse Gosset-Erard**, Mévie Didierjean, Jérôme Pansanel, Antony Lechner, Philippe Wolff, Lauriane Kuhn, Frédéric Aubriet, Emmanuelle Leize-Wagner, Patrick Chaimbault and Yannis-Nicolas François, “Nucleos’ID: new search engine enabling the untargeted identification of RNA post-transcriptional modifications from tandem mass spectrometry analyses of nucleosides”, 2023, *Analytical Chemistry*, In Press, [doi:10.1021/acs.analchem.2c04722](https://doi.org/10.1021/acs.analchem.2c04722).

PUBLICATION SOUMISE

- **Clarisse Gosset-Erard**, Frédéric Aubriet, Emmanuelle Leize-Wagner, Yannis-Nicolas François and Patrick Chaimbault, “Hyphenation of Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FTICR-MS) with separation methods: the art of compromise and the possible - a review”, submitted in *Talanta*, under minor revision

CHAPITRE DE LIVRE

- **Clarisse Gosset-Erard**, Jérémie Giorgetti, Michael Biacchi, Frédéric Aubriet, Emmanuelle Leize-Wagner, Chaimbault P. and Yannis-Nicolas François, “Complete etching process of bare-fused silica capillaries for sheathless capillary electrophoresis interface with mass spectrometry”, in: Neusüß, C., Jooß, K. (Eds.), *Capillary Electrophoresis-Mass Spectrometry: Methods and Protocols, Methods in Molecular Biology*, Springer, 2531, Humana, New York, NY, [doi:10.1007/978-1-0716-2493-7_3](https://doi.org/10.1007/978-1-0716-2493-7_3).

COMMUNICATIONS ORALES

Auteurs : Clarisse Gosset-Erard, Antony Lechner, Philippe Wolff, Frédéric Aubriet, Emmanuelle Leize-Wagner, Patrick Chaimbault, Yannis-Nicolas François

Titre : « Characterization of post-transcriptional RNA modifications by sheathless capillary electrophoresis – mass spectrometry»

Congrès : XXVèmes Rencontres du Club Jeune de la Société Française de Spectrométrie de Masse - en distanciel, le 30 mars 2021

Auteurs : Clarisse Gosset Erard, Antony Lechner, Patrick Chaimbault, Yannis-Nicolas François

Titre : « Characterization of post transcriptional RNA modifications by capillary electrophoresis - high resolution mass spectrometry »

Congrès : 5ème Journée Thématique du Club de Chromatographie Ouest Atlantique CCOA de l'Association Francophone des Sciences Séparatives : Analyse des composés polaires - en distanciel, le 8 avril 2021

Auteurs : Clarisse Gosset-Erard, Antony Lechner, Philippe Wolff, Frédéric Aubriet, Emmanuelle Leize-Wagner, Patrick Chaimbault, Yannis-Nicolas François

Titre : « Characterization of post transcriptional RNA modifications by capillary electrophoresis - high resolution mass spectrometry »

Congrès : UMR 7140 Scientific Day 2021 Edition – en distanciel, le 5 mai 2021

Liste des communications scientifiques et autres responsabilités

Auteurs : Clarisse Gosset-Erard, Antony Lechner, Philippe Wolff, Emmanuelle Leize-Wagner, Frédéric Aubriet, Yannis-Nicolas François, Patrick Chaimbault

Titre : « Characterization of post transcriptional RNA modifications by sheathless capillary electrophoresis - high resolution mass spectrometry »

Congrès : JFSM 2021 – en distanciel, le 14 juin 2021

Auteurs : Clarisse Gosset-Erard, Antony Lechner, Philippe Wolff, Emmanuelle Leize-Wagner, Frédéric Aubriet, Yannis-Nicolas François, Patrick Chaimbault

Titre : « Characterization of post-transcriptional RNA modifications by sheathless capillary electrophoresis - high resolution mass spectrometry »

Congrès : Journée des jeunes chercheurs du LCP-A2MC – Metz, le 8 Juillet 2021

Auteurs : Clarisse Gosset-Erard, Antony Lechner, Philippe Wolff, Emmanuelle Leize-Wagner, Frédéric Aubriet, Yannis-Nicolas François, Patrick Chaimbault

Titre : « Caractérisation des modifications post-transcriptionnelles de l'ARN par électrophorèse capillaire couplée à la spectrométrie de masse »

Congrès : SEP21 – Paris, le 7 Octobre 2021

Auteurs : Clarisse Gosset-Erard, Antony Lechner, Philippe Wolff, Emmanuelle Leize-Wagner, Frédéric Aubriet, Yannis-Nicolas François, Patrick Chaimbault

Titre : « Characterization of post-transcriptional RNA modifications by sheathless capillary electrophoresis - high resolution mass spectrometry »

Congrès : 69th ASMS Conference – Philadelphie, le 3 Novembre 2021

Auteurs : Clarisse Gosset-Erard, Emmanuelle Leize-Wagner, Frédéric Aubriet, Yannis-Nicolas François, Patrick Chaimbault

Titre : « Mise en place du couplage CE-FTICR-MS pour l'étude de biomolécules à visée thérapeutique »

Congrès : 2ème rencontre des doctorants du groupe CE de l'AFSEP – en distanciel, le 15 Novembre 2021

Liste des communications scientifiques et autres responsabilités

Auteurs : Clarisso Gosset-Erard, Antony Lechner, Philippe Wolff, Emmanuelle Leize-Wagner, Frédéric Aubriet, Yannis-Nicolas François, Patrick Chaimbault

