



**HAL**  
open science

# Développement de l'électrophorèse capillaire couplée à la spectrométrie de masse pour l'étude des modifications post-transcriptionnelles des ARN.

Antony Lechner

## ► To cite this version:

Antony Lechner. Développement de l'électrophorèse capillaire couplée à la spectrométrie de masse pour l'étude des modifications post-transcriptionnelles des ARN.. Chimie analytique. Université de Strasbourg, 2020. Français. NNT : 2020STRAF014 . tel-03505750

**HAL Id: tel-03505750**

**<https://theses.hal.science/tel-03505750v1>**

Submitted on 31 Dec 2021

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

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

**ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES**

**UMR 7140 - Chimie de la Matière Complexe**

# THÈSE

présentée par :

**Antony LECHNER**

soutenue le : **22 Septembre 2020**

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

Discipline/ Spécialité : **Chimie Analytique**

**Développement de l'électrophorèse capillaire couplée à  
la spectrométrie de masse pour l'étude des  
modifications post-transcriptionnelles des ARN.**

**THÈSE dirigée par :**

**D<sup>r</sup> Yannis FRANCOIS**

Maître de conférences, Université de Strasbourg

**D<sup>r</sup> Emmanuelle LEIZE-WAGNER**

Directrice de recherche, Université de Strasbourg

**RAPPORTEURS :**

**P<sup>r</sup> Jean-Luc VEUTHEY**

Rapporteur externe, University of Geneva

**P<sup>r</sup> Christine ENJALBAL**

Rapporteur externe, Université de Montpellier

---

**AUTRES MEMBRES DU JURY :**

**P<sup>r</sup> Laurence SABATIER**

Examinatrice, Université de Strasbourg

**D<sup>r</sup> Rabah GAHOUAL**

Examinateur, Université de Paris



*À ma famille*

*« Nous sommes comme des nains assis sur des épaules de géants. Si nous voyons plus de choses et plus lointaines qu'eux, ce n'est pas à cause de la perspicacité de notre vue, ni de notre grandeur, c'est parce que nous sommes élevés par eux ».*

Bernard de Chartres

*« Le travail a des racines amères, mais des fruits doux ».*

Proverbe alsacien



# REMERCIEMENTS

La définition de remercier (verbe transitif) c'est « exprimer à quelqu'un de la gratitude, de la reconnaissance pour ce qu'il fait ou a fait : Remercier un ami de son appui ». Il faudrait écrire l'équivalent d'une autre thèse pour remercier comme il se doit toutes les personnes qui ont contribué à l'accomplissement de cette thèse. Un doctorat c'est une aventure riche et aussi difficile où j'ai rencontré beaucoup de personnes formidables.

Je souhaite remercier en tout premier lieu les membres du Jury : le Professeur Laurence SABATIER, le Professeur Christine ENJALBAL, le Professeur Jean-Luc VEUTHEY ainsi que le Docteur Rabah GAHOUAL qui m'ont fait l'honneur d'évaluer mes travaux de doctorat.

Je tiens à adresser mes sincères remerciements et à exprimer mon respect le plus profond au Docteur Emmanuelle LEIZE-WAGNER, directrice du laboratoire mais aussi co-directrice de ma thèse, pour m'avoir accueilli au LSMIS mais aussi pour sa bienveillance. Tes remarques avisées ainsi que tes éclairages sur les questions scientifiques m'ont guidé et permis de mener à bien ce projet de thèse.

Je tiens à exprimer toute ma reconnaissance et ma gratitude au Docteur Yannis-Nicolas FRANÇOIS. Tu étais un directeur de thèse vraiment exceptionnel. Ce fut un honneur et un privilège de travailler sous ton aile, du master jusqu'à la thèse. Je te remercie pour la pleine confiance que tu m'as accordée, ton écoute dans les moments difficiles, ton soutien indéfectible et ton encadrement sans faille durant toutes ses années ! Tu m'as énormément apporté sur le plan scientifique mais aussi sur le plan personnel. Merci pour tout ce que tu as fait pour moi. On se passera des rebondissements qui se sont déroulés pour l'obtention du financement. Durant ces trois années de thèse tu as peaufiné ton accent alsacien. Qui l'eut cru pour un Parisien ?

Par la suite, j'aimerais également remercier tous les membres du LSMIS comme le Docteur Armelle CHARRIE, Nathalie ZORN et le Docteur Noelle POTIER. Merci de m'avoir accueilli et pour vos nombreux conseils techniques et remarques scientifiques. Merci pour votre implication et votre gentillesse.

Durant cette thèse, j'ai également eu la chance de côtoyer au quotidien deux personnes exceptionnelles... J'ai nommé : Christophe GIORGIUTTI et Jérémie GIORGETTI. On a formé un trio de choc... C'était un réel plaisir de travailler avec vous durant toutes ces années ! On a enduré les pires épreuves et surtout les blagues pas drôles de chacun ! Je n'oublierai pas votre amitié et votre bonne humeur. Alors un grand merci à vous deux, je vous souhaite que du bonheur pour la suite ! Merci aux nouveaux doctorants Mathieu et Clarisse, je suis certain que vous irez loin ! La spectrométrie de masse a encore de beaux jours devant elle ! J'espère que vous continuerez à perpétuer la traditionnelle pause-café du matin.

Un immense merci à Julien, Gaël et Yu Min, les autres thésards (et certains déjà docteurs ! Encore toutes mes félicitations ;) du même étage. Toutes les longues discussions (pas forcément très scientifiques) au pas de la porte ou bien hors laboratoire ont permis de

relâcher la pression lors des longues journées de travail. Vous êtes au top, je vous souhaite le meilleur pour l'avenir !

Je souhaite exprimer toute ma reconnaissance envers Philippe WOLFF (nom de code : PWO), LE biologiste spécialiste de l'ARN sans qui cette thèse ne serait pas ce qu'elle est. Merci énormément pour ton expertise sur les ARN, grâce à toi j'ai énormément appris en biologie moléculaire qui était une science plutôt obscure pour moi. Je te remercie pour ton appui et d'avoir répondu à mes nombreuses questions (parfois naïves je dois bien avouer !). J'en profite également pour remercier toute l'équipe de la plateforme protéomique de l'IBMC pour leur immense gentillesse (on se souviendra notamment de la transhumance du MALDI !).

Un petit mot pour Stéphan. C'était toujours un plaisir de passer les pauses midi en ta compagnie. Merci pour ton sourire et ta bonne humeur quotidienne qui ont fortement contribué à la réussite de cette thèse ! J'espère qu'un jour on arrivera à réussir les sudokus experts et les mots fléchés force 4 sans tricher (c'est pas encore gagné pour moi :)).

Je souhaiterais remercier tous les stagiaires qui sont passés au labo (et ils sont nombreux !) : Elise, Eminé, Lucas, Apolline, Louis, Ludovic, Lionel, Valentin, Thomas, Florian et Geoffrey. Grâce à vous j'ai appris à être pédagogue. Merci pour la bonne ambiance que vous avez apporté au laboratoire.

Lors de ces trois années de thèses, j'ai eu l'occasion de réaliser une mission de monitorat pour des travaux pratiques de chimie analytiques. Je remercie chaleureusement Laura BOM et Thomas WEISSENBARGER. Eh oui ! Sans votre expertise technique de la chromatographie et votre soutien dont j'ai bénéficié, rien ne se serait déroulé sans accroc ! Grâce aux nombreuses tablettes de chocolats et gâteaux (aussi au chocolat), nous avons pu mener à bien notre mission !

Les amis d'enfance Thomas et Guillaume et aussi Dam, pour les soirées, les restos et les week-ends ! Vous ne savez pas à quel point ça m'a aidé de pouvoir couper avec la thèse le temps d'un week-end. Malgré des chemins différents on est toujours resté proche et cela durera encore longtemps !

Merci également à ma famille : mon père Claude, ma mère Annie, ma grand-mère Raymonde, ma grande sœur Alexia et mon frère jumeau Nicolas. Merci de m'avoir encouragé et d'avoir toujours été présent. Merci d'avoir cru en moi, c'est aussi grâce à vous que j'ai eu la persévérance pour en arriver jusque-là.

Je terminerais ces remerciements par une personne qui m'est très chère à mes yeux : Marie. Tu es courageuse, bienveillante et tu t'es montrée très patiente envers moi. Tu as supporté mes absences et mes nombreux retards avec « mais t'es encore au labo ?! ». Moi qui suis tête en l'air (juste un peu :)). Tu as su me reconforter et être à l'écoute dans les moments difficiles. Pour tous les bons moments partagés ensemble et ton amour, je te dis un grand MERCI !

# SOMMAIRE

<b>Remerciements</b> .....	<b>5</b>
<b>Sommaire</b> .....	<b>7</b>
<b>Liste des Abréviations</b> .....	<b>10</b>
<b>Introduction Générale</b> .....	<b>13</b>
<b>Partie I : Introduction et Revue Bibliographique</b> .....	<b>17</b>
<b>CHAPITRE 1. Définition &amp; Propriétés des ARN</b> .....	<b>18</b>
<b>1.1 Introduction</b> .....	<b>18</b>
1.1.1 Découverte des acides nucléiques .....	19
<b>1.2 Propriétés des groupements constituant les ARN</b> .....	<b>19</b>
1.2.1 Groupement phosphate .....	19
1.2.2 Sucre .....	20
1.2.3 Bases nucléiques.....	20
1.2.4 Ribonucléotide.....	25
<b>1.3 Propriétés des ARN</b> .....	<b>28</b>
1.3.1 Définition et généralités .....	28
1.3.2 Propriétés chimiques .....	29
1.3.3 Considérations structurales.....	30
1.3.4 Propriété optique : absorption UV .....	34
<b>1.4 Références bibliographiques</b> .....	<b>36</b>
<b>CHAPITRE 2. Biochimie des ARN</b> .....	<b>38</b>
<b>2.1 Les types et les rôles d'ARN</b> .....	<b>38</b>
2.1.1 Les ARN messagers .....	38
2.1.2 Les ARN de transfert .....	39
2.1.3 Les ARN ribosomiques .....	40
2.1.4 Autres ARN .....	42
<b>2.2 Mécanisme général de la synthèse d'ARN</b> .....	<b>43</b>
2.2.1 L'initiation .....	43
2.2.2 L'élongation .....	44
2.2.3 La terminaison .....	44
<b>2.3 Références bibliographiques</b> .....	<b>45</b>
<b>CHAPITRE 3. Modifications Post-Transcriptionnelles</b> .....	<b>48</b>
<b>3.1 Introduction et généralités</b> .....	<b>48</b>
<b>3.2 Modifications naturelles des ARNt</b> .....	<b>50</b>
3.2.1 Modifications sur la boucle anticodon .....	50
3.2.2 Modifications influençant la structure .....	53
3.2.3 Modifications impliquées dans l'aminocyclation des ARNt .....	55
3.2.4 Autres modifications remarquables .....	56
<b>3.3 Modifications d'ARN et maladies</b> .....	<b>57</b>



3.4 Références bibliographiques .....	59
<b>CHAPITRE 4. Stratégies d'Etudes des Modifications Post-Transcriptionnelles d'ARN</b> .....	<b>64</b>
4.1 Introduction.....	64
4.2 Purification d'ARN.....	65
4.2.1 Extraction des ARN totaux.....	66
4.2.2 Isolation des différents ARN .....	67
4.3 Préparation des ARN avant analyse .....	68
4.3.1 Dessalage des échantillons.....	68
4.3.2 Premier niveau : Hydrolyse totale .....	70
4.3.3 Second niveau : Digestion partielle .....	72
4.4 Analyse des échantillons .....	73
4.4.1 Chromatographie liquide avec détection optique.....	74
4.4.2 Électrophorèse capillaire avec détection optique .....	75
4.4.3 Méthodes séparatives couplées à la spectrométrie de masse.....	76
4.5 Références bibliographiques .....	81
<b>Partie II : Mise en Place de la Méthode.....</b>	<b>87</b>
<b>CHAPITRE 5. Développement de la Méthode d'Analyse.....</b>	<b>88</b>
5.1 Introduction.....	88
5.2 Présentation du système CZE-MS/MS .....	90
5.2.1 Système d'électrophorèse capillaire .....	90
5.2.2 Interface CE-ESI-MS .....	92
5.2.3 Spectromètre de masse.....	95
5.3 Développement de la méthode pour l'étude au premier niveau .....	99
5.3.1 Détection de nucléosides .....	99
5.3.2 Mise en place de la séparation CE pour l'analyse de nucléosides.....	101
5.4 Développement de la méthode pour l'étude au second niveau .....	103
5.4.1 Détection d'oligonucléotides.....	103
5.4.2 Mise en place de la séparation CE pour l'analyse d'oligonucléotides .....	107
5.5 Références bibliographiques .....	113
<b>CHAPITRE 6. Traitement des Données.....</b>	<b>114</b>
6.1 Introduction.....	114
6.2 Identification des nucléosides .....	115
6.2.1 Simulation de l'hydrolyse totale .....	115
6.2.2 Exploitation des données d'analyse de nucléosides.....	116
6.3 Séquençage d'oligonucléotides et localisation de modifications .....	117
6.3.1 Simulation de la digestion .....	117
6.3.2 Exploitation des données d'analyse d'oligonucléotides.....	118
6.3.3 Détermination du recouvrement de séquence .....	119
6.4 Références bibliographiques .....	120
<b>Partie III : Applications et Résultats .....</b>	<b>123</b>

<b>CHAPITRE 7. Preuve de Concept.....</b>	<b>124</b>
7.1 Introduction.....	124
7.2 Publication .....	126
7.3 Conclusion .....	149
<b>CHAPITRE 8. Mise en Application de la Méthode .....</b>	<b>150</b>
8.1 Introduction.....	150
8.2 Préparation des ARN .....	151
8.2.1 Isolation des ARN par PAGE 2D .....	151
8.2.2 Digestions enzymatiques et dessalage des échantillons .....	152
8.3 Analyse et interprétation des données .....	153
8.4 Références bibliographiques .....	156
<b>Conclusion Générale et Perspectives.....</b>	<b>157</b>
<b>Listes des Communications Scientifiques .....</b>	<b>162</b>
<b>Annexes.....</b>	<b>165</b>

# LISTE DES ABREVIATIONS

A : adénine/adénosine

ADN : acide désoxyribonucléique

ARN : acide ribonucléique

ARNdb : ARN double-brin

ARNm : ARN messenger

ARNmi : micro ARN

ARNnc : ARN non-codant

ARNr : ARN ribosomique

ARNs : petit ARN

ARNsi : petit ARN interférent

ARNsn : petit ARN nucléaire

ARNsno : petit ARN nucléolaire

ARNt : ARN de transfert

ASO : oligonucléotides antisens

BAP : phosphatase alcaline bactérienne

BFS : silice vierge fondue

BGE : électrolyte support

C : cytosine/cytidine

CCM : chromatographie sur couche mince

CE : électrophorèse capillaire

CEC : électrochromatographie capillaire

CID : dissociation induite par collision

CZE : électrophorèse capillaire de zone

CZE-MS/MS : couplage de la CZE à la MS/MS

ESI : électrospray

G : guanine/guanosine

HFIP : 1,1,1,3,3,3-hexafluoro-2-propanol  
HILIC : chromatographie d'interaction hydrophile  
IP : appariement d'ions  
 $m/z$  : rapport masse/charge  
MEKC : électrochromatographie en phase micellaire  
MS : spectrométrie de masse  
MS/MS : spectrométrie de masse en tandem  
NDP : nucléotide diphosphate  
NMP : nucléotide monophosphate  
NTP : nucléotide triphosphate  
Q : quadripôle  
RMN : résonance magnétique nucléaire  
RNase : ribonucléase  
RP-HPLC : chromatographie liquide en phase inverse  
RT-PCR : transcriptase inverse-réaction en chaîne par polymérase  
SPE : extraction en phase solide  
TOF : temps de vol  
U : uracile/uridine  
UPLC : chromatographie liquide à haute performance  
UV : ultraviolet  
VIH : virus de l'immunodéficience humaine



# INTRODUCTION GENERALE

Ces dernières décennies, la spectrométrie de masse (MS) est une technique qui a révolutionné la chimie analytique. Les progrès de la MS ont permis des découvertes et des avancées majeures dans bien des domaines. Ses caractéristiques intrinsèques, telles que sa grande sensibilité ainsi que sa haute sélectivité, ont fait de cet outil une méthode de routine dans de nombreux domaines. L'électrophorèse capillaire (CE) est une technique séparative en veine liquide dont le principe repose sur l'électromigration d'espèces ioniques en solution. Cette technique permet d'accomplir des séparations à haute résolution sur de nombreuses molécules. L'un des points faibles de cette technique réside dans son manque de robustesse et de répétabilité. Toutefois les développements technologiques, avec des systèmes de plus en plus miniaturisés, ont permis de limiter ces inconvénients et de réaliser le couplage CE-MS. Ce couplage s'est longtemps heurté à des problèmes techniques avec notamment la nécessité de maintenir deux hautes tensions indépendantes. Cependant, grâce aux récents progrès technologiques réalisés sur plusieurs interfaces et la naissance de la source à électrospray (ESI), ce type de couplage a été rendu possible. Le couplage CE-ESI-MS est alors rapidement devenu une méthode puissante pour la caractérisation de diverses molécules et a été mis à contribution pour l'analyse métabolomique ou encore protéomique.

Parmi plusieurs familles de biomolécules, les acides ribonucléiques (ARN) sont des composés essentiels pour assurer la survie des cellules de la majorité des organismes vivants (eucaryotes et procaryotes). Les ARN sont des biopolymères dont les nucléotides sont les monomères. Ces molécules remplissent diverses fonctions avec notamment la transmission de l'information génétique. Dans une cellule, il existe plusieurs types d'ARN de tailles variées dont les rôles sont bien définis. Pour assurer ces différents rôles, il est très courant que certaines positions de l'ARN soient chimiquement modifiées. Ces modifications naturelles sont couramment appelées modifications post-transcriptionnelles. Toutes les classes d'ARN sont susceptibles de porter des modifications. De nombreuses découvertes ont permis de mettre en lumière l'importance de la nature et de la position des modifications sur les ARN. Ces modifications permettent d'apporter les propriétés indispensables aux ARN pour accomplir leurs rôles dans les différents processus biologiques. Ainsi, en plus des quatre nucléotides classiques qui composent un ARN, il existe une très grande variété de nucléotides modifiés qui procurent diverses nouvelles propriétés physico-chimiques à ces biopolymères.

Aujourd'hui, il y a un intérêt croissant pour l'analyse des modifications chimiques des ARN. En effet, ces modifications générées naturellement sont primordiales pour assurer les rôles de ce type de molécule dans les cellules. De plus, ces dernières décennies, plusieurs études ont signalé l'implication de certaines modifications d'ARN dans des maladies telles que le cancer. Par conséquent, l'élucidation du rôle et de l'influence de ces modifications permettent la meilleure compréhension des processus biologiques. Cependant, afin d'atteindre cet objectif, il est d'abord nécessaire de passer par l'analyse et la caractérisation

des modifications d'ARN. Plusieurs techniques ont été développées pour analyser ces modifications, mais à ce jour, la caractérisation des modifications représente toujours un défi pour les sciences analytiques. Ces techniques permettent non seulement de mieux comprendre l'influence des modifications dans un ARN mais également d'élucider les mécanismes de formation des modifications. Par ailleurs, de plus en plus de traitements à base d'ARN modifiés voient le jour. Ainsi, ces techniques d'analyses sont aussi indispensables afin de s'assurer du contrôle qualité de cette nouvelle classe de médicament.

Actuellement, la technique la plus courante pour étudier les modifications d'ARN est la chromatographie liquide couplée à la spectrométrie de masse en tandem (LC-MS/MS). En effet, ce puissant outil d'analyse a déjà prouvé son intérêt dans de très nombreux domaines, dont l'analyse de biomolécules. Dans le but de caractériser les modifications par LC-MS/MS, des étapes de préparations d'échantillons sont nécessaires avant analyse. Des stratégies basées sur des digestions enzymatiques ont été établies. Classiquement, l'analyse des modifications par LC-MS/MS est réalisée sur deux niveaux, donnant accès à des informations complémentaires sur les modifications : un premier niveau permettant d'identifier la nature des modifications présentes dans un ARN et un second niveau permettant de localiser les modifications sur la structure de l'ARN.

Dans le premier niveau, une hydrolyse totale de l'ARN est effectuée. Cette hydrolyse, réalisée par voie enzymatique, permet d'obtenir un mélange de monomères de l'ARN dont certaines portent des modifications. L'analyse de ce mélange par LC-MS/MS permet alors d'identifier la nature des modifications contenues dans l'ARN. Néanmoins, ce niveau d'analyse ne permet pas d'obtenir les informations sur les positions des modifications sur la structure de l'ARN.

C'est pourquoi l'analyse de l'ARN à un second niveau a été mise en place. Ici, c'est une approche dite « bottom-up » qui a été développée pour localiser les modifications sur l'ARN. De la même manière que la stratégie utilisée en protéomique, les ARN sont soumis à des digestions partielles accomplies à l'aide d'enzymes spécifiques. Ce type de digestion permet de produire un mélange d'oligonucléotides qui est analysé par LC-MS/MS. Les résultats d'analyses permettent alors de reconstruire les séquences et surtout de déterminer les positions des nucléotides modifiés sur l'ARN. Généralement, dans le but d'améliorer les résultats, il est courant de réaliser plusieurs digestions partielles individuelles avec des enzymes de spécificités différentes.

L'électrophorèse capillaire de zone (CZE) est souvent présentée comme une technique alternative à la LC. La CZE offre des avantages majeurs tels qu'une faible consommation d'échantillon ainsi que des séparations efficaces et rapides. La CZE permet la séparation, en solution, d'espèces ioniques et fortement polaire. Par conséquent, cette technique apparaît comme une méthode tout à fait adaptée pour analyser des ARN car ce sont des molécules ionisables. La première interface permettant le couplage de la CE à la MS utilisait un liquide additionnel afin de maintenir de façon indépendante les tensions électriques. Cependant le débit de liquide additionnel diluait de façon non-négligeable les analytes, diminuant alors fortement la sensibilité de détection. Plus tard, une autre interface sans liquide additionnel a été développée. Dans cette interface c'est un capillaire dont l'extrémité a été rendue poreuse

qui est utilisée pour maintenir les champs électriques. L'absence de liquide additionnel ainsi que les dimensions du capillaire ont permis alors d'augmenter considérablement la sensibilité de détection du couplage CE-MS.

L'objectif de cette thèse était alors de développer une méthode d'analyse permettant la caractérisation des modifications post-transcriptionnelles des ARN grâce au couplage CE-MS/MS.

Ce manuscrit de thèse est divisé en trois parties :

- Une première partie propose une introduction et une revue bibliographique sur les ARN. Les propriétés et les rôles des ARN sont décrits et expliqués ainsi que les modifications post-transcriptionnelles. Un état de l'art sur les techniques d'analyses de ces modifications basées sur la MS est également disponible.
- La deuxième partie concerne le développement de la méthode d'analyse CE-MS/MS pour caractériser les modifications d'ARN sur deux niveaux. Dans un premier temps, le système d'analyse CE-MS utilisé est décrit. Puis la méthode, de la préparation de l'échantillon à l'exploitation des résultats en passant par l'amélioration de l'analyse, est détaillée.
- La dernière partie présente l'application de la méthode CE-MS/MS pour analyser et caractériser les modifications de différents ARN et discute également des résultats obtenus.





# **PARTIE I : INTRODUCTION ET REVUE BIBLIOGRAPHIQUE**

Les travaux décrits dans cette thèse portent sur le développement d'une méthode de caractérisation des modifications des acides ribonucléotides (ARN). La première partie de ce manuscrit vise à détailler les bases théoriques et l'état de l'art de la caractérisation des modifications des ARN. Cette revue bibliographique est divisée en quatre chapitres. Le premier chapitre définit les ARN et décrit leurs propriétés. Dans ce chapitre, les ARN seront « disséqués » afin de comprendre les comportements acido-basiques de chaque constituant des ARN. L'impact des groupements chimiques sur les structures des ARN sera également discuté. Ce chapitre apportera les outils nécessaires à la compréhension du fonctionnement des ARN et permettra également de justifier les choix des méthodes d'analyses des ARN. Le deuxième chapitre introduit les rôles et le fonctionnement des ARN dans les cellules. Le but est de fournir les notions de base en biologie moléculaire indispensables à la compréhension de l'influence des modifications. Ensuite le troisième chapitre définit les modifications post-transcriptionnelles des ARN, en expliquant leurs implications et leurs conséquences sur les propriétés des ARN. Ce chapitre met en place le contexte de l'étude des modifications mais également l'intérêt et les enjeux de les étudier. Enfin, le dernier chapitre propose de dresser un état de l'art des méthodes développées pour caractériser ces modifications. Une attention particulière est portée sur les méthodes basées sur la spectrométrie de masse et ses couplages puisque c'est la technique utilisée dans cette thèse.

## CHAPITRE 1. DEFINITION & PROPRIETES DES ARN

### 1.1 Introduction

Les organismes vivants, et plus précisément les cellules, utilisent diverses biomolécules pour subsister. Il existe différentes familles de biomolécules : les lipides, les glucides, les acides nucléiques et les protéines. Ces molécules sont généralement composées d'éléments chimiques légers (C, H, N, O, S, P). Les réactions chimiques se déroulent dans des conditions physiologiques qui sont typiquement à un pH de 7,2 et une température de 37°C. L'acide ribonucléique ou oligoribonucléotide est une biomolécule d'une importance cruciale dans les cellules puisqu'elle joue plusieurs rôles essentiels dont la transmission et la traduction de l'information génétique. Ces molécules de la famille des acides nucléiques sont présentes dans les cellules de presque tous les organismes vivants (animaux, végétaux, bactéries, champignons, etc.). Ils sont les intermédiaires entre l'acide désoxyribonucléique (ADN) et la protéine (voir Figure 1). Les ARN sont des biopolymères constitués d'un enchaînement de nucléotides dont les constituants seront décrits dans ce chapitre après un bref historique sur la découverte des acides nucléiques.

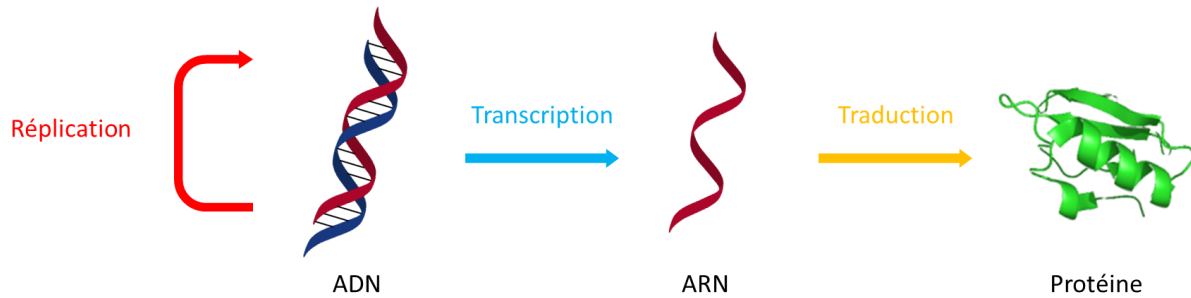


Figure 1 Dogme central de la biologie moléculaire. L'ADN, qui contient l'information génétique, se réplique. Pour produire une protéine, les informations de l'ADN sont transcrites dans une molécule d'ARN. Puis, les informations contenues dans l'ARN sont traduites en protéine.

### 1.1.1 Découverte des acides nucléiques

La découverte des acides nucléiques, ou polynucléotides, a été possible grâce à l'isolation de noyaux cellulaires et par la mise en évidence d'une substance, « la nucléine », par le biologiste Friedrich Miescher, en 1868. Il découvre également que cette substance est riche en phosphate. Ce n'est que bien plus tard, en 1939, que le terme d'acide nucléique sera employé pour la première fois, à la suite de sa découverte dans le cytoplasme de la levure. C'est en 1940 que le biologiste Jean Brachet identifie et distingue l'ARN et l'ADN. Grâce aux expériences de diffraction des rayons X, Francis Crick et James Watson [1] proposèrent en 1953 un modèle de structure en double hélice antiparallèle où les deux brins sont liés par les liaisons hydrogènes formées de façon spécifiques entre deux bases azotées complémentaires. L'hypothèse du rôle de l'ARN comme « messenger » de l'information génétique sera proposée et démontrée en 1960 par François Gros [2].

## 1.2 Propriétés des groupements constituant les ARN

Cette partie vise à définir les propriétés et les rôles des acides nucléiques avec en particulier les ARN de transfert qui ont fait l'objet d'étude dans ces travaux de thèse. La première partie se focalise sur l'ARN d'un point de vue chimique, en discutant notamment des propriétés acido-basiques et également de la géométrie adoptée par ces molécules. Les différents constituants sont décrits en remontant des « briques » jusqu'au biopolymère d'ARN. Le but est de comprendre l'influence de chaque groupements chimiques sur les propriétés d'un ARN. Ensuite, une deuxième partie s'intéresse aux rôles des différents ARN dans les cellules.

Dans cette partie, les propriétés de tous les groupements chimiques qui constituent les acides nucléiques sont décrites afin de comprendre leurs contributions dans la molécule d'ARN.

### 1.2.1 Groupement phosphate

Le groupe phosphate est issue de l'acide orthophosphorique, un triacide dont les  $pK_a$  sont 2,1 ; 7,2 et 12,4. Cette molécule adopte une géométrie tétraédrique comme schématisée sur la Figure 2.

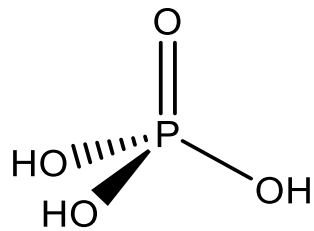


Figure 2 Structure chimique de l'acide orthophosphorique.

Cette molécule constitue, avec le ribose décrit ci-dessous, le squelette de chaque ARN.

### 1.2.2 Sucre

Les oses contenus dans les ARN sont des molécules issues d'un sucre à 5 carbones : le ribose. L'isomère dans les ARN est le  $\beta$ -D-ribofuranose : un cycle à 5 atomes (voir Figure 3). En ce qui concerne la géométrie, ce ribose peut adopter plusieurs conformations provenant du blocage des rotations des liaisons dans le cycle. Cette conformation dépend des groupements chimiques associés à cette molécule. La conformation du ribose dans les ARN est discutée plus en détail ultérieurement dans ce chapitre (cf. 1.3.3). Concernant ses propriétés acido-basiques, le ribose est une molécule constituée de groupements hydroxyles qui sont des acides très faibles. Cette faible acidité s'explique par le fait que les bases conjuguées ne sont pas stabilisées par délocalisation d'électrons ou effets inductifs. La valeur d'acidité expérimentale du D-ribose a été déterminée avec un  $pK_a \approx 12$  [3]. Le ribose est donc chargé uniquement dans des conditions très basiques.

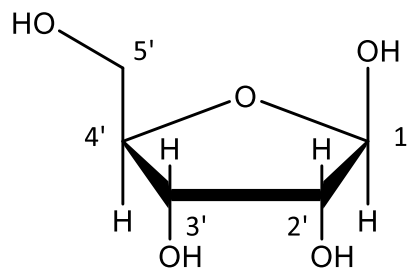


Figure 3 Structure chimique du  $\beta$ -D-ribofuranose et numérotation des carbones dans l'ARN. (recommandé par IUPAC).

### 1.2.3 Bases nucléiques

Les bases nucléiques sont des molécules organiques riches en électrons. Pour les oligoribonucléotides, il existe quatre bases, dites canoniques, qui peuvent être divisées en 2 familles. La première famille est celle des purines. Elle est composée de l'**adénine** (A) et de la **guanine** (G). Les purines sont constituées d'un hétérocycle aromatique à 6 atomes lui-même lié à un hétérocycle aromatique à 5 atomes. Quant à la famille des pyrimidines, elle est composée de la **cytosine** (C) et de l'**uracile** (U). Les pyrimidines sont simplement constituées d'un hétérocycle aromatique à 6 atomes. Les structures chimiques de ces molécules sont données dans le Tableau 1. Les  $pK_a$  de ces bases hétérocycliques sont très variables et s'expliquent par la structure et la disposition des atomes dans ces molécules.

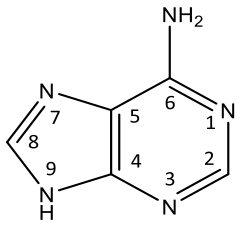
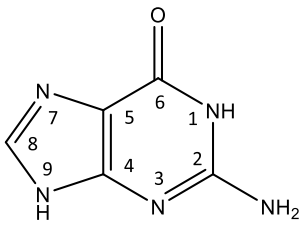
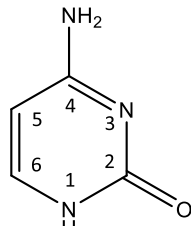
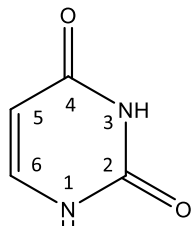
Famille	Purines		Pyrimidines	
Nom	Adénine	Guanine	Cytosine	Uracile
Code à une lettre	A	G	C	U
Formule chimique				

Tableau 1 Formules topologiques avec numérotation des atomes des bases azotées contenues dans les ARN. (recommandée par IUPAC).

L'aromaticité de ces bases azotées joue un rôle primordial sur les propriétés acido-basiques. En effet, un système de conjugaison permet de délocaliser les électrons  $\pi$  autour des cycles, ce qui rend ces composés très stables. Dans notre cas, tous les cycles sont aromatiques selon la règle de Hückel, puisqu'ici, certains atomes d'azote possèdent des doublets non-liants impliqués dans la mésomérie et que tous ces composés adoptent une géométrie plane.

De plus, les atomes d'azote présents dans ces bases forment différents groupes fonctionnels qui vont influencer les propriétés acido-basiques des noyaux puriques et pyrimidiques.

-Les **amines** sont des groupes fonctionnels composés simplement d'un atome d'azote faisant trois liaisons simples, dont au moins une avec un carbone. L'atome d'azote d'une fonction amine est hybridé  $sp^3$  et présente une géométrie pyramide trigonale. Dans ces bases hétérocycliques, il y a les amines primaires (portant 2 protons) et les amines secondaires (portant 1 proton). Ainsi, ces amines présentent au moins une liaison N-H polarisée et susceptible de se rompre. Toutefois, l'acidité de ce type d'azote est extrêmement faible puisque la base conjuguée, étant très instable, présente un  $pK_a > 30$ . Les amines sont surtout connues pour posséder un caractère basique. En effet, cette basicité est due à la présence d'un doublet non-liant sur l'atome d'azote qui peut capter un proton. Les amines secondaires sont généralement légèrement plus basiques que les amines primaires ( $pK_a$  des acides conjugués  $\approx 10$ ) car l'effet inductif donneur de deux carbones stabilise un peu plus la charge positive de l'azote protoné. Cependant, ici, les amines primaires sont liées à des cycles aromatiques. De ce fait le doublet non-liant de l'azote est délocalisable par mésomérie car ces électrons se trouvent dans une orbitale  $sp^3$ , orbitale qui est impliquée dans la conjugaison (voir Figure 4). Le doublet non-liant est alors beaucoup moins disponible pour former une liaison avec un proton. Donc cette amine aromatique (aniline) possède un caractère très faiblement basique avec un  $pK_a$  de l'acide conjugué d'environ 4,9.

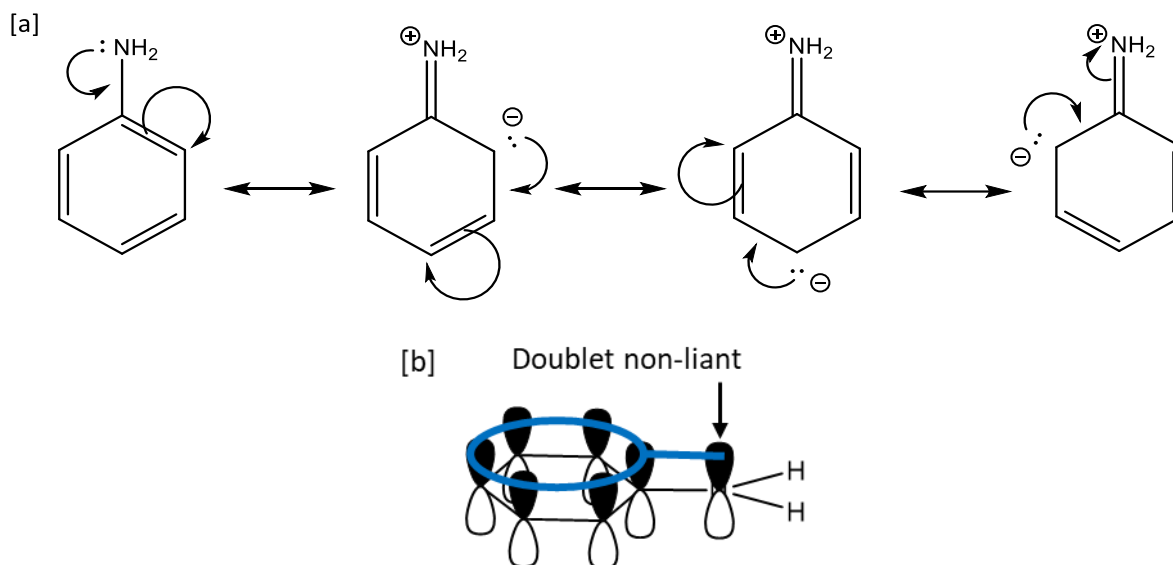


Figure 4 [a] Formes mésomères de l'aniline mettant en évidence la délocalisation du doublet non-liant de l'azote et [b] Orbitales impliquées dans le système de conjugaison d'électrons dans l'aniline.

-Certaines bases nucléiques sont également composées de **pyridine** qui sont des hétérocycles aromatiques. Contrairement à l'aniline, l'atome d'azote est inclus au sein d'un cycle aromatique et ne porte pas d'atome d'hydrogène. Toutefois, le doublet non-liant de cet atome n'est pas délocalisable car ce doublet est localisé dans une orbitale  $sp^2$  qui n'est pas impliquée dans l'aromaticité de la pyridine (voir Figure 5). Ainsi, le doublet non-liant n'étant pas délocalisable, il devrait, à priori, être disponible pour former une liaison N-H. Mais expérimentalement, les pyridines présentent un faible caractère basique avec un  $pK_a$  de l'acide conjugué d'environ 5,2. L'explication de cette faible basicité provient de l'hybridation de l'azote. En effet l'azote d'une pyridine est hybridé  $sp^2$  ce qui a pour conséquence d'accentuer l'effet inductif comparé à une hybridation  $sp^3$ . Le doublet non-liant est alors moins disponible et rend donc l'azote moins apte à se protoner. Ainsi une pyridine est une molécule présentant un caractère faiblement basique.

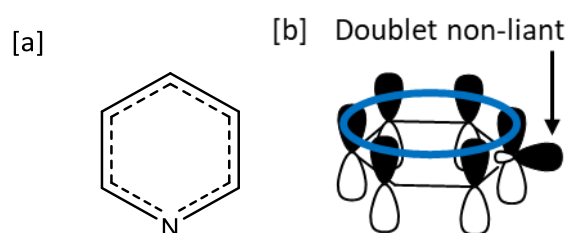


Figure 5 [a] Hybride de résonance de la pyridine et [b] Orbitales impliquées dans le système de conjugaison d'électrons dans la pyridine.

-Les amides sont des composés formés d'un carbonyle adjacent à un azote. Dans les bases nucléiques, il s'agit plus précisément d'amide cyclique connu également sous le nom de **lactame**. Le doublet non-liant de l'azote est délocalisable formant un système de conjugaison avec le carbonyle (voir Figure 6). Comme précédemment, cette délocalisation diminue la basicité de l'azote avec un  $pK_a$  de l'acide conjugué de 0,5. Le proton d'un amide se dissocie seulement avec un pH supérieur à 15. Enfin la charge partielle négative que porte l'oxygène peut former une liaison avec un proton. Néanmoins cette protonation n'est observable qu'en

condition fortement acide avec un  $pK_a$  de -1. Par conséquent, un lactame ne présente pas de propriétés acido-basiques notables en milieu aqueux.

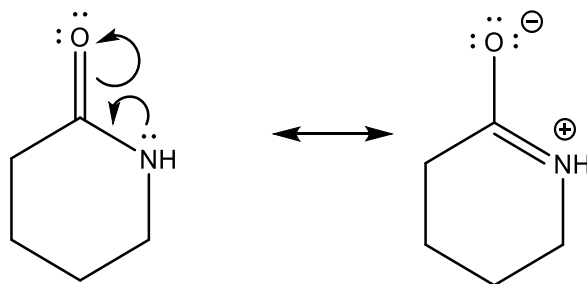


Figure 6 Structures de résonance dans un lactame.

-Dans la famille des purines, un hétérocycle à 5 atomes est présent. Ce cycle s'apparente à l'**imidazole** qui est une espèce amphotère. En effet, cette molécule possède deux atomes d'azote, un « acide » et l'autre « basique ». Concernant l'azote protoné, celui-ci est très faiblement acide puisque sa valeur de  $pK_a$  a été expérimentalement évaluée à 14,5. En effet, la base conjuguée n'est pas stabilisée par une délocalisation d'électron, d'où la faible acidité. De plus, le doublet non-liant de cette azote protoné participe à l'aromaticité du cycle, ce qui appauvrit cet azote en électron et le rend faiblement basique. En ce qui concerne le deuxième azote, il est faiblement basique puisque la valeur du  $pK_a$  de l'acide conjugué est de 6,9. C'est un azote  $sp^2$  et donc un effet inductif rend le doublet non-liant moins disponible ce qui traduit sa faible basicité [4] (voir Figure 7). L'imidazole est donc une espèce faiblement basique en solution aqueuse.

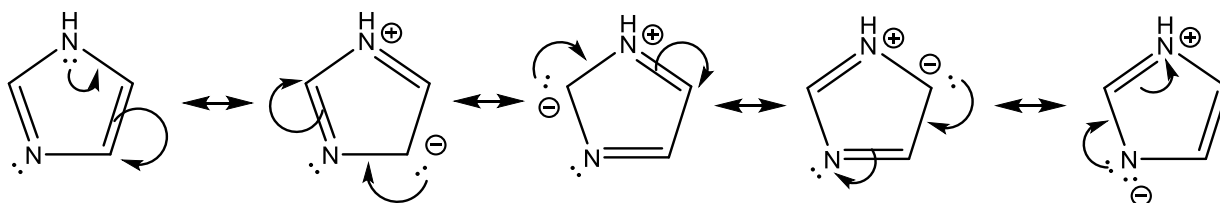


Figure 7 Formes mésomères vérifiant l'aromaticité de l'imidazole.

Enfin, la chimie des bases azotées est également impactée par la **tautomérie** [5]. La tautomérie est une réaction chimique d'une seule molécule qui va mettre en jeu un équilibre entre deux isomères de fonction. La réaction de tautomérie implique la migration d'un proton et le déplacement d'une liaison  $\pi$ . Cette réaction étant un équilibre, une forme sera favorisée par rapport à l'autre. Dans les bases nucléiques, il existe deux types d'équilibres tautomériques qui sont l'équilibre lactame-lactime et l'équilibre amino-imine (voir Figure 8).