Titre : « Characterization of post-transcriptional RNA modifications by sheathless capillary electrophoresis - high resolution mass spectrometry »

Congrès : Séminaire de l'école doctorale C2MP – Nancy, le 16 Juin 2022

Auteurs : Clarisso Gosset-Erard, Antony Lechner, Philippe Wolff, Frédéric Aubriet, Emmanuelle Leize-Wagner, Patrick Chaimbault, Yannis-Nicolas François

Titre : « Optimization of nucleotides dephosphorylation for the characterization of RNA post-transcriptional modifications »

Congrès : Journée des jeunes chercheurs du LCP-A2MC – Metz, le 28 Juin 2022

Auteurs : Clarisso Gosset-Erard, Mévie Didierjean, Lauriane Kuhn, Antony Lechner, Philippe Wolff, Jérôme Pansanel, Frédéric Aubriet, Emmanuelle Leize-Wagner, Patrick Chaimbault and Yannis-Nicolas François

Titre : « Two new search engines enabling the identification and localization of RNA post-transcriptional modifications from CE-MS/MS analyses »

Congrès : Analytics 2022 – Nantes, le 7 Septembre 2022

COMMUNICATION PAR AFFICHE

Auteurs : Clarisso Gosset-Erard, Antony Lechner, Philippe Wolff, Frédéric Aubriet, Emmanuelle Leize-Wagner, Patrick Chaimbault, Yannis-Nicolas François

Titre : « Optimization of nucleotides dephosphorylation for the characterization of RNA post-transcriptional modifications »

Congrès : Analytics 2022 – Nantes, 5-8 Septembre 2022

ENSEIGNEMENTS DELIVRES

Année universitaire 2019 – 2020 (18 HETD)

Travaux pratiques de « Chimie des solutions » - Licence 1 Physique-Chimie – Université de Lorraine

Année universitaire 2020 – 2021 (64 HETD)

Travaux pratiques de « Chimie analytique » - Licence 3 Chimie-Biologie – Université de Strasbourg

Travaux pratiques de « Chimie organique - partie analytique » - Licence 2 Chimie – Université de Strasbourg

Année universitaire 2021 – 2022 (64 HETD)

Travaux pratiques de « Chimie analytique » - Licence 3 Chimie-Biologie – Université de Strasbourg

Travaux pratiques de « Chimie organique - partie analytique » - Licence 2 Chimie – Université de Strasbourg

AUTRES RESPONSABILITES

Elue représentante des doctorants au conseil du LCP-A2MC depuis mars 2021

Co-encadrement d'une stagiaire M1 d'avril à juin 2021

Co-encadrement d'une stagiaire M2 de janvier à juin 2022

ANNEXES

ANNEXE 1 : CHAPITRE DE LIVRE SUR L'ATTAQUE DES CAPILLAIRES

Complete etching process of bare-fused silica capillaries for sheathless capillary electrophoresis interface with mass spectrometry

Clarisso Gosset-Erard, Jérémie Giorgetti, Michael Biacchi, Frédéric Aubriet, Emmanuelle Leize-Wagner, Chaimbault P. and Yannis-Nicolas François

[DOI:10.1007/978-1-0716-2493-7_3.](https://doi.org/10.1007/978-1-0716-2493-7_3)

A1.1. Abstract

HOMEMADE capillaries are a very common practice for users of capillary electrophoresis (CE), notably in CE-UV. With the advent of the capillary electrophoresis—mass spectrometry coupling since the end of the eighties, several interfaces have been developed. Among those interfaces, the porous tip sprayer allows great sensitivity at nano flow rates and has been used in numerous applications over the past few years. However, the homemade implementation of a suitable capillary for the porous tip sprayer is more challenging. The porous tip is created by etching the bare-fused silica capillary with hydrofluoric acid. Here we describe the complete process of etching bare-fused silica capillaries, from length cutting to quality control of the newly etched capillary.

A1.2. Introduction

Since the development of capillary electrophoresis (CE) in the early 1980s [1], this separation technique has been recognized for its experimental flexibility, particularly with the nature and dimensions of the capillary. In most stand-alone CE studies, capillaries are typically homemade even if some are commercially available. Whereas self-packed columns represent a tiny proportion of studies in liquid chromatography (LC) in literature, homemade CE capillaries are much more widespread. Indeed, nothing is as easy as setting up a CE-UV capillary. It only requires cutting the capillary at the right length and creating a UV window by burning the polyacrylamide sheath at a few centimeters from the capillary outlet. Since the beginning of the development of capillary electrophoresis—mass spectrometry (CE-MS) coupling at the end of the eighties [2], the design of the interface as well as the modifications of the capillary have evolved considerably [3]. The sheath-liquid design is the most common interface, notably due to its robustness and simplicity of use and implementation. The sheath-

liquid interface consists of a standard capillary surrounded by a coaxial tube filled with a conductive sheath-liquid, all within a nebulizer. The background electrolyte (BGE) and the sheath-liquid merge together at the tip of the capillary, enabling the electrical contact [4]. This interface design does not require any modification of the capillary and the only essential step is to cut the capillary at the right length and set it up directly into the interface.

Although the sheath-liquid interface is robust and easy to use, recent applications require more sensitivity due to increasing complexity and small amounts of sample. Several designs of liquid junction or sheathless interfaces have been developed such as the conductive tip coating design [5–7], the conductive wire inserted in the capillary [8,9] and the porous tip sprayer [10]. Those interfaces involve more complex physicochemical processes and therefore, their homemade implementation is not as easy as for the sheath-liquid interface. With such interfaces, due to the lack of detailed and straightforward protocols for manufacturing homemade capillaries, people tend to use commercially available capillaries for convenience.

A1.3. Materials

A1.3.1. Safety

The use of 49 % HF solution imposes strict safety rules as well as controlled experimental conditions. All experiments must be performed in Teflon or polypropylene (PP) containers, in a well-ventilated fume hood and with personal safety equipment.

1. Personal safety equipment: full-face mask, PP apron and neoprene gloves.
2. Calcium D-gluconate monohydrate was prepared as ointment in the drugstore.
3. Trivorex® was purchased from Merck (Darmstadt, Germany).

A1.3.2. Chemicals and solutions

Except for HF, prepare all solutions using ultrapure water (prepared by purifying deionized water, to reach a sensitivity of $18.2 \text{ M}\Omega/\text{cm}$ at 25°C). Chemicals used were of analytical grade or high purity grade: 49 % HF solution; acetic acid (HAc, > 99 %); sodium bicarbonate. All solutions can be stored at room temperature.

1. Etching solution: 49 % HF is placed without any treatment in a Teflon container.
2. BGE: 10 % HAc solution. 2.5 mL of HAc (> 99%) in a 25 mL volumetric flask.
3. Neutralization solution: Homogenized saturated solution of sodium bicarbonate. Add about 100 mL of water to a Teflon bottle. Add sodium bicarbonate salt until complete saturation. Homogenize.
4. Ultrapure water.

A1.3.3. Instrumentation and equipment

1. Polymicro Technologies-coil of bare fused silica capillaries (30 µm inner diameter (ID), 150 µm outer diameter (OD)).
2. A cleaving stone.
3. A bubbling system composed of a syringe filled with air and a syringe driver.
4. A PP or Teflon collecting tray (see **Note 1**).
5. A PA 800 Plus system (Sciex) was hyphenated using a sheathless nanoelectrospray (nanoESI) interface (see **Note 2**).
6. A maXis 4G mass spectrometer (Bruker Daltonics) equipped with a hybrid analyzer composed of a quadrupole followed by a time-of flight (TOF) analyzers (see **Note 3**).

A1.4. Methods

The following protocol details the different steps for etching bare-fused silica capillaries. It includes safety precautions, initial setting, etching workflow, and quality control of new etched capillaries.

A1.4.1. Safety precautions and initial setting

1. Wear a cotton smock, a PP apron, a full-face mask and neoprene gloves (see **Note 4**).
2. Keep a calcium gluconate ointment nearby during all the etching process (see **Note 5**).
3. Place the Teflon collecting tray in a well-ventilated fume hood.
4. Put a compact layer of Trivorex® in the Teflon collecting tray to cover up its entire surface. All the etching process will be done over this collecting tray (see **Note 6**).
5. In the Teflon collecting tray, place the HF solution, the saturated solution of sodium bicarbonate, and the water as shown in Figure A1.1.
6. Note the room temperature (see **Note 7**).

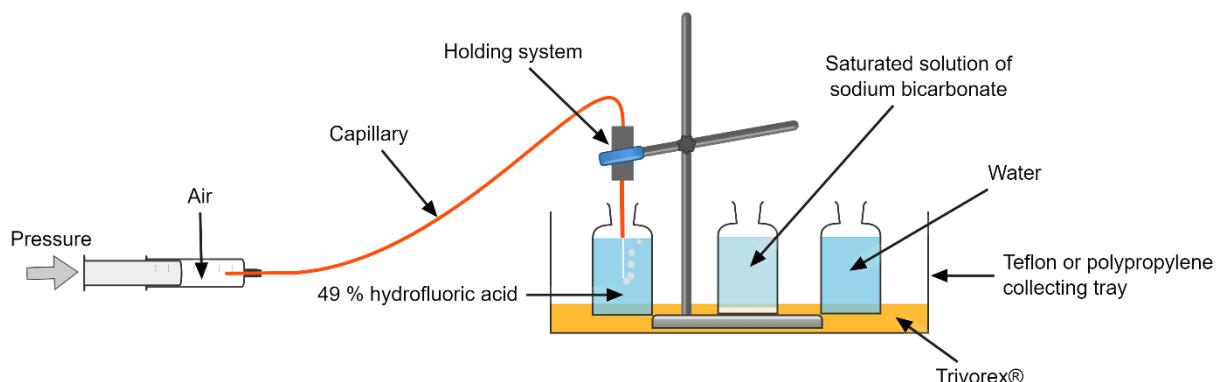


Figure A1.1: Initial etching set up.

A1.4.2. Etching workflow

The reaction of silica with HF releases heat. Thin capillaries are very sensitive to heat, which can cause weakening of the tip. Therefore, the etching must be performed in multiple

stages to prevent the tip from breaking. Between each stage, the reaction must be stopped by two neutralizing and rinsing baths.

1. Cut the desired length of the capillary (see Figure A1.2a), generally between 60 and 100 cm with the cleaving stone (see **Note 8**).
2. At approximately 2-3 cm from one end of the capillary (outlet), create a polyacrylamide-free tip (see **Note 9**) by burning the polyacrylamide sheath with a lighter as shown in Figure A1.2b (see **Note 10**). Take off the burned polyacrylamide with a low-lint wipe soaked in methanol (MeOH) or ethanol (see **Note 11**). The tip is now bare as shown in Figure A1.2c.