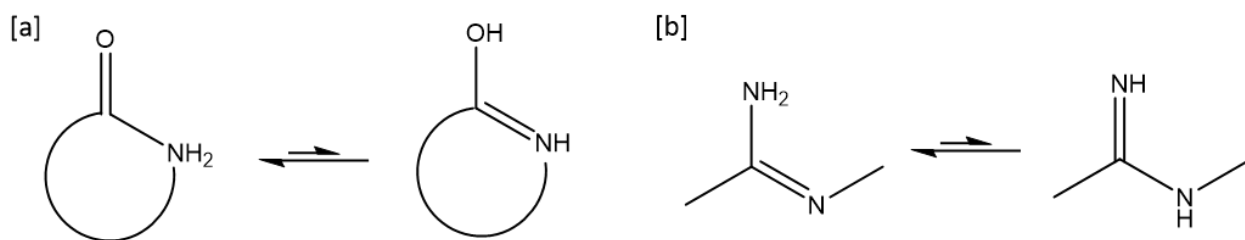


Figure 8 Les tautomères observables dans les bases azotées avec [a] équilibre lactame-lactime et [b] équilibre amino-imine

Dans ces bases hétérocycliques, il existe plusieurs formes tautomères qui se produisent en fonction de paramètres tels que le pH, la température ou bien la force ionique. [6-10]. La Figure 9 présente les différents équilibres de tautomérie et les possibles sites de tautomérie dans les quatre bases nucléiques.

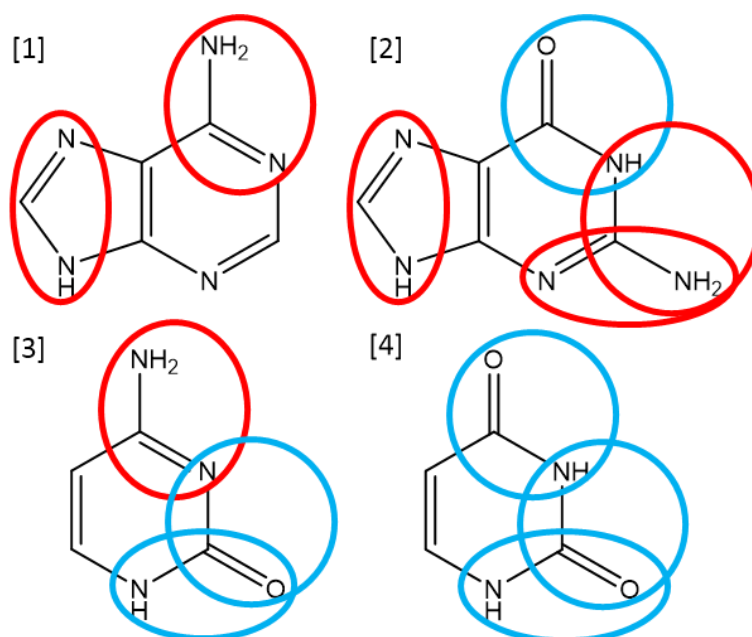


Figure 9 Sites possibles de tautomérie dans chaque base azotée. En rouge : équilibre lactame-lactime et en bleu : équilibre amino-imine. Avec [1] l'adénine, [2] la guanine, [3] la cytosine et [4] l'uracile.

Les  $pK_a$  des bases varient selon les formes tautomériques adoptées puisqu'elles induisent de nouveaux groupes fonctionnels sur ces molécules. Cependant, une forme très prédominante est observée pour chaque base azotée. En effet, les formes **lactame et amino** sont largement favorisées comme cela a été démontré par différentes expériences [11]. Par conséquent, les équilibres tautomériques impactent faiblement sur les valeurs des  $pK_a$ .

Ainsi, les bases nucléiques sont des composés présentant de nombreux groupes fonctionnels impactant plus ou moins leurs  $pK_a$ . Ces bases sont donc des groupements chimiques qui sont principalement des bases faibles [12]. Bien que ces composés présentent plusieurs sites de protonation et/ou déprotonation, seulement quelques uns sont observables en solution aqueuse. Certaines valeurs de  $pK_a$  de ces bases sont regroupées dans le Tableau 2.

Base nucléique	$pK_a$	
	Protonation	Déprotonation
Adénine	4,1-4,2	-
Guanine	3,2-3,3	9,2-9,6
Cytosine	4,4-4,5	-
Uracile	-	9,5

Tableau 2 : Valeurs de  $pK_a$  des bases azotées selon [13] ; seules les valeurs de première protonation et déprotonation sont affichées.

De ce tableau ressort le classement de basicité de ces molécules : d'abord la cytosine, suivie par l'adénine et la guanine. ( $C > A > G$ ). Malgré l'appellation de base nucléique, U n'a pas de propriété basique. Au contraire, le proton de l'azote N3 est légèrement acide ( $pK_a = 9,5$ ) en raison de la présence des groupements carbonyles adjacents. G présente également un proton faiblement acide en N3 pour la même raison que U.

Les trois groupes de molécules décrites dans cette partie sont associés pour former une « unité monomérique » des ARN qui est appelée un nucléotide.

#### 1.2.4 Ribonucléotide

Chaque nucléotides, ou plus précisément ribonucléotides dans le cas d'ARN, sont composés d'un groupe phosphate, d'un sucre et d'une des quatre bases nucléiques comme cela a été décrit précédemment. Les bases canoniques sont liées en 1' du ribose via une liaison C-N et un phosphate est lié sur le ribose après estérification pour former un nucléotide. Les nucléotides qui ne portent pas de groupement phosphate sont appelés des nucléosides. Les ribonucléotides sont désignés en fonction de la base hétérocyclique qu'ils portent, ainsi un nucléotide avec une adénine est appelé une **adénosine**, avec une guanine c'est une **guanosine**, avec une cytosine c'est une **cytosine** et avec un uracile c'est une **uridine**.

Les phosphates peuvent se fixer de différentes façons sur le ribose. Des exemples de structures chimiques de ces nucléosides phosphates sont dessinés sur la Figure 10. Les ribonucléotides dont le phosphate est situé en 5' sont l'adénosine 5'-monophosphate (5'-AMP), la guanosine 5'-monophosphate (5'-GMP), la cytosine 5'-monophosphate (5'-CMP) et l'uridine 5'-monophosphate (5'-UMP). Il est possible que le phosphate se lie en 3' des nucléosides et l'abréviation utilisée sera 3'-NMP, N désignant la base nucléique lié. On trouve également dans les cellules, des nucléotides dont le phosphate forme un cycle en se liant en 3' et 5' du pentose. Ces nucléotides sont désignés par l'abréviation NMPC. Ces composés sont d'importants régulateurs du métabolisme cellulaire [14]. De plus, deux ou trois groupes phosphates supplémentaires peuvent se lier au phosphate 5' d'un nucléotide, via des liaisons anhydrides phosphoriques, pour former les nucléotides diphosphates (NDP) ou triphosphates (NTP). Les groupes phosphates de ces molécules sont désignés par les lettres grecques  $\alpha$ ,  $\beta$  et  $\gamma$  en commençant par le phosphate  $\alpha$  qui est directement lié au ribose. Ces nucléotides sont très utilisés par les cellules dans les processus de métabolisme puisque les anhydrides phosphoriques sont d'importants réserves d'énergie chimique avec par exemple l'ATP. Dans le cas des NTP, ils sont également utilisés pour synthétiser les ARN.

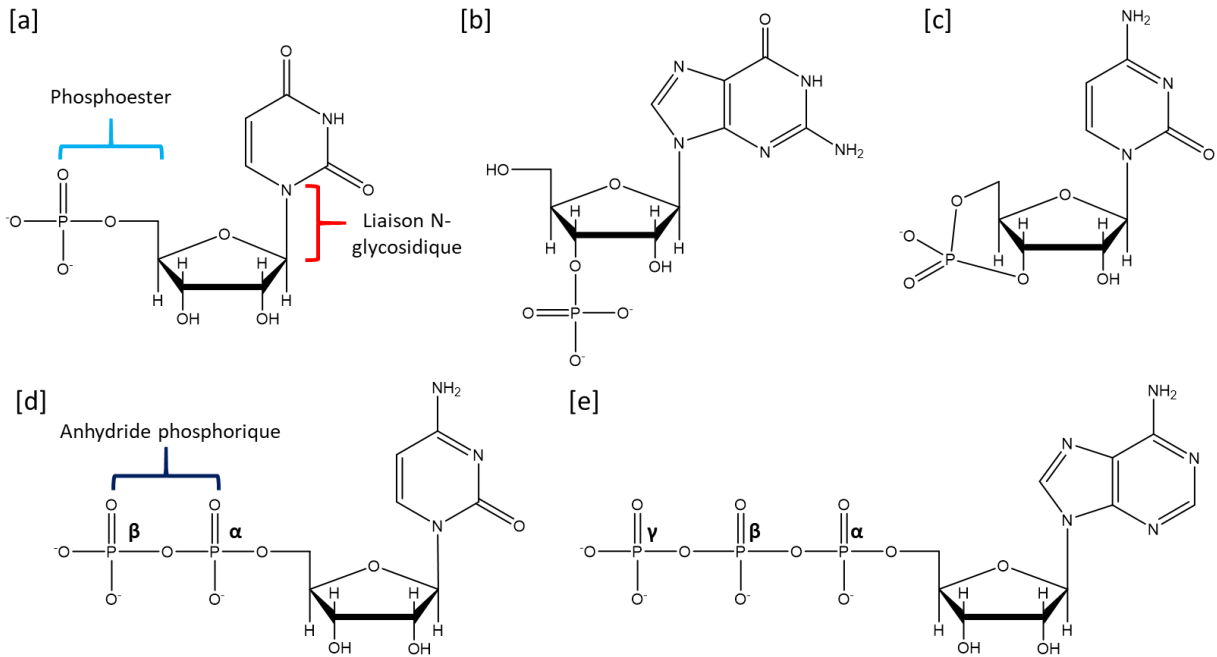


Figure 10 Structures chimiques de différents ribonucléotides avec [a] uridine 5'-monophosphate (5'-UMP) ; [b] guanosine 3'-monophosphate (3'-GMP) ; [c] cytosine 3',5'-phosphate cyclique (CMPc) ; [d] cytosine 5'-diphosphate (CDP) et [e] adénosine 5'-triphosphate (ATP).

Les propriétés acido-basiques des groupements, discutées précédemment, se retrouvent plus ou moins dans les nucléotides avec quelques nuances. Sur les bases nucléiques, certains sites de protonation ne sont plus disponibles puisqu'ils sont impliqués dans des liaisons covalentes. Par exemple, les bases nucléiques engagent un azote pour former une liaison covalente avec le pentose. Cela a pour conséquence de réduire le nombre de site de protonation mais aussi le nombre de formes tautomères des bases azotées [15]. De plus, le ribose provoque un effet inductif attracteur à la base nucléique qu'il porte, ce qui a pour conséquence une baisse des valeurs de  $pK_a$  des bases.

Les hydrogènes du ribose, étant très faiblement acides, n'impactent pas les propriétés des nucléotides.

Quant au groupe phosphate, sa première acidité est plus élevée que sous sa forme acide libre (environ 1 comparé à 2,1). À pH neutre, sa charge négative favorise alors la protonation de la base nucléique pour ainsi former une espèce zwitterionique plus stable (interactions électrostatiques entre les bases protonées et la charge négative portée par les phosphates). Cela a pour conséquence d'augmenter la basicité de A, G et C et de réduire l'acidité de G et U. Toutefois, l'ordre de basicité des nucléotides reste le même que celui des bases azotées seules comme le montre la Figure 11 où les valeurs de  $pK_a$  et les sites de protonations des différents groupements chimiques des ARN sont indiqués.

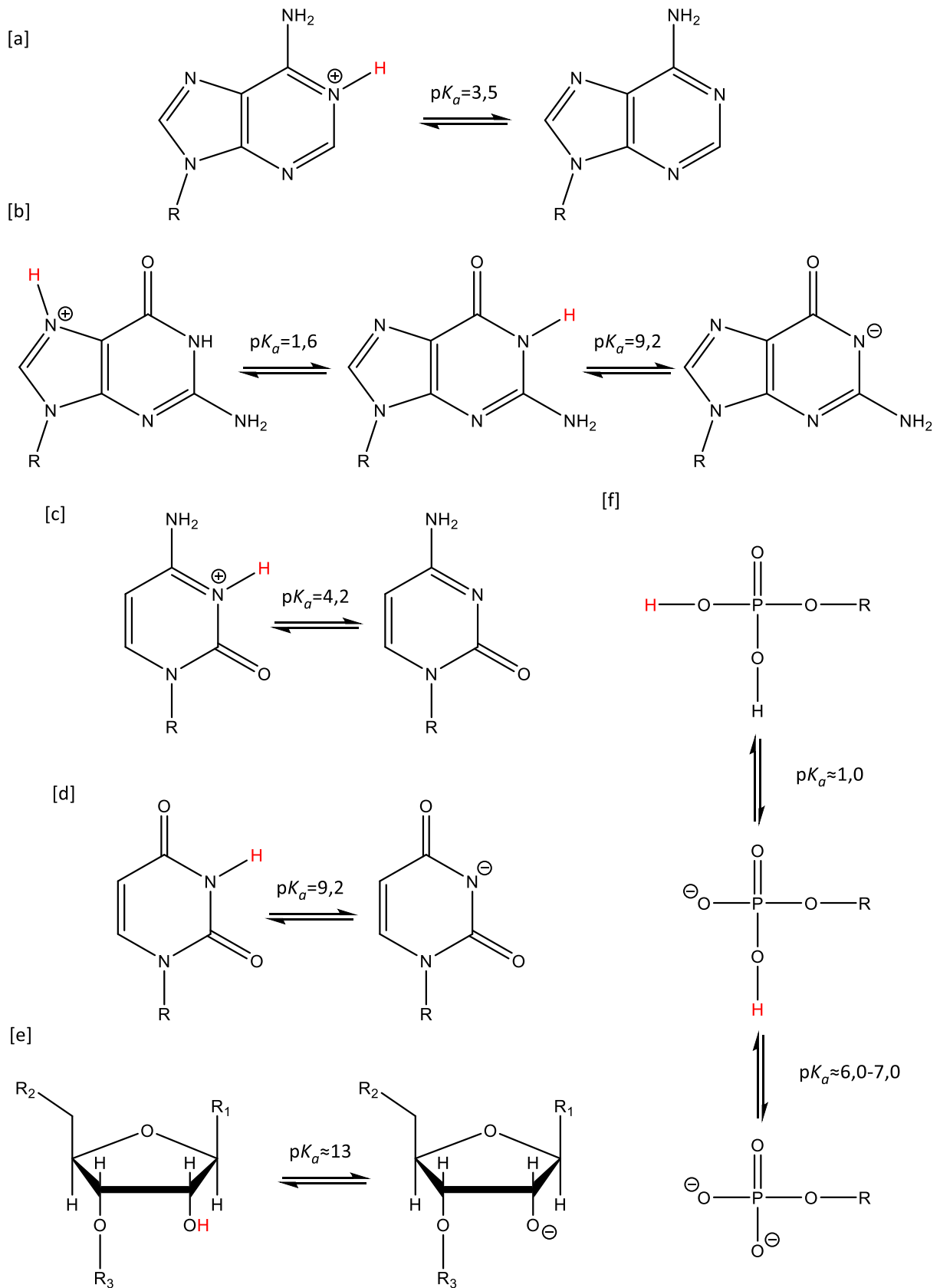


Figure 11 Sites de protonation et valeurs de  $pK_a$  des ribonucléotides avec [a] adénosine ; [b] guanosine ; [c] cytosine ; [d] uridine ; [e] 2' hydroxyle et [f] phosphate. Reproduit de [16].

Les caractères acides et basiques vont donc varier d'un nucléotide à l'autre principalement en fonction de la base hétérocyclique portée puisque les comportements acido-basiques du phosphoester et du ribose sont similaires entre les quatre nucléotides.

Les ribonucléotides sont les éléments de bases ou « briques » qui constituent les ARN. Ainsi, les ARN sont constitués simplement d'un enchaînement de squelette ribose-phosphate et de quatre bases nucléiques.

### 1.3 Propriétés des ARN

#### 1.3.1 Définition et généralités

D'un point de vue chimique, l'ARN est un copolymère linéaire constitué d'une séquence de nucléotides. Les ribonucléotides sont liés les uns aux autres de façon covalente via des liaisons phosphodiesteres en 5' et 3' du ribose formant ainsi l'ARN comme représenté sur la Figure 12.

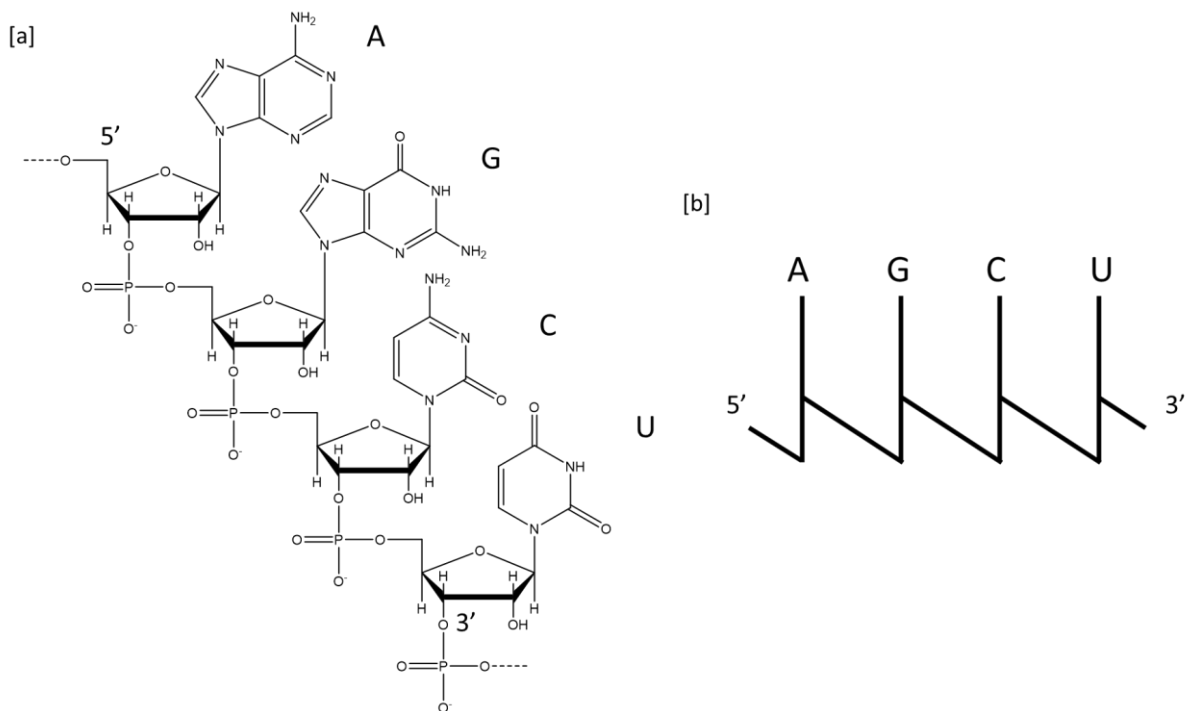


Figure 12 Représentations d'un ARN formé d'une séquence de nucléotide avec les quatre bases canoniques. Avec [a] structure chimique et [b] représentation simplifiée : un trait vertical représente un ribose et un trait horizontal représente un phosphate.

Un ARN peut avoir différentes terminaisons. Du côté 5', les ARN peuvent se terminer avec un groupe hydroxyle ou bien un groupe phosphate qui sont notés 5'OH et 5'p respectivement. À l'autre extrémité, la terminaison peut également être un groupe hydroxyle ou phosphate notés 3'OH et 3'p. Cependant, due à la présence de l'oxygène en 2' du ribose, il existe une autre forme appelée phosphate cyclique, notée 3'>p (voir Figure 13). Cette terminaison apparaît après une réaction de déshydratation entre le phosphate et l'hydroxyle 2'.

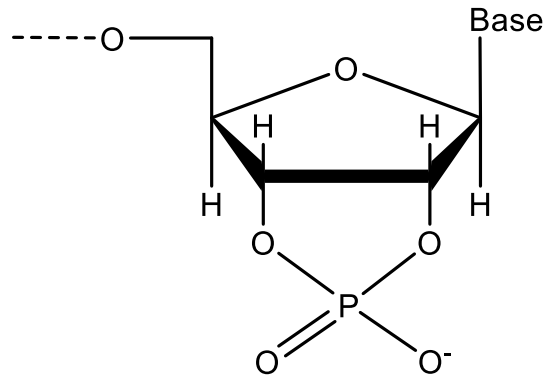


Figure 13 Structure d'une terminaison 3' phosphate cyclique d'un ARN

Par convention, la lecture des séquences d'acides nucléiques (ou structure primaire) se fait dans le sens 5' → 3'. Pour les ARN, la structure primaire s'écrit avec la séquence linéaire des 4 nucléotides en utilisant la notation simplifiée (A, G, C ou U). Si la séquence ne précise pas la composition des extrémités 5' et 3', il s'agit alors de groupements OH terminaux. Il est également possible de désigner les ARN par leur taille. Par exemple, pour un ARN composé de 120 nucléotides, il est noté ARN 120-mer.

### 1.3.2 Propriétés chimiques

En ce qui concerne les propriétés chimiques, les ARN sont acides puisque les groupes phosphates, internes ou terminaux, possèdent une valeur de  $pK_a$  d'environ 1 pour leur première déprotonation [12]. Ces espèces sont très solubles dans les solutions aqueuses et sont partiellement ou complètement insolubles dans les solvants organiques. En solution aqueuse, soit les ARN sont des espèces zwitterioniques, soit ils sont chargés négativement. Cela va dépendre de leur composition en bases azotées et du pH de la solution. À pH biologique ( $\approx 7,2$ ), les quatre bases nucléiques sont neutres ainsi que le ribose. Uniquement le phosphate porte une charge négative puisque son  $pK_a$  est d'environ 1. C'est pourquoi les ARN sont chargés négativement dans ces conditions. Dans des conditions acides, de pH 4,5 à 2, les bases C, A et G (dans cet ordre) sont protonées. En-dessous de pH 2, l'état de charge de l'ARN dépend du nombre d'uraciles portés puisque cette base n'est pas protonable. Lorsque le pH est inférieur à 1, les phosphates sont reprotonés et l'ARN peut porter une charge positive si le nombre d'uracile qu'il contient est faible. Les liaisons phosphodiesters sont relativement résistantes dans les milieux acides dilués [14]. En condition alcaline ou neutre, les oligonucléotides sont anioniques. Dans l'intervalle de pH 9-10, les bases G et U sont déprotonées et au-dessous de 12, le proton de l'hydroxyle 2' du ribose se dissocie. Cependant, avec un pH très basique, les ARN s'hydrolysent rapidement en ribonucléotides. Dans les espèces vivantes, les oligonucléotides sont sous forme de sel. En effet, les charges négatives des phosphates sont compensées par des contre-ions qui sont très généralement des métaux monovalents comme  $Na^+$  ou  $K^+$ . Les ammoniums, tels que  $NH_4^+$ ,  $(Me)_3NH^+$  ou  $(Et)_3NH^+$ , sont également des contre-ions courants pour les ARN [13].

### 1.3.3 Considérations structurales

Les groupements chimiques des acides nucléiques présentent plusieurs degrés de liberté avec des rotations de liaison et des angles de torsion entre les liaisons. Cela impacte donc la structure des ARN. Dans les nucléotides, ce sont les angles de torsions qui sont utilisés pour décrire la géométrie d'une unité nucléotidique. Si on considère 4 atomes A, B, C et D successivement liés et projetés dans un plan normal à la liaison B-C, l'angle entre le plan contenant A-B-C et le plan contenant B-C-D est appelé angle de torsion. Les angles de torsion des liaisons dans une unité nucléotidique et leurs symboles recommandés par l'IUPAC sont indiqués sur la Figure 14. Le squelette d'un ARN est constitué d'une répétition de six liaisons simples formées par les motifs phosphate-ribose donnant naissance à 6 degrés de liberté. Dans ce squelette, les angles de torsion sont notés  $\alpha \rightarrow \zeta$  dans la direction 5'  $\rightarrow$  3' de l'ARN. Ainsi l'angle  $\alpha$  de l'unité nucléotidique  $i$  représente l'angle de torsion entre les atomes O3'(i-1)-P(i)-O5'(i)-C5'(i). L'angle de torsion de la liaison N-glycosidique entre la base et le ribose est noté  $\chi$ , aussi considéré comme le 7<sup>ème</sup> degré de liberté. La séquence d'atome choisie pour définir cet angle est O4'-C1'-N9-C4 dans le cas des purines et O4'-C1'-N1-C2 dans le cas des pyrimidines. Enfin, les angles de torsion dans le sucre sont notés  $v_x$  en partant de la liaison O4'-C1' ( $v_0$ ) et suivant le sens des aiguilles d'une montre [17].

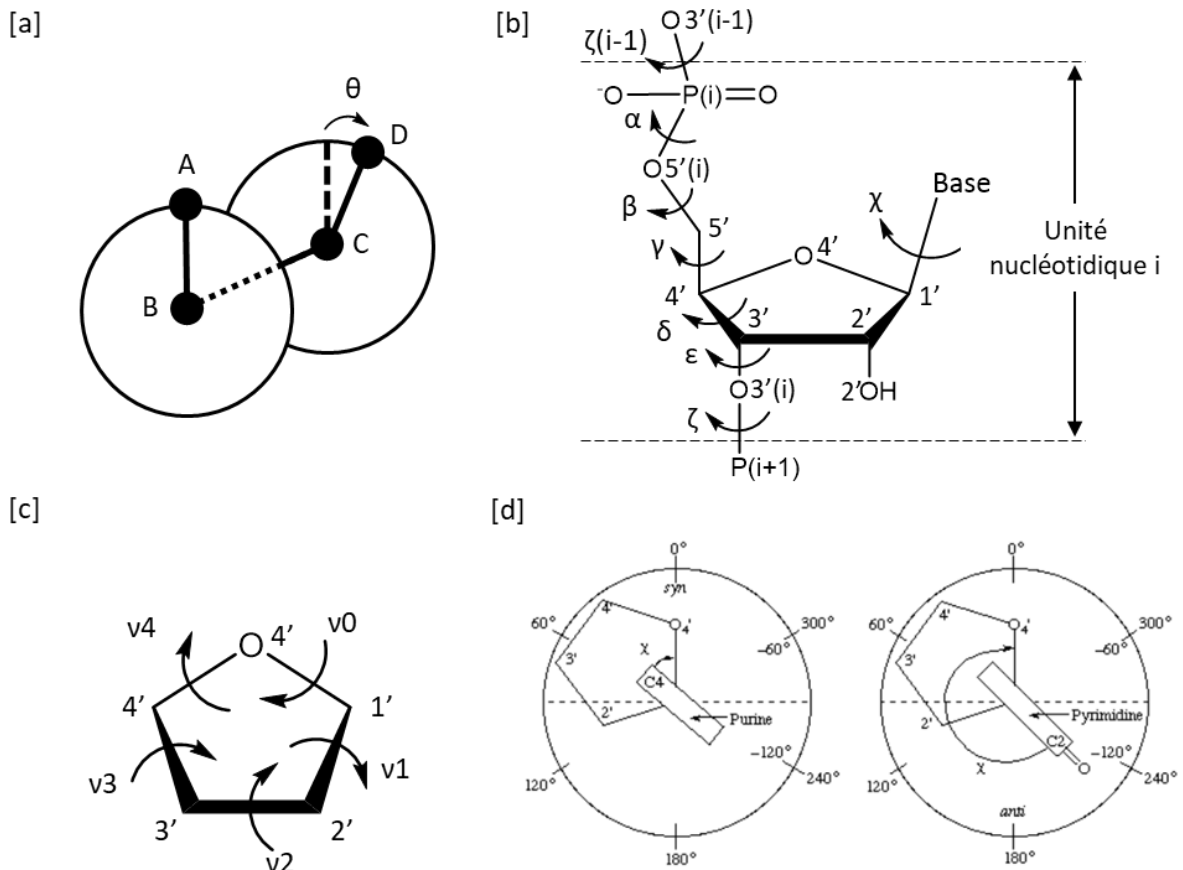


Figure 14 [a] Illustration d'un angle de torsion entre 4 atomes ; [b] Notations des angles de torsion du squelette d'ARN dans une unité nucléotidique  $i$  [c] Notations des angles de torsion dans le ribose de l'ARN et [d] Angle de torsion  $\chi$  entre le ribose et la base azotée avec à gauche une purine et à droite une pyrimidine. (recommandé par IUPAC).

La conformation des riboses influence fortement la structure des ARN puisque chaque unités nucléotidiques sont constituées d'un  $\beta$ -D-ribofuranose. Dans la structure cyclique de ce sucre, des variations de conformation surviennent avec le blocage des rotations des liaisons. Ainsi, le sucre se plie afin de limiter la torsion entre les atomes en optimisant les angles de liaison. Cela génère donc des conformations. Le plissement du sucre est dû au pentagone plat qui place tous les atomes de carbone en conformation éclipsée. Ce plissement permet donc de diminuer les effets stériques dus aux conformations éclipsées au détriment d'une légère déviation de l'angle de liaison tétraédrique ( $108^\circ$  ici au lieu de  $109.5^\circ$  pour un carbone tétraédrique). Classiquement, un furanose adopte une structure d'**enveloppe** ou de **demi-chaise** (voir Figure 15). Dans la conformation enveloppe, 4 atomes sont dans le même plan et 1 hors du plan alors que dans la conformation demi-chaise, 3 atomes sont dans le même plan et les deux autres atomes sont de part et d'autre du plan. Le plan utilisé pour référence dans ces conformations du ribose est celui formé par les atomes  $C4'-O4'-C1'$ . Le nom d'enveloppe provient de la structure rappelant une enveloppe avec le rabat ouvert et le nom de demi-chaise provient de la conformation chaise des cycles à 6.

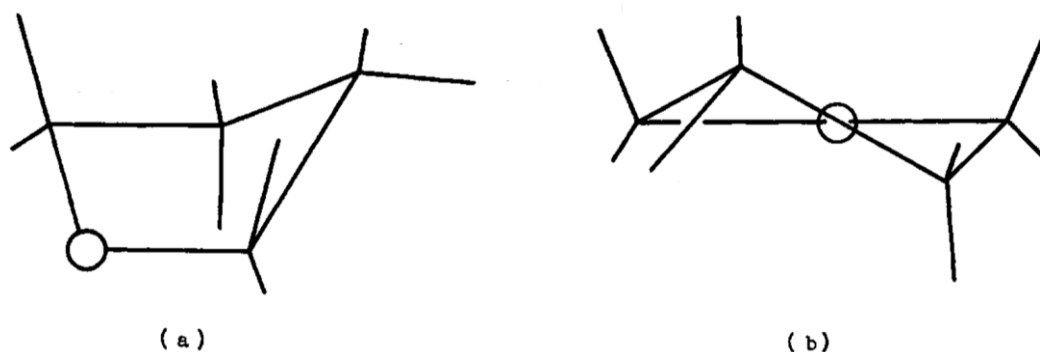


Figure 15 Conformations du ribose avec [a] la conformation enveloppe (E) (enveloppe en anglais) et [b] la conformation demi-chaise (T) (twist en anglais) Issue de [18].

En réalité, il est rare que le ribose des ARN adopte parfaitement une conformation enveloppe ou demi-chaise. Le ribose peut donc adopter plusieurs conformations intermédiaires comme décrit sur la Figure 16. Le pli est décrit en fonction de l'atome de carbone le plus éloigné du plan. Lorsqu'un atome adjacent est légèrement déplacé vers le côté opposé, cela donne des torsions avec des motifs nommés S (south) ou N (north) en fonction de l'orientation [19]. Les notation E et T désignent respectivement la conformation enveloppe et demi-chaise. Le chiffre en exposant indique le carbone qui est du même côté que  $C5'$  et la base azotée, tandis que le chiffre en indice indique le carbone qui du côté opposé à  $C5'$  et la base azotée. Dans la conformation E des riboses, seul le carbone le plus éloigné du plan est désigné puisque les autres atomes sont dans le plan. De plus, ces conformations dépendent, bien entendu, des groupements chimiques liés au pentose.



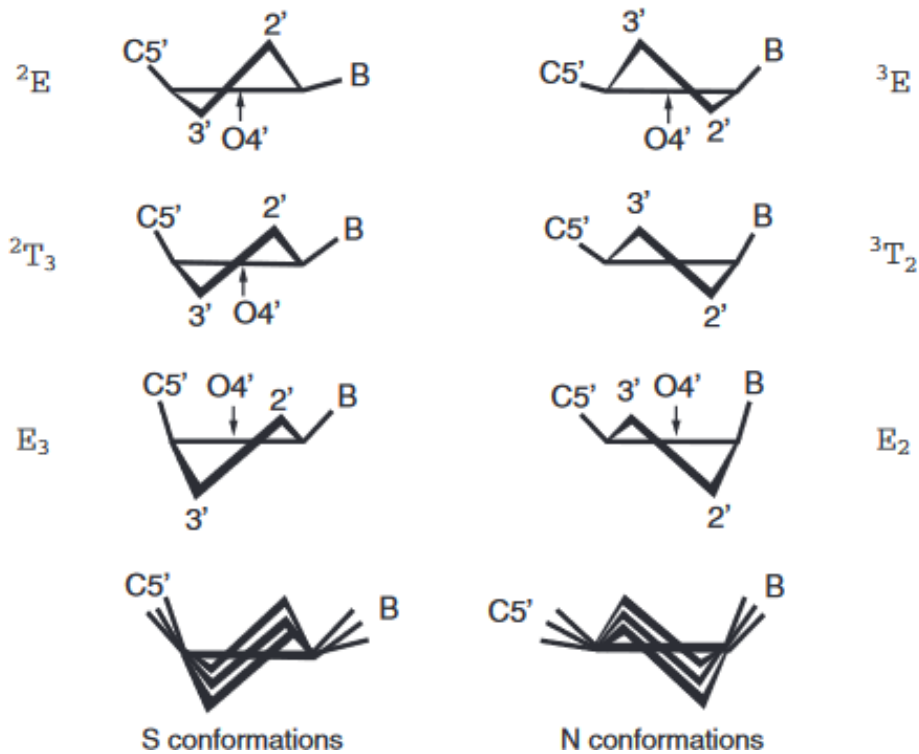


Figure 16 Conformations possibles adoptées par le ribose d'un ARN. Reproduit de [19].

Dans les ARN, il a été montré que la conformation favorisée par les  $\beta$ -D-ribofuranoses se rapproche de la conformation  $^3E$ , où le carbone C3' pointe du même côté que le carbone C5' et la base azotée (voir la Figure 17). Cette conformation est couramment nommée sous le nom de **C3'-endo**. C'est donc la conformation de plus faible énergie qui est fortement favorisée bien que des interconversions entre les conformations C3'-endo et C2'-endo se produisent [18].

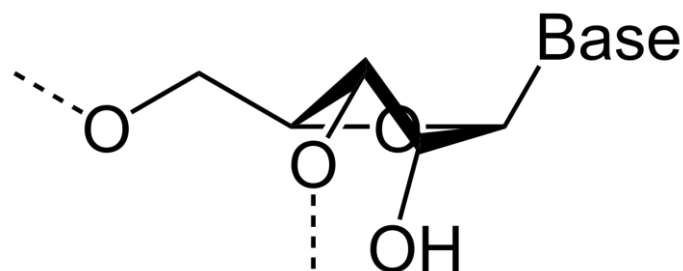


Figure 17 Conformation C3'-endo du ribose observée chez les ARN.

Les phosphates donnent une flexibilité à la structure de l'ARN grâce à la liberté de rotation des liaisons simples de ce groupement. Cette flexibilité est très importante pour les ARN car elle autorise les nucléotides à former des appariements de façon intramoléculaires qui permettent de stabiliser les structures.

Une autre caractéristique fondamentale et incontournable pour comprendre les structures des acides nucléiques est l'appariement des bases azotées. En effet, les bases azotées des nucléotides peuvent créer, entre elles, des liaisons faibles et former ce que l'on appelle des paires de bases canoniques selon le même principe énoncé par Watson et Crick sur les ADN [1]. Ces liaisons faibles entre les bases sont des ponts hydrogènes. Chez les ARN,

on compte **deux paires de bases principales**. Avec un pH biologique, l'adénosine s'apparie avec l'uridine en formant 2 liaisons hydrogènes (**A=U**) et la guanosine s'apparie avec la cytidine via 3 liaisons hydrogènes (**G≡C**) (voir Figure 18). Ces paires de bases sont formées grâce à la disposition géométrique des atomes donneurs de protons et accepteurs de protons dans les bases nucléiques. Ces appariements se font de façon coplanaire entre les deux bases azotées des nucléotides. La paire GC est énergétiquement plus stable que la paire AU puisqu'elle met en jeu trois liaisons hydrogènes [20]. Ces appariements peuvent aussi bien se produire au sein d'une même molécule (intramoléculaire) qu'entre deux molécules (intermoléculaire). Dans les ARN, il est possible d'observer des paires de bases appelées « wobble » ou appariements bancals. Un appariement wobble courant chez les ARN est celui entre la guanosine et l'uridine par le biais de deux ponts hydrogènes (G=U) [21]. Cependant, cette paire de base est énergétiquement moins stable que les paires canoniques. Il existe aussi des paires de bases non-canoniques avec par exemple une paire G-C qui est formée avec deux liaisons hydrogènes au lieu de trois. Il faut noter que des formes tautomères peuvent se produire dans les bases nucléiques et créer de nouvelles paires de bases. Cependant, ces formes sont défavorisées avec des constantes de vitesse de l'ordre de  $100 \text{ s}^{-1}$  et une fraction molaire des tautomères minoritaires d'environ 0.1 [22, 23].

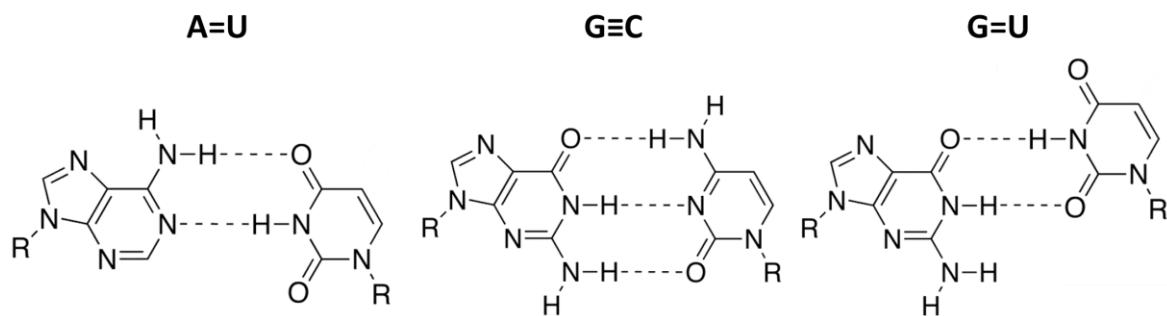


Figure 18 Paires de bases dans les ARN présentant les appariements Watson-Crick (adénosine-uridine et guanosine-cytidine) et wobble (guanosine-uridine) avec les liaisons hydrogènes représentées en pointillées.

Les ARN adoptent, dans la très grande majorité des cas, une structure de chaîne simple (aussi appelée monocaténaire ou simple brin). Autrement, dit les nucléotides ne forment pas de paires de bases. Cependant, les ARN peuvent se replier sur eux-mêmes pour que les nucléotides s'apparient de façon intramoléculaire. L'empilement des bases permet de former différentes structures très stables en double hélice stabilisées par la force des ponts hydrogènes. Ainsi, les structures secondaires des molécules d'ARN montrent les appariements entre les bases. La succession des bases appariées forment, dans les structures secondaires, ce qu'on appelle des tiges. Lorsque dans une région de l'ARN il n'y a pas de complémentarité de base, c'est-à-dire une absence d'appariement, il y a apparition de motifs. Dans les structures secondaires, plusieurs motifs élémentaires sont remarquables. La Figure 19 montre les principaux motifs observables dans les ARN.

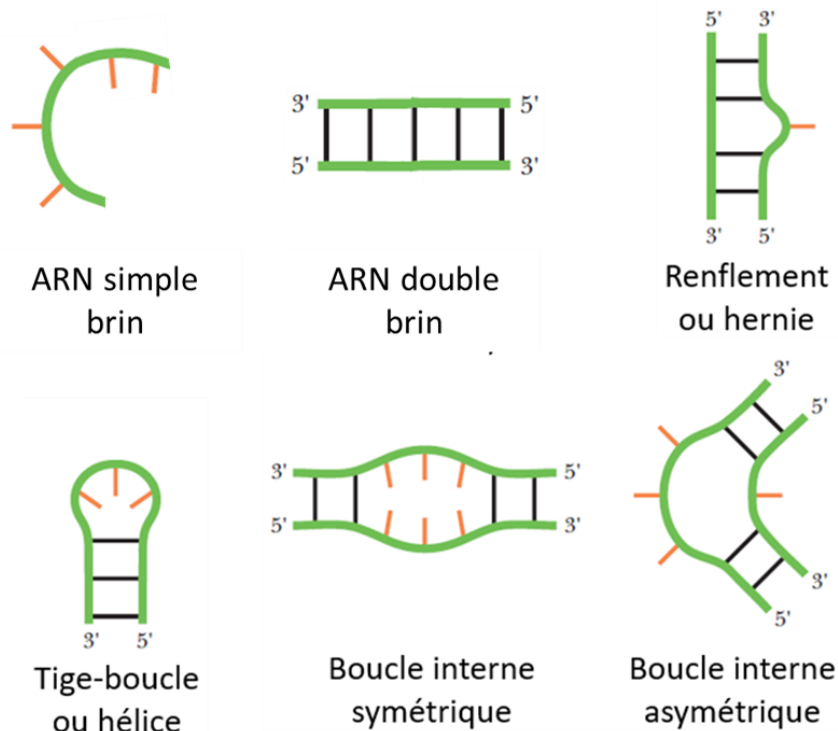


Figure 19 Motifs observables dans les structures secondaires d'ARN induites par l'empilement de base complémentaires. Les traits verts représentent l'enchaînement ribose-phosphate, les traits noirs représentent un appariement entre deux bases et les traits orange représentent des bases non-appariées. Adaptée de [14].

Ces motifs ont des tailles variables selon le nombre de nucléotides non appariés. Le nombre de motif dans un ARN dépend de sa séquence. Ainsi, plusieurs motifs peuvent se dessiner dans une même molécule d'ARN formant des structures complexes qui affectent considérablement la structure 3D (aussi appelée structure tertiaire). En effet, la structure tertiaire des ARN a une importance cruciale car elle permet d'assurer les fonctions de l'ARN dans une cellule et notamment le repliement correct de la chaîne [24], comme cela est discutée plus tard dans le chapitre suivant (cf. 2.1). Par conséquent, les structures tertiaires des ARN sont régies par les appariements des bases nucléiques.