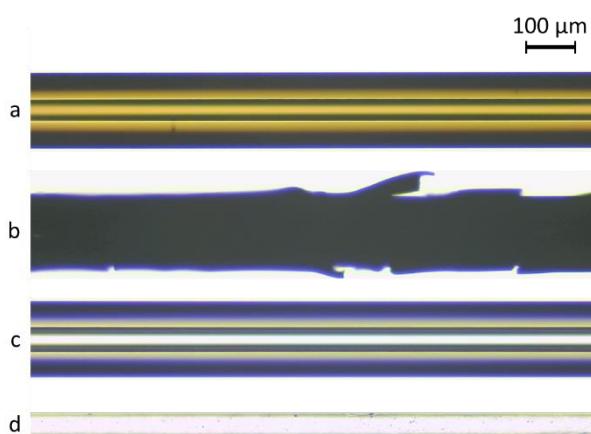


Figure A1.2: Capillaries of 30 μm ID and 150 μm OD (a) with the polyacrylamide sheath, (b) with the burned polyacrylamide sheath, (c) without the polyacrylamide sheath, and (d) after complete etching. All were observed under a microscope.

3. Place the polyacrylamide-free tip in a holding system as shown in Figure A1.1.
4. The other end of the capillary (inlet) must be attached with a tight seal to a bubbling system (see **Note 12**).
5. Under bubbling condition, dive the outlet bare tip of the capillary in HF for the first etching stage (see **Note 13**). The duration of each etching stage is depending on the temperature, refer to Table A1.1 (see **Note 14**).

Table A1.1: Duration of each etching stage in HF depending on the temperature in the fume hood.

Temperature ($^{\circ}\text{C}$)	Duration of stage 1 (min)	Duration of stage 2 (min)	Duration of stage 3 (min)
16	30	10	5
17	28	10	5
18	27	10	5
19	26	10	4
20	26	9	4
21	25	9	4
22	23	9	4

6. While preserving the bubbling, transfer the outlet bare tip of the capillary to the neutralization solution (saturated sodium bicarbonate) for 3 min (see **Note 15**).
7. While preserving the bubbling, transfer outlet bare tip of the capillary to the deionized water for 3 min as a washing step.
8. For the second etching stage, repeat the protocol detailed in step 5 to 7, by taking into consideration the duration of stage 2 (see **Note 13**) detailed in Table A1.1 (see **Note 14**).
9. For the third etching stage, repeat the protocol detailed in step 5 to 7, by taking into consideration the duration of stage 3 (see **Note 13**) detailed in Table A1.1 (see **Note 14**).
10. Rinse thoroughly the porous tip (see Figure A1.2d) with a squeezed bottle of deionized water (see **Note 16**). Take off the bubbling and the holding systems carefully from the capillary.
11. Install the porous tip capillary in the CE-MS cartridge (see **Note 17**).

A1.4.3. Quality control

To control the quality of the newly etched capillary, its performance (current/voltage delivery and CE-MS electrospray formation) must be tested. For this purpose, 10 % HAc is used as BGE and voltage and current values are given for a capillary of 30 µm ID, 150 µm OD and 90 cm total length (see **Note 18**).

1. Install the cartridge in your CE system and position the porous tip in front of the MS inlet (see **Note 19**).
2. Fill up the conductive line with BGE by applying a reversed pressure of 50 psi for 5 min. A droplet of BGE must appear at the end of the stainless-steel reservoir as shown in Figure A1.3 (see **Note 20**).

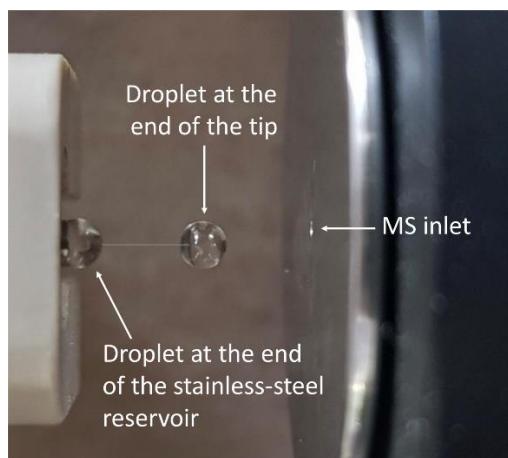


Figure A1.3: Porous tip in front of a MS inlet with droplets at the end of the porous tip and at the end of the stainless-steel reservoir.

3. Fill up the separation capillary with BGE (see **Note 21**) by applying a forward pressure of 50 psi for 3 min. A droplet of BGE must appear at the end of the porous tip as shown in Figure A1.3. (see **Note 22**).
4. Set the CE voltage to +20 kV.
5. Check the CE current. It must be stable with a value around 2.0-2.5 µA for a capillary of 90 cm total length (see **Note 18**).
6. Check the presence of the droplet at the end of the porous tip under the CE electrical field when the electrospray source voltage is set to 0 V (see **Note 23**) as shown in Figure A1.3.
7. Set the electrospray source voltage to 1500 V. The droplet should disappear, replaced by a spray (see **Note 24**). Check the stability of the electrospray (see **Note 25**).
8. If all the previous steps are passed, it can be considered that the etching was performed well enough to use the capillary for future analyses (see **Note 26**).
9. For overnight storage, keep the cartridge in the CE system but dive the porous tip in water (see **Note 27**). For long term storage, empty the separation capillary by applying a pressure of 50 psi for 3 min with air and take off the cartridge of the CE system.

A1.5. Notes

1. Never use glass material when working with HF. As glass is composed of silica, HF will etch any glass material. Teflon or PP materials are resistant to HF.
2. Any other CE system adapted to CE-MS hyphenation can be used.
3. Any other MS instrument can be used as long as it is equipped with a nanoESI source.
4. Due to the dangerousness of HF, safety precautions must be rigorously considered. Indeed, HF gnaws bones by reacting with calcium, inducing severe damages. If there is any contact with HF, no matter the size and amount, immediately seek help and visit an emergency doctor.
5. The calcium gluconate ointment must be stored in the refrigerator and used in case HF contacts the skin to prevent bones gnawing while waiting for medical advice.
6. Trivorex® is an absorbent used for neutralization of HF in case of spill in the Teflon collecting tray or on the bench.
7. The temperature must be precisely measured in the fume hood where the etching process is taking place. This parameter is essential for determining the reaction kinetics of acid etching.
8. It is also possible to cut very precisely and accurately the capillary without damaging using a diamond cutter. However, for a capillary of 150 µm OD, it is more complicated than using a cleaving stone.
9. HF only reacts with silica. Therefore, removing the polyacrylamide sheath of the capillary enables HF to access the silica.
10. Burn the polyacrylamide sheath in the high part of the lighter flame. As soon as the sheath becomes red, stop the flame. If the burning is too long, silica will be weakened, leading to potential breaking of the tip.

11. When removing the burned polyacrylamide, be careful to remove all the burned acrylamide parts to ensure homogeneity of the etching.
12. The airflow from the bubbling prevents HF from going into the capillary and thus etching the inner wall.
13. The duration for each etching stage has been optimized for a capillary of 30 µm inner diameter and 150 µm outer diameter, which corresponds to a good compromise between the low outer diameter value and the robustness of the capillary. However, if capillaries with a different outer diameter are used, the timing of each step must be adapted and optimized accordingly.
14. The duration for each etching step has been optimized using fresh HF. However, the reaction between silica and HF releases water over time. Thus, dilution effects can affect the kinetic of the reaction. Consequently, after six months, it can be necessary to extend each etching step duration from 30 s to 1 min or to replace the HF solution.
15. The saturated solution of sodium bicarbonate is used to neutralize the HF and therefore stop the etching process.
16. The water jet must be in the same axis as the porous tip to prevent breaking of the porous tip.
17. CE-MS cartridges with porous tip capillaries are commercially available by Sciex (Brea, CA, USA). However, commercial cartridges are closed systems, making it difficult to change the capillary. Otherwise, you can use your typical CE cartridge and adapt it with a stainless-steel reservoir, in which the porous tip of the separation capillary meets a conductive line. The stainless-steel reservoir can be made of a classical ESI needle emitter [10] or you can build your own system with a T-junction. The conductive line is typically a bare-fused silica capillary of 50 µm ID and 375 µm OD, and 10–20 cm shorter than the separation capillary.
18. The stated current values are expected for the setup described in our protocol. Other BGEs, capillary length, or voltage can be used to perform quality control of the etching process. Consequently, the expected current value will differ accordingly. Note that a current over 10 µA can damage the porous tip due to joule heating effects.
19. The porous tip must be placed a few millimeters away from the MS inlet. Use the three ultra-fine adjustment screws to precise control of the distance between the porous tip and the MS inlet. Depending on the MS system, the tip can be placed in the same axis as the MS inlet or slightly angled.
20. If no droplet appears when filling the conductive capillary, three diagnostic tests can be performed: (i) the conductive capillary might be incorrectly installed in the CE-MS cartridge. Check the interface between the conductive capillary and the stainless-steel reservoir; (ii) the conductive capillary might be clogged. Generally, the clogging process appears at the extremity of the capillary. Cut the inlet of the separation capillary (1 or 2 mm) with the cleaving stone and fill up the conductive line with BGE by applying a reversed pressure of 50 psi for 5 min. If still no droplet appears, cut the outlet of the capillary (1 or 2 mm) with the cleaving stone and repeat the test procedure; (iii) if no

droplet appears after cutting both ends, it means that the capillary is broken. Put a new one in the cartridge and then repeat the test procedure from the beginning.

21. To maintain electrical contact, both capillaries must be filled with the same BGE.
22. If no droplet appears when filling the separation capillary, two diagnostic tests can be performed: (i) the separation capillary might be clogged. Generally, the clogging process appears at the extremity of the capillary. Cut the inlet of the separation capillary (1 or 2 mm) with the cleaving stone and fill up the capillary with BGE by applying a forward pressure of 50 psi for 3 min. If still no droplet appears, cut the porous tip of the capillary (1 or 2 mm) with the cleaving stone and repeat the test procedure; (ii) if no droplet appears after cutting both ends, it means that the capillary is broken. Put a new one in the cartridge and then repeat the test procedure from the beginning.
23. If the droplet at the end of the porous tip disappears and electrospray is observed when the CE voltage is applied (+20 kV) and ESI voltage is set to 0 V, the porous tip is not etched enough. Indeed, the porosity is not sufficient to allow the two distinct CE and MS electric fields. The appearance of the spray can be explained by the fact that the ground earth of the CE electric field is shared with the MS one. Therefore, a fourth etching step of 30 s to 1 min, followed by a neutralization of 3 min in saturated sodium bicarbonate and a rinse of 3 min in water, is necessary.
24. If no electrospray is observed with 1500 V, optimize the position of the porous tip in front of the MS inlet and gradually increase the electrospray source voltage up to 2000 V until the formation of a Taylor cone. As soon as a stable spray is observed, stop increasing the electrospray source voltage.
25. If the electrospray is not stable, perform a conditioning protocol by successive flushes with MeOH, 0.1 M NaOH, water, 0.1 M HCl, water and then with BGE under a pressure of 50 psi for 10 min each. After conditioning, repeat the test procedure.
26. Measuring the wall thickness of the porous tip with a microscope is an alternative way to confirm the quality of the etching. The capillary wall is porous when its thickness is approximately 5–10 µm [10]. It is also possible to observe the porous tip wall surface using atomic force microscopy and to monitor ion transfer through the porous wall using inverted fluorescence microscopy [20].
27. Keeping the porous tip in water overnight prevent clogging phenomenon.