### 1.3.4 Propriété optique : absorption UV

Les bases azotées des acides nucléiques sont des groupements chimiques riches en électrons présentant des systèmes conjugués comme vu auparavant. Cette caractéristique permet à ces macromolécules d'avoir la propriété d'absorber les photons dont les longueurs d'ondes se situent dans le domaine de l'ultraviolet (UV). Bien que cette capacité à absorber la lumière n'est pas impliquée dans les processus biologiques, elle est exploitée afin de caractériser les ARN. Les bases nucléiques possèdent des doublets non-liants (électrons  $n$ ), des liaisons simples (électrons  $\sigma$ ) et aussi des liaisons doubles (électrons  $\pi$ ) susceptibles d'absorber la lumière UV. L'absorption de la lumière UV par ces molécules se traduit par plusieurs transitions électroniques, c'est-à-dire que des électrons vont passer d'un niveau énergétique  $E_1$  à un niveau énergétique  $E_2$  tel que  $E_1 < E_2$ . Les transitions qui peuvent être observées dans ces molécules sont, par exemple, les transitions  $n \rightarrow \sigma^*$ ,  $n \rightarrow \pi^*$  ou encore  $\pi \rightarrow \pi^*$ . C'est essentiellement cette dernière qui sera utilisée dans la pratique. En effet, la

délocalisation des électrons  $\pi$  permet de diminuer l'énergie nécessaire aux transitions  $\pi \rightarrow \pi^*$  par effet bathochrome, il est donc possible d'observer cette transition avec une longueur d'onde plus grande. Dans la pratique, cette capacité à absorber la lumière UV est très utilisée afin de détecter les ARN. La détection des ARN est classiquement réalisée autour de 260 nm puisque les 4 bases azotées présentent un maximum d'absorbance autour de cette longueur d'onde [25, 26].

De plus, d'après la loi de Beer-Lambert, l'absorption d'une longueur d'onde par un échantillon est reliée à sa concentration suivant l'équation (1) :

$$A = \varepsilon \cdot l \cdot c \quad (1)$$

$A$  : absorbance à une longueur d'onde donnée (sans unité)

$\varepsilon$  : coefficient d'extinction molaire à une longueur d'onde donnée ( $L \cdot mol^{-1} \cdot cm^{-1}$ )

$l$  : longueur du trajet optique (cm)

$c$  : concentration ou molarité ( $mol \cdot L^{-1}$ )

Cette équation est valable pour une onde monochromatique et des solutions diluées pour éviter les phénomènes de complexation ou d'association des molécules qui peuvent changer de manière significative les coefficients d'extinction. Ainsi, une simple mesure spectrophotométrique permet de déterminer la concentration d'une solution en ARN.

Dans les cas des acides nucléiques, c'est la concentration massique qui est déterminée avec l'équation de Beer-Lambert modifiée suivante (2) :

$$\rho = \frac{A \cdot e}{l} \quad (2)$$

$\rho$  : concentration massique ( $ng \cdot \mu L^{-1}$ )

$e$  : coefficient d'extinction massique à une longueur d'onde donnée ( $ng \cdot cm \cdot \mu L^{-1}$ )

Le pH de la solution est un paramètre important à prendre en compte lors des mesures d'absorbance de solution d'ARN car l'intensité mais aussi les longueurs d'ondes des bandes d'absorption varient en fonction de celui-ci [27]. (Voir les spectres ci-dessous).

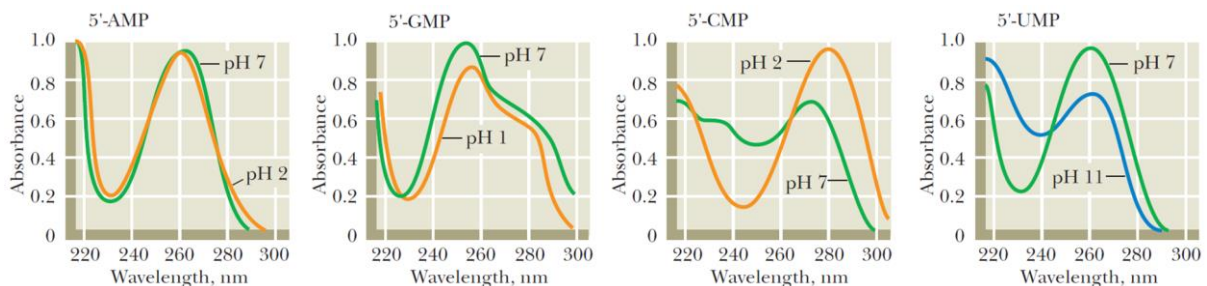


Figure 20 Spectres d'absorption UV des quatre nucléotides canoniques mesurés à différents pH. Issue de [14].

Dans la très grande majorité des cas, les solutions d'ARN sont proches du pH neutre. Par ailleurs, à ce pH, les absorptions maximales de ces quatre nucléotides sont approximativement observées autour de la longueur d'onde  $\lambda=260$  nm. Cette longueur d'onde

est ainsi couramment utilisée pour déterminer la concentration d'une solution en oligomère. Cependant, le coefficient d'extinction  $\epsilon$  dépend de la séquence d'ARN puisque les quatre bases présentent des coefficients d'extinction molaire différents à 260 nm. A titre indicatif, les coefficients d'extinction molaire des 5'NMP :

- AMP : 15200 L.mol<sup>-1</sup>.cm<sup>-1</sup>
- GMP : 12010 L.mol<sup>-1</sup>.cm<sup>-1</sup>
- CMP : 7050 L.mol<sup>-1</sup>.cm<sup>-1</sup>
- UMP : 10000 L.mol<sup>-1</sup>.cm<sup>-1</sup>

Dans la pratique, il n'est pas obligatoire de calculer systématiquement les coefficients d'extinction spécifiques pour chaque acide nucléique. Pour les ARN, un coefficient d'extinction moyen, accepté par la communauté scientifique, est généralement utilisé pour estimer les concentrations avec une valeur de 40 ng.cm.μL<sup>-1</sup>. Néanmoins, avec des ARN dont la séquence est très courte (<25 nucléotides), la contribution de chaque nucléotide a une plus grande influence sur le coefficient  $\epsilon$ . Une estimation de ce coefficient est alors nécessaire pour éviter de sous-évaluer ou surévaluer la concentration de l'ARN en solution. Pour des valeurs plus précises, il est possible d'effectuer une comparaison avec un étalon de concentration connue. Une mesure de l'absorbance de cet échantillon permet de déterminer sa valeur  $\epsilon$ . Il est aussi possible de tracer une gamme d'étalonnage à partir d'une solution étalon. Cette dernière méthode est plus précise car elle permet, en plus de déterminer  $\epsilon$ , de vérifier la linéarité entre absorbance et concentration en ARN.

Ici ont été décrits et définis les ARN d'un point de vue moléculaire. Toutes les caractéristiques, telles que la chimie ou la structure tertiaire des ARN, sont nécessaires pour assurer les différents rôles essentiels des ARN dans la cellule. Les rôles ainsi que les différents types des ARN sont décrits dans le chapitre suivant.

#### 1.4 Références bibliographiques

1. Watson, J.D. and F.H.C. Crick, *Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid*. Nature, 1953. **171**(4356): p. 737-738.
2. Gros, F., et al., *Unstable ribonucleic acid revealed by pulse labelling of Escherichia coli*. Nature, 1961. **190**: p. 581-5.
3. Sen, S., U. Pal, and N.C. Maiti, *pK(a) Determination of D-Ribose by Raman Spectroscopy*. Journal of Physical Chemistry B, 2014. **118**(4): p. 909-914.
4. Grimmett, M.R., *Advances in Imidazole Chemistry*, in *Advances in Heterocyclic Chemistry*, A.R. Katritzky and A.J. Boulton, Editors. 1970, Academic Press. p. 103-183.
5. Nowak, M.J., et al., *Spectroscopic studies on vapor-phase tautomerism of natural bases found in nucleic-acids*. Zeitschrift Fur Naturforschung C-a Journal of Biosciences, 1978. **33**(11-1): p. 876-883.
6. Wong, Y.P., *Tautomerism of nucleic acid bases. Guanine and cytosine*. Journal of the American Chemical Society, 1973. **95**(11): p. 3511-3515.
7. Dreyfus, M., et al., *Tautomerism of purines. 1. N(7)H N(9)H equilibrium in adenine*. Journal of the American Chemical Society, 1975. **97**(9): p. 2369-2376.
8. Brown, R.D., et al., *Tautomers of Cytosine by Microwave Spectroscopy*. Journal of the American Chemical Society, 1989. **111**(6): p. 2308-2310.

9. Leszczynski, J., *Tautomerism of uracil - The final chapter - 4th-order electron correlation contributions to the relative energies of tautomers*. Journal of Physical Chemistry, 1992. **96**(4): p. 1649-1653.
10. Nir, E., et al., *Guanine tautomerism revealed by UV-UV and IR-UV hole burning spectroscopy*. Journal of Chemical Physics, 2001. **115**(10): p. 4604-4611.
11. Katritzky, A.R., M. Karelson, and P.A. Harris, *Prototropic tautomerism of heteroatomic-compounds*. Heterocycles, 1991. **32**(2): p. 329-369.
12. Cantor, C.R., et al., *Biophysical Chemistry: Part I: The Conformation of Biological Macromolecules*. 1980: W. H. Freeman.
13. Nordhoff, E., F. Kirpekar, and P. Roepstorff, *Mass spectrometry of nucleic acids*. Mass Spectrometry Reviews, 1996. **15**(2): p. 67-138.
14. Walsh, E.J., *Biochemistry (Garrett, Reginald H.; Grisham, Charles M.)*. Journal of Chemical Education, 1997. **74**(2): p. 189.
15. Sheina, G.G., et al., *Prototropic Tautomerism of Nucleic-Acid Purine-Bases*. Studia Biophysica, 1986. **114**(1-3): p. 123-131.
16. Thaplyal, P. and P.C. Bevilacqua, *Chapter Nine - Experimental Approaches for Measuring pKa's in RNA and DNA*, in *Methods in Enzymology*, D.H. Burke-Aguero, Editor. 2014, Academic Press. p. 189-219.
17. Pullman, B. and J. Jortner. *IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN)*. in *Nucleic Acids: The Vectors of Life*. 1983. Dordrecht: Springer Netherlands.
18. Sundaralingam, M., *Stereochemistry of nucleic acids and their constituents. IV. Allowed and preferred conformations of nucleosides, nucleoside mono-, di-, tri-, tetraphosphates, nucleic acids and polynucleotides*. Biopolymers, 1969. **7**(6): p. 821-860.
19. Saenger, W., *Structure and function of nucleosides and nucleotides*. Angewandte Chemie International Edition in English, 1973. **12**(8): p. 591-601.
20. Cieplak, P. and P.A. Kollman, *Calculation of the Free-Energy of Association of Nucleic-Acid Bases in Vacuo and Water Solution*. Journal of the American Chemical Society, 1988. **110**(12): p. 3734-3739.
21. Varani, G. and W.H. McClain, *The G center dot U wobble base pair - A fundamental building block of RNA structure crucial to RNA function in diverse biological systems*. Embo Reports, 2000. **1**(1): p. 18-23.
22. Ruterjans, H., et al., *Evidence for Tautomerism in Nucleic-Acid Base-Pairs - H-1-Nmr Study of N-15 Labeled Transfer-Rna*. Nucleic Acids Research, 1982. **10**(21): p. 7027-7039.
23. Westhof, E., *Isostericity and tautomerism of base pairs in nucleic acids*. Febs Letters, 2014. **588**(15): p. 2464-2469.
24. Sundaralingam, M., et al., *Mechanisms of chain folding in nucllic-acids - (Omega',Omega) Plot and its correlation to nucleotide geometry in yeast transfer-RNA-PHE1*. Nucleic Acids Research, 1976. **3**(10): p. 2471-2484.
25. Huberman, J.A., *Importance of measuring nucleic-acid absorbency at 240 nm as well as at 260 and 280 nm*. Biotechniques, 1995. **18**(4): p. 636-636.
26. Manchester, K.L., *Value of A(260)/A(280) ratios for measurement of purity of nucleic-acids*. Biotechniques, 1995. **19**(2): p. 208-210.
27. Wilfinger, W.W., K. Mackey, and P. Chomczynski, *Effect of pH and ionic strength on the spectro-photometric assessment of nucleic acid purity*. Biotechniques, 1997. **22**(3): p. 474-481.

## CHAPITRE 2. BIOCHIMIE DES ARN

Dans les cellules, les ARN sont des composés très importants dans la synthèse de protéine. Ici, les notions de bases sont exposées afin de comprendre les mécanismes de production des protéines et également celui de la formation des ARN.

### 2.1 Les types et les rôles d'ARN

Les ARN remplissent plusieurs fonctions biologiques et leur principal rôle est le **support d'informations génétiques**. Afin d'assurer leurs rôles, ces acides nucléiques effectuent diverses fonctions chimiques telles que la catalyse, la reconnaissance moléculaire et le repliement [1]. En effet, les caractéristiques de ces oligomères permettent de synthétiser les protéines dont les cellules ont besoin. Le processus permettant de produire les protéines dans la cellule est appelé la **traduction**. Plusieurs procédés biochimiques sont nécessaires afin de traduire l'information génétique en protéine.

Ici sont cités quelques rôles des ARN dans une cellule :

- Transmettre l'information génétique : les ARN sont utilisés comme support pour transmettre les informations génétiques nécessaires à la production de protéines (processus de traduction).

- Catalyser des réactions : les ARN peuvent également adapter leur structure afin de catalyser des réactions biochimiques.

- Régulateurs de l'expression génétique : Des ARN non codants ont pour rôle de réprimer l'expression de certains gènes.

Ainsi, il existe plusieurs types d'ARN dans les cellules qui accomplissent ces différentes tâches. Les principales fonctions et caractéristiques de ces types d'ARN sont décrites ci-dessous.

#### 2.1.1 Les ARN messagers

Les ARN messagers (ARNm) correspondent aux séquences complémentaires et antiparallèles de l'ADN, c'est-à-dire des copies de l'ADN. Ce sont les **supports de l'information génétique** car ils apportent les informations utiles à la production des protéines. Ce sont donc des ARN codants. Les ARNm sont très hétérogènes selon les cellules et sont, la plupart du temps, sous la forme de chaîne individuelle. Leur taille peut varier significativement d'une cellule à une autre. Les ARNm contiennent, dans leur séquence de nucléotides, un code (les codons) permettant de synthétiser spécifiquement une protéine dans la cellule. Les codons, constitués chacun de trois nucléotides (triplets de nucléotides), respectent le code génétique : un codon correspond à un acide aminé (mais un acide aminé peut correspondre à plusieurs codons). C'est pour cela que les ARNm sont dits « codants ». Le message encodé est lu par le ribosome afin de traduire cette séquence en protéine nécessaire à la cellule. C'est donc cet

ARN qui orchestre la synthèse de la protéine en indiquant dans quel ordre doivent se succéder chaque acides aminés. Dans les eucaryotes, un ARNm permet de produire un unique polypeptide.

### 2.1.2 Les ARN de transfert

Les ARN de transfert (ARNt), sont les ARN dont le rôle est de **reconnaître spécifiquement l'ARNm** pour **apporter les acides aminés** servant à former une protéine. Ce sont des ARN fonctionnels car ils sont directement impliqués dans la production de protéines. Ils font partie des ARN non-codant (ARNnc) puisqu'ils ne sont pas traduits en protéines. Les ARNt, sont des petits oligomères composés en moyenne de 75 résidus [2]. Les ARNt sont couramment désignés par l'acide aminé qu'ils portent. Plus rarement, la séquence de leur anticodon est indiquée. Ainsi, l'ARNt provenant de la levure et portant un phénylalanine est noté ARNt<sup>Phe</sup><sub>GAA</sub>.

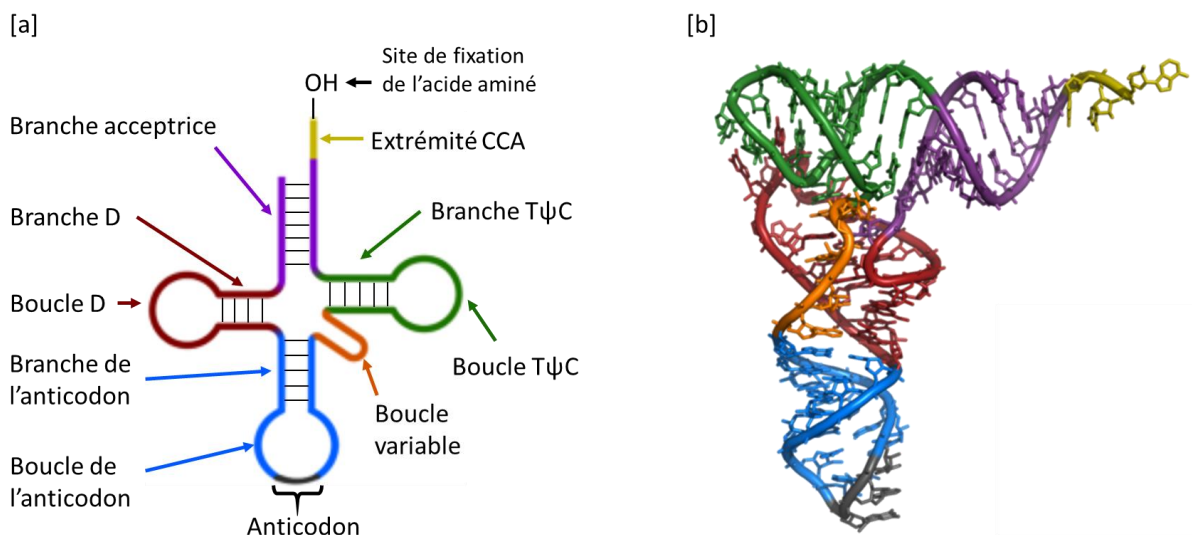


Figure 21 Représentation schématique de la structure de l'ARNt [a] Structure secondaire typique d'un ARNt en forme de feuille de trèfle. Les noms des régions sont indiqués et distingués avec différentes couleurs et [b] structure 3D typique d'un ARNt en forme de L inversé.

Les appariements intramoléculaires des nucléotides d'un ARNt induisent une structure secondaire particulière dite en forme de feuille de trèfle en raison des motifs tige-boucles comme le montre la Figure 21. La boucle D est constituée de nucléotides non-appariés et s'appelle ainsi car elle porte souvent des dihydrouridines (une uridine dont la base est hydrogénée). Quant à la boucle TψC, elle est constituée de sept nucléotides non-appariés comprenant systématiquement la séquence constante de nucléotide TψC. Enfin la constitution de la boucle variable va différer d'un ARN à un autre. Ces boucles interagissent entre elles via l'appariements de certains nucléotides. Ces interactions rapprochent spatialement la boucle D des boucles T et variable. L'ARNt est donc contraint de se replier sur lui-même. L'ARNt adopte alors une structure en forme de L inversé (cf. 3.2.2 pour plus de détails). Cette structure 3D, typique et très stable des ARNt, permet d'assurer le rôle de l'ARNt lors de la traduction [3].



Afin de remplir leurs fonctions, les ARNt possèdent toujours une extrémité 3' invariante sur la branche acceptrice. En effet, les séquences des ARNt se terminent toujours par les nucléotides CCA-3'-OH. L'acide aminé à livrer est lié à l'ARNt par l'intermédiaire d'une liaison ester sur l'adénine de cette extrémité. Cette liaison est aussi appelée un ester d' aminoacyle. Cet ester d' aminoacyle est réalisée entre l'acide carboxylique de l'acide aminé et l'extrémité 3'OH ou, plus rarement, 2'OH de l'adénine grâce à l'enzyme aminoacyl-ARNt synthétase [4]. Dans un premier temps, cette enzyme reconnaît spécifiquement l'ARNt puis elle incorpore l'acide aminé sur l'ARNt en catalysant la formation de la liaison. La molécule ainsi formée est appelée aminoacyl-ARNt. Il faut noter qu'un ARNt est conçu pour porter un acide aminé spécifique. Il existe donc un ARNt spécifique pour chaque acides aminés. La reconnaissance de l'ARNt par la synthétase se fait via un second code génétique. Pour interpréter ce second code génétique, la synthétase doit discriminer les 20 acides aminés et un grand nombre d'ARNt, et choisit uniquement ses substrats appropriés [5].

Une autre caractéristique fondamentale de l'ARNt se trouve dans l'anticodon. La boucle de l'anticodon est un segment à double-hélice avec sept nucléotides non-appariés dont trois forment l'anticodon. Pour produire une protéine, un codon de l'ARNm est reconnu spécifiquement par l' aminoacyl-ARNt portant l'acide aminé correspondant à ce codon (processus de décodage). Cette reconnaissance se fait via l'anticodon dont les trois nucléotides reconnaissent l'information génétique de l'ARNm qui spécifie un acide aminé. Ces nucléotides (positions 34, 35 et 36) sont ceux qui s'apparient de façon complémentaire et antiparallèle aux nucléotides du codon (interaction codon-anticodon) (voir Figure 22).

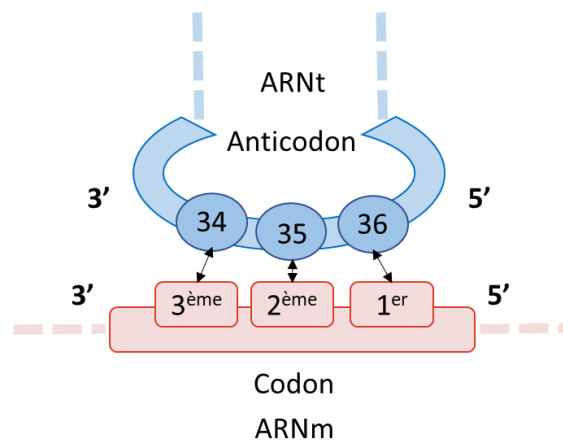


Figure 22 Représentation schématique de l'appariement anticodon-codon des nucléotides dans le processus de décodage.

Ainsi, pour former une protéine, l'ARNt assure deux fonctions : reconnaître le codon de l'ARNm et livrer l'acide aminé. L'ARNt est donc « l'interprète » de l'ARNm. Une fois la reconnaissance accomplie, l'acide aminé livré est inséré dans la chaîne polypeptidique, grâce au ribosome.

### 2.1.3 Les ARN ribosomiques

Les ARN ribosomiques (ARNr) font partie d'un très grand complexe ribonucléoprotéique appelé le ribosome. Les ARNr sont les oligomères les plus abondants

dans la cellule et contiennent un grand nombre de nucléotides. Comme les ARNt, ces macromolécules sont des ARNnc et sont essentiels dans le processus de traduction. Afin de caractériser ces types d'ARN, il est courant de les désigner par le coefficient de sédimentation dont l'unité est le Svedberg (S). Par exemple, le 5S ARNr correspond à l'ARN ribosomique possédant un coefficient de sédimentation de 5 Svedberg. En plus de leur grande taille, les ARNr présentent un très grand nombre de motifs dans leur structure secondaire [5] (voir Figure 23). Chez les eucaryotes, on retrouve les ARNr 5S ; 5,8S ; 28S (grande sous-unité) et 18S (petite sous-unité).

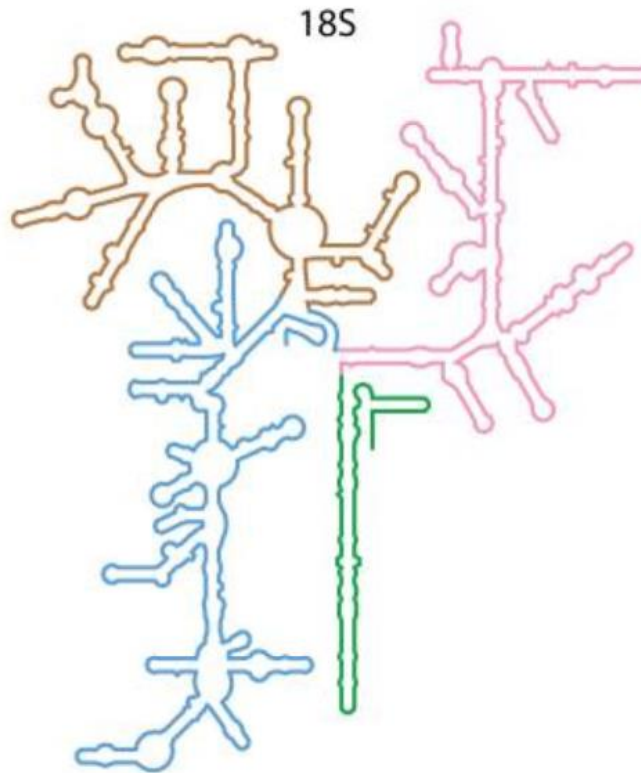


Figure 23 Structure secondaire de l'ARNm 18S d'une levure (*S. cerevisiae*). Extrait de [6].

Le complexe ribosomique est formé de plusieurs protéines et d'ARNr avec une masse totale dépassant  $10^6$  Da (environ 1/3 de protéines et 2/3 de nucléotides). **C'est dans cet assemblage supramoléculaire**, constitué de plusieurs sous-unités, **que la traduction est accomplie** à l'aide de plusieurs enzymes et cofacteurs [7]. Pour cela, le ribosome se déplace le long de l'ARNm et l'une des sous-unités dirige les interactions entre les codons successifs et les anticodons correspondants des aminoacyl-ARNt. Ainsi, puisqu'il permet la reconnaissance entre ARNm et ARNt, le ribosome a également une fonction de catalyseur dans la production des polypeptides. En effet, l'appariement de l'ARNt sur l'ARNm dans l'une des sous-unités facilite la livraison de l'acide aminé sur la protéine en croissance qui est réalisée dans une autre sous-unité. La réaction qui permet de transférer l'acide aminé porté par l'ARNt sur le polypeptide est appelée transpeptidation. Cette réaction est catalysée dans une sous-unité du ribosome par une enzyme : la peptidyltransférase. Une fois l'acide aminé incorporé dans la protéine, le ribosome se déplace sur l'ARNm dans le sens  $5' \rightarrow 3'$  jusqu'au codon suivant et libère l'ARNt qui a livré l'acide aminé (phénomène de translocation). Ensuite, un autre aminoacyl-ARNt reconnaît ce nouveau codon et le processus est répété jusqu'à ce que le

ribosome atteint le codon stop de l'ARNm. Ce codon stop est reconnu par des protéines spécifiques qui terminent la synthèse de la protéine. Cela permet de libérer l'ARNm, la protéine et le dernier ARNt du ribosome. Le phénomène de traduction est schématiquement représenté sur la Figure 24.

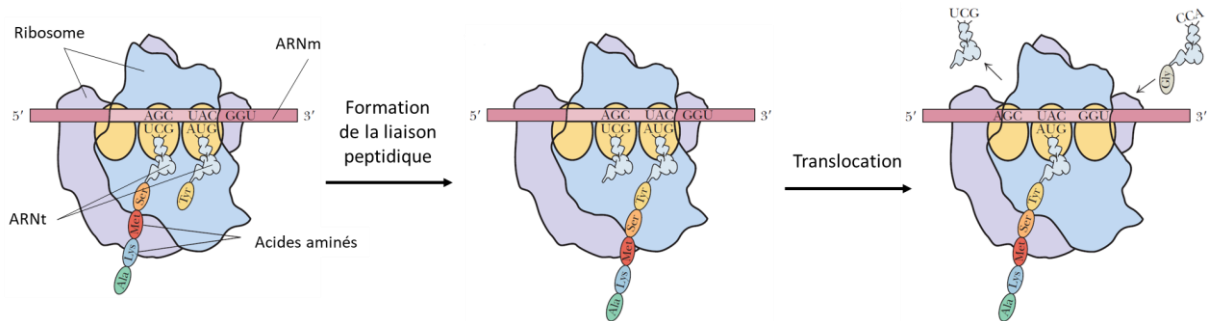


Figure 24 Représentation simplifiée des étapes de la traduction de l'ARNm en protéine. Le ribosome contient trois sites de liaison d'ARNt (représentés en jaune) : le premier site A pour accepteur (à droite), le deuxième P pour peptidyl (au milieu) et le troisième E (exit) pour sortie (à gauche). Adaptée de [5].

Ici ont été décrits les principaux ARN permettant d'assurer la synthèse des protéines. Comme cela a été montré, ces ARN aux caractéristiques uniques travaillent de concert pour produire des polypeptides dont leur séquence est bien définie.

### 2.1.4 Autres ARN

Il existe d'autres classes d'ARN [8-10] avec notamment les petits ARN (ARNs), appelés ainsi car ils ne sont composés que d'une vingtaine de nucléotides, qui sont impliqués dans un grand nombre de fonctions biologiques. Il est également courant de les appeler ARNnc puisqu'ils ne contiennent pas d'informations génétiques. Ces autres ARN, bien qu'ils ne soient pas directement impliqués dans le processus de traduction, sont indispensables notamment dans la régulation de l'expression de gène [5, 11]. Ces ARN peuvent cibler spécifiquement des ADN ou ARN via l'appariements de bases complémentaires [12]. Ces ARNs sont divisés en sous-classes en fonction de leurs rôles, avec par exemple les petits ARN interférents (ARNsi).

Les ARNsi perturbent l'expression des gènes en inhibant la production de protéines spécifiques. Cet ARN bloque la synthèse de protéine en s'appariant de façon complémentaire à un ARNm cible formant ainsi des régions d'ARN double brin (ARNdb). Le duplex formé est ensuite spécifiquement dégradé, empêchant par conséquent le processus de traduction de cet ARNm [13].

Une autre sous-classe, les micros ARN (ARNmi), permet le ralentissement de l'expression génétique [14]. Ce sont des petits ARN d'une vingtaine de nucléotides. Ces ARN empêchent également la synthèse de polypeptide en s'appariant avec des ARNm cible dont ils sont partiellement complémentaires. En revanche, contrairement aux ARNsi, les ARNmi ne provoquent pas la dégradation de l'ARNm. Dans ce cas précis, on dit que l'ARNm est réduit au silence.

Dans certains ARNm, des riboswitchs sont présents sur la partie 5' non-traduite et jouent un rôle dans la régulation de l'expression des gènes [15]. Les riboswitchs sont

constitués de 30 à 150 nucléotides et possèdent deux parties : l'aptamère et la plateforme d'expression. L'aptamère peut se chélater de façon très spécifique à un ligand. Cette liaison induit une modification de la structure de la plateforme d'expression et entraîne généralement le blocage de l'expression du gène mais peut aussi, dans certains cas, l'activer.

De plus, ces dernières années, de nouvelles classes d'ARN sont régulièrement mises en lumière avec par exemple les petits ARN nucléolaires (ARNsno) qui sont des catalyseurs pour modifier les ARNr ou encore les petits ARN nucléaires (ARNsn) qui permettent de maturer les ARNm ainsi que les ARNr afin de les rendre fonctionnels [16, 17].

Les différentes classes d'ARN sont disponibles dans la figure ci-dessous :

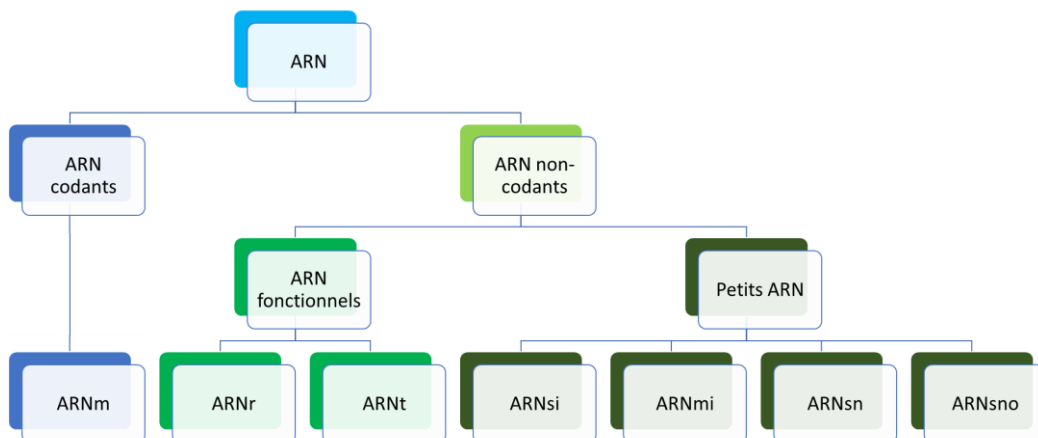


Figure 25 Classification des principaux ARN.

Les caractéristiques et les rôles des ARN ont été décrits dans cette partie. En effet, ces différents ARN jouent un rôle précis dans la production de protéines dans les cellules. Afin de créer les ARN, les cellules utilisent les ADN. En effet, les ADN contiennent les informations nécessaires pour produire les protéines. Les ARN sont donc les intermédiaires entre l'ADN et la protéine. La partie suivante décrit la production d'ARN dans une cellule.

## 2.2 Mécanisme général de la synthèse d'ARN

Le processus de production d'ARN à partir d'un ADN est appelé la **transcription**. De la même manière que pour la synthèse de polymère, la synthèse d'ARN implique trois étapes principales : l'initiation, l'élongation et la terminaison. Cette synthèse, catalysée par l'ARN polymérase, permet de créer une copie de l'ADN dont seul l'un des deux brins est copié. Un bref rappel du mécanisme de cette synthèse, chez les eucaryotes, est disponible ci-dessous [5].

### 2.2.1 L'initiation

La première étape de la transcription débute à l'aide d'une enzyme appelée l'ARN polymérase. Cette polymérase se fixe sur l'ADN à transcrire après la reconnaissance d'une séquence spécifique appelée le promoteur. Ils forment ainsi un complexe d'initiation nécessaire au bon déroulement de la deuxième étape. Dans le cas de la synthèse d'ARNt chez

les eucaryotes, c'est l'ARN polymérase III qui est sollicité. L'enzyme requiert la présence de plusieurs facteurs de transcription pour assurer ses fonctions.

### 2.2.2 L'élongation

L'élongation est l'étape qui consiste à produire l'ARN en liant les unités ribonucléotides entre elles. Pour cela, l'ARN polymérase commence à ouvrir la double hélice d'ADN sur laquelle elle est fixée. L'enzyme se déplace alors le long de l'ADN, sur l'unité de transcription, afin de l'ouvrir et de catalyser la formation des liaisons entre les nucléotides pour produire l'ARN. L'ouverture de l'ADN permet à des nucléotides triphosphates (NTP) de s'apparier aux bases de cette ADN de façon complémentaire. Les nucléotides s'apparient uniquement sur un seul des deux brins d'ADN. Le morceau d'ADN copié est appelé brin matrice. Ensuite, une liaison phosphodiester est créée entre les nucléotides. Cette liaison est formée par attaque nucléophile de l'hydroxyle 3' d'un nucléotide de l'ARN en croissance sur le phosphate  $\alpha$  d'un nucléotide triphosphate adjacent (voir Figure 26). De cette façon, les liaisons phosphodiester des ARN sont toujours en 5' et 3' des nucléotides. L'oligomère grandit au fur et à mesure de l'apport en nucléotide triphosphate avec une vitesse d'environ 30 nucléotides par seconde à 37°C et les chaînes d'ADN reforment le double brin en fonction de l'avancée du processus. L'enchaînement des ribonucléotides constituant l'ARN est alors complémentaire à la séquence d'ADN, excepté la substitution des thymines par les uraciles, d'où le terme de transcription.

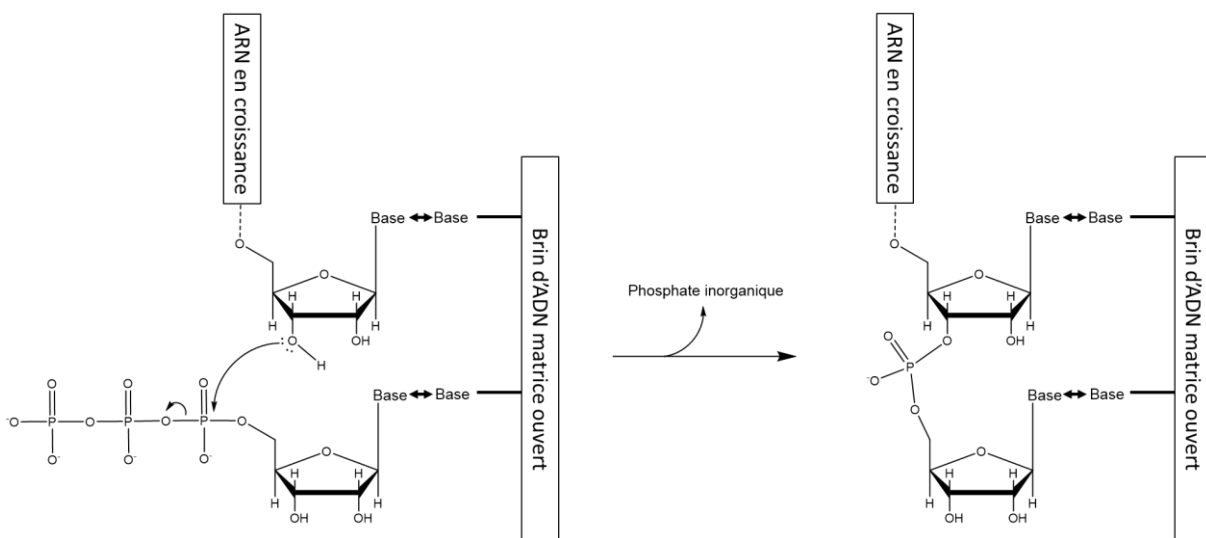


Figure 26 Principe de l'élongation d'un ARN à partir du brin matrice d'un ADN dans le processus de transcription.

### 2.2.3 La terminaison

La terminaison est l'étape de la transcription qui consiste en la reconnaissance d'une séquence de l'ADN (codon stop) à partir de laquelle l'élongation de l'ARN est stoppée. À partir de ce point, appelé le site de terminaison, l'ARN polymérase est bloquée, achevant ainsi la transcription. Enfin, le biopolymère transcrit et la polymérase sont libérés et l'ADN retrouve sa structure hélicoïdale.

Les différentes étapes d'une transcription sont représentées schématiquement sur la Figure 27.

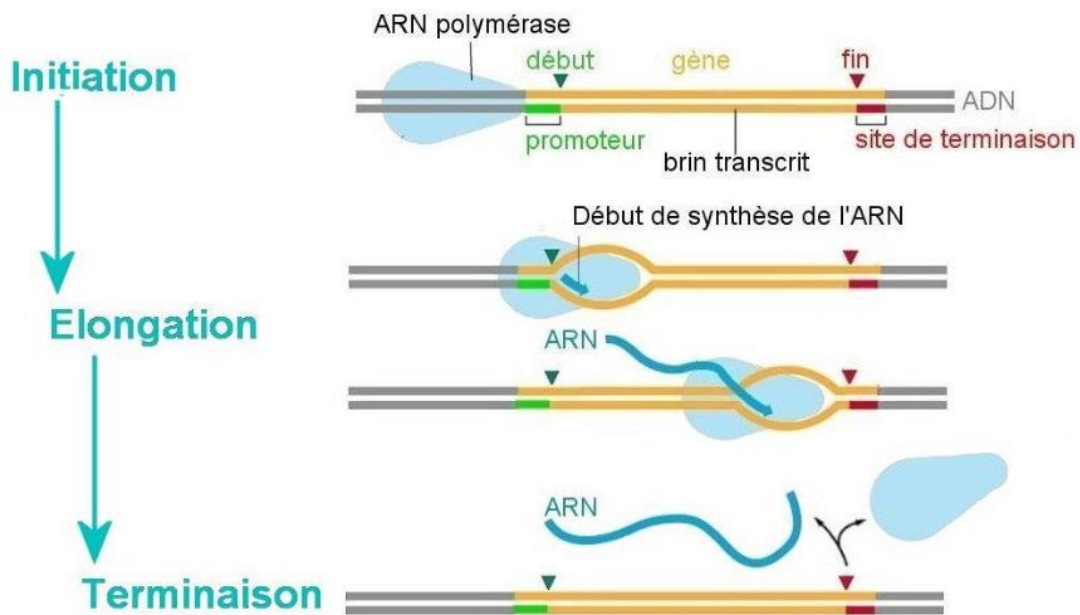


Figure 27 Schéma simplifié des différentes étapes impliquées dans le processus de transcription.

Ainsi, les molécules d'ARN sont produites, contenant les caractéristiques nécessaires pour synthétiser les protéines. Cependant, les ARN formés en tant que tels ne sont pas encore fonctionnels et nécessitent une étape supplémentaire : la maturation de l'ARN. Cette étape permet de modifier un ARN, après sa transcription, afin de lui donner toutes les propriétés essentielles pour assurer son bon fonctionnement. Cette étape fait partie intégrante de la transcription et dépend du type d'ARN. L'ensemble des modifications issu de ce processus sont appelées des modifications post-transcriptionnelles. Plusieurs réactions biochimiques complexes sont mises en œuvre, avec notamment l'aide d'enzymes, pour former ces modifications et obtenir un ARN mature et fonctionnel.

### 2.3 Références bibliographiques

1. Holley, R.W., et al., *Structure of a ribonucleic acid*. Science, 1965. **147**(3664): p. 1462-5.
2. Sharp, S.J., et al., *Structure and Transcription of Eukaryotic tRNA Gene*. Critical Reviews in Biochemistry, 1985. **19**(2): p. 107-144.
3. Yu, Q. and C.D. Morrow, *Identification of critical elements in the tRNA acceptor stem and T psi C loop necessary for human immunodeficiency virus type 1 infectivity*. Journal of Virology, 2001. **75**(10): p. 4902-4906.
4. Ibba, M. and D. Soll, *Aminoacyl-tRNA synthesis*. Annual Review of Biochemistry, 2000. **69**: p. 617-650.
5. Walsh, E.J., *Biochemistry (Garrett, Reginald H.; Grisham, Charles M.)*. Journal of Chemical Education, 1997. **74**(2): p. 189.
6. Petrov, A.S., et al., *Secondary Structures of rRNAs from All Three Domains of Life*. Plos One, 2014. **9**(2): p. 6.

7. Cech, T.R., *Structural biology - The ribosome is a ribozyme*. Science, 2000. **289**(5481): p. 878-879.
8. Eddy, S.R., *Non-coding RNA genes and the modern RNA world*. Nature Reviews Genetics, 2001. **2**(12): p. 919-929.
9. Argaman, L., et al., *Novel small RNA-encoding genes in the intergenic regions of Escherichia coli*. Current Biology, 2001. **11**(12): p. 941-950.
10. Storz, G., *An expanding universe of noncoding RNAs*. Science, 2002. **296**(5571): p. 1260-1263.
11. Morris, K.V. and J.S. Mattick, *The rise of regulatory RNA*. Nature Reviews Genetics, 2014. **15**(6): p. 423-437.
12. Wagner, E.G.H. and P. Romby, *Small RNAs in Bacteria and Archaea: Who They Are, What They Do, and How They Do It*, in *Advances in Genetics, Vol 90*, T. Friedmann, J.C. Dunlap, and S.F. Goodwin, Editors. 2015, Elsevier Academic Press Inc: San Diego. p. 133-208.
13. Bernstein, E., et al., *Role for a bidentate ribonuclease in the initiation step of RNA interference*. Nature, 2001. **409**(6818): p. 363-366.
14. Bartel, D.P., *MicroRNAs: Genomics, biogenesis, mechanism, and function*. Cell, 2004. **116**(2): p. 281-297.
15. Winkler, W.C., *Riboswitches and the role of noncoding RNAs in bacterial metabolic control*. Current Opinion in Chemical Biology, 2005. **9**(6): p. 594-602.
16. Allmang, C., et al., *Functions of the exosome in rRNA, snoRNA and snRNA synthesis*. Embo Journal, 1999. **18**(19): p. 5399-5410.
17. Bachellerie, J.P., J. Cavaille, and A. Huttenhofer, *The expanding snoRNA world*. Biochimie, 2002. **84**(8): p. 775-790.