A1.6. References

- [1] Jorgenson, J. *et al.*, “Capillary Zone Electrophoresis.”, *Science*, **1983**, 222 (4621), 266–272.
- [2] Smith, R. D. *et al.*, “Improved Electrospray Ionization Interface for Capillary Zone Electrophoresis-Mass Spectrometry.”, *Anal. Chem.*, **1988**, 60 (18), 1948–1952.
- [3] Stoltz, A. *et al.*, “Recent Advances in Capillary Electrophoresis-Mass Spectrometry: Instrumentation, Methodology and Applications.”, *ELECTROPHORESIS*, **2019**, 40 (1), 79–112.

- [4] Hommerson, P. *et al.*, "Ionization Techniques in Capillary Electrophoresis-Mass Spectrometry: Principles, Design, and Application.", *Mass Spectrometry Reviews*, **2011**, *30* (6), 1096–1120.
- [5] Nilsson, S. *et al.*, "A Simple and Robust Conductive Graphite Coating for Sheathless Electrospray Emitters Used in Capillary Electrophoresis/Mass Spectrometry.", *Rapid Communications in Mass Spectrometry*, **2001**, *15* (21), 1997–2000.
- [6] Ramsey, R. S. *et al.*, "Capillary Electrophoresis/Electrospray Ionization Ion Trap Mass Spectrometry Using a Sheathless Interface.", *Journal of Microcolumn Separations*, **1995**, *7* (5), 461–469.
- [7] Zamfir, A. D. *et al.*, "Copper-Coated Microsprayer Interface for on-Line Sheathless Capillary Electrophoresis Electrospray Mass Spectrometry of Carbohydrates.", *Journal of Separation Science*, **2006**, *29* (3), 414–422.
- [8] Fang, L. *et al.*, "Online Time-of-Flight Mass Spectrometric Analysis of Peptides Separated by Capillary Electrophoresis.", **1994**.
- [9] Cao, P. *et al.*, "A Novel Sheathless Interface for Capillary Electrophoresis/Electrospray Ionization Mass Spectrometry Using an in-Capillary Electrode.", *Journal of the American Society for Mass Spectrometry*, **1997**, *8* (5), 561–564.
- [10] Moini, M., "Simplifying CE-MS Operation. 2. Interfacing Low-Flow Separation Techniques to Mass Spectrometry Using a Porous Tip.", *Anal. Chem.*, **2007**, *79* (11), 4241–4246.
- [11] Belov, A. M. *et al.*, "Analysis of Proteins, Protein Complexes, and Organellar Proteomes Using Sheathless Capillary Zone Electrophoresis - Native Mass Spectrometry.", *J. Am. Soc. Mass Spectrom.*, **2017**, *28* (12), 2614–2634.
- [12] Giorgetti, J. *et al.*, "Combination of Intact, Middle-up and Bottom-up Levels to Characterize 7 Therapeutic Monoclonal Antibodies by Capillary Electrophoresis – Mass Spectrometry.", *Journal of Pharmaceutical and Biomedical Analysis*, **2020**, *182*, 113107.
- [13] Faserl, K. *et al.*, "Investigating Capillary Electrophoresis-Mass Spectrometry for the Analysis of Common Post-Translational Modifications.", *ELECTROPHORESIS*, **2018**, *39* (9–10), 1208–1215.
- [14] Ibrahim, M. *et al.*, "Improvement of Mitochondria Extract from *Saccharomyces Cerevisiae* Characterization in Shotgun Proteomics Using Sheathless Capillary Electrophoresis Coupled to Tandem Mass Spectrometry.", *J Chromatogr Sci*, **2016**, *54* (4), 653–663.
- [15] Gulersonmez, M. C. *et al.*, "Sheathless Capillary Electrophoresis-Mass Spectrometry for Anionic Metabolic Profiling: Liquid Phase Separations.", *ELECTROPHORESIS*, **2016**, *37* (7–8), 1007–1014.
- [16] Drouin, N. *et al.*, "Capillary Electrophoresis-Mass Spectrometry at Trial by Metabo-Ring: Effective Electrophoretic Mobility for Reproducible and Robust Compound Annotation.", *Anal. Chem.*, **2020**, *92* (20), 14103–14112.
- [17] Lechner, A. *et al.*, "Characterization of Post-Transcriptional RNA Modifications by Sheathless Capillary Electrophoresis-High Resolution Mass Spectrometry.", *Anal. Chem.*, **2020**, *92* (10), 7363–7370.
- [18] Maxwell, E. J. *et al.*, "Twenty Years of Interface Development for Capillary Electrophoresis-Electrospray Ionization-Mass Spectrometry.", *Analytica Chimica Acta*, **2008**, *627* (1), 25–33.

- [19] Gahoual, R. *et al.*, “Revealing the Potential of Capillary Electrophoresis/Mass Spectrometry: The Tipping Point.”, *Rapid Communications in Mass Spectrometry*, **2019**, 33 (S1), 11–19.
- [20] Tie, C. *et al.*, “Study of the Electrical Connection Mechanism of Sheathless Interface for Capillary Electrophoresis-Electrospray Ionization-Mass Spectrometry.”, *Journal of Mass Spectrometry*, **2012**, 47 (11), 1429–1434.

ANNEXE 2 : POSTER PRÉSENTE À ANALYTICS 2022 (NANTES)



Optimization of nucleotides dephosphorylation for the characterization of RNA post-transcriptional modifications

Clarisse Gosset-Erard^{1,2}, Antony Lechner^{3,4}, Philippe Wolff^{3,4}, Frédéric Aubriet¹, Emmanuelle Leize-Wagner², Patrick Chaimbault¹ and Yannis-Nicolas François²

¹Laboratoire de Chimie et Physique—Approche Multi-échelles des Milieux Complexes (LCP-A2MC), Université de Lorraine, Metz, France

²Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS) UMR 7140, CNRS, Université de Strasbourg, Strasbourg, France

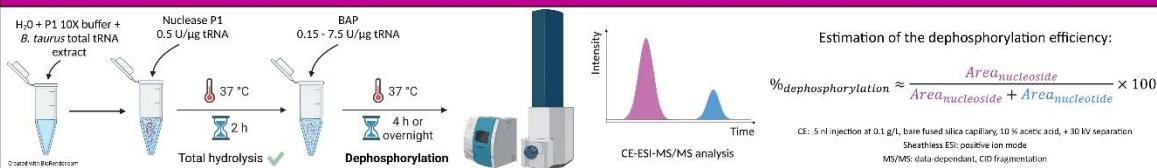
³Architecture et Réactivité de l'ARN (ARN) UPR 9002, CNRS, Université de Strasbourg, Strasbourg, France

⁴Plateforme Protéomique Strasbourg Esplanade FR 1589, CNRS, Strasbourg, France

Introduction

As part of RNA characterization, the identification of post-transcriptional modifications can be performed by coupling separation methods with mass spectrometry. To highlight these RNA modifications with such methods, a first total digestion followed by a dephosphorylation step is usually required to convert RNA into simple nucleosides. Effective digestion and dephosphorylation are essential to avoid further complications in the analysis and the data interpretation. Therefore, the aim of this work was to optimize the dephosphorylation step using a total extract of transfer RNA (tRNA). The protocol was developed on *Bos taurus* as its tRNAs are well known and described in the literature¹. Two protocols leading to a complete dephosphorylation were established based on time and bacterial alkaline phosphatase (BAP) consumptions.

Method

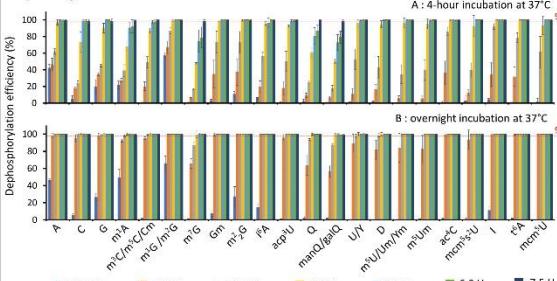


Verification of the total digestion step

- ♦ Miscleavages of hypermodifications previously reported when using only Nuclease P1².
- ♦ P1 digestion was considered **complete** as no dinucleotides were detected.
- ⇒ Therefore, the following study can be focused only on the dephosphorylation step.

Full dephosphorylation for all nucleotides

- ♦ Dephosphorylation efficiencies for all nucleotides depending on incubation time and quantity of BAP:

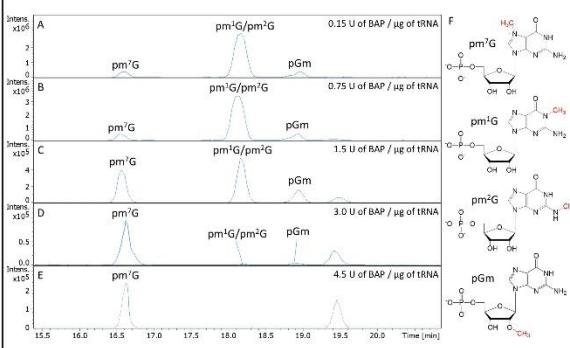


Dephosphorylation efficiencies estimated for 29 nucleotides after (A) 4-hour and (B) overnight incubation at 37 °C with seven quantities from 0.15 to 7.5 U of BAP for 1 μg of RNA. Standard errors were obtained from triplicates.

- ♦ Sample considered **fully dephosphorylated** when %efficiency ≥ 98 % for all nucleotides.

Differences in dephosphorylation efficiencies

- ♦ Expected differences in dephosphorylation efficiencies among large modifications.
- ♦ Significant differences in dephosphorylation efficiencies for small-sized isomeric modifications with close structures.
- ♦ Example of m¹G/m²G, m⁷G and Gm after a 4-hour incubation :



Extracted ion electropherograms (m/z 378.08) for the analysis total tRNA extract from *B. taurus* after digestion with nuclease P1 and dephosphorylation after a 4-hour incubation at 37 °C with (A) 0.15 U of BAP per μ g of RNA, (B) 0.75 U of BAP per μ g of RNA, (C) 1.5 U of BAP per μ g of RNA, (D) 3.0 U of BAP per μ g of RNA and (E) 4.5 U of BAP per μ g of RNA. (F) Structures of the four isomers of methylated monophosphate guanosine.