## CHAPITRE 3. MODIFICATIONS POST-TRANSCRIPTIONNELLES

### 3.1 Introduction et généralités

Il existe une grande variété de modifications observable chez les ARN. On peut notamment citer les modifications des extrémités comme l'ajout d'un acide aminé en 3' des ARNt ou encore un changement de la séquence de nucléotide. Les ARNr, qui sont les ARN les plus abondants, présentent plusieurs types de modifications chimiques localisées dans des séquences et des contextes structurels différents. Plusieurs modifications ont aussi été découvertes dans les ARN moins abondants incluant les ARNm [1], les ARNsn [2] et aussi les ARNmi [3]. Dans cette thèse, nous nous sommes particulièrement focalisés sur les **modifications chimiques naturelles** des nucléotides. Bien que toutes les classes d'ARN de la cellule puissent contenir des nucléotides modifiés, la plus grande fraction est présente dans les ARNt. En effet, dans les ARNt, les nucléotides modifiés sont omniprésents et représentent environ 10 à 15 % de leur séquence totale et peut monter jusqu'à 25 % [4]. Ces modifications post-transcriptionnelles se traduisent par l'ajout ou la substitution d'un ou plusieurs atomes dans les nucléotides. Ces modifications chimiques covalentes peuvent survenir aussi bien sur la base nucléique que sur le ribose. Ainsi, hormis les quatre nucléotides canoniques, les ARNt sont couramment constitués d'une grande variété de nucléotides modifiés.

À ce jour, il y a plus de 160 modifications naturelles connues [5]. Parmi, ces modifications on peut citer l'ajout d'un groupement chimique plus ou moins complexe, la réduction/oxydation et même l'isomérisation de la base nucléique. La production d'une modification requiert, dans la très grande majorité des cas, plusieurs enzymes impliquant des réactions complexes. Chaque modifications nécessitent des enzymes spécifiques et présentent des voies de synthèse différentes qui ne sont pas encore toutes connues à ce jour [6, 7]. De plus, des nucléotides peuvent porter plusieurs modifications ou une modification complexe (hypermodification) qui ont nécessité une biosynthèse multi-étapes orchestrée dans un ordre précis pour les produire. La base de données Modomics (<https://iimcb.genesilico.pl/modomics/> [5]) répertorie les modifications d'ARN et contient diverses informations telles que le nom, la structure et également la voie de synthèse de ces modifications. C'est une base de données régulièrement vérifiée et constamment mise à jour.

Une nomenclature spéciale existe pour désigner un nucléotide modifié, selon la nature de la modification portée et aussi de sa localisation sur le nucléotide. Le Tableau 3 regroupe les abréviations des modifications (Remarque : Les abréviations des modifications situées en bas à droite du tableau sont couramment utilisées au même titre que les quatre bases canoniques A, G, C et U). Ainsi, pour écrire qu'un nucléotide est modifié, il faut préciser le nucléotide canonique avec la modification qu'il porte en utilisant l'abréviation correspondante. Si, dans l'écriture, l'abréviation est à gauche du nucléotide canonique, alors cela signifie que la modification se situe sur la base azotée. Dans le cas contraire, la modification est localisée sur l'hydroxyle en 2' du ribose. Un nombre en exposant après l'abréviation de la modification indique la position de celle-ci sur le nucléotide (numérotation

purine/pyrimidine) et un nombre en indice indique le nombre de cette modification présente sur cette position. Une base portant plusieurs modifications est désignée avec plusieurs abréviations [8] (voir l'exemple illustré sur la Figure 28). Dans une séquence d'ARN, il est possible d'écrire les nucléotides modifiés entre crochets.

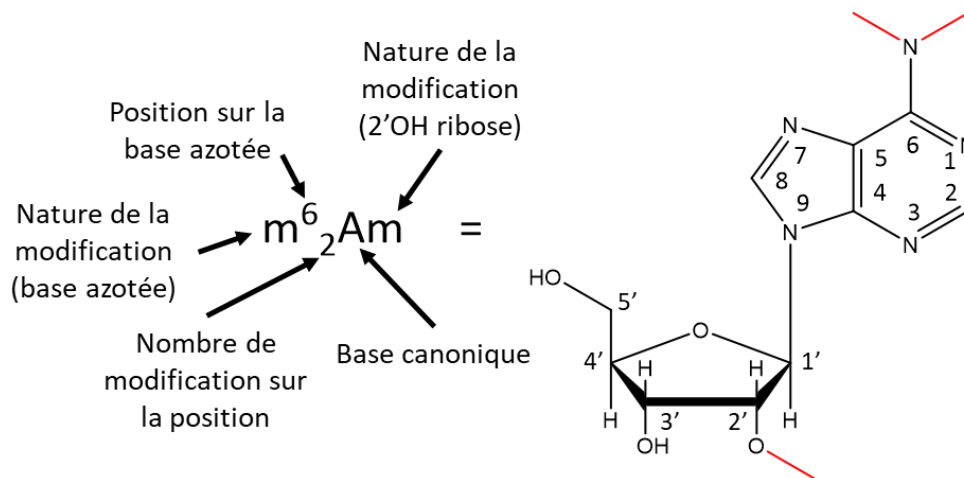


Figure 28 Exemple d'écriture simplifiée d'un nucléotide modifié ( $N^6, N^6, 2'-O$ -triméthyladenosine) et correspondance avec la structure chimique.

Abbr.	full description	Abbr.	full description
m	methyl	se	seleno
ac	acetyl	cmo	5-oxyacetic acid
t	threonylcarbamoyl	mcmo	5-oxyacetic acid methyl ester
rm	taurinomethyl	mo	methoxy
i	isopentenyl	ho	hydroxy
io	cis-hydroxyisopentenyl	f	formyl
s	thio	gal	galactosyl
ms	methylthio	glu	glutamyl
inm	isopentylaminomethyl	man	mannosyl
g	glycylcarbamoyl	r	2'-O-ribosyl
hn	hydroxynorvalylcarbamoyl	l	inosine
hm	hydroxymethyl	Ψ	pseudouridine
nm	aminomethyl	Q	queuosine
acp	(3-amino-3-carboxypropyl)	yW	wybutosine
cm	carboxymethyl	OHyW	hydroxywybutosine
chm	carboxyhydroxymethyl	C <sup>+</sup>	agmatidine
cmnm	carboxymethylaminomethyl	k <sup>2</sup> C	lysidine
ncm	carbamoylmethyl	D	dihydrouridine
mnm	methylaminomethyl	G <sup>+</sup>	archaeosine

Tableau 3 Liste non exhaustive des abréviations des modifications chimiques présentes dans les ARN Reproduit de [9].

Il existe également un code d'une lettre/caractère pour désigner un nucléotide modifié. Ce code est utilisé notamment par les logiciels informatiques. Par exemple, la modification  $m^6_2Am$  est symbolisée par  $\eta$ .

Les modifications d'ARN sont impliquées dans un grand nombre de processus biologiques. **La nature** mais aussi **la localisation des modifications** post-transcriptionnelles

sur la séquence de l'ARN sont d'une importance capitale puisqu'elles permettent d'assurer le bon fonctionnement de ces ARN. Ces modifications sont essentielles dans plusieurs aspects des ARNt car elles permettent de maintenir sa structure tertiaire, d'augmenter sa stabilité et de garantir le bon déroulement du processus de décodage comme cela est discuté dans la partie suivante [6, 10, 11].

### 3.2 Modifications naturelles des ARNt

Les nucléotides des ARNt peuvent subir une très grande variété de modifications chimiques. Ces modifications de nucléotides permettent d'apporter de nouvelles propriétés physico-chimiques à l'ARN essentielles à son bon fonctionnement. Le Tableau 4 résume les contributions physico-chimiques induites par des modifications de nucléotides dans les ARN.

1	Introduction de charges partielles dépendant de la protonation
2	Introduction d'une ou plusieurs charge positive ou/et négative permanentes
3	Altération ou changement de la conformation d'un nucléotide
4	Inhibition d'appariements non canoniques
5	Perturbation d'appariements canoniques
6	Amélioration des interactions d'appariements de bases
7	Réorganisation des molécules d'eau autour de la structure de l'ARN
8	Simplification de la coordination avec des ions métalliques
9	Limitation des conformations des liaisons phosphodiester et/ou du ribose
10	Production de nouvelles conformations et propriétés chimiques entre nucléotides modifiés

Tableau 4 Tableau regroupant les contributions physico-chimiques apportées par les modifications de nucléotides dans un ARN. Adapté de [12].

Dans les ARNt, les modifications sont distinguées en trois groupes suivant leurs fonctionnalités. Le premier groupe inclut les modifications situées sur ou adjacentes à l'anticodon, qui participent généralement au **processus de reconnaissance codon-anticodon**. Une seconde classe regroupe les modifications situées hors de la boucle de l'anticodon et qui contrôlent la **stabilité et le repliement de la structure** de l'ARNt. Enfin, le dernier groupe comprend les modifications qui **déterminent l'identité** de l'ARNt. Ces dernières jouent un rôle critique au regard de la reconnaissance entre l'ARNt et la synthétase correspondante car elles assurent le chargement de l'acide aminé spécifique sur l'extrémité de l'ARNt.

#### 3.2.1 Modifications sur la boucle anticodon

La position 34 ou wobble (premier nucléotide de l'anticodon) et la position 37 (nucléotide adjacent à l'anticodon) des ARNt sont fortement susceptibles de contenir des nucléotides modifiés [9]. Le nucléotide 37 est toujours une base de la famille des purines et est souvent hypermodifié. Les modifications les plus fréquentes sur cette position sont la  $N^1$ -méthylguanosine ( $m^1G$ ) et la wybutosine ( $\gamma W$ ) pour les guanosines ainsi que la  $N^6$ -thréonylcarbamoyladénosine ( $t^6A$ ) et la  $N^6$ -isopentenyladénosine ( $i^6A$ ) pour les adénosines [13, 14]. Ces modifications réduisent la flexibilité de la boucle anticodon permettant de rendre la structure plus rigide et donc d'améliorer la précision des interactions codon-anticodon. Plus important encore, des études structurales ont démontré l'implication de la base modifiée

$ms^2i^6A37$  à renforcer l'interaction faible codon-anticodon. En effet, le premier nucléotide du codon forme une paire de bases canonique avec le nucléotide 36 de l'anticodon. L'hypermodification de l'adénosine en position 37 apporte une interaction supplémentaire avec le premier nucléotide du codon via des liaisons de van der Waals (voir Figure 29). Cette interaction d'empilement d'une base supplémentaire sur le premier codon de l'ARNm permet alors d'augmenter la fidélité du décodage [15]. Cet apport de stabilisation via le nucléotide adjacent à l'anticodon a aussi été observé pour plusieurs autres ARNt avec les hypermodifications telles que  $\gamma W$ ,  $i^6A$  et  $t^6A$  [14, 16]. De ce fait, la plupart des nucléotides hypermodifiés appartiennent au premier groupe. Une autre étude a montré que l'utilisation d'un mutant d'ARNt avec un nucléotide non-modifié en position 37 induit une forte baisse du taux de croissance d'une chaîne polypeptidique *in vivo* lors du processus de traduction, suggérant ainsi le rôle critique de la modification en position 37 dans le décodage [17]. Dans certains cas, l'affaiblissement de l'appariement de bases par des modifications sert à discriminer l'interaction à des codons proches apparentés [4]. Ainsi, les modifications de la position 37 augmente la précision du décodage. Cependant pour les ARNt lisant les codons débutant par C ou G, ils ne présentent généralement pas de purine modifiée ou une modification simple telle qu'une méthylation ( $m^1G$  ou  $m^2G$ ). En effet, l'appariement C-G est énergétiquement plus stable comparée à la paire A-U et donc la présence d'une modification en 37 n'est pas nécessaire [18]. De cette façon, la modification du nucléotide de la position 37 permet donc d'améliorer le processus de décodage.

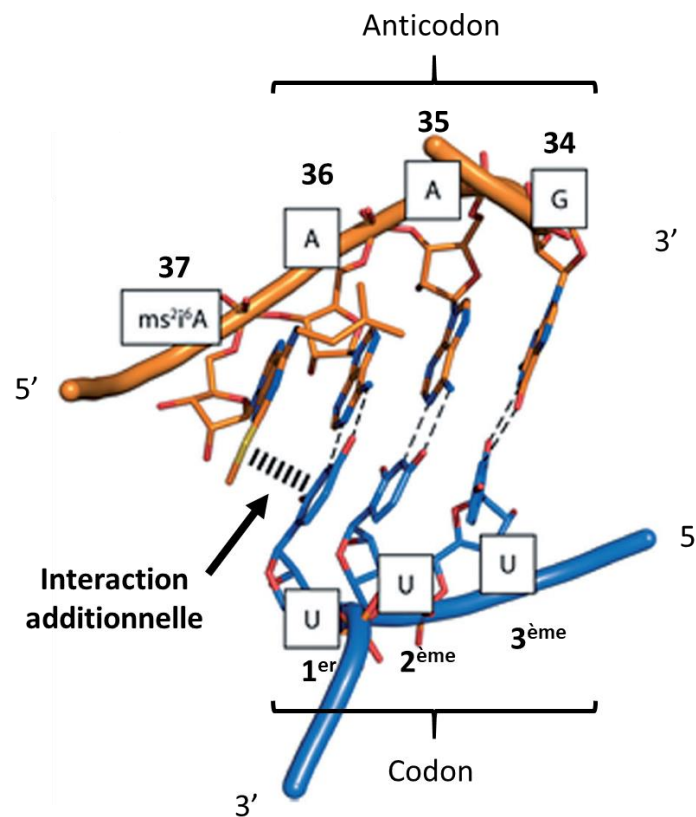


Figure 29 Structure 3D d'une interaction codon-anticodon montrant l'implication de l'hypermodification  $ms^2i^6A$  en position 37 dans le processus de décodage. En orange : anticodon de l'ARNt; en bleu : codon de l'ARNm. Adaptée de [9].

Les modifications de nucléotides situées sur la première position de l'anticodon (34) peuvent totalement changer les règles d'appariements. Pour rappel, l'adénine forme une paire avec l'uridine. Lorsqu'une adénine subit une désamination, elle se transforme en inosine (notée I). L'amine, en position C6 de l'adénine, est alors substituée par un carbonyle. L'amine qui participait à l'interaction entre A et U en tant que donneur de proton n'est donc plus disponible pour assurer cet appariement. Cependant, l'oxygène de l'inosine permet de faire de nouveaux appariements wobble thermodynamiquement stables en tant qu'accepteur de proton. Cette modification peut former des paires de bases telles que I=C, I=A mais aussi I=U [19]. Ainsi, les anticodons contenant une inosine en 1<sup>ère</sup> position sont capables de créer une paire avec trois différents codons [20]. Cette modification en position wobble de l'anticodon influence donc significativement le processus de traduction [21]. Cette modification se retrouve dans un grand nombre d'ARNt chez les bactéries et les eucaryotes. Il a été suggéré que la conversion de A34 en I34 améliore la fidélité et l'efficacité du décodage [22, 23]. Un autre exemple : le nucléotide G de la position 34 est souvent modifié par la queuosine (Q) et ses dérivés ou bien Gm, formant des paires de bases wobble avec le troisième nucléotide du codon. Les uridines de cette position sont couramment modifiées (mnm<sup>5</sup>U ou cmo<sup>5</sup>U) et forment également des paires de bases de type wobble. Ces nucléotides modifiés changent alors profondément le décodage de l'information génétique. Mais, en apportant une spécificité à l'anticodon des ARNt, les modifications participent à améliorer la fidélité du décodage en assurant une discrimination. Les différentes modifications observables sur la position wobble sont représentées sur la Figure 30. Les mécanismes d'appariements sont souvent complexes et semblent spécifiques aux ARNt. Des hypothèses d'appariements de type wobble impliquant des nucléotides modifiés ont été formulés afin de prédire un autre code génétique dépendant de ces modifications [13, 24, 25]. Outre le fait que les modifications de la positions 34 optimisent le processus de décodage, elles permettent également de maintenir la structure 3D de la boucle anticodon pour assurer un décodage plus efficace [24].

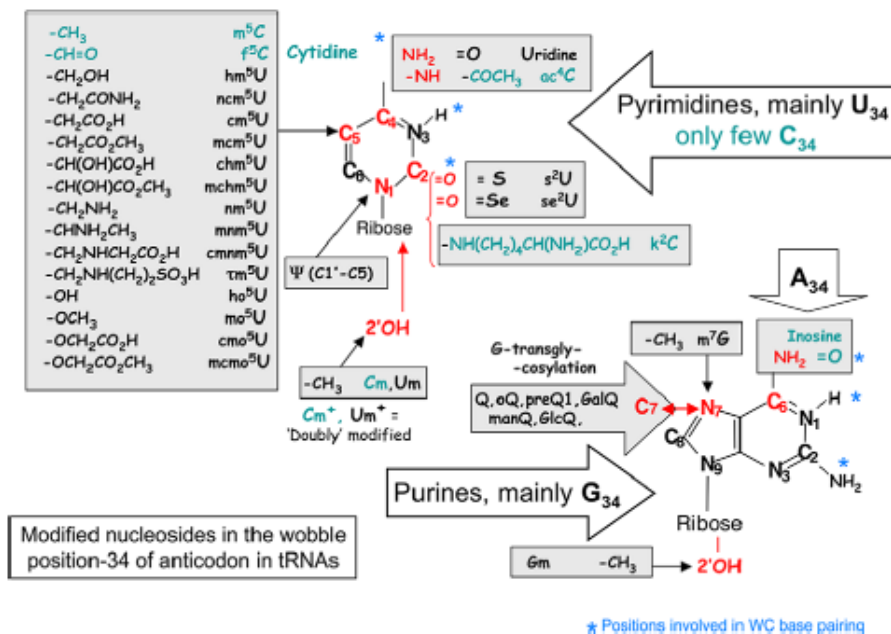


Figure 30 Résumé des modifications chimiques des nucléotides situées sur la position wobble (n°34) des ARN. Les rectangles listent les modifications, et leurs abréviations, pouvant être fixées sur les atomes d'une pyrimidine (à gauche) ou d'une purine (à droite). Issue de [25].

### 3.2.2 Modifications influençant la structure

Les modifications du deuxième groupe sont celles qui permettent le repliement correct de l'ARNt et aussi d'augmenter la stabilité de la structure tridimensionnelle [26]. Les modifications chimiques influencent la structure de plusieurs façons comme cela est décrit dans les paragraphes ci-dessous.

Les méthylations sont des modifications omniprésentes chez les ARNt car elles permettent de façonner la structure 3D de différentes manières [2]. Les nucléotides modifiés par une méthylation d'hydroxyle 2' sont connus pour favoriser la conformation C3'-endo du ribose. En effet, la présence de cette méthylation accroît la gêne stérique, renforçant ainsi cette conformation et rigidifiant la structure des duplex d'oligonucléotides. Par exemple, Kawai *et al* ont montré que cette modification favorise la conformation C3'-endo de l'UMP par 0,8 kcal/mol [27]. Le nucléotide de la position 4 des ARNt est souvent modifié par l'ajout d'un méthyle en 2'OH. Cette 2'-O-méthylation participe à rigidifier la structure double-hélice interne du bras accepteur de l'ARNt permettant d'améliorer le transfert de l'acide aminé [28]. D'autres nucléotides modifiés tels que la 5-méthyl-2-thiouridine ( $m^5s^2U$ ) ou la  $N^4$ -acétylcytidine ( $ac^4C$ ) stabilisent également cette conformation avec environ 1 kcal/mol [29]. De plus, lorsque la  $N^4$ -acétylcytidine est combinée avec la méthylation de l'hydroxyle 2' ( $ac^4Cm$ ), cela augmente la stabilité de la conformation C3'-endo par 1,53 kcal/mol [27]. Les nucléotides G de la position 26, formant la jonction entre la boucle D et la boucle anticodon, sont souvent modifiés par la modification  $N^2,N^2$ -diméthylguanosine ( $m^2_2G$ ). En effet, l'azote dont un des protons était impliqué dans la formation d'une paire de base est substitué par des méthyles. Ce nucléotide modifié ne peut donc plus former une paire de base canonique avec un autre nucléotide car les protons de l'azote N2 ne sont plus disponibles. Néanmoins, ce nucléotide forme tout de même une paire de base wobble avec une adénosine. Cette perturbation d'appariement Watson-Crick via une modification permet alors de donner la structure tertiaire correcte à l'ARNt [30]. Dans le cas des modifications 7-méthylguanosine ( $m^7G$ ) et 1-méthyladénosine ( $m^1A$ ), ce n'est pas l'hydrophobicité ou la gêne stérique du groupement méthyle qui contraint la structure de l'ARN mais plutôt la charge positive induite par celui-ci. En effet, ces méthylations forment des azotes quaternaires qui possèdent alors une charge positive permanente. Les charges impactent donc la structure de l'ARN via des contraintes électrostatiques [31].

La pseudouridine ( $\psi$ ), un isomère de l'uridine, est une modification post-transcriptionnelle très fréquente chez les ARNt. Considérée parfois comme le « 5<sup>ème</sup> nucléotide », elle peut se retrouver à différentes positions sur les ARNt [32]. En effet, cette modification est très souvent présente sur la boucle T permettant, en s'appariant avec d'autres nucléotides, de maintenir leur structure 3D en forme de L [28, 33-35]. Les  $\psi$  sont souvent localisées sur les motifs d'ARN tels que les jonctions des régions simple brin et hélicoïdales. L'isomérisation de U en  $\psi$  permet à la base de se lier à une molécule d'eau via l'azote N1-H. Cette capacité renforce l'appariement des bases ce qui permet de rigidifier le squelette phosphate-ribose adjacent et donc d'améliorer la stabilité de la structure 3D des ARN [36]. Dans le cas de l'ARNt<sup>Lys</sup>, il a été observé que lorsque cette modification est absente en position 39, la température de fusion diminue de 5°C. Après des expériences de RMN <sup>1</sup>H,

les auteurs expliquent que c'est un tautomère, favorisé par la modification de U en  $\psi$ , qui renforce l'appariement avec le nucléotide A31 et stabilise la structure de l'ARN [37]. La pseudouridylation est donc une modification permettant de stabiliser localement la structure des ARNt. Lorsque cette modification se trouve sur l'anticodon, elle renforce l'appariement codon-anticodon pour améliorer la précision du décodage [38].

Au même titre que  $\psi$ , le nucléotide D est une modification post-transcriptionnelle commune à une grande majorité des ARNt aussi bien chez les bactéries que chez les eucaryotes [39]. La dihydrouridine est formée par réduction de la double liaison C5=C6 de l'uridine. A l'inverse des modifications stabilisant la structure des ARN, les analyses conformationnelles d'une dihydrouridine indiquent que cette modification déstabilise la conformation C3'-endo des sucres. Une étude montre que la substitution du carbone C5 d'une uridine influence l'interaction de liaison entre le doublet non-liant de l'atome O4' et l'orbitale  $\pi^*$  de la double liaison C5=C6 et change les préférences de conformation du ribose [40]. Lorsque la double liaison est hydrogénée, cette interaction est alors défavorisée et modifie la préférence conformationnelle du sucre. L'effet de D favorise alors la conformation C2'-endo du ribose avec une enthalpie estimée à quelques kcal/mol [41]. La production d'une dihydrouridine est donc une stratégie biologique permettant le changement conformationnelle du ribose ajoutant ainsi une flexibilité dans certaines régions de la structure des ARNt. Cette flexibilité s'avère nécessaire pour faciliter l'interaction entre la boucle D avec la boucle T et donner la forme de L aux ARNt, d'où la présence de cette modification sur la boucle D.

Les dihydrouridines -situées sur la boucle D-, la thymine ( $m^5U$  ou T) et la pseudouridine ( $\psi$ ) -situées sur la boucle T- sont des modifications post-transcriptionnelles qui sont presque systématiquement présentes dans les ARNt. Comme cela a déjà été évoqué, certains nucléotides des boucles D et T forment des paires de bases entre elles et replient la structure des ARNt. Ces interactions sont permises car les nucléotides D permettent de favoriser la conformation 2'endo des riboses. Il faut noter que plusieurs paires de bases ne sont pas canoniques entre les boucles. Par exemple dans le cas de l'ARNt<sup>Phe</sup>, la  $\psi$  de la position 55 (située sur la boucle T) forme une seule liaison hydrogène avec le nucléotide G18 (situé sur la boucle D). Le nucléotide T54 de la boucle T forment une paire de bases wobble avec le nucléotide  $m^1A$  situé sur la même boucle. Un autre exemple, dans cet ARNt, montre que les nucléotides A9, U12 et A23 sont en interactions, bien qu'ils ne soient pas modifiés : A23 et U12 forment une paire de base canonique et A9 s'apparie de façon non-canonique à A23 via deux ponts hydrogènes. Les exemples d'appariements des nucléotides sont illustrés Figure 31. Ces modifications permettent donc de régir l'interaction entre les deux boucles et assurent l'empilement correct des nucléotides entre les boucles. Ces nombreuses interactions sont toutes essentielles afin de donner la forme de L typique des ARNt nécessaire dans le processus de traduction [42].

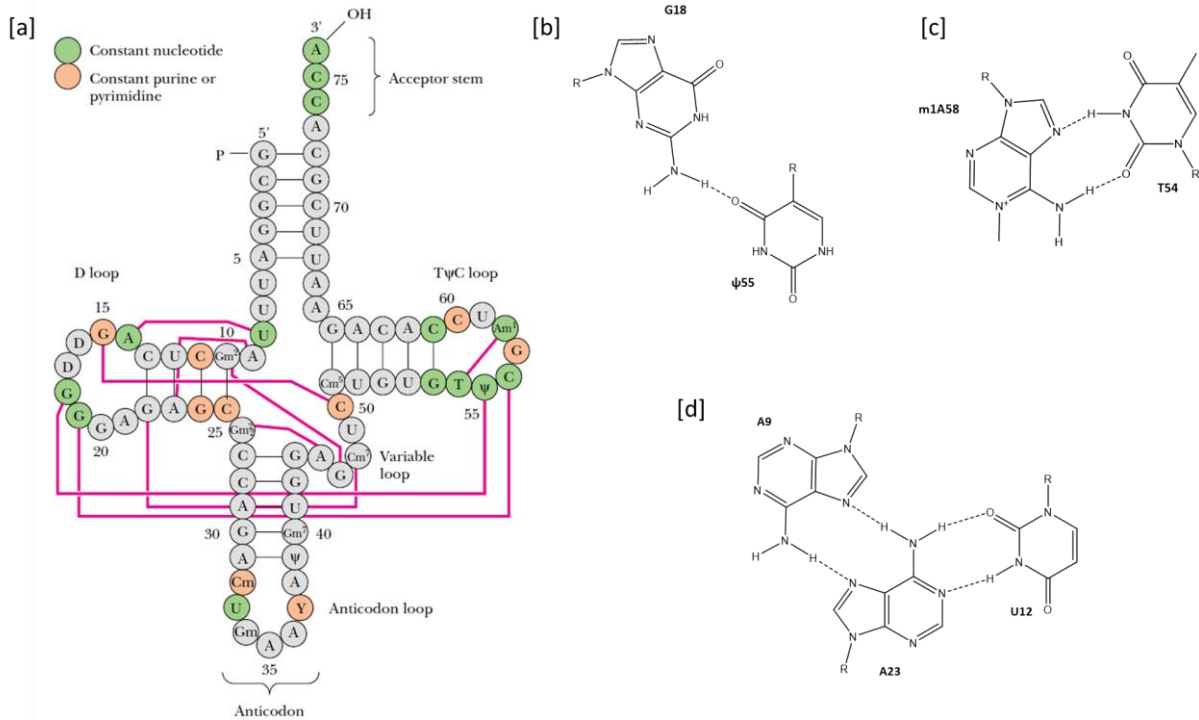


Figure 31[a] Structure secondaire de l'ARNt<sup>Phe</sup> montrant les interactions tertiaires. Reproduit de [43]. Et exemples d'appariements non-canoniques induisant la forme de L inversé avec appariement [b] entre G18 et ψ55, [c] entre m1A58 et T54 et [d] entre A9 A23 et U12.

Afin de démontrer le rôle de ces modifications sur la stabilité des molécules d'ARN, plusieurs expériences ont été menées notamment en étudiant la stabilité thermique des ARN. Ainsi, des études ont montré que la température de fusion d'un ARNt non-modifié est significativement inférieure à celle d'un ARNt normalement modifié. Dans certains cas, cette température de fusion peut diminuer jusqu'à 20°C [11, 44]. Une autre étude a démontré que les ARN sous-modifiés se dégradent beaucoup plus rapidement que les ARN naturellement modifiés [45].

Les modifications chimiques des nucléotides de ce groupe permettent alors de former la structure des ARNt de différentes manières. Certaines modifications modifient les appariements entre les nucléotides et donc influencent le repliement des ARN. D'autres favorisent une conformation du ribose et ainsi rigidifient les structures. Enfin, certaines structures sont impactées par des modifications changeant la charge électronique des nucléotides.

### 3.2.3 Modifications impliquées dans l'aminoacylation des ARNt

Le dernier groupe réunit les modifications qui donnent une identité spécifique à l'ARNt. Après synthèse de l'ARNt, celui-ci n'est pas encore tout à fait fonctionnel car il ne porte pas l'acide aminé protéinogène sur la terminaison 3'. Pour que l'acide aminé correspondant à l'ARNt soit fixé, l'aminoacyl-ARNt synthétase doit reconnaître spécifiquement cet ARNt. La reconnaissance entre l'enzyme et l'ARNt s'effectue de façons variées et complexes suivant l'ARNt et l'acide aminé à fixer. Les modifications chimiques des ARNt sont bien souvent impliqués dans la reconnaissance spécifique de l'aminoacyl-ARNt synthétase [46]. En



l'absence des modifications « identitaires », l'estérification de l'acide aminé sur l'ARNt est souvent moins efficace. Des modifications chimiques simples, telles qu'une thiolation d'uridine ( $s^2U$ ), sont suffisantes pour assurer le rôle critique de l'acylation des ARNt. De plus, il a même été démontré qu'un acide aminé ne correspondant pas à l'ARNt est fixé lorsqu'une modification de ce type est absente. En effet, l'arginyl-ARNt synthétase fixe par erreur une arginine sur l'ARNt<sup>Asp</sup> s'il manque une méthylation sur une guanine ( $m^1G$ ) [47]. Ainsi, ce groupe de modification permet aussi d'éviter que de mauvais acides aminés soient fixés sur les ARNt.

### 3.2.4 Autres modifications remarquables

La lysidine ( $k^2C$ ) est une modification intéressante puisqu'elle joue deux rôles dans le cas de l'ARNt<sup>Ile</sup><sub>CAU</sub>. L'anticodon de cet ARNt est constitué des nucléotides CAU qui décode, d'après les règles d'appariement Watson-Crick, pour le codon méthionine AUG. Néanmoins, la cytidine en position 34 de cet ARN est modifiée par l'ajout d'une lysine sur la base azotée. Cette cytidine modifiée s'apparie alors de façon wobble avec une adénosine. Ainsi, l'anticodon  $[k^2C]AU$  s'apparie avec le codon isoleucine AUA. En plus de son rôle dans le décodage, il a été démontré que cette modification est aussi utile pour la reconnaissance de l'ARN par l'acyl-ARNt synthétase fournissant l'isoleucine. En effet, lorsque cette modification est absente, l'acylation de cet ARNt est beaucoup plus lente [48, 49]. Le double rôle d'une modification située en position 34 a aussi été observé pour l'agmatidine ( $C^+$ ), une cytidine modifiée par l'ajout d'une arginine [50].

Par conséquent, les nucléotides non-canoniques, leurs natures et leurs localisations, sont primordiaux sous plusieurs aspects pour garantir le bon fonctionnement des ARNt. En effet, elles donnent une identité unique à l'ARNt afin que l'acyl-ARNt synthétase fixe l'acide aminé correspondant et elles participent à la stabilité de la structure tridimensionnelle de l'ARNt. Les nucléotides modifiés de la position 34 apportent une spécificité additionnelle afin d'améliorer la fidélité du décodage. Les modifications de la position 37 permettent d'augmenter et d'affiner la précision du processus de décodage en apportant des interactions additionnelles entre le codon et l'anticodon. De ce fait, en plus des trois nucléotides de l'anticodon, les ARNt décodent les ARNm par le biais d'un « codon supplémentaire », ce qui ajoute une information additionnelle [51]. La complexité, la diversité et la richesse des modifications de nucléotide permettent également d'augmenter l'information contenue dans les différents ARN [9, 52]. La Figure 32 offre un panorama des modifications connues, et leurs positions, dans les ARNt de la levure. Récemment, il a été observé qu'il existe une dynamique dans les modifications, ce qui complexifie encore les mécanismes biologiques [53].

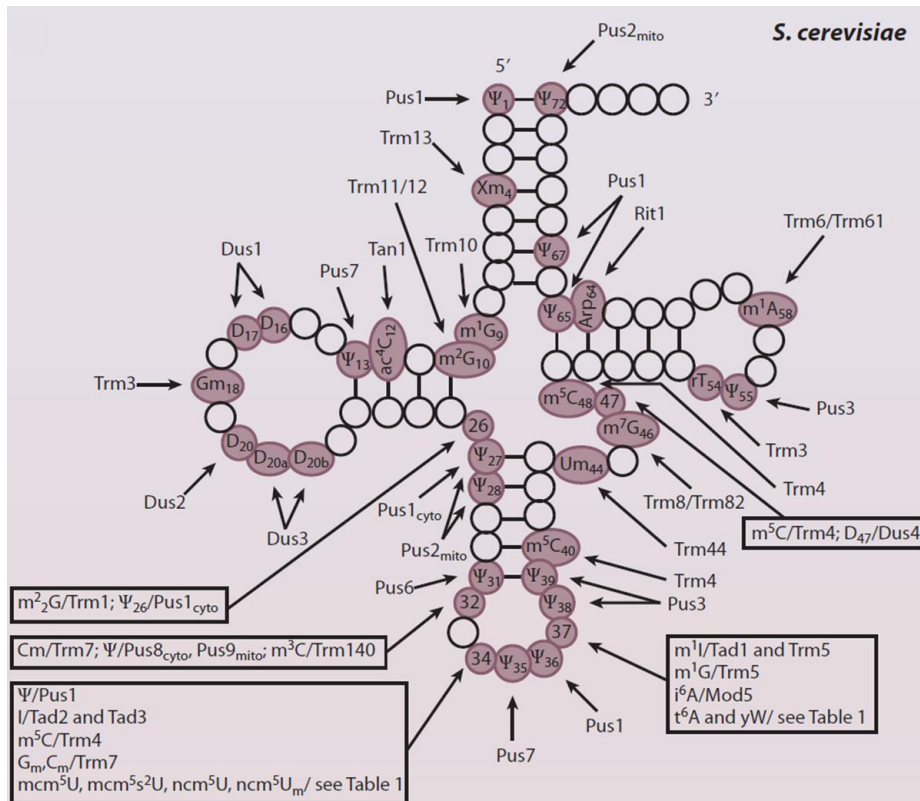


Figure 32 Structure secondaire d'un ARNt (*S. cerevisiae*) présentant les positions susceptibles de porter une modification et les enzymes responsables de leur formation (non-exhaustive). Issue de [6].

### 3.3 Modifications d'ARN et maladies

Les modifications post-transcriptionnelles ont longtemps été considérées comme des caractéristiques intrinsèques des ARN permettant le réglage fin de leur structure pour assurer ses fonctions dans la cellule. Toutefois, ces dernières années, plusieurs études ont mis en évidence que **certaines modifications** des ARN sont **impliquées dans des maladies** comme le diabète de type II, des maladies mitochondriales, l'obésité ou encore le cancer [54-57]. Plusieurs articles mettent en lumière l'implication de nucléotides modifiés dans les cancers aussi bien sur les ARNt que sur d'autres types d'ARN [58]. Dans les années 90, Huang *et al* ont observé une baisse de niveau de queuosine dans des tissus tumoraux. Cette modification est utilisée en tant que biomarqueur pour le cancer des poumons [59]. Des études plus récentes, ont montré que la modification m<sup>6</sup>A est fortement impliquée dans une grande variété de cancers [60-62]. Puisque des modifications sont impliquées dans des maladies, elles peuvent alors servir à les diagnostiquer. Le niveau élevé de nucléosides modifiés dans les échantillons d'urines sont des biomarqueurs potentiels du cancer. Par conséquent, de nos jours, ces modifications sont des biomarqueurs puissants puisqu'une simple analyse quantitative des nucléosides modifiés présents dans les urines permet le diagnostic précis de différent type de cancer [63, 64]. Ces analyses permettent un diagnostic précoce des cancers et également un suivi de la thérapie. Cependant l'origine et le rôle de ces modifications dans ces maladies ne sont, à ce jour, pas totalement élucidés. Néanmoins, ces récentes découvertes ont démontré que les modifications participent de différentes façons à la régulation de l'expression génétique et ont ainsi inspiré et accéléré l'intérêt de la recherche de médicaments à base d'ARN pour traiter les cancers ou autres maladies [65].

Des chercheurs, en s'inspirant des ARNm susceptibles d'entraver un ARNm cible, ont proposé d'utiliser ces petits oligomères comme nouvel outil biologique thérapeutique. En effet, l'idée a été d'empêcher l'expression du message nocif d'un ARNm par le biais d'ARN synthétiques. À la fin des années 70, les premiers oligonucléotides antisens (ASO), une nouvelle classe de médicament, sont synthétisés en laboratoire et utilisés dans des buts thérapeutiques [66]. Le principe d'un traitement par la stratégie antisens est relativement simple : l'utilisation d'un oligodésoxyribonucléotide capable de s'apparier sur un ARNm cible pour inhiber son expression et donc induire un blocage dans le transfert d'informations génétique de l'ADN vers la protéine [67]. Le duplex ARNm-ASO est reconnu par une enzyme qui dégrade l'ARNm, bloquant alors le processus de traduction. Cependant, les acides nucléiques non-modifiés sont rapidement dégradés par des enzymes, les nucléases, naturellement présentes dans les matrices biologiques et ont donc une utilité limitée en tant qu'agent thérapeutique.

Ainsi, les ASO sont des petits acides nucléiques généralement constitués d'une vingtaine de nucléotides et sont volontairement modifiés pour augmenter leur résistance à la dégradation. Les oligonucléotides thérapeutiques peuvent être modifiés à plusieurs niveaux suivant les propriétés pharmacocinétiques voulues. Il existe une large variété de modifications non-naturelles et naturelles pouvant être incorporées notamment sur le phosphate pour limiter l'hydrolyse. Par exemple, il est possible de substituer les phosphodiester par des phosphorothioates où un atome d'oxygène est remplacé par un atome de soufre. Ce soufre perturbe l'action de dégradation des nucléases et permet donc d'augmenter significativement la durée de demi-vie de l'acide nucléique sans perturber la reconnaissance de l'enzyme de dégradation du duplex. La base et le ribose peuvent aussi être modifiés afin de « régler » les propriétés physico-chimiques et pharmaceutiques de ces médicaments dans le but d'améliorer leur action thérapeutique [68]. Les ASO ont montré leur efficacité pour traiter par exemple l'atrophie musculaire spinale ou encore certains cancers [69, 70]. Le premier ASO commercialisé était le Fomiversen qui a été approuvé en 1998 par la Food and Drug Administration et en 1999 par l'European Medicines Agency. Même si seulement 9 oligonucléotides thérapeutiques ont été approuvés, une accélération a été observée entre 2016 et 2019 avec 6 nouvelles approbations [71]. En effet, la conception de ce genre de drogue se heurte à de nombreux obstacles dont notamment leur stabilité et leur spécificité. À ce jour, il y a plus de 900 oligonucléotides thérapeutiques en développement, montrant le fort potentiel que peut apporter cette nouvelle thérapie (GlobalData 15/05/2020).

Dans les années 80, des drogues basées sur des nucléosides modifiés pour traiter le virus de l'immunodéficience humaine (VIH) ont également été développées. Ces nucléosides agissent en inhibant l'action de la transcriptase employée par le VIH pour réaliser un brin d'ADN à partir de son ARN. Les combinaisons de ces nucléosides avec d'autres molécules permettent de limiter le développement du VIH et ainsi d'augmenter significativement l'espérance de vie des personnes atteintes de ce virus [72, 73]. Ce sont des traitements qui sont encore très utilisés à ce jour. En effet, sept inhibiteurs nucléosidiques sont commercialisés et plusieurs sont en cours d'essais cliniques. L'amélioration de la synthèse des ARN, des nucléosides et des bases modifiées devraient permettre l'apparition de nouveaux médicaments [74, 75].

Ainsi, bien que les modifications post-transcriptionnelles soient essentielles au bon fonctionnement des ARN, elles peuvent également déréguler les cellules et participer au développement de maladies lorsqu'elles sont mal exprimées. Des outils basés sur la détection de ces modifications ont été développés afin de diagnostiquer les maladies liées à ces modifications. De plus, des médicaments basés sur des ARN modifiés ont vu le jour afin de traiter des maladies [76].

À ce jour, les rôles de certaines de ces modifications restent inconnus. La compréhension de l'influence des modifications d'ARN est donc cruciale dans de nombreux aspects. Ainsi, depuis la dernière décennie, il y a un intérêt croissant pour l'analyse des modifications d'ARN. L'analyse de ces modifications en réponse au stress, aux changements environnementaux et aux infections causées par des agents pathogènes a alimenté de nouveaux domaines de recherche [77]. La nécessité de découvrir les rôles de ces modifications dans les ARN passe d'abord par leur caractérisation. Il y a peu d'informations sur les modifications de l'ARN et pour accroître les connaissances de la communauté, il est important de développer des méthodes sensibles pour les explorer en profondeur, mais cela reste un défi analytique. Le chapitre suivant décrit les outils analytiques disponibles avec les avantages et inconvénients de chacune de ces méthodes.

### 3.4 Références bibliographiques

1. Cowling, Victoria H., *Regulation of mRNA cap methylation*. Biochemical Journal, 2010. **425**(2): p. 295.
2. Motorin, Y. and M. Helm, *RNA nucleotide methylation*. Wiley Interdisciplinary Reviews-Rna, 2011. **2**(5): p. 611-631.
3. Yu, B., et al., *Methylation as a crucial step in plant microRNA biogenesis*. Science, 2005. **307**(5711): p. 932-935.
4. Duechler, M., et al., *Nucleoside modifications in the regulation of gene expression: focus on tRNA*. Cellular and Molecular Life Sciences, 2016. **73**(16): p. 3075-3095.
5. Boccaletto, P., et al., *MODOMICS: a database of RNA modification pathways. 2017 update*. Nucleic Acids Research, 2018. **46**(D1): p. D303-D307.
6. El Yacoubi, B., M. Bailly, and V. de Crecy-Lagard, *Biosynthesis and Function of Posttranscriptional Modifications of Transfer RNAs*. Annual Review of Genetics, Vol 46, 2012. **46**: p. 69-95.
7. Wang, X.X., et al., *Enzymatic Synthesis of Modified Oligonucleotides by PEAR Using Phusion and KOD DNA Polymerases*. Nucleic Acid Therapeutics, 2015. **25**(1): p. 27-34.
8. Cantara, W.A., et al., *The RNA modification database, RNAMDB: 2011 update*. Nucleic Acids Research, 2011. **39**: p. D195-D201.
9. Carell, T., et al., *Structure and Function of Noncanonical Nucleobases*. Angewandte Chemie International Edition, 2012. **51**(29): p. 7110-7131.
10. Bjork, G.R., et al., *Transfer-Rna Modification*. Annual Review of Biochemistry, 1987. **56**: p. 263-287.
11. Kowalak, J.A., et al., *The Role of Posttranscriptional Modification in Stabilization of Transfer-Rna from Hyperthermophiles*. Biochemistry, 1994. **33**(25): p. 7869-7876.
12. Agris, P.F., *The importance of being modified: Roles of modified nucleosides and Mg<sup>2+</sup> in RNA structure and function*. Progress in Nucleic Acid Research and Molecular Biology, Vol 53, 1996. **53**: p. 79-129.

13. Agris, P.F., F.A.P. Vendeix, and W.D. Graham, *tRNA's wobble decoding of the genome: 40 years of modification*. Journal of Molecular Biology, 2007. **366**(1): p. 1-13.
14. Thiaville, P.C., et al., *Essentiality of threonylcarbamoyl adenosine (t(6)A), a universal tRNA modification, in bacteria*. Molecular Microbiology, 2015. **98**(6): p. 1199-1221.
15. Jenner, L.B., et al., *Structural aspects of messenger RNA reading frame maintenance by the ribosome*. Nature Structural & Molecular Biology, 2010. **17**(5): p. 555-U48.
16. Murphy, F.V., et al., *The role of modifications in codon discrimination by tRNA(Lys) UUU*. Nature Structural & Molecular Biology, 2004. **11**(12): p. 1186-1191.
17. Ericson, J.U. and G.R. Bjork, *Pleiotropic Effects Induced by Modification Deficiency Next to the Anticodon of Transfer-Rna from Salmonella-Typhimurium-Lt2*. Journal of Bacteriology, 1986. **166**(3): p. 1013-1021.
18. Cieplak, P. and P.A. Kollman, *Calculation of the Free-Energy of Association of Nucleic-Acid Bases in Vacuo and Water Solution*. Journal of the American Chemical Society, 1988. **110**(12): p. 3734-3739.
19. Torres, A.G., E. Batlle, and L.R. de Pouplana, *Role of tRNA modifications in human diseases*. Trends in Molecular Medicine, 2014. **20**(6): p. 306-314.
20. Novoa, E.M., et al., *A Role for tRNA Modifications in Genome Structure and Codon Usage*. Cell, 2012. **149**(1): p. 202-213.
21. Lustig, F., et al., *Codon Discrimination and Anticodon Structural Context*. Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(18): p. 6873-6877.
22. Schaub, M. and W. Keller, *RNA editing by adenosine deaminases generates RNA and protein diversity*. Biochimie, 2002. **84**(8): p. 791-803.
23. Rafels-Ybern, A., et al., *The Expansion of Inosine at the Wobble Position of tRNAs, and Its Role in the Evolution of Proteomes*. Molecular Biology and Evolution, 2019. **36**(4): p. 650-662.
24. Weixlbaumer, A., et al., *Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines*. Nature Structural & Molecular Biology, 2007. **14**(6): p. 498-502.
25. Grosjean, H., V. de Crecy-Lagard, and C. Marck, *Deciphering synonymous codons in the three domains of life: Co-evolution with specific tRNA modification enzymes*. Febs Letters, 2010. **584**(2): p. 252-264.
26. Sampson, J.R., et al., *Role of the Tertiary Nucleotides in the Interaction of Yeast Phenylalanine Transfer-Rna with Its Cognate Synthetase*. Biochemistry, 1990. **29**(10): p. 2523-2532.
27. Kawai, G., et al., *Conformational Rigidity of Specific Pyrimidine Residues in Transfer-Rna Arises from Posttranscriptional Modifications That Enhance Steric Interaction between the Base and the 2'-Hydroxyl Group*. Biochemistry, 1992. **31**(4): p. 1040-1046.
28. Wilkinson, M.L., et al., *The 2'-O-methyltransferase responsible for modification of yeast tRNA at position 4*. Rna-a Publication of the Rna Society, 2007. **13**(3): p. 404-413.
29. Watanabe, K., P. Davanloo, and M. Sprinzl, *Effect of the modification of the U-54 residue on the structure and function of transfer-RNA*. Hoppe-Seylers Zeitschrift Fur Physiologische Chemie, 1979. **360**(3): p. 396-396.
30. Steinberg, S. and R. Cedergren, *A correlation between N-2-dimethylguanosine presence and alternate tRNA conformers*. Rna-a Publication of the Rna Society, 1995. **1**(9): p. 886-891.

31. Agris, P.F., H. Sierzputowskagracz, and C. Smith, *Transfer-Rna Contains Sites of Localized Positive Charge - Carbon Nmr-Studies of [C-13]Methyl-Enriched Escherichia-Coli and Yeast Transfer Rnaphe*. *Biochemistry*, 1986. **25**(18): p. 5126-5131.
32. Li, X., S. Ma, and C. Yi, *Pseudouridine: the fifth RNA nucleotide with renewed interests*. *Current Opinion in Chemical Biology*, 2016. **33**: p. 108-116.
33. Arnez, J.G. and T.A. Steitz, *Crystal-Structure of Unmodified Trna(Gln) Complexed with Glutaminyl-Transfer-Rna Synthetase and Atp Suggests a Possible Role for Pseudo-Uridines in Stabilization of Rna Structure*. *Biochemistry*, 1994. **33**(24): p. 7560-7567.
34. Kierzek, E., et al., *The contribution of pseudouridine to stabilities and structure of RNAs*. *Nucleic Acids Research*, 2014. **42**(5): p. 3492-3501.
35. Westhof, E., *Pseudouridines or how to draw on weak energy differences*. *Biochemical and Biophysical Research Communications*, 2019. **520**(4): p. 702-704.
36. Davis, D.R., *Stabilization of RNA stacking by pseudouridine*. *Nucleic Acids Research*, 1995. **23**(24): p. 5020-5026.
37. Durant, P.C. and D.R. Davis, *Stabilization of the anticodon stem-loop of tRNA(Lys,3) by an A(+)-C base-pair and by pseudouridine*. *Journal of Molecular Biology*, 1999. **285**(1): p. 115-131.
38. Davis, D.R., C.A. Veltri, and L. Nielsen, *An RNA model system for investigation of pseudouridine stabilization of the codon-anticodon interaction in tRNA(Lys), tRNA(His) and tRNA(Tyr)*. *Journal of Biomolecular Structure & Dynamics*, 1998. **15**(6): p. 1121-1132.
39. Steinberg, S., A. Misch, and M. Sprinzl, *Compilation of Transfer-Rna Sequences and Sequences of Transfer-Rna Genes*. *Nucleic Acids Research*, 1993. **21**(13): p. 3011-3015.
40. Egert, E., et al., *Influence of Substituents at the 5 Position on the Structure of Uridine*. *Journal of the American Chemical Society*, 1980. **102**(11): p. 3707-3713.
41. Dalluge, J.J., et al., *Posttranscriptional modification of tRNA in psychrophilic bacteria*. *Journal of Bacteriology*, 1997. **179**(6): p. 1918-1923.
42. Giege, R., et al., *Structure of transfer RNAs: similarity and variability*. *Wiley Interdisciplinary Reviews-Rna*, 2012. **3**(1): p. 37-61.
43. Walsh, E.J., *Biochemistry (Garrett, Reginald H.; Grisham, Charles M.)*. *Journal of Chemical Education*, 1997. **74**(2): p. 189.
44. Perret, V., et al., *Conformation in Solution of Yeast Transfer Rnaasp Transcripts Deprived of Modified Nucleotides*. *Biochimie*, 1990. **72**(10): p. 735-744.
45. Alexandrov, A., et al., *Rapid tRNA decay can result from lack of nonessential modifications*. *Molecular Cell*, 2006. **21**(1): p. 87-96.
46. Phizicky, E.M. and A.K. Hopper, *tRNA biology charges to the front*. *Genes & Development*, 2010. **24**(17): p. 1832-1860.
47. Putz, J., et al., *Single Methyl-Group Prevents the Mischarging of a Transfer-Rna*. *Nature Structural Biology*, 1994. **1**(9): p. 580-582.
48. Muramatsu, T., et al., *Codon and Amino-Acid Specificities of a Transfer-Rna Are Both Converted by a Single Post-Transcriptional Modification*. *Nature*, 1988. **336**(6195): p. 179-181.
49. Grosjean, H. and G.R. Bjork, *Enzymatic conversion of cytidine to lysidine in anticodon of bacterial tRNA(Ile) - an alternative way of RNA editing*. *Trends in Biochemical Sciences*, 2004. **29**(4): p. 165-168.
50. Mandal, D., et al., *Agmatidine, a modified cytidine in the anticodon of archaeal tRNA(Ile), base pairs with adenosine but not with guanosine*. *Proceedings of the*

- National Academy of Sciences of the United States of America, 2010. **107**(7): p. 2872-2877.
51. Gustilo, E.M., A.P.F. Franck, and P.F. Agris, *tRNA's modifications bring order to gene expression*. Current Opinion in Microbiology, 2008. **11**(2): p. 134-140.
  52. Song, J.H. and C.Q. Yi, *Chemical Modifications to RNA: A New Layer of Gene Expression Regulation*. Acs Chemical Biology, 2017. **12**(2): p. 316-325.
  53. Chen, K., B.X.S.M. Zhao, and C.A. He, *Nucleic Acid Modifications in Regulation of Gene Expression*. Cell Chemical Biology, 2016. **23**(1): p. 74-85.
  54. Arragain, S., et al., *Identification of Eukaryotic and Prokaryotic Methylthiotransferase for Biosynthesis of 2-Methylthio-N-6-threonylcarbamoyladenine in tRNA*. Journal of Biological Chemistry, 2010. **285**(37): p. 28425-28433.
  55. Suzuki, T., et al., *Taurine as a constituent of mitochondrial tRNAs: new insights into the functions of taurine and human mitochondrial diseases*. Embo Journal, 2002. **21**(23): p. 6581-6589.
  56. Kirino, Y., et al., *Codon-specific translational defect caused by a wobble modification deficiency in mutant tRNA from a human mitochondrial disease*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(42): p. 15070-15075.
  57. Gerken, T., et al., *The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase*. Science, 2007. **318**(5855): p. 1469-1472.
  58. Haruehanroengra, P., et al., *RNA modifications and cancer*. Rna Biology, 2020.
  59. Huang, B.S., R.T. Wu, and K.Y. Chien, *Relationship of the queuine content of transfer-ribonucleic-acids to histopathological grading and survival in human lung-cancer*. Cancer Research, 1992. **52**(17): p. 4696-4700.
  60. Liebich, H.M., et al., *Capillary electrophoresis of urinary normal and modified nucleosides of cancer patients*. Journal of Chromatography A, 1998. **793**(2): p. 341-347.
  61. Deng, X., et al., *RNA N6-methyladenosine modification in cancers: current status and perspectives*. Cell Research, 2018. **28**(5): p. 507-517.
  62. Zhang, Y., et al., *m(6)A modification-mediated CBX8 induction regulates stemness and chemosensitivity of colon cancer via upregulation of LGR5*. Molecular Cancer, 2019. **18**(1).
  63. Liebich, H.M., et al., *Quantitation of urinary nucleosides by high-performance liquid chromatography*. Journal of Chromatography A, 1997. **763**(1-2): p. 193-197.
  64. Jiang, Y.Q. and Y.F. Ma, *A Fast Capillary Electrophoresis Method for Separation and Quantification of Modified Nucleosides in Urinary Samples*. Analytical Chemistry, 2009. **81**(15): p. 6474-6480.
  65. Jonkhout, N., et al., *The RNA modification landscape in human disease*. Rna, 2017. **23**(12): p. 1754-1769.
  66. Paterson, B.M., B.E. Roberts, and E.L. Kuff, *Structural Gene Identification and Mapping by DNA-Messenger-Rna Hybrid-Arrested Cell-Free Translation*. Proceedings of the National Academy of Sciences of the United States of America, 1977. **74**(10): p. 4370-4374.
  67. Dias, N. and C.A. Stein, *Antisense oligonucleotides: Basic concepts and mechanisms*. Molecular Cancer Therapeutics, 2002. **1**(5): p. 347-355.
  68. Bennett, C.F., *Therapeutic Antisense Oligonucleotides Are Coming of Age*. Annual Review of Medicine, Vol 70, 2019. **70**: p. 307-321.

69. Passini, M.A., et al., *Antisense Oligonucleotides Delivered to the Mouse CNS Ameliorate Symptoms of Severe Spinal Muscular Atrophy*. *Science Translational Medicine*, 2011. **3**(72).
70. Sahin, U., et al., *Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer*. *Nature*, 2017. **547**(7662): p. 222-+.
71. Stein, C.A. and D. Castanotto, *FDA-Approved Oligonucleotide Therapies in 2017*. *Molecular Therapy*, 2017. **25**(5): p. 1069-1075.
72. 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*. *New England Journal of Medicine*, 1999. **341**(25): p. 1865-1873.
73. Fletcher, C.V., et al., *Concentration-controlled zidovudine therapy*. *Clinical Pharmacology & Therapeutics*, 1998. **64**(3): p. 331-338.
74. Ubiali, D., et al., *Synthesis of 2'-deoxynucleosides by transglycosylation with new immobilized and stabilized uridine phosphorylase and purine nucleoside phosphorylase*. *Advanced Synthesis & Catalysis*, 2004. **346**(11): p. 1361-1366.
75. Thompson, M.A., et al., *Antiretroviral Treatment of Adult HIV Infection 2010 Recommendations of the International AIDS Society-USA Panel*. *Jama-Journal of the American Medical Association*, 2010. **304**(3): p. 321-333.
76. Fichou, Y. and C. Ferec, *The potential of oligonucleotides for therapeutic applications*. *Trends in Biotechnology*, 2006. **24**(12): p. 563-570.
77. Frye, M., et al., *RNA modifications: what have we learned and where are we headed?* *Nature Reviews Genetics*, 2016. **17**(6): p. 365-372.



# CHAPITRE 4. STRATEGIES D'ETUDES DES MODIFICATIONS POST-TRANSCRIPTIONNELLES D'ARN

## 4.1 Introduction

Dans l'optique d'étudier les modifications chimiques des nucléotides, plusieurs techniques et stratégies ont été développées et imaginées pour les caractériser. Les modifications d'ARN impliquent des différences au niveau moléculaire qui changent généralement les propriétés physico-chimiques des ARN. Par conséquent, plusieurs méthodes basées sur ces changements ont été mises en place. Le but de ces techniques est multiple : il s'agit de **déterminer les structures chimiques** des nucléotides modifiés présents sur l'ARN, de **localiser** ces modifications sur les structures primaires et également de les **quantifier** [1]. L'élucidation des divers rôles biologiques de ces modifications chimiques covalentes repose sur leur détection précise [2]. Ces informations sont précieuses afin de comprendre les intérêts de ces modifications sur les oligomères. Dans la pratique, pour obtenir toutes ces informations, il est nécessaire d'utiliser plusieurs méthodes et stratégies. Afin d'obtenir ces informations, les méthodes les plus courantes sont les méthodes de biologie moléculaire et notamment les méthodes de séquençage. La plupart des techniques de séquençage d'ARN se basent sur l'amplification en chaîne par polymérase (RT-PCR) [3]. Brièvement, la RT-PCR est une technique qui consiste à dupliquer la séquence d'un ARN à analyser afin de permettre sa détection. Cette duplication est accomplie par une enzyme qui copie les séquences d'ARN via l'appariement de bases complémentaires. Toutefois, cette technique n'est pas adaptée pour l'analyse de certaines modifications. En effet, les modifications qui n'influencent pas l'appariement des bases ne sont pas détectées et certaines modifications peuvent bloquer l'enzyme de PCR résultant par une terminaison prématurée. Ainsi, la RT-PCR est une technique qui suggère les positions des nucléotides modifiés sur la séquence sans être capable de déterminer avec précision leur nature [4]. Par conséquent, des techniques alternatives ont été développées en parallèle pour étudier les modifications et notamment la MS [5]. Généralement, les ARN purifiés doivent dans un premier temps subir une préparation d'échantillon pour obtenir des informations exploitables par MS. Ainsi, la stratégie de caractérisation des modifications post-transcriptionnelles par MS se fait très généralement sur deux niveaux :

- Caractérisations structurales des nucléotides modifiés présents dans un ARN.
- Séquençage des ARN et localisation des nucléotides modifiés sur la structure primaire.

Dans ce chapitre, le protocole classique d'analyse d'ARN par MS est détaillé (voir Figure 33). Dans un premier temps, la purification des ARN est décrite. En effet, la purification et l'isolation des ARN permettent l'étude précise des modifications dans chaque ARN. Par la suite, les méthodes de préparation des ARN sont détaillées pour l'analyse par MS. La préparation d'échantillon est effectuée à deux niveaux : les méthodes de préparation

permettant la détection globale des modifications et les méthodes permettant de localiser ces modifications. Dans le premier niveau, il s'agit d'identifier les types de nucléotides modifiés présents dans les ARN. Pour cela, les échantillons d'ARN sont totalement hydrolysés en nucléotides ou nucléosides. Cette étape permet de faciliter l'analyse et de détecter les modifications chimiques. Le deuxième niveau permet de déterminer les emplacements des nucléotides modifiés sur les séquences d'ARN. Pour ce faire, différentes digestions enzymatiques ont été mises en place. Ensuite, afin d'analyser ces échantillons, les techniques classiques de MS pour détecter et localiser de façon précise les modifications seront décrites. Toutefois, il est également possible d'analyser directement les ARN totaux préalablement digéré à l'aide de plusieurs enzymes [6].

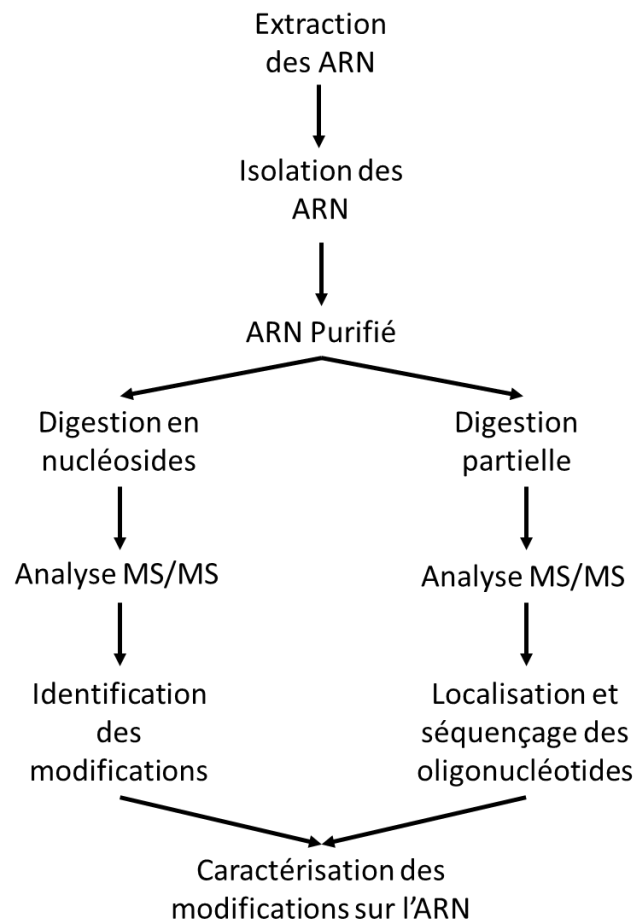


Figure 33 Stratégie globale pour la caractérisation des modifications d'un ARN par MS.

## 4.2 Purification d'ARN

Avant de pouvoir effectuer l'analyse d'ARN, il faut dans un premier temps **les extraire** de leur matrice biologique. En effet, ces biomolécules sont très nombreuses dans les cellules et dans le but d'étudier les modifications sur un seul ARN, il faut d'abord **les isoler**. Ainsi, la première section présente les grandes lignes de la purification des ARN qui sont par la suite étudiés individuellement par les stratégies décrites dans les parties suivantes. Ces purifications sont essentielles pour l'analyse des modifications d'ARN afin de s'assurer que l'étude soit faite uniquement sur un seul ARN. Ces méthodes ont longtemps été améliorées

et aujourd'hui d'excellentes purifications d'ARN sont possibles, suivant le type d'ARN. De plus, il est important qu'aucune de ces étapes ne dégradent les nucléotides afin d'éviter de perdre des informations sur les modifications.

#### 4.2.1 Extraction des ARN totaux

L'extraction des ARN est l'étape qui permet de récupérer les ARN présents dans les cellules. Il y a **quatre étapes principales** dans le protocole d'extraction des acides nucléiques [7] :

- La lyse des cellules.
- La dénaturation des protéines et des complexes nucléoprotéiques.
- L'inactivation des nucléases.
- La purification des acides nucléiques.

La première étape de la purification d'ARN consiste à détruire les membranes des cellules afin de libérer tous les constituants de celle-ci, dont les ARN. Cette étape, appelée lyse cellulaire, peut être accomplie de différentes manières : par des moyens chimiques, biologiques ou encore mécaniques [8, 9]. La lyse des cellules par la voie chimique est réalisée à l'aide de détergents ioniques avec notamment le dodécylsulfate de sodium (SDS) ou le bromure de céthyltriméthyl ammonium (CTAB). En effet, la partie hydrophobe des détergents interagit sur la membrane lipidique des cellules. Cette interaction perturbe la membrane, ce qui permet de l'éclater. L'utilisation d'enzymes permet également la lyse des cellules. Le lysozyme, la protéinase K ou encore la zymolyase sont des enzymes qui catalysent différentes réactions et dégradent la paroi des cellules. Enfin, il est possible de lyser les cellules en utilisant des broyeurs à billes, l'ultracentrifugation ou encore des ultrasons. Dans ce cas, ce sont des forces qui permettent de percer la paroi des cellules [10]. Les produits résultants de cette lyse sont appelés lysats.

Après la lyse des cellules, l'étape suivante consiste à désactiver les protéines et les complexes ribonucléoprotéiques (par exemple : le ribosome) présents dans le lysat. Les protéines sont souvent dénaturées à l'aide de composés chimiques. On peut citer l'utilisation des détergents avec par exemple le SDS, des agents réducteurs comme le béta-mercaptoéthanol ou encore les agents chaotropiques tels que le chlorure de guanidine. La protéinase K peut aussi être utilisée afin de digérer les protéines.

Ensuite les nucléases (enzymes qui digèrent les acides nucléiques) sont inactivées. Pour cela, des agents chaotropiques sont couramment utilisés comme l'urée, le chlorure de guanidine ou le thiocyanate de guanidine [11].

À la fin de ces trois étapes, les acides nucléiques sont présents dans un mélange constitué de débris cellulaires, de protéines, de lipides, d'oligosaccharides etc. Ainsi, la dernière étape consiste à séparer les acides nucléiques des impuretés. Pour cela, il existe différentes techniques telles que l'ultracentrifugation en gradient, la séparation en phase solide, ou l'extraction liquide-liquide. Cette dernière méthode est actuellement la plus

couramment utilisée notamment par extraction au phénol-chloroforme. En effet, c'est la technique la plus simple pour purifier les acides nucléiques [12]. De plus, cette méthode a l'avantage de pouvoir séparer les ARN des ADN. Les acides nucléiques sont extraits avec un mélange équimolaire de phénol:chloroforme en présence d'un chaotrope (couramment le thiocyanate de guanidinium). Les acides nucléiques, étant chargés négativement, sont insolubles dans la phase organique et se retrouvent dans la phase aqueuse. Cependant, lorsque la solution est acidifiée, les ARN restent solubles dans la phase aqueuse mais les ADN deviennent insolubles dans cette phase et se retrouvent donc dans la phase organique [13]. Les ARN sont ainsi séparés des ADN avec une bonne pureté et un bon rendement. Par la suite, d'autres étapes, comme une précipitation et un lavage, peuvent être effectuées afin d'améliorer la pureté des ARN. Néanmoins, à la fin de ce protocole, un mélange des différents ARN est obtenu. L'étude des modifications dans un mélange d'ARN totaux est difficile. Par conséquent, une autre étape permettant de séparer les différents ARN est nécessaire.

Dans le cas des ARNt, il est possible de les récupérer directement avec une simple extraction phénolique. En effet, le phénol permet de lyser les cellules et ainsi de libérer les ARN. L'extraction phénolique permet la récupération des ARNt dans la phase aqueuse. Après une étape de purification (précipitation), les ARNt sont déacylés, c'est-à-dire que les acides aminés liés aux ARNt sont retirés. Cette étape est réalisée en milieu alcalin (pH 9,0) ce qui permet d'hydrolyser les esters d'aminoacyle. Enfin, des étapes de lavages sont effectuées.

#### **4.2.2 Isolation des différents ARN**

Ainsi, après avoir obtenu le mélange d'ARN totaux, l'isolation de chaque ARN est réalisée. Il existe de nombreuses techniques permettant la séparation des biomolécules telles que les techniques chromatographiques et électrophorétiques, cependant celle qui est la plus utilisée est l'électrophorèse sur gel de polyacrylamide (PAGE). L'électrophorèse sur gel consiste en une plaque de gel réticulé baignant dans un électrolyte support. Les échantillons à séparer sont déposés sur l'une des extrémités du gel. La séparation des composés est effectuée en appliquant une tension électrique de part et d'autre du gel. Les acides nucléiques chargés négativement migrent alors à travers le gel de polyacrylamide réticulé. Le déplacement des composés à travers le gel permet ainsi leur séparation en fonction de leur taille. La PAGE est donc une technique tout à fait adaptée à la séparation des biomolécules. De plus, avec la PAGE, il est possible d'estimer la masse des composés séparés grâce à l'utilisation de standards dont les poids moléculaires sont connus. Une fois la séparation achevée, les ARN sont révélés aux UV après coloration au bromure d'éthidium.

Toutefois, afin d'améliorer la séparation des ARN, et notamment les ARNt, il est possible d'effectuer deux séparations successives. Dans une première dimension, la séparation des ARN totaux est effectuée en condition totalement dénaturante (12.5 % acrylamide/bis-acrylamide (19:1), urée 8 M, tampon 1x Tris Borate EDTA). Puis, la bande de gel de la 1<sup>ère</sup> dimension est placée sur le gel de la 2<sup>ème</sup> dimension et la séparation est réalisée en condition semi-dénaturante (20 % acrylamide/bis-acrylamide (19:1), 4 M urée, 1x Tris Borate EDTA) [14] (voir exemple Figure 34).

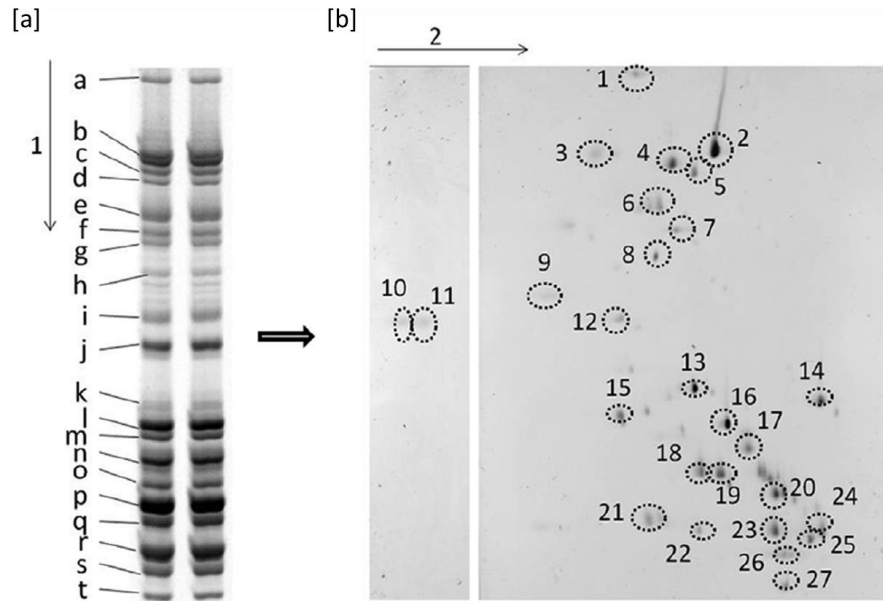


Figure 34 Séparation 2D PAGE d'un mélange d'ARNt avec [a] 1<sup>ère</sup> dimension de séparation en condition totalement dénaturante et [b] 2<sup>ème</sup> dimension de séparation en condition semi-dénaturante. Les flèches indiquent les sens de migration. Reproduit de [14].

Cette méthode, communément appelée 2D PAGE, permet d'améliorer significativement la résolution de séparation des ARN. Les ARN sont donc ainsi isolés et l'étude individuelle des modifications sur chaque ARNt est alors possible. Cependant, l'étude des modifications sur un ARN intact reste délicate par MS car ce sont des composés qui restent complexes. En effet, l'identification et la localisation des nucléotides modifiés sont des informations difficiles à obtenir en MS si l'ARN n'est pas préalablement digéré. Ainsi, des étapes permettant de digérer les ARN sont nécessaires avant de les analyser.

### 4.3 Préparation des ARN avant analyse

Pour étudier les modifications par MS, les ARN sont analysés à deux niveaux : le premier étant l'**hydrolyse totale** et le deuxième la **digestion spécifique** des ARN. La réduction préalable des ARN permet d'obtenir des fragments d'ARN qui sont beaucoup plus facilement analysables par MS. Avant de pouvoir réaliser ces étapes, il est très important de dessaler les échantillons. En effet, les ARN ont une forte tendance à former des adduits avec les cations potassium et surtout sodium. Ces adduits alcalins sont problématiques car ils réduisent fortement la qualité des spectres, il est donc important d'éliminer ces ions avant analyse MS. Ainsi, une première partie décrit brièvement les méthodes utilisées pour dessaler les ARN.

#### 4.3.1 Dessalage des échantillons

Afin d'éliminer les cations nuisibles à la détection ESI-MS, il existe de nombreuses techniques de dessalage des ARN. Cependant les plus communes sont détaillées ici.

-Précipitation :

Le principe de dessalage de cette technique repose sur l'échange des ions alcalins avec des ions ammonium qui sont compatibles avec la MS. Pour réaliser le changement d'ions, les

oligonucléotides sont dissous dans une solution fortement concentrée en sel d'ammonium. Ensuite, le mélange est précipité à  $-20^{\circ}\text{C}$  en ajoutant un large excès d'éthanol. Ainsi, les ions  $\text{Na}^+$  et  $\text{K}^+$  sont échangés avec des ions  $\text{NH}_4^+$ . L'opération peut être répétée afin d'améliorer le dessalage. Bien que cette méthode soit très simple à mettre en œuvre, elle est très chronophage car la précipitation est réalisée overnight. De plus, malgré plusieurs cycles de précipitation, tous les cations  $\text{Na}^+$  ne sont pas retirés [15].

-Extraction en phase solide (SPE) :

Le principe du dessalage par extraction en phase solide repose sur l'immobilisation des analytes sur une phase solide puis leur rinçage. Dans un premier temps, la phase solide doit être activée avec un solvant. Cette phase solide est greffée avec des chaînes carbonées (la plupart du temps  $\text{C}_{18}$ ). Après conditionnement, l'échantillon est ajouté et les ARN sont piégés sur cette phase solide. L'étape suivante consiste à laver les oligonucléotides en utilisant des solutions concentrées en sel d'ammonium. Cela va permettre de retirer les cations alcalins et de les remplacer par des ions ammoniums. Enfin, les analytes dessalés sont récupérés en les éluant de la phase stationnaire (voir Figure 35). Pour dessaler les ARN, il est très courant d'utiliser des pointes de pipette avec un lit de résine  $\text{C}_{18}$  aussi connus sous le nom de « ZipTips » [16]. Cette technique est beaucoup moins chronophage que la précipitation tout en permettant un dessalage efficace. Il est aussi important de noter que le dessalage SPE peut être réalisé après digestion enzymatique des ARN.

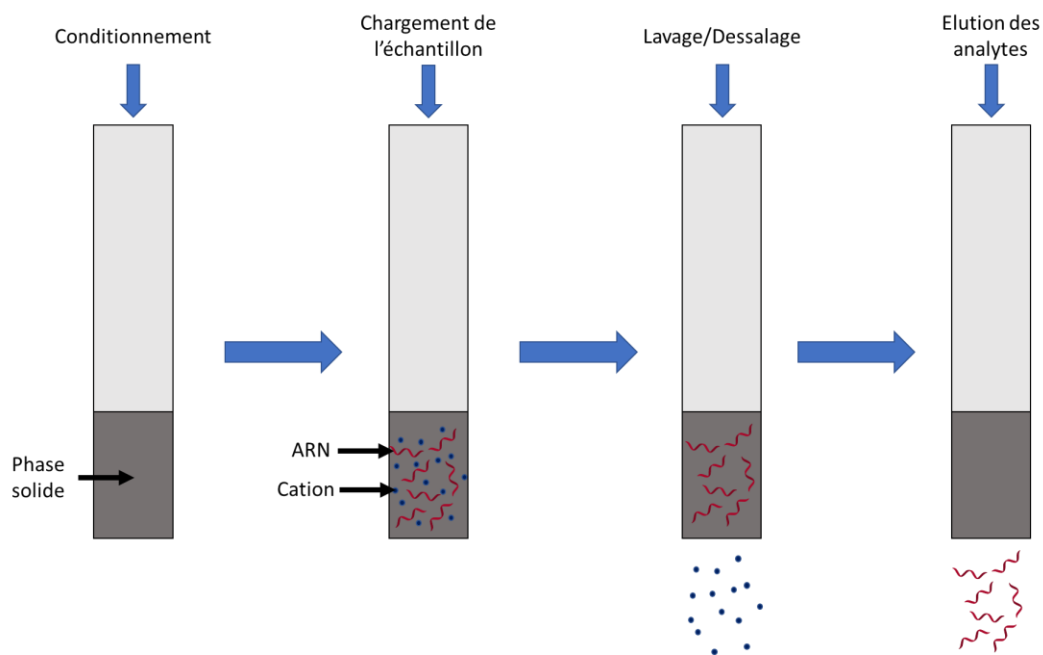


Figure 35 Principe de dessalage par extraction en phase solide.

Ainsi, en retirant les cations alcalins, ces techniques permettent d'améliorer significativement la qualité des données spectrales obtenues par ESI-MS.

### 4.3.2 Premier niveau : Hydrolyse totale

Dans ce premier niveau d'analyse, il s'agit d'identifier les modifications présentes sur un ARN donné. Pour cela, les ARN sont dans un premier temps réduits par clivages de toutes les liaisons phosphodiester entre les nucléotides qui constituent l'ARN. Les produits d'hydrolyses formés sont des nucléotides ( $\approx 300$  Da vs 20 kDa pour un ARNt). Dans le but d'hydrolyser totalement les ARN, il est possible d'effectuer une hydrolyse chimique en jouant sur le pH. Les ARN sont relativement résistants à pH acide dilué, cependant les oligonucléotides soumis à des conditions drastiques, c'est-à-dire en milieu fortement acide et chauffage, se transforment en un mélange d'oses, de phosphates et de bases nucléiques. Les ARN sont beaucoup plus sensibles en milieu alcalin puisque l'hydroxyle 2' du ribose est susceptible de céder son proton ( $\text{pH} > 11$ ). L'alcoolate formé est ensuite capable d'effectuer une attaque nucléophile sur le phosphate 3' adjacent, ce qui permet de former un phosphate cyclique en brisant la liaison 5'. Ensuite, le phosphate cyclique peut être hydrolysé de deux façons donnant alors un mélange de 2'- et de 3'-nucléoside monophosphate (voir figure ci-dessous) [17].

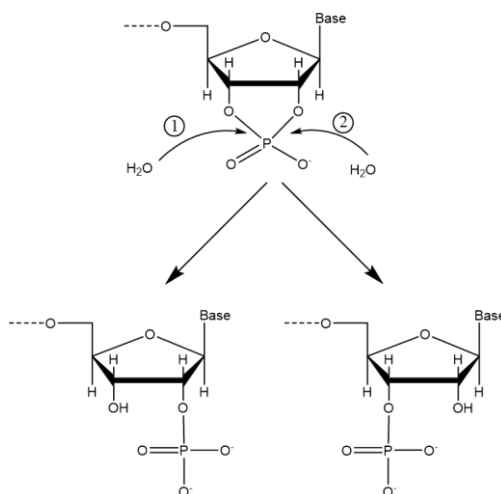


Figure 36 Hydrolyse alcaline des ARN. Hydrolyse aléatoire du 2',3'-phosphate cyclique produisant un mélange de 2'- NMP et 3'-NMP.

Bien que l'hydrolyse des ARN par voie chimique soit très simple à mettre en place, elle présente des inconvénients majeurs. En hydrolyse acide, il peut y avoir une perte d'information si un nucléotide est modifié à la fois sur la base et sur le ribose puisque ce type d'hydrolyse coupe la liaison *N*-glycosidique. En condition basique, l'obtention d'un mélange de 2'-NMP et 3'-NMP complexifie le mélange. Mais surtout, dans ces conditions de pH drastiques, les modifications risquent fortement d'être dégradées ce qui impliquerait des pertes d'informations essentielles. Ainsi, d'autres méthodes plus douces, basées sur l'utilisation d'enzymes, sont couramment utilisées pour digérer les ARN en mononucléosides/mononucléotides [18].

Il existe plusieurs enzymes permettant de réaliser l'hydrolyse totale des ARN. Ces protéines sont des phosphodiesterases puisqu'elles catalysent la réaction de coupure des toutes les liaisons phosphodiester. Ces enzymes sont aussi couramment appelées nucléases. Elles permettent le clivage des liaisons phosphodiester dans des conditions plus douces préservant ainsi les modifications portées par les nucléotides.

Les nucléases présentent différents niveaux de spécificité et sont classées par :

- Leur mode d'attaque de la chaîne : extrémité (exonucléase) ou intérieur (endonucléase).
- Leurs spécificité vis-à-vis du substrat : ADN (DNase), ARN (RNase) ou les deux et de la structure, simple ou double brin.
- Leur spécificité de reconnaissance des sites : nucléotide(s) ou leur enchaînement.
- Le type de coupure de la liaison phosphodiester (voir Figure 37).

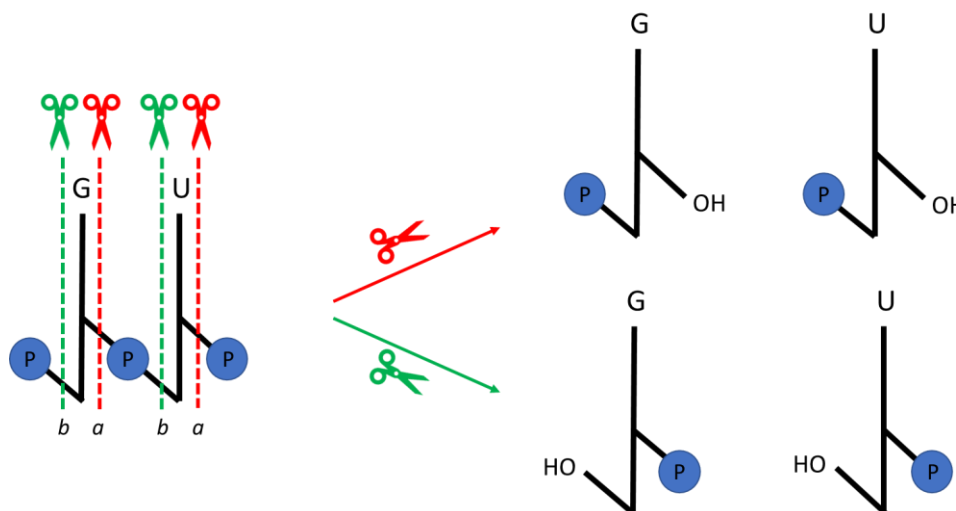


Figure 37 Sites de coupure d'une phosphodiesterase sur une chaîne d'ARN. La coupure sur le site a laisse le phosphate sur le côté 5' du nucléotide adjacent alors que la coupure sur le site b produit des 3'-NMP.

Dans le Tableau 5 sont regroupées les principales nucléases utilisées pour réaliser l'hydrolyse totale des ARN.

Nom de l'enzyme	Résidus	Site de coupure	Mode d'action (sens)	Produits formés
Nucléase P <sub>1</sub> ( <i>Penicillium citrinum</i> )	A, G, C et U	a	Endo	5'-NMP
Benzonase ( <i>Serratia marcescens</i> )	A, G, C et U	a	Endo	5'-NMP
Phosphodiesterase I (venin de <i>Crotalus adamanteus</i> )	A, G, C et U	a	Exo (3' → 5')	5'-NMP
Phosphodiesterase II (rate bovine)	A, G, C et U	b	Exo (5' → 3')	3'-NMP
Phosphatase alcaline bactérienne ( <i>E. coli</i> C90)*	A, G, C et U	a et b	Phosphates terminaux	5'-OH-[RNA]-3'-OH

Tableau 5 Liste non-exhaustive des nucléases utilisées pour l'hydrolyse totale des ARN avec leurs spécificités. \*La phosphatase alcaline bactérienne (BAP) catalyse exclusivement le clivage des phosphates terminaux.

Pour réaliser l'hydrolyse enzymatique totale des ARN, l'utilisation de plusieurs nucléases est souvent nécessaire. Par exemple, la nucléase P<sub>1</sub> est une phosphodiesterase qui



est classiquement utilisée pour hydrolyser les ARN. Cette enzyme digère les ARN en 5'-nucléosides monophosphates indépendamment de la nature du sucre (ribose/désoxyribose) et de la base. Cependant, une légère résistance à l'hydrolyse a été observée pour certains ribonucléotides chimiquement modifiés [19]. Ainsi, dans le but d'assurer une digestion complète des ARN, la nucléase P<sub>1</sub> est donc souvent utilisée avec une autre nucléase comme la phosphodiesterase I (venin de *Crotalus adamanteus*) [18, 20]. De cette manière, les ARN sont totalement hydrolysés. Par la suite, il est très courant de digérer les nucléotides en nucléosides dans le but d'améliorer la détection MS en mode positif. Cette étape est réalisée à l'aide d'alkaline phosphatase et notamment la phosphatase alcaline bactérienne (BAP) qui est thermiquement stable (à 65°C pendant au moins 1h) [21, 22].