Conclusion

Two protocols can be established for the nucleotides dephosphorylation in routine for CE-MS/MS but also LC-MS/MS analysis purposes. According to individual needs:

- ♦ Quick but requiring higher amounts of enzyme: 7.5 U of BAP for 1 μ g of RNA incubated at 37 °C for 4 hours,
- ♦ Long but more economical in enzyme: 3.0 U of BAP for 1 μ g of RNA incubated overnight at 37 °C.

Acknowledgements

This work was supported by a PhD grant from the RESEX project funded by the European Regional Development Fund (FEDER), the general council of Moselle, Region Grand Est, Metz Métropole and the University of Lorraine.

References

- ¹Boccaletto *et al.*, *Nucleic Acids Research* (2022), 50, D231-D235.
- ²Crain *et al.*, *Methods in Enzymology* (1990), 193, 782-790.

Based on the publication: Gosset-Erard *et al.*, *Journal of Chromatography B* (2022), 1208, 123396.



Développement méthodologique en CE-FTICR-MS pour l'étude de biomolécules à visée thérapeutique

Résumé

La spectrométrie de masse (MS) s'est imposée depuis quelques années comme une méthode essentielle pour la caractérisation des biomolécules. Le développement de nouveaux analyseurs à ultra-haute résolution et grande précision de mesure de masse, tels que la résonance cyclotronique des ions à transformées de Fourier (FTICR), ont amélioré la sélectivité de la méthode. Afin de faciliter la caractérisation d'échantillons complexes dans de nombreux domaines, la MS peut être couplée à des méthodes séparatives telles que l'électrophorèse capillaire (CE). La CE est la méthode électrophorétique la plus performante en termes de résolution, d'efficacité et de capacité de pics tout en étant très rapide. Le couplage CE-MS a été développé pour un grand nombre d'analytes et permet d'obtenir une sensibilité optimale de la MS grâce au recours d'une interface sheathless permettant l'utilisation de nano-débit. Cependant, pour pouvoir combiner les performances de séparation de la CE, la haute sélectivité et sensibilité de l'interface sheathless avec l'ultra-haute résolution du FTICR-MS, il est nécessaire de relever les défis techniques liés aux propriétés intrinsèques de la CE et du FTICR telles que la rapidité de la séparation, la haute efficacité des pics et le temps d'acquisition de la MS. Les travaux présentés dans ce manuscrit présentent donc la mise en place du couplage CE-FTICR-MS, nouveau au laboratoire, et son application dans l'étude de biomolécules, et notamment la caractérisation des modifications post-transcriptionnelles des acides ribonucléiques (ARN). Une étude du couplage a tout d'abord été réalisée à l'aide de standards, puis un premier transfert de méthode a été effectué sur des échantillons biologiques déjà décrits dans la littérature. Divers développements de méthodes sont ensuite présentés tels que l'optimisation de la préparation d'échantillon, et le développement de nouveaux outils bio-informatiques permettant d'aller plus loin dans la caractérisation de ces modifications et dans la complexité des échantillons. Enfin, le couplage CE-FTICR-MS ainsi que des derniers développements de méthode ont été mis en application sur des échantillons plus complexes et dont les modifications post-transcriptionnelles ne sont pas décrites dans la littérature. L'utilisation du couplage CE-FTICR-MS pour la caractérisation des ARN a notamment permis d'identifier une modification post-transcriptionnelle inconnue dans un de ces échantillons complexes et peu étudiés dans la littérature.

Mots clés : Spectrométrie de masse, Electrophorèse capillaire, CE-FTICR-MS, Ultra-haute résolution, Biomolécules, Acides ribonucléiques

Summary

Mass spectrometry (MS) has emerged in recent years as benchmark method for the characterization of biomolecules. The development of new analyzers with ultra-high resolution and high mass accuracy, such as Fourier Transform Ion Cyclotron Resonance (FTICR), has improved the selectivity of the method. To ease the characterization of complex samples, MS can be coupled with separative methods such as capillary electrophoresis (CE). CE is the most powerful electrophoretic method in terms of resolution, efficiency and peak capacity, and is very fast. The CE-MS coupling has been developed for a large number of analytes and allows to obtain an optimal sensitivity of MS thanks to the use of a sheathless interface allowing the use of nanoflow rates. However, in order to combine the separation performances of CE, the high selectivity and sensitivity of the sheathless interface with the ultra-high resolution and the high mass accuracy of FTICR-MS, it is necessary to address the technical challenges related to the intrinsic properties of CE and FTICR such as the speed of separation, high peak efficiency and MS acquisition time. The work presented in this manuscript presents the implementation of the CE-FTICR-MS hyphenation and its application for the study of biomolecules, and in particular the characterization of post-transcriptional modifications of ribonucleic acids (RNA). A study of the CE-FTICR-MS hyphenation was first performed using standards, then a first method transfer was performed on biological samples already described in the literature. Various method developments are then presented such as the optimization of the sample preparation, and the development of new bioinformatics tools allowing to go further in the characterization of these modifications and in the complexity of the samples. Finally, the CE-FTICR-MS hyphenation as well as the latest method developments were applied on more complex samples, and whose post-transcriptional modifications are not described in the literature. The use of the CE-FTICR-MS hyphenation for the characterization of RNA enabled the identification of an unknown post-transcriptional modification in one of these complex samples that has not been previously studied in the literature.

Keywords : Mass Spectrometry, Capillary electrophoresis, CE-FTICR-MS, Ultra-high resolution, Biomolecules, Ribonucleic acids