Après avoir totalement hydrolysé les ARN, un mélange des différents nucléosides, comprenant les nucléosides modifiés, est obtenu. Ces fragments sont par la suite analysés afin d'identifier les modifications présentes dans l'échantillon. La stratégie d'hydrolyse totale permet non seulement d'identifier les nucléosides modifiés présents sur les ARN mais aussi de les quantifier. Toutefois, afin de comprendre le rôle des modifications, il est crucial de déterminer la localisation de ces ARN. Ainsi, des stratégies ont été développées pour pouvoir localiser les modifications sur les structures primaires des ARN.

### 4.3.3 Second niveau : Digestion partielle

Le second niveau d'analyse consiste à localiser les modifications sur la séquence d'ARN. Grâce à la digestion spécifique des ARN, la production et l'analyse par MS des petits fragments d'ARN permet la localisation des modifications [23]. Dans ce cas, seulement quelques liaisons phosphodiesters sont coupées. Les oligoribonucléotides issus d'une digestion partielle sont constitués généralement de un jusqu'à une dizaine de nucléotides (≈300 Da jusqu'à ≈3 kDa). Afin de produire ces oligonucléotides, l'utilisation d'enzymes spécifiques est requise. Ces nucléases, de très haute spécificité, sont capables de catalyser sélectivement le clivage de certaines liaisons phosphodiesters dans des conditions douces. Cette spécificité se traduit par la capacité de la ribonucléase (RNase) à cliver sélectivement la liaison phosphodiester de certains nucléotides canoniques [24]. De plus, ces enzymes sont souvent thermorésistantes (par exemple, la RNase T<sub>1</sub> résiste 10 min à 90 °C). Le Tableau 6 liste les principales RNases qui catalysent spécifiquement la coupure de liaisons phosphodiesters.

Nom de l'enzyme	Résidus	Site de coupure	Mode d'action (sens)	Fragments formés
RNase T <sub>1</sub>	G	b	Endo	5'-OH-[RNA]-3'-Gp/G>p
RNase A	C et U	b	Endo	5'-OH-[RNA]-3'-Cp/C>p, 5'-OH-[RNA]-3'-Up/U>p
RNase U <sub>2</sub>	G et préférentiellement A	b	Endo	5'-OH-[RNA]-3'-Ap/A>p, 5'-OH-[RNA]-3'-Gp/G>p
RNase MC1	Préférence pour U	a	Endo	5'-Up-[RNA]-3'-OH
Cusativin	Préférence pour C	b	Endo	5'-OH-[RNA]-3'-Cp/C>p

Tableau 6 Liste non-exhaustive des nucléases utilisées pour la digestion partielle des ARN avec leurs spécificités. RNase T<sub>1</sub> [24, 25] RNase A [26, 27] RNase U<sub>2</sub> [28, 29] RNase MC1 [30] Cusativin [31].

L'utilisation d'enzymes de très haute spécificité permet de générer des fragments dont certains portent des nucléotides modifiés [32]. La RNase T<sub>1</sub> et la RNase A sont les nucléases de références pour digérer les ARN [33, 34]. En effet, ces enzymes sont extrêmement spécifiques, permettant l'absence de « miscleavage » et ainsi d'obtenir des digestions très reproductibles. Il faut cependant noter que les RNases peuvent également cliver au niveau de certains nucléotides modifiés. C'est le cas de la RNase T<sub>1</sub> qui coupe, par exemple, les N<sup>2</sup>-méthylguanosines (m<sup>2</sup>G). La RNase A hydrolyse également les liaisons phosphodiester après la dihydrouridine (D) et la pseudouridine (ψ) [25]. Les ARN sont souvent digérés individuellement avec la RNase T<sub>1</sub> et la RNase A pour obtenir des oligonucléotides dont les séquences peuvent partiellement se recouvrir. Cela permet de compléter et éventuellement de confirmer la séquence d'un oligonucléotide et la position d'un nucléotide modifié. Toutefois, il est parfois nécessaire d'utiliser d'autres RNases afin de compléter les éventuels trous dans les séquences. Les autres RNases montrent une plus faible spécificité et cela génère donc une plus grande hétérogénéité des fragments. Cette hétérogénéité augmente la complexité de l'échantillon et donc son analyse. De plus, les fragments formés avec les RNases moins spécifiques dépendent des conditions de la digestion. Par exemple, une digestion plus longue forme un plus grand nombre de fragments qui sont aussi plus petits. Ainsi, pour avoir une digestion reproductible, il est impératif de maîtriser les paramètres de température et de temps d'incubation [35]. Un autre inconvénient de la digestion partielle des ARN est la redondance de séquence. En effet, les ARN ne sont constitués que de quatre bases nucléiques. De ce fait, il est fortement probable que les ARN contiennent des séquences redondantes. Malgré la possibilité d'utiliser plusieurs enzymes, les oligonucléotides aux séquences redondantes ne peuvent pas être discriminés par MS. La comparaison des résultats obtenus avec les banques de données est conseillée pour s'assurer d'identifier correctement l'ARN. Une fois la digestion d'ARN effectuée, l'échantillon est analysé pour localiser les modifications sur l'ARN.

#### 4.4 Analyse des échantillons

Afin d'identifier les modifications contenues dans l'ARN, les échantillons obtenus par digestions sont analysés à l'aide d'une technique séparative. La séparation préalable des fragments permet de décomplexifier le mélange et facilite l'identification des analytes. Plusieurs techniques sont disponibles pour séparer les molécules. Ces méthodes sont basées sur divers principes physico-chimiques afin de permettre la séparation des composés. Les principales techniques sont des méthodes chromatographiques et électrophorétique.

Une revue des applications des techniques séparatives avec détection optique pour l'analyse des échantillons est proposée. Dans un premier temps, uniquement les analyses d'échantillons totalement hydrolysés seront décrites car les détections optiques apportent très peu d'information sur les oligonucléotides modifiés. L'application de ces techniques séparatives à l'aide de détecteurs autres que la MS est décrite pour l'analyse des digests d'ARN.

#### 4.4.1 Chromatographie liquide avec détection optique

La chromatographie liquide est la méthode la plus répandue pour la séparation et l'analyse de molécules. Cette technique a également été utilisée dans le but de caractériser les modifications d'ARN. Gehrke *et al.* ont utilisé la HPLC en phase inverse (RP-HPLC), équipée d'un détecteur UV à barrette de diodes, pour séparer et identifier plus de 65 ribonucléosides en une seule analyse après hydrolyse enzymatique d'ARNt. La détection a été effectuée à 254 nm et 280 nm. L'identification des nucléosides a été réalisée à l'aide de standards. La méthode a également permis la quantification de nucléosides présents dans deux ARNr avec d'excellentes courbes de calibration [36]. Le groupe de Liebich a développé et optimisé une méthode basée sur la RP-HPLC avec détection UV permettant la quantification des nucléosides. Cette méthode, basée sur la quantification de nucléosides modifiés dans les urines, a été testée pour des tumeurs du sein et de la bouche. La séparation de tous les ribonucléosides a été réalisée, permettant une quantification de 16 nucléosides. Les recouvrements obtenus pour les 16 nucléosides étaient compris entre 83 % et 100 % sauf pour deux nucléosides. En 2006, Zhao *et al.* ont comparé la RP-HPLC avec la chromatographie liquide ultra-haute performance en phase inverse (RP-UPLC) pour l'analyse des nucléosides modifiés dans les urines. Par détection UV, 79 pics ont été détectés en HPLC-UV contre 92 en UPLC-UV suggérant la meilleure efficacité de cette dernière [37]. Ainsi, grâce aux performances offertes par la RP-HPLC, cette technique est devenue la technique de référence pour l'analyse des produits d'hydrolyse des ARN. D'autres systèmes chromatographiques ont également été développés telles que la chromatographie d'interaction hydrophile (HILIC) pour l'analyse de nucléosides par Marrubini *et al.* Afin d'optimiser la séparation, les paramètres de température, pH, concentration de sel et % d'acétonitrile ont été évalués. La méthode optimisée a permis de totalement séparer un mélange de 12 analytes, incluant nucléosides et bases azotées, en 60 min [38]. La chromatographie d'échange d'ions a également été mise en œuvre pour étudier les ARN. En 1975, ce mode chromatographique a été mis à contribution afin d'étudier la méthylation d'un ARNm. Après hydrolyse alcaline de l'ARNm et analyse, les auteurs ont déterminé la composition moyenne en nucléosides méthylés dans cette ARNm avec 3,0 de m<sup>6</sup>A, 1,0 de m<sup>7</sup>G et 1,7 de 2'-O-méthyl nucleoside [39]. La RP-HPLC avec appariement d'ions (IP-RP-HPLC) est une technique largement utilisée pour caractériser les modifications d'ARN. Ryll *et al.* ont aussi utilisé la technique d'appariement d'ions dans le but de séparer et de quantifier des nucléotides. En utilisant l'hydrogénosulfate de tétrabutylammonium en tant qu'agent d'appariement, la séparation de 31 nucléotides mono-, di- et triphosphates ainsi que des nucléotide-sucres ont été séparés avec une grande efficacité en moins de 25 min [40]. L'équipe de Lazzarino a amélioré cette technique pour réaliser la séparation de 38 composés incluant les ribonucléotides et également des désoxyribonucléotides [41]. L'IP-RP-HPLC a aussi été utilisée pour étudier des nucléosides thérapeutiques. Grâce à cette technique, Ahluwalia *et al.* ont étudié les deux nucléosides thérapeutiques : le didésoxyadénosine et le didésoxyinosine. Ils ont également pu étudier les impuretés [42]. Enfin, la chromatographie sur couche mince (CCM) en deux dimensions, une technique devenue obsolète mais qui a eu une importance cruciale dans l'étude des modifications, a été très utilisée pour caractériser les modifications d'ARN. Rogg *et al.* ont développé et optimisé cette technique, aussi appelée chromatographie planaire, pour la

séparation et la quantification de nucléosides modifiés. La CCM 2D a permis aux auteurs de séparer 40 nucléotides [43]. Cette technique a été mise en œuvre pour l'étude d'un mutant de l'ARNt<sup>Tyr</sup>. Après digestion enzymatique de l'ARNt<sup>Tyr</sup> en nucléosides, les analyses CCM ont mis en évidence l'absence de certaines modifications sur ce mutant [44]. En 1995, Keith *et al* ont publié les mobilités d'une grande variété de nucléotides modifiés en CCM 2D [45].

#### 4.4.2 Électrophorèse capillaire avec détection optique

L'électrophorèse capillaire (CE) est une technique alternative à la chromatographie liquide. En effet, ici, la séparation des analytes n'est pas basée sur des phénomènes de rétention mais des phénomènes de migration. L'électrophorèse capillaire de zone (CZE) permet la séparation d'espèces polaires et chargées en solution. En CZE, les analytes se séparent en fonction de la taille et de la charge. Ainsi, l'analyse par CZE des biomolécules, qui sont généralement ionisables et de tailles différentes, s'avère tout à fait adaptée. Par conséquent, plusieurs groupes de chercheurs ont reporté l'utilisation de la CZE pour étudier les ARN. Plusieurs études ont aussi reporté l'utilisation de la CZE-UV pour analyser les nucléosides ou nucléotides modifiés [46, 47]. Les BGE utilisés sont souvent à pH alcalin car les composés, étant acides, seront sous leurs formes chargés permettant d'assurer leur migration. Une méthode simple, basée sur l'utilisation de la CZE-UV pour la détection et la quantification des modifications de nucléosides dans des échantillons d'urine, a été reportée par Jiang *et al*. La CZE-UV a permis la séparation et la quantification de nucléosides canoniques et modifiés avec un temps d'analyse de seulement 10 min. La limite de détection a été estimée de l'ordre du  $\mu\text{mol/L}$ . Les auteurs suggèrent l'utilisation de cette technique en routine pour le diagnostic de cancer puisque cette méthode est rapide et fiable [48]. Une autre méthode a même permis de séparer les 2'-MNP des 3'-MNP grâce à l'ajout de cyclodextrine dans le BGE [49]. Mechref *et al* ont développé une méthode de CE permettant de digérer spécifiquement les ARN et de séparer les oligonucléotides simultanément [50]. Toutefois, bien que la CZE-UV soit une technique adaptée pour le diagnostic, elle souffre d'une faible sensibilité pour l'étude des modifications contenues dans les ARN qui sont parfois en faible quantité. Ainsi, la CZE couplée avec un détecteur LIF a été développée pour améliorer la sensibilité de détection. Les nucléotides n'étant pas fluorescents, il est nécessaire de les conjuguer avec une molécule fluorescente pour les rendre détectables par LIF. Tseng *et al* ont développé cette méthode pour la détection de faible concentration de nucléoside. Le chloroacétaldéhyde a été utilisé pour dériver spécifiquement les composés à base d'adénine et les rendre fluorescents [51]. Plus tard, un autre agent dérivant, le BODIPY, a été développé pour rendre les nucléotides 5'-monophosphates fluorescents [52]. Cornelius *et al* ont développé la CZE-LIF pour détecter et identifier les modifications. Après digestion à la nucléase P<sub>1</sub>, les nucléotides sont conjugués avec le BODIPY. Par la suite, l'analyse de ces composés leur a permis de détecter plusieurs nucléotides modifiés avec notamment l'inosine, la pseudouridine et la 2'-O-méthyladénosine et aussi de les quantifier. Grâce à cette méthode, la sensibilité est fortement améliorée avec une LOD d'environ 100 pmol/L [53].

D'autres modes de séparation électrophorétique ont également été développés pour analyser les nucléosides. Helboe *et al* ont proposé de séparer les nucléosides par électrochromatographie capillaire (CEC), une technique à mi-chemin entre HPLC et CZE. La

séparation de six nucléosides a été optimisée par Helboe *et al.* Une séparation complète de ces nucléosides a été obtenue en 13 min [54]. La méthode de « short-end injection » a été adaptée par Cahours *et al* pour l'analyse très rapide de nucléosides par CEC. La séparation de 5 nucléosides a été achevée en moins de 3 min [55]. L'électrochromatographie en phase micellaire (MEKC), une technique permettant de séparer des molécules non-chargées grâce à une phase pseudo-stationnaire, a été utilisée par Ruta *et al* afin de séparer des énantiomères de l'adénosine. La phase pseudo-stationnaire a permis la séparation de la  $\beta$ -D-adénosine de la  $\beta$ -L-adénosine [56].

Bien que ces techniques soient capables de séparer efficacement les nucléosides modifiés, elles présentent un inconvénient majeur car le détecteur ne fournit pas d'informations structurales permettant d'identifier les modifications. Afin d'attribuer les pics, il est nécessaire d'utiliser des standards qui ne sont pas toujours disponibles. De plus, ce genre de détecteur ne permet pas de caractériser une éventuelle modification inconnue. Depuis l'invention de la source électrospray (ESI) [57], le couplage de techniques séparatives à la MS a ouvert la voie à de nombreuses possibilités d'analyses dont celle des ARN.

#### 4.4.3 Méthodes séparatives couplées à la spectrométrie de masse

La spectrométrie de masse (MS) est une méthode permettant de détecter, d'identifier et aussi de quantifier les molécules à analyser grâce à la mesure de leur rapport masse sur charge ( $m/z$ ). Le principe de la MS repose sur la séparation des molécules à analyser, qui ont été préalablement volatilisées et ionisées, en fonction de leur  $m/z$ . Pour ce faire, un spectromètre de masse (MS) doit assurer trois fonctions :

- La production d'ions en phase gazeuse.
- L'analyse des  $m/z$  des analytes.
- La détection des ions.

Pour réaliser ces fonctions, un MS est généralement constitué de quatre parties. Premièrement, une source par laquelle les molécules sont chargées et rendues gazeuses. Il y a également une interface assurant la transmission des ions de la source jusqu'à l'analyseur. L'analyseur de masse est le système permettant de discriminer les ions selon leur  $m/z$ . Enfin, le détecteur qui permet de générer des signaux électriques qui, après retraitement, fournissent les spectres de masse. Il existe plusieurs technologies de MS qui sont adaptées pour répondre à diverses problématiques [58]. De plus, la MS ne cesse de se perfectionner, en proposant des mesures toujours plus précises et également d'apporter des informations supplémentaires notamment avec la spectrométrie de masse en tandem (MS/MS). Les expérimentations MS/MS consistent à mesurer le  $m/z$  d'une molécule (ion parent) puis de la fragmenter dans une cellule de collision pour mesurer les rapports  $m/z$  des fragments produits (ions fils). Cette technique permet ainsi d'obtenir des informations structurales sur les molécules fragmentées. Il existe différents modes de fragmentation mais la plus courante est la dissociation induite par collision (CID). En CID, les ions traversent une cellule remplie d'un gaz inerte ( $N_2$ , He ou Ar). Les analytes entrent alors en collision avec ce gaz. L'énergie de la collision induit la fragmentation des analytes et donne naissance aux ions fils.

Les modifications chimiques des ARN engendrent quasiment systématiquement une augmentation de la masse moléculaire des nucléosides canoniques. Par conséquent, la MS est devenue rapidement une méthode puissante permettant l'identification et la caractérisation structurale des modifications d'ARN [16, 59, 60]. Bien qu'il existe différentes technologies d'ionisation, la plus courante est l'ESI. En effet, l'ESI est une source d'ionisation permettant le couplage de la MS à des techniques séparatives. La LC-MS/MS est un outil extrêmement précis mais aussi sensible et son importance pour l'analyse des modifications de l'ARN est mise en évidence par les nombreuses publications récentes [7, 20, 61, 62]. Ce couplage permet de combiner la séparation des molécules offerte par la LC et également d'obtenir des informations structurales sur les molécules préalablement séparées. De plus, la source ESI, permettant le couplage LC-MS, est une méthode d'ionisation douce qui permet de préserver les modifications chimiques portées par les nucléotides.

La LC-MS/MS en phase inverse est ainsi rapidement devenue la technique de référence pour la détection mais aussi pour la quantification de nucléosides modifiés [27, 63]. Avant analyse par LC-MS/MS, les produits d'hydrolyse totale subissent un traitement permettant de transformer les nucléotides en nucléosides. Le bénéfice d'éliminer les phosphates est triple :

- Les nucléosides ont plus d'affinité avec la phase stationnaire C<sub>18</sub>, puisqu'un groupement fortement polaire est retiré, et sont donc mieux séparés.
- Cela permet de limiter les adduits causés par les phosphates. En effet, ces adduits sont préjudiciables pour la détection ESI-MS.
- La détection MS en mode positif des nucléosides est plus sensible puisqu'un groupement acide est enlevé.

La LC permet de séparer les nucléosides modifiés qui sont par la suite analysés par MS. La MS mesure le  $m/z$  des nucléosides et cette détection est réalisée en mode positive puisque les nucléosides sont chargés en milieu légèrement acide. La différence de masse entre un nucléoside canonique et le nucléoside modifié donne ainsi une information sur la nature de la modification. Toutefois, cette seule analyse ne permet pas de déterminer si la modification est située sur le ribose ou la base nucléique. C'est pourquoi une analyse MS/MS du nucléoside modifié est requise pour obtenir cette information. Lorsque les nucléosides subissent de la fragmentation CID, ils sont généralement fragmenter au niveau de la liaison N-glycosidique. Les ions fils produits sont donc le ribose et la base nucléique. Il est donc possible de déterminer, grâce aux spectres MS/MS, si la modification est liée au ribose ou à la base nucléique en comparant les différences de masse entre les composés modifiés et non-modifiés. En outre, la quantification totale des nucléosides est possible par cette technique [20, 61, 64, 65]. Afin d'identifier les nucléosides, les MS de type triple quadripôle (QqQ) sont très courantes car cette technologie fournit un pouvoir résolutif suffisant et permet d'avoir le maximum de sensibilité pour des études quantitatives. Une caractérisation structurale plus poussée de ces modifications est possible grâce au marquage isotopique [66-68] et à la dérivation chimique [69-71]. La sophistication du couplage LC-MS/MS a permis la découverte d'un grand nombre de modifications. Par exemple, cette dernière décennie, l'agmatidine (C<sup>+</sup>), la 2-méthylthioadénosine ou encore la 5-cyanométhyl-2-thiouridine (c<sub>nm</sub><sup>5s2U</sup>) ont été découvertes [6].

Les modifications contenues dans les ARN sont donc identifiées par trois critères :

- Leur temps de rétention.
- La masse de l'ion parent.
- Les masses des ions fils.

Ainsi, ce premier niveau d'analyse permet d'identifier les modifications présentes sur les ARN. Toutefois, la localisation des modifications sur la séquence de l'ARN n'est pas possible à ce niveau car les ARN sont totalement hydrolysés. Par conséquent un deuxième niveau d'analyse des ARN est nécessaire pour obtenir cette information.

La LC-MS/MS est également une technique mise à contribution pour localiser les modifications. En 1993, l'équipe de McCloskey a démontré la puissance de cette technique pour l'analyse d'oligonucléotides [23, 72]. Cette méthode, dite « bottom-up » permet de reconstituer la séquence d'un ARN à partir de ses fragments. Pour cela, les ARN sont d'abord partiellement digérés à l'aide de RNases puis les échantillons sont analysés par LC-MS/MS. La LC permet de séparer les oligonucléotides pour faciliter la détection et la MS mesure le  $m/z$  des oligonucléotides. Cependant, pour déterminer l'emplacement des modifications, la mesure MS n'est pas suffisante. Les oligonucléotides sont donc fragmentés par MS/MS de façon à obtenir des informations de séquençage. Dans ce cas, l'énergie de fragmentation CID est plus faible comparée à la fragmentation des nucléosides afin de rompre spécifiquement les liaisons phosphodiester [73]. Ces liaisons peuvent se briser à différents niveaux et une nomenclature proposée par McLuckey existe pour désigner ces différents ions [74] (voir Figure 38). Cette nomenclature est similaire à celle utilisée pour désigner les fragments des polypeptides.

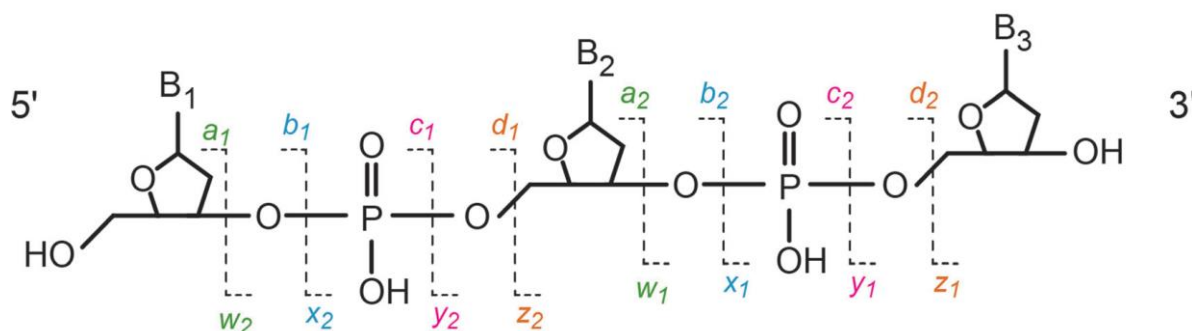


Figure 38 Nomenclature des ions formés par fragmentation en phase gazeuse des oligonucléotides.

En réalisant des expériences de fragmentation CID sur des ARN, les ions fils préférablement formés sont les ions c et y (rupture de la liaison P-O5') [75, 76]. Les ions (a-B) (perte de la base hétérocyclique et rupture de la liaison C3'-O3') et w (rupture de la liaison C3'-O3') sont aussi observables mais beaucoup moins favorables [77]. Par conséquent, grâce aux spectres MS/MS, il est possible de reconstituer les séquences des oligonucléotides grâce à ces ions. Le séquençage est réalisé en suivant une série d'ions fils. La différence de masse entre chaque ions successifs permet de déterminer l'enchaînement des nucléotides dans un oligonucléotide [78] (voir Tableau 7). Ce type de séquençage est également connu sous le nom de séquençage *de Novo*.

Nucléotide	$\Delta$ de masse (Da)
C	305,18
U	306,17
A	329,21
G	345,21

Tableau 7 Correspondances entre le nucléotide et les différences de masse entre les fragments d'ARN de longueur de 1 nucléotide.

Pour séquencer les oligonucléotides, le MS doit être suffisamment résolutif pour distinguer une C d'une U car le  $\Delta$  de masse est de 1 Da. Les MS hybrides classiquement utilisés pour analyser les oligonucléotides sont le quadripôle-temps de vol (Q-TOF), la trappe ionique (IT) ou encore le quadripôle-Orbitrap (Q-Orbitrap).

La localisation d'un nucléotide modifié se fait de la même manière. Si, pendant le séquençage *de Novo*, la différence de masse entre les ions ne correspond pas à l'un des quatre nucléotides alors il s'agit d'un nucléotide modifié. Cela permet ainsi de localiser un nucléotide modifié sur la séquence d'ARN. Des expériences de marquage isotopique [26, 79] et de dérivations chimiques [80] peuvent être réalisées pour faciliter la localisation de certaines modifications sur les spectres MS/MS. En effet, l'ajout d'une masse supplémentaire permet de facilement confirmer, sur les spectres MS/MS, la position d'une modification sur l'ARN. Ces techniques de marquages sont également couramment utilisées pour quantifier les modifications post-transcriptionnelles [68]. Ici, contrairement à la quantification au niveau des nucléosides, la quantification des modifications est réalisable sur une seule modification à la fois. Autrement dit, si une même modification est présente à plusieurs positions sur la séquence d'ARN, ce niveau d'analyse permet la quantification individuelle de chacune de ces modifications.

Ainsi, le couplage LC-MS/MS est également la méthode de référence qui permet non seulement de localiser les modifications d'ARN sur leur structure primaire mais également de séquencer les ARN [81]. Avec cette technique, la détection des oligonucléotides est réalisée en mode négative. Cette technique a été mise en œuvre avec succès pour l'identification et la localisation de modifications d'ARN [82, 83] et aussi d'impuretés d'ARN thérapeutiques phosphorothioatés [84]. À ce jour, le mode chromatographique le plus courant pour analyser les oligonucléotides est l'IP-RP-HPLC. En effet, les agents d'appariements d'ions (par exemple la triéthylamine) permettent, en s'associant avec les analytes, de masquer la polarité des phosphodiesters. Cette association permet alors de donner un comportement hydrophobe aux analytes et donc une meilleure interaction avec la phase stationnaire  $C_{18}$  apolaire. Les agents d'appariements permettent la rétention des oligonucléotides polaires et offrent ainsi des séparations en phase inverse, facilitant grandement l'interprétation des spectres. Pour assurer une correcte ionisation des oligonucléotides, il est nécessaire d'utiliser un additif : le 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP) [85]. Néanmoins, l'utilisation de cet additif est problématique pour des raisons techniques car la forte concentration en HFIP peut boucher la colonne de chromatographie et la durée de vie du détecteur est diminuée car il est saturé par la détection constante de HFIP [86].



Toutefois, bien que la localisation d'un nucléotide modifié sur la séquence soit possible, ce niveau d'analyse ne permet pas de déterminer si la modification chimique est située sur le ribose ou sur la base du nucléotide. De plus, il est très rare d'obtenir un recouvrement de séquence total d'ARN en MS et, de ce fait, des modifications peuvent ne pas être localisables. Afin de confirmer les résultats obtenus expérimentalement, il est très courant de les confronter à une banque de données. Une digestion *in silico* peut être réalisée afin de simuler les oligonucléotides formés et leur masse correspondante ainsi qu'une simulation des masses des ions fils. Ensuite, les résultats expérimentaux sont comparés aux résultats théoriques. L'ARN est confirmé si les masses des ions parents et des fragments correspondent. Modomics est une banque de données qui contient une grande variété d'informations sur les ARN comme les séquences et aussi les modifications (natures et positions). <https://iimcb.genesilico.pl/modomics/> [87]. Mongo Oligo Mass Calculator est un outil de calcul des masses d'oligonucléotides et de fragments obtenus par (CID) et/ou digestions enzymatiques qui a été développé par l'équipe du Professeur Limbach <https://mods.rna.albany.edu/masspec/Mongo-Oligo>. Ces outils ont été utilisés dans le cadre des travaux de thèse.

En ce qui concerne le couplage CE-MS, plusieurs groupes ont décrits son utilisation pour l'analyse d'ARN. En 1995, la CE-MS a été utilisée pour analyser des nucléotides monophosphates [88]. L'utilisation de la CE-MS a été développée pour l'analyse des NMP, NDP et NTP dans les nourritures pour bébés. La méthode a été validée en évaluant l'effet de matrice, les limites, la justesse et la fidélité [89]. La détection et la quantification de nucléotides modifiés par CE-MS/MS ont été réalisées par Yu *et al.* La séparation et la détection de 12 nucléotides ont été réalisées en moins de 15 min [90]. En 2016, Khan *et al.* ont développé une méthode basée sur la CE-MS pour caractériser directement les modifications d'ARNmi dans des sérums de patients cancéreux. Les auteurs ont également réalisé la quantification de l'ARNmi à l'aide de cette technique [91]. La caractérisation d'oligonucléotides thérapeutiques par CZE-MS a été réalisée par Demelenne *et al.* [84]. Néanmoins, aucun groupe n'a développé la CE-MS pour le séquençage et la localisation des modifications d'ARN.

L'amélioration des méthodes analytiques favorisera certainement l'élucidation des rôles des modifications de l'ARN. À l'avenir, cela pourra offrir de nouvelles opportunités aux chercheurs pour manipuler les modifications et développer de petites molécules permettant d'altérer les fonctions de l'ARN à des fins de recherche fondamentale et thérapeutique.

Ce chapitre a dressé un état de l'art des méthodes classiques basées sur l'utilisation de la MS pour caractériser les modifications post-transcriptionnelles. Aujourd'hui, bien que le couplage CZE-MS/MS soit disponible, aucune méthode n'a été décrite dans la littérature pour la caractérisation des modifications d'ARN. En effet, la CZE est une technique alternative à la LC qui pourrait ainsi apporter des informations complémentaires à la LC. De plus, la CZE est une technique séparative d'espèces chargées en solution. Or, dans le premier chapitre, il a été énoncé que les quatre nucléotides canoniques ont des comportements acido-basiques différentes qui peuvent être exploités dans le cadre d'une séparation électrophorétique. C'est pourquoi, dans ces travaux de thèse, une méthode basée sur le couplage CZE-MS/MS a été développée afin d'apporter une alternative et une complémentarité à la LC-MS/MS.

## 4.5 Références bibliographiques

1. Chen, B., B.-F. Yuan, and Y.-Q. Feng, *Analytical Methods for Deciphering RNA Modifications*. Analytical Chemistry, 2019. **91**(1): p. 743-756.
2. Helm, M. and Y. Motorin, *Detecting RNA modifications in the epitranscriptome: predict and validate*. Nature Reviews Genetics, 2017. **18**(5): p. 275-291.
3. Motorin, Y., et al., *Identification of modified residues in RNAs by reverse transcription-based methods*. Rna Modification, 2007. **425**: p. 21-53.
4. Limbach, P.A., *Indirect mass spectrometric methods for characterizing and sequencing oligonucleotides*. Mass Spectrometry Reviews, 1996. **15**(5): p. 297-336.
5. Kellner, S., J. Burhenne, and M. Helm, *Detection of RNA modifications*. Rna Biology, 2010. **7**(2): p. 237-247.
6. Yu, N.X., et al., *tRNA Modification Profiles and Codon-Decoding Strategies in Methanocaldococcus jannaschii*. Journal of Bacteriology, 2019. **201**(9).
7. Cai, W.M., et al., *A Platform for Discovery and Quantification of Modified Ribonucleosides in RNA: Application to Stress-Induced Reprogramming of tRNA Modifications*. Rna Modification, 2015. **560**: p. 29-71.
8. Kim, J., et al., *Microfluidic sample preparation: cell lysis and nucleic acid purification*. Integrative Biology, 2009. **1**(10): p. 574-586.
9. Zhang, J., et al., *Rapid and label-free nanomechanical detection of biomarker transcripts in human RNA*. Nature Nanotechnology, 2006. **1**(3): p. 214-220.
10. Boom, R., et al., *Rapid and Simple Method for Purification of Nucleic-Acids*. Journal of Clinical Microbiology, 1990. **28**(3): p. 495-503.
11. Akhtar, S., et al., *A method to extract intact and pure RNA from mycobacteria*. Analytical Biochemistry, 2011. **417**(2): p. 286-288.
12. Chomczynski, P. and N. Sacchi, *The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on*. Nature Protocols, 2006. **1**(2): p. 581-585.
13. Chomczynski, P. and N. Sacchi, *Single-Step Method of Rna Isolation by Acid Guanidinium Thiocyanate Phenol Chloroform Extraction*. Analytical Biochemistry, 1987. **162**(1): p. 156-159.
14. Antoine, L., et al., *Mapping post-transcriptional modifications in Staphylococcus aureus tRNAs by nanoLC/MSMS*. Biochimie, 2019. **164**: p. 60-69.
15. Potier, N., et al., *Negative Electrospray-Ionization Mass-Spectrometry of Synthetic and Chemically-Modified Oligonucleotides*. Nucleic Acids Research, 1994. **22**(19): p. 3895-3903.
16. Thomas, B. and A.V. Akoulitchev, *Mass spectrometry of RNA*. Trends in Biochemical Sciences, 2006. **31**(3): p. 173-181.
17. Markham, R. and J.D. Smith, *The structure of ribonucleic acid. I. Cyclic nucleotides produced by ribonuclease and by alkaline hydrolysis*. The Biochemical journal, 1952. **52**(4): p. 552-557.
18. Crain, P.F., *Preparation and Enzymatic-Hydrolysis of DNA and Rna for Mass-Spectrometry*. Methods in Enzymology, 1990. **193**: p. 782-790.
19. Edmonds, C.G., M.L. Vestal, and J.A. McCloskey, *Thermospray Liquid Chromatography-Mass Spectrometry of Nucleosides and of Enzymatic Hydrolysates of Nucleic-Acids*. Nucleic Acids Research, 1985. **13**(22): p. 8197-8206.

20. Thüring, K., et al., *Analysis of RNA modifications by liquid chromatography–tandem mass spectrometry*. *Methods*, 2016. **107**: p. 48-56.
21. Pang, H., et al., *Identification of Nucleosides in Hydrolysates of Transfer-Rna by High-Resolution Mass-Spectrometry*. *European Journal of Biochemistry*, 1982. **127**(3): p. 459-471.
22. Gehrke, C.W., et al., *Quantitative Enzymatic-Hydrolysis of Transfer-Rnas - Reversed-Phase High-Performance Liquid-Chromatography of Transfer-Rna Nucleosides*. *Journal of Chromatography*, 1982. **230**(2): p. 297-308.
23. Kowalak, J.A., et al., *A Novel Method for the Determination of Posttranscriptional Modification in Rna by Mass-Spectrometry*. *Nucleic Acids Research*, 1993. **21**(19): p. 4577-4585.
24. Steyaert, J., *A Decade of Protein Engineering on Ribonuclease T1— Atomic Dissection of the Enzyme-Substrate Interactions*. *European Journal of Biochemistry*, 1997. **247**(1): p. 1-11.
25. Ross, R.L., X.Y. Cao, and P.A. Limbach, *Mapping Post-Transcriptional Modifications onto Transfer Ribonucleic Acid Sequences by Liquid Chromatography Tandem Mass Spectrometry*. *Biomolecules*, 2017. **7**(1).
26. Waghmare, S.P. and M.J. Dickman, *Characterization and Quantification of RNA Post-transcriptional Modifications Using Stable Isotope Labeling of RNA in Conjunction with Mass Spectrometry Analysis*. *Analytical Chemistry*, 2011. **83**(12): p. 4894-4901.
27. Gaston, K.W. and P.A. Limbach, *The identification and characterization of non-coding and coding RNAs and their modified nucleosides by mass spectrometry*. *Rna Biology*, 2014. **11**(12): p. 1568-1585.
28. Puri, P., et al., *Systematic identification of tRNAome and its dynamics in Lactococcus lactis*. *Molecular Microbiology*, 2014. **93**(5): p. 944-956.
29. Houser, W.M., et al., *Combining recombinant ribonuclease U2 and protein phosphatase for RNA modification mapping by liquid chromatography-mass spectrometry*. *Analytical Biochemistry*, 2015. **478**: p. 52-58.
30. Addepalli, B., N.P. Lesner, and P.A. Limbach, *Detection of RNA nucleoside modifications with the uridine-specific ribonuclease MC1 from Momordica charantia*. *Rna*, 2015. **21**(10): p. 1746-1756.
31. Addepalli, B., et al., *Novel ribonuclease activity of cusativin from Cucumis sativus for mapping nucleoside modifications in RNA*. *Analytical and Bioanalytical Chemistry*, 2017. **409**(24): p. 5645-5654.
32. Pomerantz, S.C., J.A. Kowalak, and J.A. McCloskey, *Determination of Oligonucleotide Composition from Mass Spectrometrically Measured Molecular-Weight*. *Journal of the American Society for Mass Spectrometry*, 1993. **4**(3): p. 204-209.
33. Pace, C.N., et al., *Ribonuclease-T1 - Structure, Function, and Stability*. *Angewandte Chemie-International Edition*, 1991. **30**(4): p. 343-360.
34. Raines, R.T., *Ribonuclease A*. *Chemical Reviews*, 1998. **98**(3): p. 1045-1065.
35. Solivio, B., et al., *Improving RNA modification mapping sequence coverage by LC-MS through a nonspecific RNase U2-E49A mutant*. *Analytica Chimica Acta*, 2018. **1036**: p. 73-79.
36. Gehrke, C.W. and K.C. Kuo, *Ribonucleoside Analysis by Reversed-Phase High-Performance Liquid-Chromatography*. *Journal of Chromatography*, 1989. **471**: p. 3-36.

37. Zhao, X.J., et al., *Urinary profiling investigation of metabolites with cis-diol structure from cancer patients based on UPLC-MS and HPLC-MS as well as multivariate statistical analysis*. Journal of Separation Science, 2006. **29**(16): p. 2444-2451.
38. Marrubini, G., B.E.C. Mendoza, and G. Massolini, *Separation of purine and pyrimidine bases and nucleosides by hydrophilic interaction chromatography*. Journal of Separation Science, 2010. **33**(6-7): p. 803-816.
39. Desrosiers, R.C., K.H. Friderici, and F.M. Rottman, *Characterization of Novikoff Hepatoma Messenger-Rna Methylation and Heterogeneity in Methylated 5' Terminus*. Biochemistry, 1975. **14**(20): p. 4367-4374.
40. Ryll, T. and R. Wagner, *Improved Ion-Pair High-Performance Liquid-Chromatographic Method for the Quantification of a Wide Variety of Nucleotides and Sugar Nucleotides in Animal-Cells*. Journal of Chromatography-Biomedical Applications, 1991. **570**(1): p. 77-88.
41. Dipierro, D., et al., *An Ion-Pairing High-Performance Liquid-Chromatographic Method for the Direct Simultaneous Determination of Nucleotides, Deoxynucleotides, Nicotinic Coenzymes, Oxypurines, Nucleosides, and Bases in Perchloric-Acid Cell-Extracts*. Analytical Biochemistry, 1995. **231**(2): p. 407-412.
42. Ahluwalia, G., et al., *Initial Studies on the Cellular Pharmacology of 2',3'-Dideoxyinosine, an Inhibitor of Hiv Infectivity*. Biochemical Pharmacology, 1987. **36**(22): p. 3797-3800.
43. Rogg, H., et al., *Improved method for separation and quantitation of modified nucleosides transfer-RNA*. Nucleic Acids Research, 1976. **3**(1): p. 285-295.
44. Ericson, J.U. and G.R. Bjork, *Pleiotropic Effects Induced by Modification Deficiency Next to the Anticodon of Transfer-Rna from Salmonella-Typhimurium-Lt2*. Journal of Bacteriology, 1986. **166**(3): p. 1013-1021.
45. Keith, G., *Mobilities of Modified Ribonucleotides on 2-Dimensional Cellulose Thin-Layer Chromatography*. Biochimie, 1995. **77**(1-2): p. 142-144.
46. Shao, X.W., et al., *Analysis of Nucleotide Pools in Human Lymphoma-Cells by Capillary Electrophoresis*. Journal of Chromatography A, 1994. **680**(2): p. 463-468.
47. Oneill, K., et al., *Capillary Electrophoresis of Nucleotides on Ucon-Coated Fused-Silica Columns*. Analytical Biochemistry, 1994. **222**(1): p. 185-189.
48. Jiang, Y.Q. and Y.F. Ma, *A Fast Capillary Electrophoresis Method for Separation and Quantification of Modified Nucleosides in Urinary Samples*. Analytical Chemistry, 2009. **81**(15): p. 6474-6480.
49. Tadey, T. and W.C. Purdy, *Capillary Electrophoretic Separation of Nucleotide Isomers Via Complexation with Cyclodextrin and Borate*. Journal of Chromatography B-Biomedical Applications, 1994. **657**(2): p. 365-372.
50. Mechref, Y. and Z. Elrassi, *Capillary Enzymophoresis of Nucleic-Acid Fragments Using Coupled Capillary Electrophoresis and Capillary Enzyme Microreactors Having Surface-Immobilized Rna-Modifying Enzymes*. Electrophoresis, 1995. **16**(11): p. 2164-2171.
51. Tseng, H.C., R. Dadoo, and R.N. Zare, *Selective Determination of Adenine-Containing Compounds by Capillary Electrophoresis with Laser-Induced Fluorescence Detection*. Analytical Biochemistry, 1994. **222**(1): p. 55-58.
52. Wang, P.G. and R.W. Giese, *Phosphate-specific fluorescence labeling with BO-IMI: reaction details*. Journal of Chromatography A, 1998. **809**(1-2): p. 211-218.
53. Cornelius, M.G. and H.H. Schmeiser, *RNA analysis by MEKC with LIF detection*. Electrophoresis, 2007. **28**(21): p. 3901-3907.

54. Helboe, T. and S.H. Hansen, *Separation of nucleosides using capillary electrochromatography*. Journal of Chromatography A, 1999. **836**(2): p. 315-324.
55. Cahours, X., et al., *Fast separation of nucleosides by capillary electrochromatography on non-encapped phenyl-bonded silica phase using short-end injection method*. Hrc-Journal of High Resolution Chromatography, 2000. **23**(2): p. 138-142.
56. Ruta, J., et al., *Aptamer-Modified Micellar Electrokinetic Chromatography for the Enantioseparation of Nucleotides*. Analytical Chemistry, 2009. **81**(3): p. 1169-1176.
57. Fenn, J.B., et al., *Electrospray Ionization for Mass-Spectrometry of Large Biomolecules*. Science, 1989. **246**(4926): p. 64-71.
58. Cohen, A., *Mass Spectrometry, Review of the Basics: Electrospray, MALDI and Commonly Used Mass Analyzers*. Applied Spectroscopy Reviews, 2009. **44**(4): p. 362-362.
59. Crain, P.F., *Mass-Spectrometric Techniques in Nucleic-Acid Research*. Mass Spectrometry Reviews, 1990. **9**(5): p. 505-554.
60. Banoub, J.H., et al., *Recent developments in mass spectrometry for the characterization of nucleosides, nucleotides, oligonucleotides, and nucleic acids*. Chemical Reviews, 2005. **105**(5): p. 1869-1915.
61. Su, D., et al., *Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry*. Nature Protocols, 2014. **9**(4): p. 828-841.
62. Takakura, M., et al., *Biogenesis and functions of aminocarboxypropyluridine in tRNA*. Nature Communications, 2019. **10**.
63. Esmans, E.L., et al., *Liquid chromatography mass spectrometry in nucleoside, nucleotide and modified nucleotide characterization*. Journal of Chromatography A, 1998. **794**(1-2): p. 109-127.
64. Fan, H., et al., *Qualitative and quantitative determination of nucleosides, bases and their analogues in natural and cultured Cordyceps by pressurized liquid extraction and high performance liquid chromatography-electrospray ionization tandem mass spectrometry (HPLC-ESI-MS/MS)*. Analytica Chimica Acta, 2006. **567**(2): p. 218-228.
65. Russell, S.P. and P.A. Limbach, *Evaluating the reproducibility of quantifying modified nucleosides from ribonucleic acids by LC-UV-MS*. Journal of Chromatography B, 2013. **923-924**: p. 74-82.
66. Kellner, S., et al., *Absolute and relative quantification of RNA modifications via biosynthetic isotopomers*. Nucleic Acids Research, 2014. **42**(18).
67. Kellner, S., et al., *Profiling of RNA modifications by multiplexed stable isotope labelling*. Chemical Communications, 2014. **50**(26): p. 3516-3518.
68. Borland, K., et al., *Production and Application of Stable Isotope-Labeled Internal Standards for RNA Modification Analysis*. Genes, 2019. **10**(1).
69. Patteson, K.G., L.P. Rodicio, and P.A. Limbach, *Identification of the mass-silent post-transcriptionally modified nucleoside pseudouridine in RNA by matrix-assisted laser desorption/ionization mass spectrometry*. Nucleic Acids Research, 2001. **29**(10).
70. Jiang, H.P., et al., *Determination of formylated DNA and RNA by chemical labeling combined with mass spectrometry analysis*. Analytica Chimica Acta, 2017. **981**: p. 1-10.
71. Li, Q.-Y., B.-F. Yuan, and Y.-Q. Feng, *Mass Spectrometry-based Nucleic Acid Modification Analysis*. Chemistry Letters, 2018. **47**(12): p. 1453-1459.
72. Taoka, M., et al., *An analytical platform for mass spectrometry-based identification and chemical analysis of RNA in ribonucleoprotein complexes*. Nucleic Acids Research, 2009. **37**(21): p. e140-e140.

73. Phillips, D.R. and J.A. McCloskey, *A Comprehensive Study of the Low-Energy Collision-Induced Dissociation of Dinucleoside Monophosphates*. International Journal of Mass Spectrometry and Ion Processes, 1993. **128**(1-2): p. 61-82.
74. McLuckey, S.A. and S. Habibigoudarzi, *Decompositions of multiply-charged oligonucleotide anions*. Journal of the American Chemical Society, 1993. **115**(25): p. 12085-12095.
75. Tromp, J.M. and S. Schürch, *Gas-Phase Dissociation of Oligoribonucleotides and their Analogs Studied by Electrospray Ionization Tandem Mass Spectrometry*. Journal of the American Society for Mass Spectrometry, 2005. **16**(8): p. 1262-1268.
76. Andersen, T.E., F. Kirpekar, and K.F. Haselmann, *RNA fragmentation in MALDI mass spectrometry studied by H/D-exchange: Mechanisms of general applicability to nucleic acids*. Journal of The American Society for Mass Spectrometry, 2006. **17**(10): p. 1353-1368.
77. Wu, J. and S.A. McLuckey, *Gas-phase fragmentation of oligonucleotide ions*. International Journal of Mass Spectrometry, 2004. **237**(2-3): p. 197-241.
78. Joyner, J.C., K.D. Keuper, and J.A. Cowan, *Analysis of RNA cleavage by MALDI-TOF mass spectrometry*. Nucleic Acids Research, 2013. **41**(1).
79. Li, S. and P.A. Limbach, *Identification of RNA sequence isomer by isotope labeling and LC-MS/MS*. Journal of Mass Spectrometry, 2014. **49**(11): p. 1191-1198.
80. Durairaj, A. and P.A. Limbach, *Improving CMC-derivatization of pseudouridine in RNA for mass spectrometric detection*. Analytica Chimica Acta, 2008. **612**(2): p. 173-181.
81. Wetzel, C. and P.A. Limbach, *Global identification of transfer RNAs by liquid chromatography-mass spectrometry (LC-MS)*. Journal of Proteomics, 2012. **75**(12): p. 3450-3464.
82. Guymon, R., et al., *Influence of Phylogeny on Posttranscriptional Modification of rRNA in Thermophilic Prokaryotes: The Complete Modification Map of 16S rRNA of Thermus thermophilus*. Biochemistry, 2006. **45**(15): p. 4888-4899.
83. Kirpekar, F., et al., *Mapping of ribosomal 23S ribosomal RNA modifications in Clostridium sporogenes*. RNA Biology, 2018. **15**(8): p. 1060-1070.
84. 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*. Journal of Chromatography A, 2020. **1614**.
85. Apffel, A., et al., *Analysis of oligonucleotides by HPLC-electrospray ionization mass spectrometry*. Analytical Chemistry, 1997. **69**(7): p. 1320-1325.
86. Lobue, P.A., et al., *Oligonucleotide analysis by hydrophilic interaction liquid chromatography-mass spectrometry in the absence of ion-pair reagents*. Journal of Chromatography A, 2019. **1595**: p. 39-48.
87. Boccaletto, P., et al., *MODOMICS: a database of RNA modification pathways. 2017 update*. Nucleic Acids Research, 2018. **46**(D1): p. D303-D307.
88. Zhao, Z.X., et al., *Online Capillary Electrophoresis Electrospray-Ionization Mass-Spectrometry of Nucleotides*. Electrophoresis, 1995. **16**(3): p. 389-395.
89. 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 Chemistry, 2017. **233**: p. 38-44.

90. Yu, Y., et al., *Ultrasensitive and simultaneous determination of RNA modified nucleotides by sheathless interfaced capillary electrophoresis-tandem mass spectrometry*. Chemical Communications, 2019. **55**(53): p. 7595-7598.
91. Khan, N., G. Mironov, and M.V. Berezovski, *Direct detection of endogenous MicroRNAs and their post-transcriptional modifications in cancer serum by capillary electrophoresis-mass spectrometry*. Analytical and Bioanalytical Chemistry, 2016. **408**(11): p. 2891-2899.

## PARTIE II : MISE EN PLACE DE LA METHODE



Cette partie du manuscrit vise à décrire, dans un premier temps, le développement de la méthode d'analyse CE-MS/MS des modifications post-transcriptionnelles puis, dans un second temps, l'interprétation des données. Pour cela, dans le premier chapitre de cette partie, l'équipement utilisé et ses caractéristiques sont décrits puis les travaux réalisés dans l'optique de séparer et de détecter les analytes sont détaillés. Dans le chapitre suivant, c'est la méthode utilisée pour exploiter et interpréter les données qui est décrite. Ici, des outils permettant de simuler les résultats d'analyses ont été utilisés puis, ces résultats ont été confrontés avec les résultats expérimentaux.

## CHAPITRE 5. DEVELOPPEMENT DE LA METHODE D'ANALYSE

### 5.1 Introduction

Les ARN sont des acides nucléiques dont la taille est très variable selon leurs fonctions et l'organisme dans lequel ils se trouvent. De plus, il est courant que les nucléotides des ARN subissent des modifications chimiques. Ces modifications jouent un rôle prépondérant dans un grand nombre de processus biochimiques. En effet, les modifications, en apportant de nouvelles propriétés, permettent de stabiliser les structures des ARN et aussi d'assurer leur bon fonctionnement. Il existe une grande variété de modifications chimiques et notamment dans les ARN de transfert. Ces modifications sont situées à des endroits stratégiques sur les séquences d'ARN afin d'apporter les propriétés nécessaires. La caractérisation préliminaire des modifications post-transcriptionnelles d'ARN permet l'étude et la compréhension des mécanismes biochimiques complexes impliquant ces espèces.

Les outils d'analyse ne cessent de s'améliorer au bénéfice d'une meilleure compréhension de l'impact des modifications sur les ARN. Dans l'optique d'étudier ces modifications, la MS s'est rapidement imposée comme une méthode de référence grâce à ses caractéristiques propres comme son excellente spécificité, sensibilité et la possibilité d'obtenir des informations structurales. Par conséquent la MS peut être utilisée dans l'étude et la découverte de nouvelles modifications chimiques de nucléotides ou encore pour le contrôle qualité d'ARN thérapeutiques. Pour cela, la MS doit couvrir les besoins tels que les séquences d'ARN ainsi que la structure chimique et la position dans les séquences des modifications. Toutefois, l'analyse par MS d'échantillons complexes requiert une technique séparative en amont dans le but de faciliter l'interprétation des spectres MS. De plus, il est primordial que ces méthodes analytiques ne dégradent pas les échantillons au risque de perdre des informations essentielles. Depuis son invention, la source électrospray ESI permet de coupler des techniques séparatives en veine liquide à un MS. L'ESI est une source d'ionisation douce capable de transférer en phase gazeuse des biomolécules polaires de haute masse moléculaire. Cette source a permis de coupler la MS à la RP-HPLC qui est aujourd'hui utilisée en routine pour l'analyse d'ARN. Plus récemment, différentes interfaces ont été développées pour permettre le couplage de la CE à la MS. La CE est une technique alternative à la LC car son principe de séparation repose sur la migration des analytes sous un champ électrique.

Ici, un nouveau protocole simple, basé sur le couplage CZE-MS/MS, a été développé dans le but d'analyser les modifications d'ARN. Dans ces travaux de thèse, le développement d'un protocole basé sur plusieurs digestions individuelles d'ARN est détaillé. Par la suite, l'analyse de ces digestats par le couplage CZE-MS/MS en mode positif permet d'obtenir les informations spectrales donnant accès à la caractérisation des modifications chimiques. Ces différentes étapes de préparation d'échantillon permettent d'analyser les ARN à différents niveaux afin d'obtenir des informations sur les modifications [1]. Dans un premier temps, un échantillon d'ARN est totalement hydrolysé à l'aide d'enzymes pour obtenir un mélange de nucléosides. L'analyse de cet échantillon par CZE-MS/MS permet de déterminer les différents nucléotides modifiés présents sur l'ARN. Ensuite, l'ARN est digéré par des endonucléases. Ces digestions permettent de produire des oligonucléotides dont l'analyse CZE-MS/MS permet de fournir les informations de séquençage mais également de déterminer l'emplacement des modifications sur la structure primaire de l'ARN. Dans le but d'améliorer les recouvrements de séquences, deux enzymes spécifiques ont été utilisées (voir figure ci-dessous).

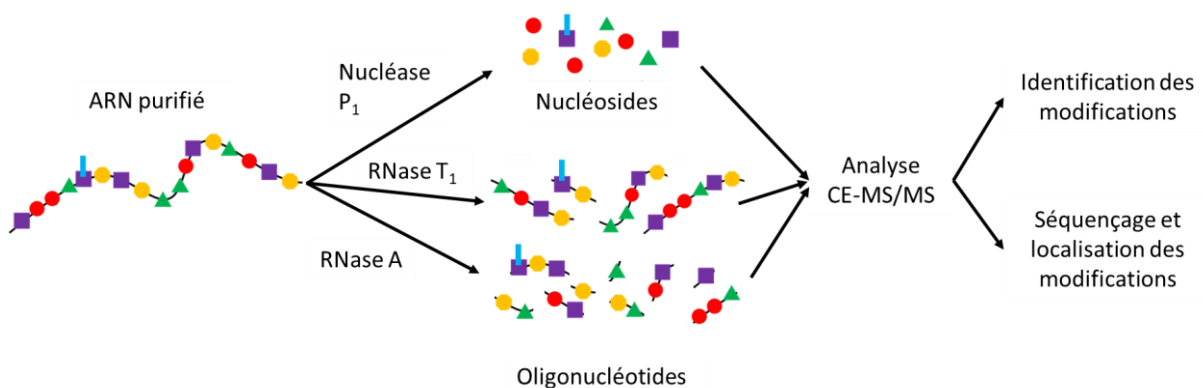


Figure 39 Protocole adopté pour la caractérisation des modifications d'ARN par CZE-MS/MS. Les deux stratégies d'études sont présentées ainsi que les informations obtenues.

Dans un premier temps, les données obtenues par l'analyse de l'échantillon d'ARN totalement hydrolysé indiquent quelles sont les modifications présentes sur l'ARN. L'hydrolyse totale des ARN est effectuée par digestion enzymatique à l'aide de la nucléase  $P_1$ . Cette digestion permet d'obtenir un mélange de ribonucléotides, dont ceux qui portent les modifications, issus de l'ARN. Ensuite, les nucléotides sont déphosphorylés à l'aide de la BAP pour permettre d'améliorer la détection en mode positif. L'hydrolysate est ensuite analysé par CZE-MS/MS. Les modifications sont identifiées lorsque des transitions spécifiques à ces modifications sont détectées. Grâce à ce premier niveau d'analyse, il est donc possible d'identifier quels sont les nucléotides modifiés présents sur un ARN. Toutefois, ce niveau de caractérisation ne permet pas de localiser les modifications sur les séquences d'ARN.

Par conséquent, une stratégie a été mise en place pour déterminer les positions des nucléotides modifiés sur les séquences d'ARN. Pour cela, une approche dite « bottom-up », directement inspirée de la stratégie couramment utilisée en protéomique, a été mise en place. Cette approche permet d'obtenir des mélanges de petits oligonucléotides qui sont ensuite facilement analysables par CZE-MS/MS. Ici, la reconstruction des séquences et la localisation de modifications (« modification mapping ») sont réalisées à l'aide de deux digestions individuelles d'ARN. En effet, les spécificités des deux RNases utilisées permettent d'obtenir

différents fragments à partir d'un même ARN. L'analyse de ces digests d'ARN permet alors de compléter ou de confirmer le séquençage et la localisation d'une modification. La RNase T<sub>1</sub> et la RNase A sont les enzymes qui ont été utilisées pour produire les fragments d'ARN. Durant ces travaux, il a été remarqué que lorsque les phosphates terminaux des oligonucléotides sont retirés, cela permet d'améliorer grandement la qualité de séparation électrophorétique. Ainsi, la BAP est également utilisée pour déphosphoryler les phosphates terminaux des oligonucléotides avant analyse CZE-MS/MS. Les informations fournies par les spectres MS et MS/MS permettent de reconstituer les séquences des ARN et, avec le même jeu de données MS/MS, de localiser sur les séquences d'ARN les modifications précédemment identifiées dans le premier niveau d'analyse. Le séquençage et la localisation des modifications sont alors effectuées simultanément par séquençage *de Novo* en suivant les séries d'ions c et y (règles de fragmentation de McLuckey cf. 4.4.3). Ce second niveau d'analyse permet donc de séquencer un ARN et également de situer les nucléotides modifiés sur sa structure primaire.

Dans ce chapitre, après une présentation du système d'analyse utilisé, le développement de la séparation électrophorétique des ARN et l'optimisation des paramètres de détection par MS sont présentés et discutés. Afin de développer cette méthode, des digests d'ARN ont dans un premier temps été infusés pour optimiser la détection puis la séparation électrophorétique a été évaluée. La mise en place de l'analyse au premier niveau est d'abord décrite puis au second niveau. L'analyse de l'échantillon totalement hydrolysé par CZE-MS/MS a montré la séparation des 4 nucléosides canoniques. L'attribution des pics a été effectuée sans ambiguïté grâce aux transitions spécifiques des nucléosides. Les oligonucléotides déphosphorylés, issus des digestions partielles, ont été analysés et séparés par CZE-MS/MS dans les mêmes conditions que les nucléosides. La séparation des oligonucléotides non-modifiés par CZE a également été très efficace, permettant de simplifier grandement l'exploitation des spectres MS et MS/MS. Les oligonucléotides de l'ARN ont ainsi été facilement séquencés par MS/MS. La séparation d'oligonucléotides isobariques avec des séquences différentes a également été réalisée, démontrant ainsi la puissance de la CZE pour la caractérisation des ARN.

## 5.2 Présentation du système CZE-MS/MS

Dans cette partie, les descriptions ainsi que les caractéristiques des appareils utilisées pour l'analyse des modifications d'ARN sont disponibles. Les paramètres sur lesquelles peut jouer l'utilisateur, dans le but d'optimiser une analyse, sont également décrits.

### 5.2.1 Système d'électrophorèse capillaire

Dans le but de séparer les analytes, l'instrument d'électrophorèse capillaire utilisé dans ces travaux de thèse était le système d'électrophorèse capillaire CESI8000 (Sciex Separations). Grâce à cette CE, il est possible de réaliser des séparations en appliquant une différence de potentiel de 1 jusqu'à 30 kV entre les électrodes avec une intensité du courant électrique entre 1 et 300  $\mu$ A. Il est également possible de choisir la polarité du champ électrique. Par défaut, la polarité dite « forward » signifie que l'électrode d'entrée correspond à l'anode et donc que l'électrode de sortie est la cathode.

Les séparations sont réalisées dans des capillaires de faible diamètre interne dont les parois internes sont constituées de silice vierge fondue (BFS). Le faible diamètre interne de ces capillaires permet de limiter l'échauffement par effet Joule qui peut apparaître en appliquant une haute tension électrique. Ce capillaire de séparation est installé dans une cartouche permettant d'assurer son bon positionnement vis-à-vis des électrodes. Cet instrument est équipé d'un système de détection optique par UV. Une lampe au deutérium permet d'émettre des rayonnements UV qui peuvent être captés en utilisant un détecteur UV ou alors un détecteur à barrettes de diodes. Néanmoins, ce système n'a pas été utilisé dans le cadre de cette thèse. En effet, les quantités d'échantillons injectées ne sont pas détectables par détection UV car ce système n'est pas assez sensible étant donné le trajet optique très court dans les capillaires (30  $\mu\text{m}$  i.d.).

En ce qui concerne l'injection des échantillons dans le capillaire de séparation, celle-ci est assurée par une électrovanne en utilisant la méthode d'injection hydrodynamique. Cette électrovanne permet d'appliquer des pressions allant de 0,1 jusqu'à 100 psi (soit de 6,9 à 6895 mbar). Ce système est également capable de créer une aspiration avec des valeurs comprises entre 0,1 et 5 psi. L'injection est effectuée en appliquant une pression dans le vial échantillon (pour une injection hydrodynamique, la pression ne peut excéder 5 psi afin de garantir la quantité injectée). L'application de cette pression force le liquide contenu dans le vial à circuler dans le capillaire et permet d'injecter un volume précis d'échantillon. L'équation permettant de déterminer le volume d'injection est la suivante :

$$V_{inj} = \frac{\pi \cdot d_c^4 \cdot \Delta P \cdot t}{128 \cdot \eta \cdot L} \cdot 10^{12} \quad (3)$$

$V_{inj}$  : Volume injecté (nL)

$d_c$  : Diamètre interne (m)

$\Delta P$  : Pression appliquée (Pa)

$t$  : Temps d'injection (s)

$\eta$  : Viscosité dynamique (Pa.s)

$L$  : Longueur totale du capillaire (m)

Les volumes d'injections ont été calculés avec l'application CEToolbox. <https://play.google.com/store/apps/details?id=com.github.cetoolbox&hl=fr>

Dans le but d'optimiser une séparation en CZE, l'utilisateur peut jouer sur différents paramètres tels que :

- Électrolyte support (nature, pH et force ionique).
- Tension de séparation.
- Volume d'injection.

Le choix de la composition de l'électrolyte support (BGE) est extrêmement important pour obtenir une séparation efficace des analytes. Néanmoins, dans le cadre du couplage de la CE à la MS, le choix des BGE est limité. En effet, le BGE doit notamment assurer le contact

électrique pour la séparation électrophorétique mais doit également être compatible ESI-MS. Ainsi, les BGE utilisés doivent être volatiles et cela restreint les possibilités de séparation. Les BGE les plus communément utilisés en CE-MS sont l'acide acétique et l'acide formique pour des pH bas et leur sel d'ammonium respectif pour des pH neutres. Pour des milieux alcalins, ce sont le carbonate d'ammonium ou l'hydroxyde d'ammonium qui sont utilisés. En plus de la volatilité du BGE, la force ionique du BGE doit être considérée. En effet, une force ionique trop élevée engendre une forte intensité du courant électrique pouvant créer un chauffage par effet Joule et risque de dégrader les analytes ou bien le capillaire de séparation [2].

La tension de séparation doit être réglée en fonction du BGE utilisé. En effet, les caractéristiques du BGE influent sur la valeur de la tension électrique à appliquer. Dans la pratique, il faut veiller à ce que la tension n'induisse pas un courant électrique trop intense au risque de produire un effet Joule. Généralement, il est recommandé d'appliquer des fortes valeurs de tension afin d'assurer des séparations efficaces et rapides. En augmentant la tension de séparation, les différences de mobilités entre les composés vont aussi augmenter et ainsi permettre d'améliorer la résolution des séparations. Il s'agit donc de trouver une valeur de tension à appliquer pour obtenir une séparation efficace et rapide tout en évitant un échauffement du capillaire.

Enfin, le volume d'échantillon injecté doit également être considéré pour réaliser des séparations en CZE. Si le volume d'injection est trop grand, les pics détectés sur l'électrophérogramme seraient étêtés et perdraient donc fortement en résolution de séparation. Cependant, il ne faut pas que le volume d'injection soit trop faible au risque de perdre en sensibilité de détection. En règle générale, le volume d'échantillon injecté ne doit pas dépasser 1 % de la longueur totale du capillaire.

Afin de réaliser le couplage CE-ESI-MS, plusieurs concepts ont été reportés dans la littérature. Dans la partie suivante, les considérations techniques et certaines interfaces mises au point pour ce type de couplage sont discutées.

### 5.2.2 Interface CE-ESI-MS

Le couplage CE-ESI-MS permet de bénéficier de l'excellent pouvoir résolutif de la CE et de l'excellente sensibilité de la MS mais offre aussi la possibilité d'obtenir des informations structurales sur les analytes préalablement séparés. Cependant, ce type de couplage s'est longtemps heurté à une limitation majeure. En effet, dans ce type de couplage, il est primordial de maintenir deux hautes tensions de façon indépendante :

- Une différence de potentiel, appliquée aux extrémités du capillaire, nécessaire pour réaliser une séparation électrophorétique.

- Une tension, appliquée entre la sortie du capillaire de séparation et l'entrée du MS, utilisée pour assurer l'ionisation ESI.

La valeur de la tension de séparation peut aller jusqu'à 30 kV et pour la tension ESI cette valeur est comprise entre 900 et 5000 V. Les grandes valeurs de ces tensions, ainsi que leur proximité, peuvent induire la formation d'arcs électriques. L'indépendance de ces hautes

tensions permet d'assurer des séparations électrophorétiques efficaces et une ionisation correcte des analytes tout en évitant la formation d'arcs électriques. Par conséquent, il est très important de comprendre le fonctionnement de ce type de couplage afin de limiter les risques. En effet, la conception des sources ESI varie selon les constructeurs, il convient donc de s'assurer de bien comprendre le fonctionnement de la source avant de mettre en place ce type de couplage.

Dans ces travaux, un spectromètre de la société Bruker a été utilisé. Cet appareil est équipé d'une source ESI. Ici, le nébuliseur de l'ESI est relié à la terre et un potentiel est appliqué à l'entrée du spectromètre afin de former un électrospray. Pour appliquer un champ électrique permettant la séparation des analytes dans le capillaire, un potentiel est appliqué sur l'électrode située à l'entrée du capillaire et l'électrode de sortie est reliée à la terre. Ainsi, pour appliquer ces deux tensions de façon indépendante, c'est donc l'électrode de sortie qui sert de retour à la terre pour la tension de séparation et également la tension ESI (voir le schéma Figure 40). Ce retour à la terre commun permet de maintenir les tensions de façon indépendante afin d'assurer les bonnes séparations électrophorétiques et la bonne ionisation ESI tout en prévenant la formation d'arc électrique.

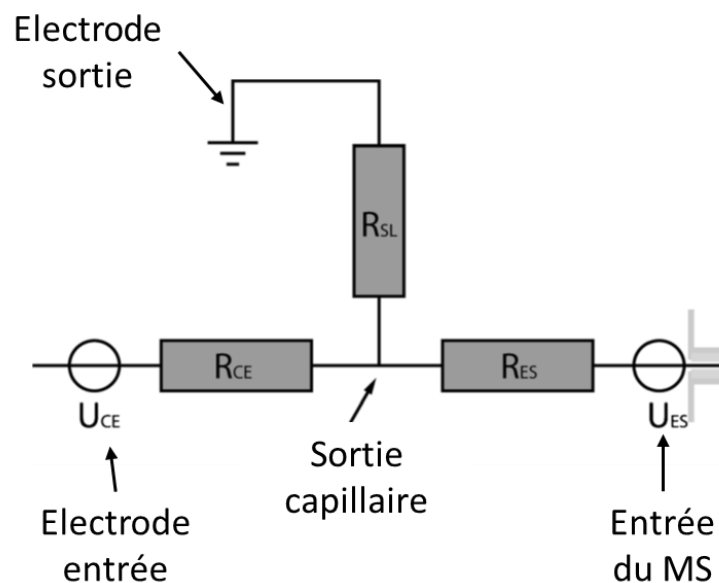


Figure 40 Schéma électrique pour le couplage CE-MS. Correspondances :  $U_{ce}$  = potentiel de séparation,  $U_{es}$  = potentiel ESI,  $R_{ce}$  = résistance du capillaire,  $R_{sl}$  = résistance BGE,  $R_{es}$  = résistance entre la sortie du capillaire et l'entrée du MS. Adapté de [3].

Ainsi, durant ces trois dernières décennies, plusieurs interfaces permettant de coupler la CE à la MS ont été développées [4]. La première, dite « sheath liquid », a été introduite en 1988 par Smith *et al* [5]. Dans cette interface, le capillaire de séparation est inséré dans un nébuliseur. Afin de réaliser le retour à la terre, un capillaire coaxiale fourni un débit d'une dizaine de  $\mu\text{L}/\text{min}$  de liquide additionnel (« sheath liquid ») à l'extrémité du capillaire de séparation. Ce liquide additionnel doit être conducteur et, suivant l'application, ce liquide peut être aqueux ou constitué d'un mélange eau-solvant. Ce liquide a deux fonctions car il permet de créer un contact électrique et également de stabiliser l'ESI. Ainsi, une jonction est

réalisée entre l'extrémité du capillaire de séparation et du capillaire coaxial assurant le contact électrique. L'autre extrémité du capillaire coaxial est plongée dans le réservoir dans lequel plonge l'électrode reliée à la terre. C'est l'interface la plus répandue pour réaliser le couplage CE-MS car elle est très robuste et facile à mettre en œuvre. Dans le but de faciliter la formation du spray, un gaz de nébulisation est utilisé. Bien que le débit de liquide additionnel permette de fournir le contact électrique, il dilue de manière non-négligeable les analytes. De plus, les débits générés par cette interface (de 5 à 10  $\mu\text{L}/\text{min}$ ) réduisent l'efficacité de désolvatation des analytes dans le processus ESI. Par conséquent, cette interface fournit une faible sensibilité de détection.

Plus tard, un autre type d'interface dite sans liquide additionnel ou « sheathless » a été développée. Ici, c'est plus précisément l'interface utilisant un capillaire poreux qui a été utilisée. Cette interface a été introduite pour la première fois par Moini *et al* [6]. Dans ce type d'interface, une courte longueur de l'extrémité du capillaire de séparation est rendue poreuse par attaque acide (acide fluorhydrique). Cette porosité permet aux électrons et ions de faibles diamètres de traverser cette paroi. Ainsi, la jonction électrique est donc réalisée via l'extrémité poreuse de ce capillaire. De plus, c'est à cette extrémité poreuse que sera généré l'ESI. Un capillaire coaxial, ou capillaire de contact, amène un liquide conducteur statique (souvent un BGE) non pas à l'extrémité du capillaire de séparation mais sur la partie poreuse de celui-ci. Ce capillaire de contact est plongé dans un réservoir de liquide conducteur qui contient l'électrode reliée à la terre. Ainsi, cette interface permet de maintenir le contact électrique par le biais de l'extrémité poreuse du capillaire de séparation et du liquide conducteur statique. Cette interface est celle qui fournit la meilleure sensibilité parmi les interfaces CE-MS disponibles. En effet, aucun liquide additionnel ne dilue les analytes et les faibles débits du capillaire de séparation (< 20-30 nL/min) permettent de générer un nanoESI. Ainsi, ce faible débit dispense l'utilisation d'un gaz de nébulisation. Ce sont les faibles diamètres internes du capillaire de séparation qui produisent un faible débit et favorisent la formation de fines gouttelettes, améliorant de façon significatif le phénomène de désolvatation pour l'ESI. Un schéma de ce type d'interface est disponible sur la Figure 41.

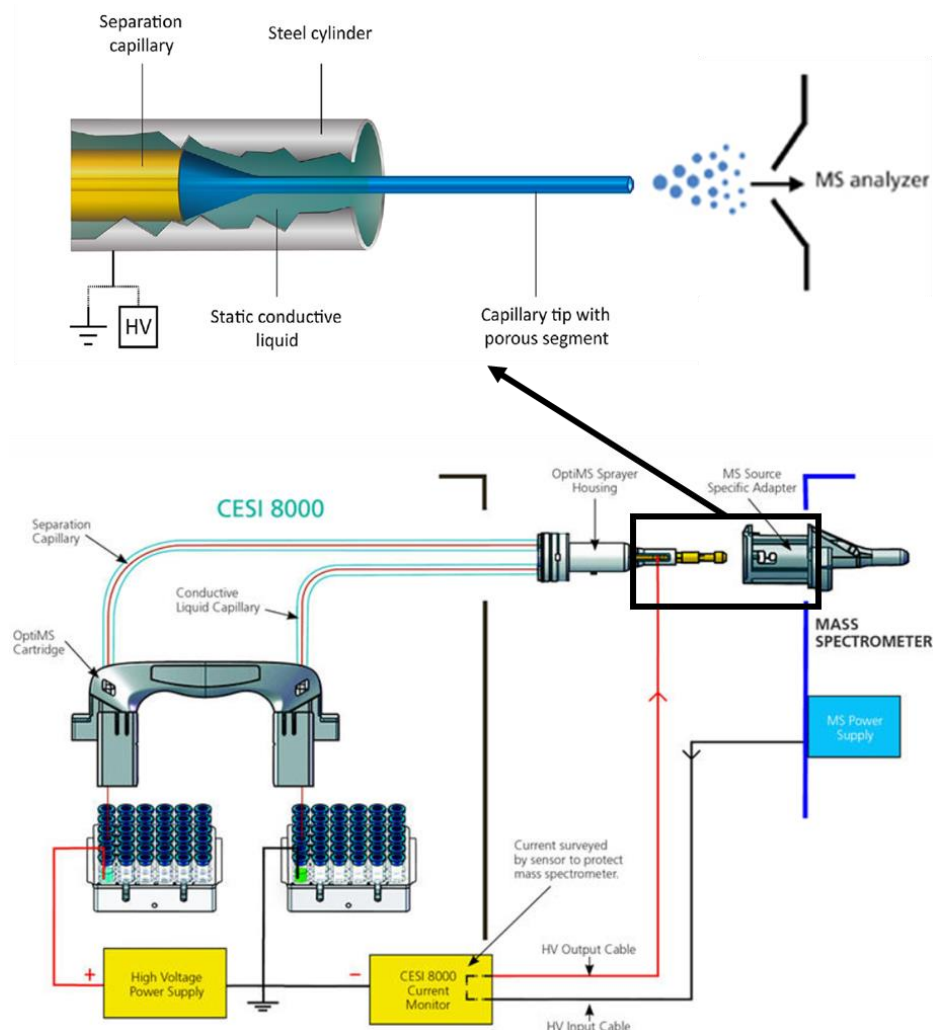


Figure 41 Schéma du système d'interface sheathless utilisant un capillaire poreux pour le couplage CE-MS permettant d'appliquer la tension de séparation et la tension ESI de façon indépendante. Adaptée de [2].

Ainsi, dans le cadre de ces travaux de thèse, c'est l'interface « sheathless » qui a été mise en application pour réaliser le couplage CE-MS car c'est celle qui offre la meilleure sensibilité. Cette sensibilité est nécessaire pour détecter des nucléotides modifiés présents en faible quantité.

Pour réaliser ce type de couplage, deux capillaires sont nécessaires. Les dimensions du capillaire de séparation sont classiquement 100 cm de longueur, 30  $\mu\text{m}$  de diamètre interne et 175  $\mu\text{m}$  de diamètre externe. Une extrémité de ce capillaire a été attaquée à l'acide fluorhydrique, sur environ 3 cm, afin de rendre la pointe poreuse. Quant au capillaire de contact, les dimensions sont de 80 cm de longueur, 75  $\mu\text{m}$  de diamètre interne et 365  $\mu\text{m}$  de diamètre externe. Les parois internes de ces deux capillaires sont en BFS.

### 5.2.3 Spectromètre de masse

Afin de développer cette méthode, le MS utilisé dans le cadre de ces travaux était le maXis 4G (Bruker Daltonics). C'est un spectromètre de masse hybride de type ESI-Qq-TOF. Avec ce spectromètre, il est possible de changer la source ESI en nanoESI. Cet instrument offre



la possibilité de réaliser des analyses MS et des expériences MS/MS. Le maXis 4G est doté d'un premier analyseur qui est un quadripôle permettant de sélectionner, filtrer ou balayer les ions parents dans le cadre d'analyses MS/MS. La fragmentation de l'ion parent est réalisée par CID dans une cellule de collision située en aval du quadripôle. La dissociation de l'ion parent est obtenue par collision avec un gaz inerte comme l'argon ou l'azote. Ces ions sont ensuite analysés par TOF (voir schéma du MS Figure 42).

Les performances de ce spectromètre de masse annoncées par le constructeur sont les suivantes :

- Résolution : jusqu'à 60 000 (FWHM).
- Précision de mesure de masse : 600 ppb.
- Gamme de masse : 20 – 20000 m/z.
- Fréquence maximale d'acquisition : 30 Hz.

Cet instrument de haute résolution permet la mesure précise des  $m/z$ , tout en donnant accès à la possibilité d'effectuer des expériences MS/MS. Par conséquent, cet appareil est tout à fait adapté pour l'étude des modifications d'ARN.

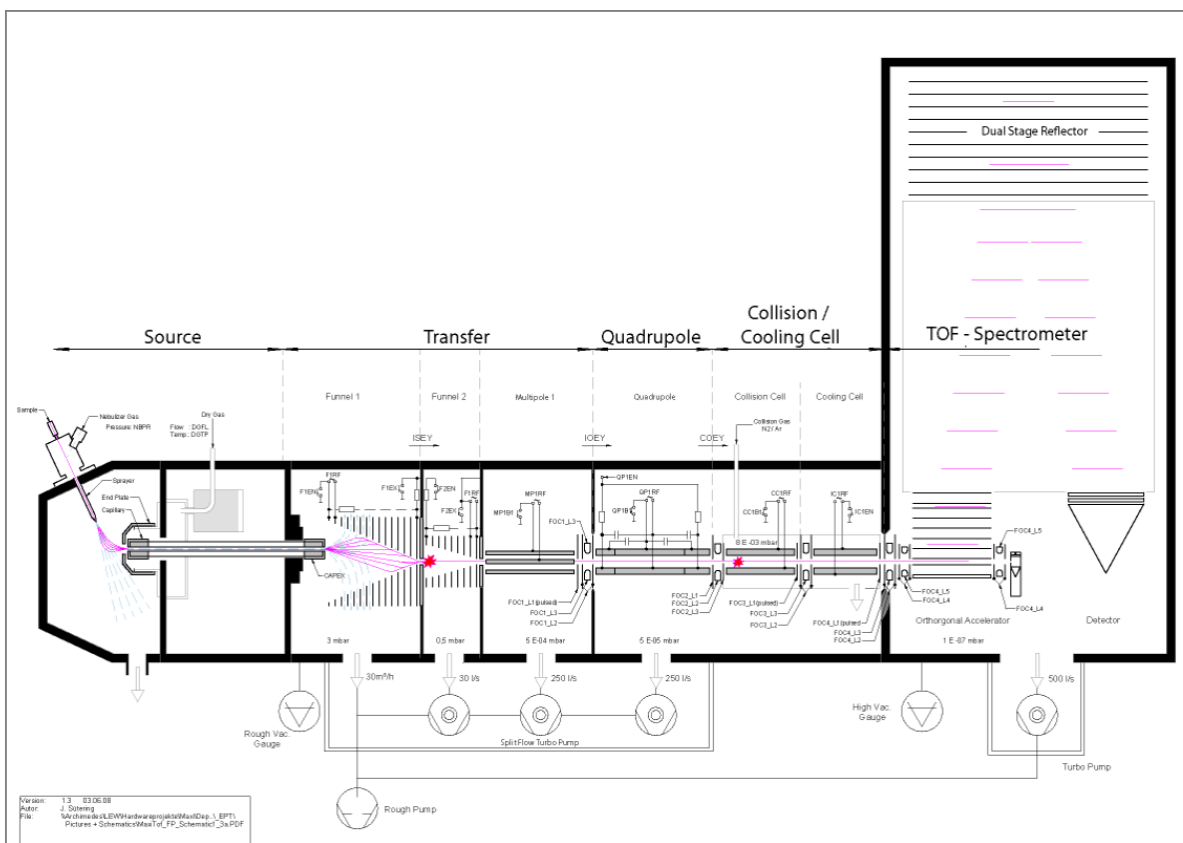


Figure 42 Représentation schématique du spectromètre de masse maXis 4G (Bruker Daltonics) détaillant les différentes parties de cet instrument.

Avec cet instrument plusieurs paramètres sont réglables, de la source jusqu'à la fragmentation en passant par le transfert des ions, afin d'optimiser le signal de détection des analytes.

Dans un premier temps, les paramètres de la source ESI sont réglés. Le réglage des paramètres de la source permet d'assurer la bonne ionisation des analytes. Dans le cadre du couplage CE-MS, la source du maXis 4G doit être mise en mode nanoESI online. En effet, les faibles débits générés en nanoESI rendent inutile l'utilisation d'un gaz de nébulisation. Ainsi ce mode élimine le gaz de nébulisation. Ensuite, plusieurs paramètres sont réglables pour optimiser le processus ESI avec notamment la tension ESI, le débit de gaz séchant et la température de la source. La tension appliquée en nanoESI est bien inférieure à celle appliquée en ESI classique ( $\approx 1,5$  kV vs  $\approx 4$  kV). La valeur de cette tension dépend notamment de la distance entre l'entrée du MS et de l'extrémité du capillaire [7]. La distance est généralement de l'ordre de quelques millimètres. Plus cette distance est grande, plus la tension à appliquer est grande. Le faible débit de la nanoESI ainsi que la proximité entre l'extrémité et l'entrée du MS permettent d'obtenir un spray stable avec une tension significativement plus faible. Toutefois, il est important d'éviter l'application de trop fortes tensions car cela accélère l'usure du capillaire. La polarité (positive ou négative) de cette tension peut être modifiée suivant la nature des analytes à détecter. Ensuite le débit de gaz séchant et la température de la source sont modifiables, permettant d'assurer l'évaporation complète des gouttelettes générées par l'ESI. Ces paramètres dépendent du BGE utilisé pour la séparation électrophorétique. Cependant, un débit trop fort de gaz séchant risque de diminuer la sensibilité du signal MS.

Ensuite, il est également possible de régler les paramètres de transfert d'ions. En MS, le réglage fin des paramètres de transfert d'ion permet d'obtenir le maximum de signal pour les analytes. Le tableau ci-dessous regroupe les paramètres de transfert d'ions, avec les valeurs possibles et leurs significations.

Nom	Valeurs (unité)	Signification
Funnel RF	0,0-400,0 (V)	Tension RF appliquée pour le « funnel ». La valeur est ajustée suivant les ions à analyser.
isCID Energy	0-200 (eV)	Valeur d'énergie appliquée entre la sortie du « funnel 1 » et l'entrée du « funnel 2 ». Attention, il ne s'agit pas de l'énergie utilisée pour fragmenter les ions en isCID. Dans certains cas, il est parfois intéressant que cette valeur ne soit pas nulle pour améliorer le signal de détection.
Multipole RF	0-2000 (V)	Tension RF appliquée pour le multipôle. La valeur est ajustée suivant les ions à analyser.
Ion Energy	0-200 (eV)	Différence d'énergie appliquée entre le multipôle et le quadripôle. Ce paramètre doit être ajusté suivant les ions à analyser tout en évitant de les fragmenter.
Low Mass	20,0-3000,0(m/z)	Rapport m/z minimum pour la fragmentation. Les ions possédant un rapport m/z inférieur à cette valeur ne traversent pas le quadripôle.
Collision Energy	0-200 (eV)	Différence d'énergie entre le quadripôle et la cellule de collision. Attention, il ne s'agit pas de l'énergie utilisée pour fragmenter les ions en CID. Néanmoins, une énergie trop grande risque de fragmenter les molécules.

Collision RF	0-2000 (V)	Tension RF appliquée dans la cellule de collision. La valeur est ajustée suivant les ions à analyser.
Transfer time	1,0-199,0 (µs)	Définit le temps accordé pour le transfert des ions au TOF. Généralement, plus le rapport m/z d'un ion est grand, plus ce temps doit être long.
Pre-pulse storage	1,0-128,5 (µs)	Définit le délai pour la collecte des ions entre le « transfer time » et l'activation du pulseur TOF. De la même manière qu'avant, plus le rapport m/z d'un ion est grand, plus ce temps doit être long. Remarque : le temps accordé pour le « pre-pulse storage » ne peut être inférieur au « transfer time ».

Tableau 8 Paramètres ajustables de transfert d'ions du maXis 4G avec les valeurs et leurs significations.

De la même manière, l'utilisateur a la possibilité d'ajuster les paramètres pour les expériences de MS/MS. Il existe différents modes de MS/MS suivant les informations voulues mais dans notre cas, nous nous focaliserons sur le mode d'acquisition MS/MS en mode automatique. Dans ce mode, le MS effectue un premier balayage de la gamme de masse choisie. Ensuite, une sélection des ions parents est réalisée par le logiciel en fonction des paramètres imposés par l'utilisateur. Puis, les ions parents sont fragmentés dans la cellule de collision et les ions fils sont analysés et détectés. Une fois que tous les ions parents ont été analysés, le MS effectue un nouveau cycle. Les paramètres de sélections d'ions mais aussi de fragmentation pour l'analyse MS/MS sont disponibles dans le Tableau 9.

Nom	Valeurs (unité)	Signification
Spectra rate	Valeurs dépendant des paramètres MS (Hz)	Fréquence d'acquisition des spectres. Ce paramètre permet de définir le temps d'acquisition pour obtenir un spectre. D'une manière générale, plus il y a d'ions parents à sélectionner, plus cette valeur est faible. Toutefois si la valeur est trop faible, les valeurs d'intensités des ions parents risquent d'être peu intenses ainsi que leur spectre MS/MS.
Precursor ions	0-30	Nombre d'ions parents qui sont automatiquement sélectionnés pour la fragmentation lors d'un cycle.
Threshold	0-1 000 000 (Abs.) 0-100(Rel.)	Définit le seuil d'intensité minimal d'un ion précurseur pour qu'il soit sélectionné en tant qu'ion parent. Ce paramètre permet donc d'éviter de fragmenter des ions de faibles intensités qui ne sont pas informatifs. Attention, si ce paramètre est trop élevé, la sélection d'ions d'intérêts risque de ne pas être réalisée.
Active exclusion		Ce paramètre permet d'exclure temporairement des ions dans un cycle. Il est possible d'exclure un ion pour une certaine durée ou après un certain nombre de fragmentation. Ce paramètre permet d'éviter la sélection d'ions de forte intensité qui ont déjà été analysés et permet donc d'analyser des ions de plus faible intensités. Il est aussi possible de créer des listes d'exclusion ou d'inclusion.

Charge states	1-6	Possibilité de sélectionner les ions en fonction de leur état de charge (inclusion ou exclusion). Ce paramètre est intéressant pour l'utilisateur qui connaît les ions d'intérêts.
CID	0-200 (eV)	L'utilisateur peut paramétrer l'énergie de fragmentation CID en fonction de la masse et de l'état de charge des ions parents. Si l'énergie est trop faible, l'ion précurseur risque de ne pas être fragmenter mais, au contraire, si l'énergie est trop forte cela risque de trop fragmenter et engendrer de la perte d'informations.

Tableau 9 Paramètres de sélection et de fragmentation d'ion du maXis 4G avec les valeurs et leurs significations.

Dans cette partie ont été décrits les instruments utilisés pour réaliser l'analyse des modifications post-transcriptionnelles d'ARN ainsi que les paramètres ajustables de ces appareils. Par la suite, les différents paramètres de ces appareils ont été optimisés dans le but de fournir des analyses de la meilleure qualité possible afin de caractériser les modifications de façon précise. Les méthodes d'analyses ont été adaptées pour les deux niveaux de caractérisation car les analytes sont de nature différentes et que les informations à obtenir ne sont pas les mêmes.

### 5.3 Développement de la méthode pour l'étude au premier niveau

L'objectif de l'étude des ARN au premier niveau est d'identifier les nucléotides modifiés présents dans un échantillon. Pour cela, l'ARN est totalement hydrolysé pour obtenir un mélange de nucléosides. Avant de tenter d'analyser des digests d'hydrolyse totale d'ARN à l'aide du couplage CE-MS/MS, il est nécessaire d'optimiser les paramètres du MS afin de pouvoir détecter les modifications d'ARN avec une bonne précision de mesure. L'optimisation des paramètres de détection a été effectuée par infusion d'échantillon d'ARN. Après optimisation de la détection MS, les conditions de séparation électrophorétique ont été mises en place. Dans cette partie, les différents choix effectués pour les paramètres d'analyse des nucléosides sont discutés et justifiés.

#### 5.3.1 Détection de nucléosides

Dans le but d'optimiser la détection des nucléosides, un échantillon d'un extrait d'ARN total de levure (*Saccharomyces cerevisiae*), préalablement dessalé par précipitation, a été totalement hydrolysé selon un protocole classique à l'aide de la nucléase P<sub>1</sub> et de la BAP [8]. Cette digestion a permis d'obtenir un mélange de nucléosides qui a été utilisé pour optimiser les paramètres de détection MS. La détection des nucléosides par MS est réalisée classiquement en mode positif [9]. En effet, les nucléosides sont dépourvus de fonction acide et les bases azotées sont capables d'accepter un proton, ainsi le mode positif est recommandé et a été choisi pour les détecter. Après digestion, l'échantillon est dilué à 0,05 g/L dans un tampon volatile d'acétate d'ammonium (AcONH<sub>4</sub>) 200 mM à pH 4 dans le but de favoriser la protonation des analytes dans le processus ESI. Le système d'électrophorèse capillaire a été utilisé afin d'infuser cet échantillon. Pour effectuer l'infusion, une pression de 5 psi a été

appliquée afin d'obtenir un débit constant d'échantillon et une tension ESI de 1500 V a également été appliquée dans le but de former un spray stable. L'optimisation de la détection a été réalisée en considérant les pics correspondant aux nucléosides canoniques. La gamme de masse utilisée était de 80 à 800 m/z dans le but de détecter les nucléosides modifiés. Après optimisation des paramètres de transfert, le spectre MS ci-dessous a été obtenu (Figure 43).

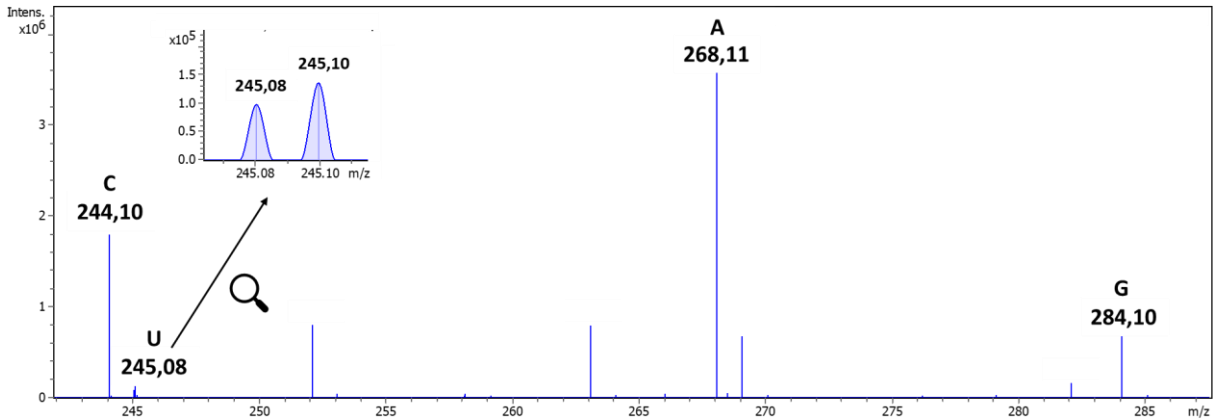


Figure 43 Spectre MS de l'échantillon d'ARN total (levure), digéré en nucléosides, obtenu par infusion sur le maxis 4G en mode positif avec les paramètres de transfert optimisés et zoom.

Sur ce spectre, les pics correspondant aux 4 nucléosides canoniques ont été détectés avec une forte intensité et une bonne précision de masse ( $\pm 0,02$  m/z). De plus, le pouvoir résolutif de cet instrument est suffisamment élevé pour permettre la distinction du premier isotope de U (245,08 m/z) et du deuxième isotope de C (245,10 m/z) avec une résolution d'environ 45 000. Enfin, l'absence des pics correspondant aux nucléotides canoniques démontre l'efficacité de la BAP pour totalement hydrolyser les phosphates terminaux. Néanmoins, l'étude des nucléosides par infusion ne permet pas de distinguer des nucléosides possédant la même masse moléculaire. De plus, la masse d'un pic ne permet pas d'attribuer avec certitude un nucléoside.

De ce fait, dans le but d'améliorer la confiance en l'attribution des pics, il est possible de mettre en œuvre des expériences de fragmentation. En effet, la fragmentation d'un ion précurseur permet de produire des ions fils qui sont souvent spécifiques à une molécule. Ainsi, lorsque les transitions ion parent  $\rightarrow$  ions fils correspondants à la molécule sont détectées, cela permet de fournir une preuve supplémentaire de la présence de cette molécule. Lorsque les nucléosides sont soumis à de la fragmentation CID, c'est en générale la liaison N-glycosidique qui est rompue produisant les fragments ribose et base azotée [10]. Par conséquent, une méthode MS/MS a été mise en place pour produire des ions fils à partir des ions parents par CID. Une énergie de collision trop forte résulte en la formation de fragments trop courts et non-spécifiques. En revanche, une énergie trop faible ne permet pas de fragmenter les nucléosides. L'énergie de collision a été ajustée afin de fragmenter les nucléosides au niveau de la liaison N-glycosidique. Expérimentalement, il a été observé qu'une énergie de collision de 20 eV était un bon compromis pour observer les transitions spécifiques aux quatre nucléosides non-modifiés comme cela est montré sur la Figure 44.

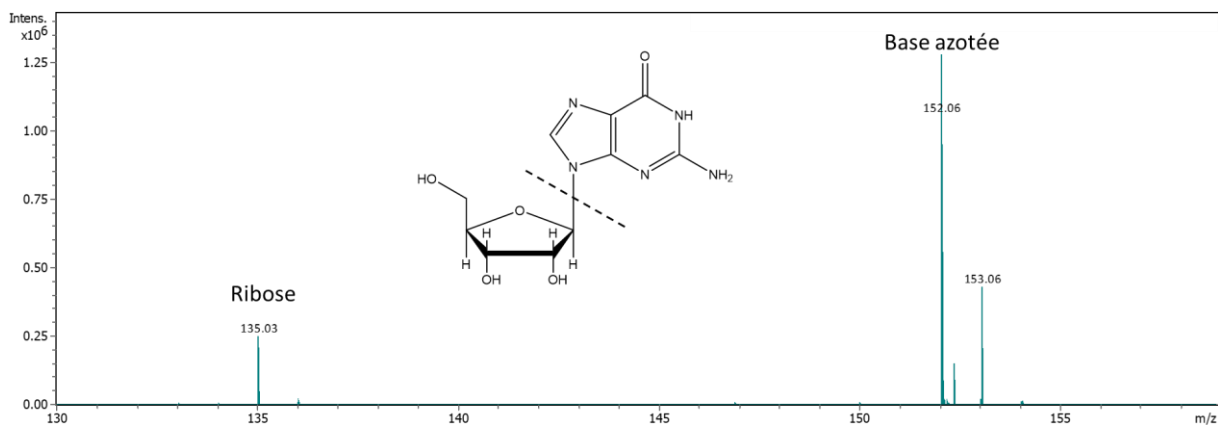


Figure 44 Fragmentation CID de la guanosine : Spectre MS/MS de 284,10 m/z, à 20 eV, montrant les fragments correspondants au ribose (135,03 m/z) et à la base azotée de la guanosine (152,06 m/z) issus de la rupture de la liaison N-glycosidique.

Ainsi, en plus d'apporter une confirmation supplémentaire de la présence du nucléoside, la fragmentation CID des nucléosides modifiés permet également de déterminer si la modification se situe sur le ribose ou sur la base nucléique voir même si ces deux composés portent des modifications. Cette méthode permet donc de discriminer des nucléosides isomériques qui n'ont pas les mêmes voies de fragmentation (par exemple, il est possible de distinguer U avec la transition 245→113 de  $\psi$  avec la transition 245 → 125) [11].

La détection des nucléosides a été réglée pour avoir le maximum d'intensité. La fragmentation CID a été mise en place dans le but de confirmer la présence des nucléosides mais aussi d'obtenir plus d'informations sur la localisation des modifications sur le nucléoside. Cependant, l'exploitation d'un spectre MS issu d'une infusion d'un mélange complexe reste difficile. De plus, la MS/MS ne permet pas de distinguer des isomères qui ont la même voie de fragmentation (par exemple  $m^1G$  et  $m^2G$  qui possèdent tous les deux un méthyle sur leur base azotée mais à des positions différentes). Par conséquent, le couplage d'une méthode séparative en amont du MS est indispensable pour surmonter ces obstacles. La méthode séparative permet de décomplexifier l'échantillon et facilite donc l'interprétation des spectres. Dans ce manuscrit de thèse, la CZE a été couplée à la MS dans le but de d'évaluer la capacité de cette méthode électrophorétique à surmonter ces problématiques.

### 5.3.2 Mise en place de la séparation CE pour l'analyse de nucléosides

Pour mettre en place la méthode de séparation des nucléosides, plusieurs paramètres ont été ajustés comme le BGE, la tension de séparation et aussi le volume d'échantillon injecté.

La composition du BGE est primordiale pour réussir une séparation en CZE. En effet, le choix va dépendre de plusieurs facteurs et surtout de la nature des analytes à séparer. La CZE est une méthode qui permet la séparation d'espèces chargées en solution. Il est donc essentiel que la composition du BGE fasse apparaître une ou plusieurs charges sur les analytes dans le but de les faire migrer. Comme cela a été évoqué dans la première partie de ce manuscrit (cf. 1.2.4 et 1.3.2), les nucléosides sont des espèces faiblement basiques, c'est-à-dire susceptibles de capter un proton. Il est donc nécessaire que le BGE possède un pH

suffisamment acide pour leur conférer une charge. C'est pourquoi le choix du BGE s'est orienté vers l'acide acétique. En effet, l'acide acétique à 10% est le BGE le plus couramment utilisé pour analyser diverses molécules en CZE-MS/MS. Ainsi, ce BGE a été testé pour séparer les nucléosides en CZE.

Ensuite, un test courant a été réalisé avec ce BGE. Ce test permet d'évaluer si le capillaire subit un échauffement par effet Joule. Pour cela, le courant induit par le BGE, en appliquant une tension, est mesuré. Le capillaire de séparation est d'abord rempli de BGE puis une différence de potentiel est appliquée entre les deux électrodes. Si, avec un capillaire de 30  $\mu\text{m}$  i.d. et 100 cm de longueur, la tension appliquée induit un courant supérieur à 10  $\mu\text{A}$ , alors il y a un risque d'échauffement. Avec la valeur de tension la plus élevée (30 kV soit 300 V/cm), le courant électrique mesuré a été de 3,2  $\mu\text{A}$  avec un BGE d'acide acétique 10%. Par conséquent, l'effet Joule est limité dans ces conditions. C'est donc cette tension qui a été sélectionnée pour mettre en œuvre les séparations. Concernant la polarité, avec un BGE acide et un capillaire BFS, il est nécessaire d'appliquer une tension de façon à ce que l'électrode de sortie soit la cathode. De cette manière, le flux électroosmotique est dirigé vers la sortie du capillaire et permet d'obtenir un débit de BGE stable permettant d'assurer le processus ESI.

Une fois ces paramètres sélectionnés, la séparation des nucléosides a été réalisée en CE-MS/MS. Le même digest d'ARN total (*S. cerevisiae*) a été utilisé pour évaluer la séparation. Une première analyse a été réalisée en injectant un volume de 7,5 nL, correspondant à 1 % de la longueur totale du capillaire. Cependant, des pics étêtés ont été obtenus. Ainsi, une autre séparation a été effectuée avec un volume d'injection inférieur (5 nL) et une meilleure qualité de séparation a été obtenue, comme le montre l'électrophérogramme de la Figure 45.

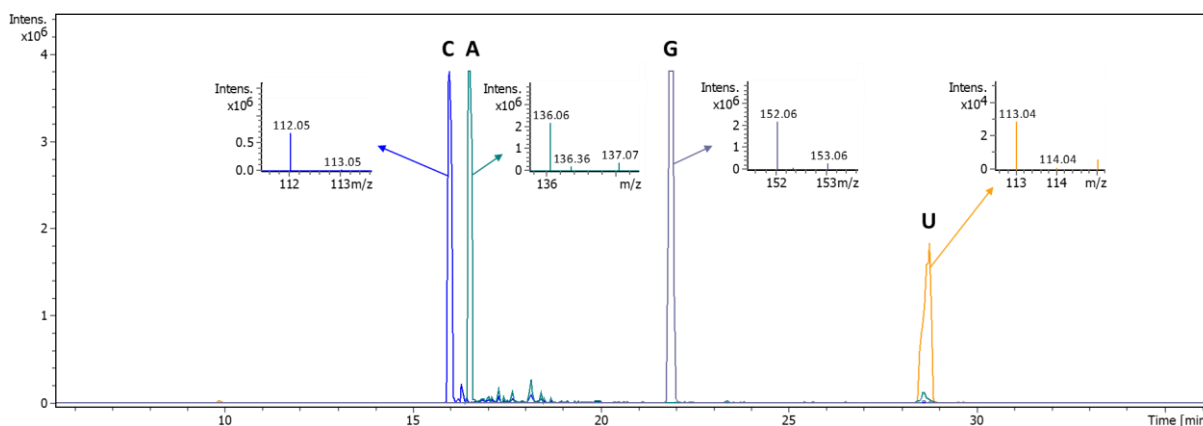


Figure 45 Électrophérogramme, obtenu par CZE-MS/MS, d'un extrait d'ARN total digéré à la nucléase  $P_1$  et la BAP. Les pics correspondant aux quatre nucléosides non-modifiés sont montrés ainsi que les spectres MS/MS associés.

Avec la méthode mise en place, la détection des quatre nucléosides canoniques monochargés a été achevée en moins de 30 minutes. De plus, les nucléosides ont été séparés avec une grande efficacité et une excellente résolution. L'attribution des pics a été réalisée avec une tolérance de  $\pm 0,02$  m/z. Les spectres de fragmentation, associés à ces pics, ont permis d'apporter une preuve additionnelle pour l'attribution des pics. Dans le cas des nucléosides non-modifiés, c'est la masse de la base azotée qui permet de valider l'attribution d'un pic (avec également une tolérance de  $\pm 0,02$  m/z), puisque le ribose est commun aux quatre nucléosides canoniques.

Ainsi, la méthode d'analyse CZE-MS/MS a été développée dans le but d'étudier les nucléosides. Grâce à ce couplage, l'identification des nucléosides est basée sur les trois critères suivants :

- Le temps de migration.
- La masse du composé.
- Les masses des ions fils.

Les performances de séparation de la CE associées à la détection par MS/MS sont donc tout à fait adaptées pour l'analyse et l'identification de nucléosides. La méthode permet donc d'identifier les modifications présentes dans un ARN. Cependant, cette méthode n'est pas adaptée pour déterminer les positions des modifications sur les structures primaires des ARN. Ainsi, une méthode CZE-MS/MS a été développée dans le but de fournir ces informations.

## 5.4 Développement de la méthode pour l'étude au second niveau

L'étude des modifications au second niveau permet de situer les nucléotides modifiés sur les séquences d'ARN. Pour cela, les ARN sont d'abord digérés à l'aide d'enzymes spécifiques pour produire un mélange d'oligonucléotides. Les oligonucléotides sont des molécules plus complexes que les nucléosides du fait de leur taille mais aussi de leurs propriétés chimiques. De plus, des informations de séquençage sont nécessaires pour l'étude à ce niveau. Par conséquent, la méthode CZE-MS/MS doit être adaptée dans le but d'analyser les oligonucléotides. De la même manière que précédemment, les paramètres de transfert ont d'abord été optimisés afin de détecter ces composés et la fragmentation MS/MS a été réglée dans le but d'obtenir des informations de séquençage. Ensuite, la séparation d'oligonucléotides par CE a été mise en place. Les choix des conditions d'analyses sont détaillés et justifiés.

### 5.4.1 Détection d'oligonucléotides

L'optimisation des paramètres de détection des oligonucléotides a d'abord été réalisée à l'aide d'un petit ADN synthétique. Cette molécule, qui est chimiquement très proche de l'ARN, a permis dans un premier temps de déterminer des paramètres tels que la polarité de détection, la concentration et la nature de la solution tampon. La séquence de cet ADN était CCCCT avec une masse moléculaire de 1398,3 Da.

Dans la littérature, il est décrit que, d'une façon générale, la détection des acides nucléiques en mode négatif est plus sensible que le mode positif du fait du caractère acide de ces composés [12]. Ainsi, dans le but de favoriser la détection des composés en mode négatif, une base faible est ajoutée dans le tampon d'échantillon. La triéthylamine (TEA) est une base classiquement utilisée pour détecter les acides nucléiques par MS en mode négatif [13]. Ainsi, l'échantillon d'ADN a d'abord été dilué à 0,075 g/L dans une solution de MeOH/TEA 0,1% 50/50 (v/v). Ensuite, l'échantillon a été infusé via le système d'électrophorèse capillaire. L'infusion a aussi été réalisée en appliquant une pression de 5 psi et une tension ESI de 1500 V. Les paramètres de transfert ont été réglés afin d'obtenir le maximum de signal.



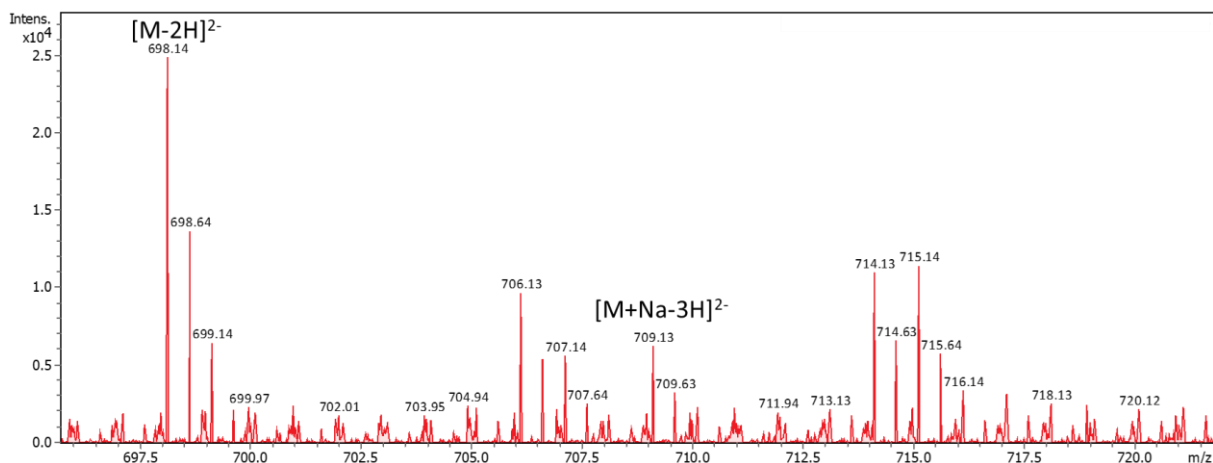


Figure 46 Spectre de masse de l'ADN synthétique obtenu sur le maXis 4G en mode négatif. ADN à 0,075 g/L dans MeOH/TEA 0,1 % 50/50 (v/v).

Après optimisation des paramètres de transfert, on observe très bien les  $m/z$  correspondant à l'ADN comme le montre la Figure 46. Toutefois, les adduits sodiques sont également très présents.

Par la suite, la détection en mode positif a été évaluée. De la même manière, afin d'améliorer l'intensité du signal, l'ADN a été dilué à 0,075 g/L dans un tampon  $\text{AcONH}_4$  200 mM pH 4. Un léger ajustement des paramètres de transfert a été réalisé et les spectres ont également montrés la présence d'ADN mais aussi de ses adduits alcalins (voir Figure 47).

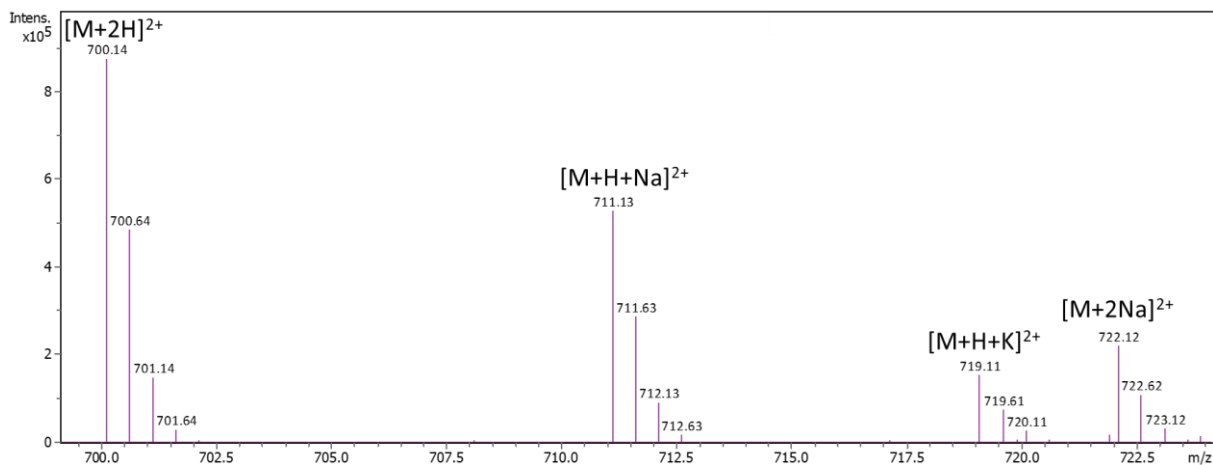


Figure 47 Spectre de masse de l'ADN synthétique obtenu sur le maXis 4G en mode positif. ADN à 0,075 g/L dans  $\text{AcONH}_4$  200 mM pH 4.

Cependant, les spectres obtenus en mode positif montrent une intensité de signal plus grande que pour le mode négatif. Cela peut s'expliquer par la nature de l'ADN qui est constitué principalement de nucléotides basiques (C) mais également par la meilleure capacité d'ionisation de la solution tampon.

Pour optimiser les paramètres de détection, une infusion d'un digest d'un ARN synthétique a été réalisée. Cet ARN 80-mer de séquence connue (GGGCUCAGACCACUCUAGACGGUGUAAAAUCUCUAGCAGUGGCGCCCGAACAGGGACUUGAA

AGUGAAAGUACAGCCC), qui ne porte pas de modification, a préalablement été digéré à la RNase T<sub>1</sub> afin d'obtenir les petits oligonucléotides. Une digestion *in silico* de cet ARN a été réalisée afin de simuler les fragments produits par une digestion T<sub>1</sub>. Les produits de digestion et leur masse sont disponibles dans le Tableau 10. Les données de ce tableau ont été obtenues par le calculateur MongoOligo.

Séquence	Masse (Da)
Gp	364,07
Gp	364,07
Gp	364,07
CUCAGp	1608,22
ACCACUCUAGp	3182,43
ACGp	997,15
Gp	364,07
UGp	669,08
UAAAAUCUCUAGp	4171,55
CAGp	997,15
UGp	669,08
Gp	364,07
CGp	668,10
CCCGp	1278,18
AACAGp	1655,25
Gp	364,07
Gp	364,07
ACUUGp	1609,20
AAAGp	1350,21
UGp	669,08
AAAGp	1350,21
UAACAGp	1961,28
CCC	853,17

Tableau 10 Tableau regroupant les séquences et les masses moléculaires des oligonucléotides issus de l'ARN 80-mer après digestion à la RNase T<sub>1</sub>.

Par la suite, le digest a été dilué dans un mélange MeOH/TEA 0,1 % 50/50 pour la détection en mode négatif et dans AcONH<sub>4</sub> 200 mM pH 4 pour la détection en mode positif. L'optimisation de la détection des oligonucléotides a été réalisée en infusant ces digests, via un capillaire BFS. Les paramètres ont été légèrement ajustés par rapport à la détection de l'ADN. Les spectres MS correspondants aux échantillons analysés en mode négatif et positif sont disponibles Figure 48 et Figure 49.

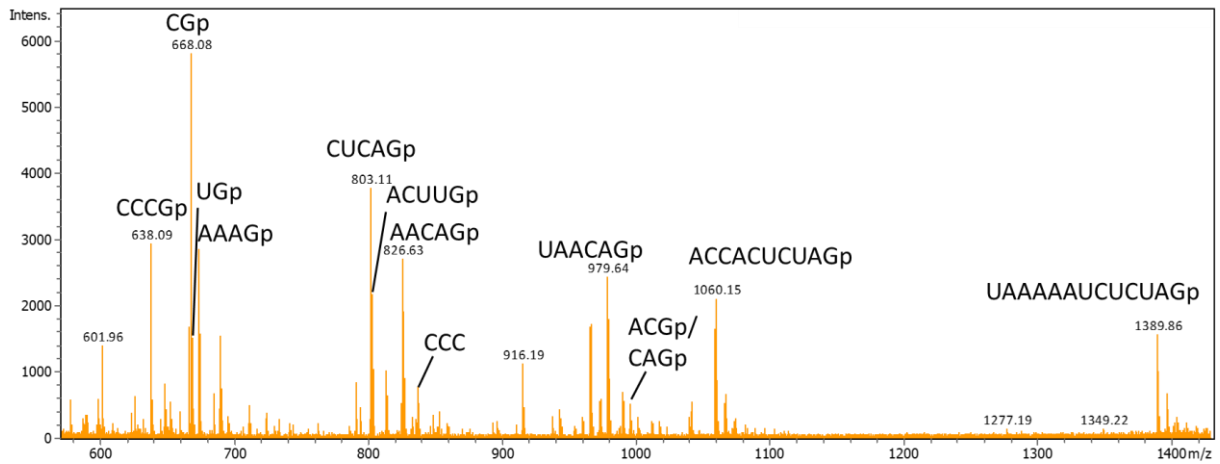


Figure 48 Spectre de masse de l'ARN 80 synthétique digéré à la RNase T<sub>1</sub> obtenu sur le maXis 4G en mode négatif, Digest d'ARN à 0,025 g/L dans MeOH/TEA 0,1 % 50/50 (v/v).

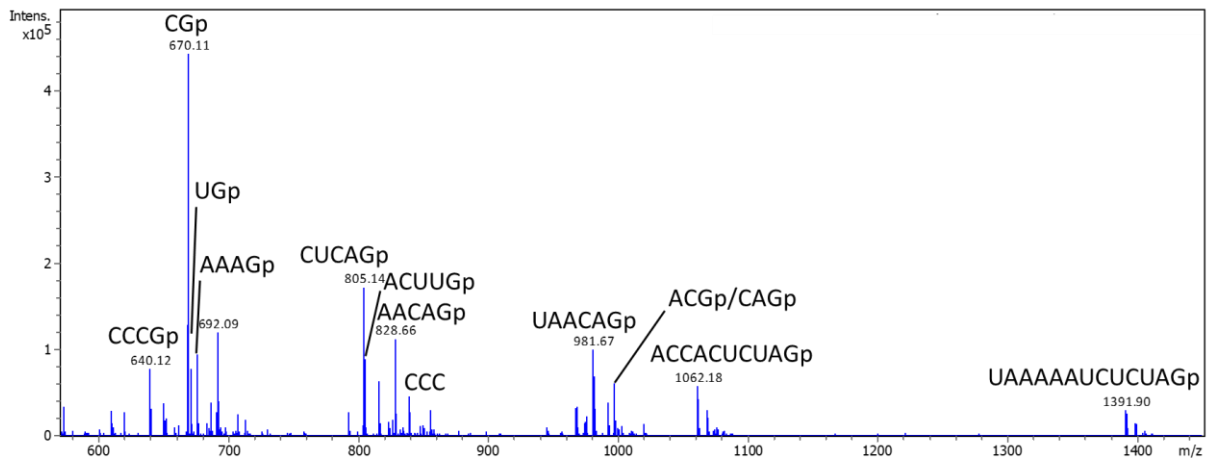


Figure 49 Spectre de masse de l'ARN 80 synthétique digéré à la RNase T<sub>1</sub> obtenu sur le maXis 4G en mode positif, Digest d'ARN à 0,025 g/L dans AcNH<sub>4</sub> 200 mM pH 4.

Sur les deux spectres, tous les oligonucléotides simulés ont été détectés avec  $\pm 0,02$  m/z. Les états de charge des composés variaient entre 1 et 3 et plus rarement 4. Des pics correspondant à des adduits de sodium ont néanmoins été détectés. Peu de différences ont été observées entre les deux polarités de détection mis à part l'intensité du signal. Ainsi, la détection en mode positif a été sélectionnée pour analyser les oligonucléotides puisque cela permet de réaliser des analyses dans les mêmes conditions que les nucléosides.

Par la suite, une méthode MS/MS a été mise en place dans le but de séquencer ces oligonucléotides. Selon le même principe que pour la fragmentation des nucléosides, il s'agit de trouver une énergie de collision ni trop forte ni trop faible. Cependant, dans le cas des oligonucléotides, l'objectif est d'obtenir des informations de séquence. Ainsi, les ruptures des liaisons qui permettent d'obtenir des fragments utiles à la reconstitution de la séquence sont nécessaires. En CID, ce sont généralement les liaisons les plus faibles qui se rompent en premier. Dans une chaîne d'ARN, les liaisons les plus faibles se trouvent au niveau des groupements phosphodiester. En effet, les phosphodiester sont riches en atomes électronégatifs, fragilisant de ce fait les liaisons covalentes de ces groupements. C'est pourquoi les énergies de collision ont été optimisées pour casser ces liaisons et produire les

ions permettant le séquençage (les ions c et y selon les règles de McLuckey). Expérimentalement, il a été observé que l'énergie de fragmentation devait être ajustée suivant la taille de l'oligonucléotide et également sa charge. D'une manière générale, plus la masse de l'oligonucléotide est grande, plus l'énergie de collision doit être grande et plus l'état de charge est élevé, plus cette énergie doit être faible. L'énergie de collision a donc été ajustée en fonction du  $m/z$  des ions précurseurs avec des valeurs variant de 6 à 48 eV. Le spectre de la Figure 50 montre un exemple de séquençage *de Novo*.

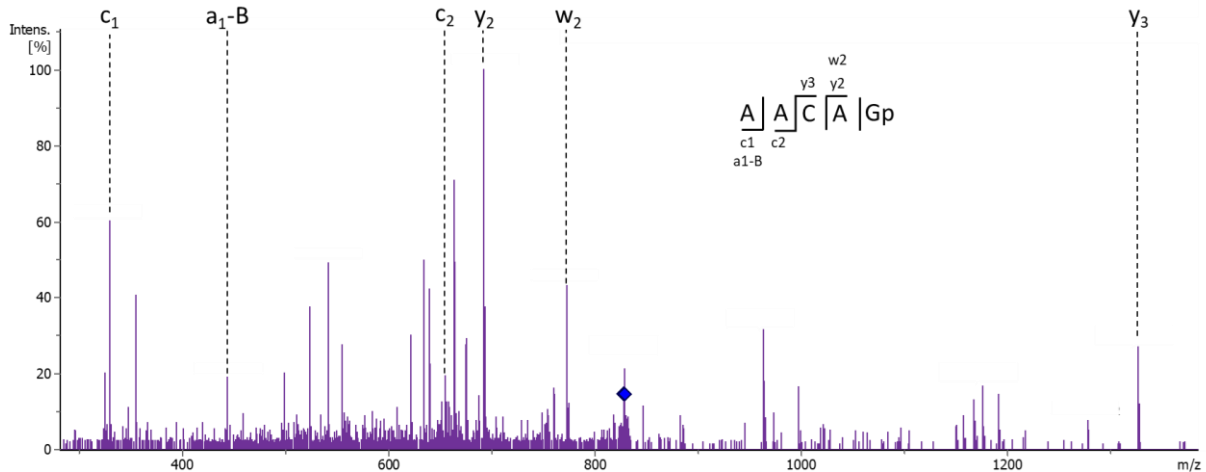


Figure 50 Spectre CID de 828,64  $m/z$  montrant le séquençage de l'oligonucléotide AACAGp.

Ainsi, la méthode de détection des oligonucléotides a été optimisée en mode positif. Dans le but de faciliter l'interprétation des spectres, les conditions d'analyses CE ont été sélectionnées afin d'obtenir une séparation des analytes en amont du MS.

#### 5.4.2 Mise en place de la séparation CE pour l'analyse d'oligonucléotides

Dans un premier temps, la séparation électrophorétique des oligonucléotides a été évaluée dans les mêmes conditions que les nucléosides. L'échantillon d'ARN synthétique, préalablement digéré à la RNase T<sub>1</sub>, a été utilisé pour évaluer la méthode. Une fois l'analyse effectuée, les masses des oligonucléotides ont été extraites. L'analyse a été réalisée en 90 min. Cependant, malgré un temps d'acquisition long, certains produits de digests n'ont pas été détectés en analyse CZE-MS. Il a été observé que les oligonucléotides manquants sont les plus grands et ceux qui possèdent un grand nombre de U. Il a également été observé que plus les oligonucléotides possèdent un grand nombre de U, plus leur temps de migration est retardé. Ainsi, la même séparation a été effectuée sauf qu'une pression de 3 psi a été appliquée à partir de 80 min dans le but d'observer les composés non-détectés. L'électrophérogramme ci-dessous a été obtenu (Figure 51).

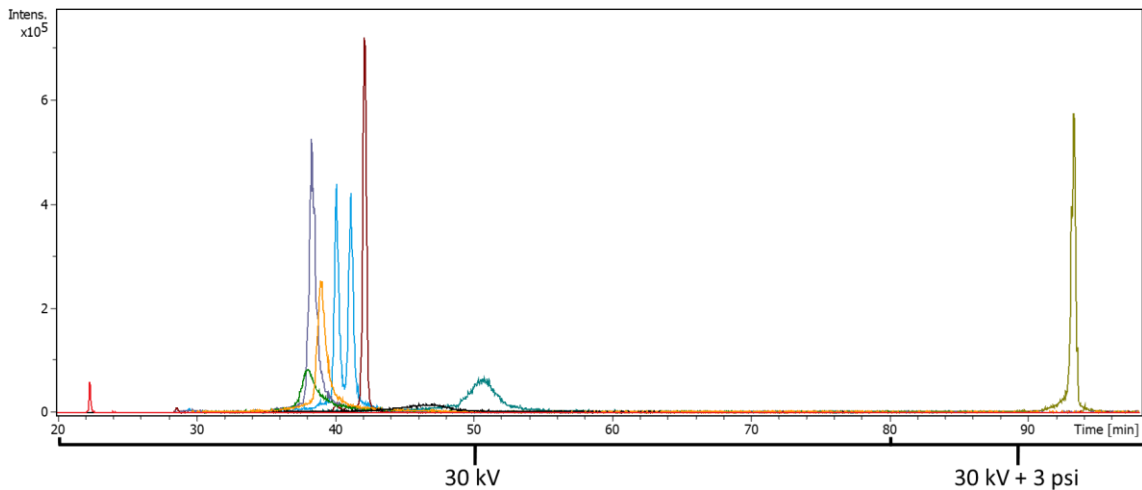


Figure 51 Électrophérogramme, obtenu par CZE-MS/MS, de l'ARN synthétique digéré à la RNase T<sub>1</sub>. L'analyse a été réalisée en appliquant une tension de 30 kV pendant 80 min puis 30 kV et 3 psi. Les rapports m/z simulés ont été extraits.

Une séparation des oligonucléotides a été obtenue dans ces conditions. Des pics de très bonne efficacité (jusqu'à 250 000 plateaux théoriques) ont été obtenus pour les oligonucléotides de petites et moyennes tailles. De plus, l'excellent pouvoir résolutif de la CE a permis de séparer des espèces isobariques. En effet, lors de la digestion de l'ARN 80-mer avec de la RNase T<sub>1</sub>, les oligonucléotides ACGp et CAGp sont produits. La MS, sans technique séparative en amont, ne permet pas de discriminer ces deux composés puisqu'ils ont la même masse et que leurs spectres MS/MS sont confondus. Cependant, ici, la CE avec les conditions mises en place ont permis de séparer ces deux composés. Ensuite, l'attribution des pics a été réalisée sans ambiguïté grâce aux spectres MS/MS correspondant comme le montre l'exemple sur la figure ci-dessous.

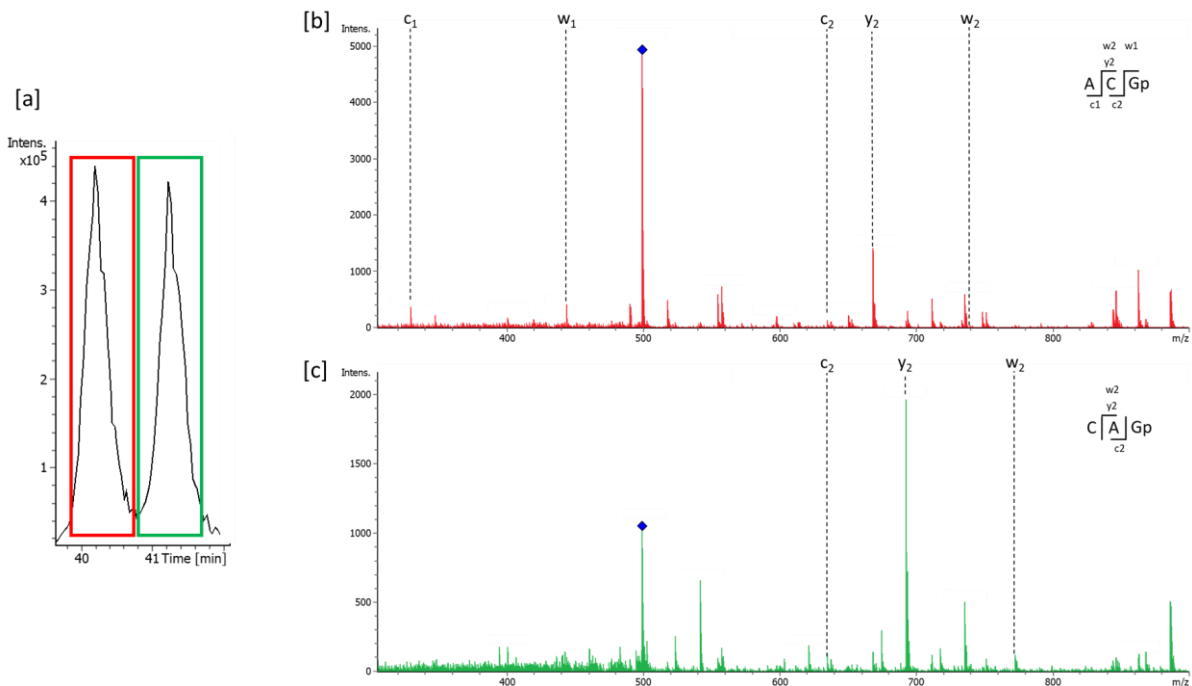


Figure 52 Exemple de séparation et d'identification d'isomères avec [a] électrophérogramme de 499,58 m/z de l'analyse du digest de RNase T<sub>1</sub> de l'ARN 80-mer correspondant aux produits ACGp et CAGp (état de charge 2+) [b] spectre de masse CID de 499,58 m/z @ 40,2 min permettant de reconstruire la séquence ACGp et [c] spectre de masse CID de 499,58 m/z @ 41,2 min permettant de reconstruire la séquence CAGp.

Cette possibilité de pouvoir séparer, au préalable, des espèces isobariques avant la détection MS est un avantage majeur car en digérant les ARN, il est fréquent d'observer des oligonucléotides de même masse moléculaire mais de séquences différentes. La séparation de ces espèces permet donc d'augmenter les recouvrements de séquence. De plus, grâce au couplage CE-MS/MS, les expériences de MS/MS ont permis de séquencer et d'attribuer avec certitude les pics électrophorographiques.

Bien que les conditions mises en place ont permis d'obtenir une très bonne efficacité de séparation sur les oligonucléotide de séquences courtes, cette efficacité diminue de manière significative suivant la longueur de l'analyte. De plus, les oligonucléotides acides n'ont pas été détectés en appliquant uniquement une tension. Cependant, grâce à l'application d'une faible pression additionnelle, le pic correspondant à UGp a été détecté à 93 min mais les oligonucléotides les plus long n'ont toujours pas été détectés même après 120 min d'analyse. L'absence des pics correspondant aux oligonucléotides de grande taille peut être due à une efficacité de pic médiocre qui empêcherait la détection ou bien à un mauvais sens de migration de ces composés dans le capillaire de séparation qui serait dû au caractère trop acide de ces composés.

Ainsi, dans le but d'améliorer les séparations des oligonucléotides, il serait possible de changer le BGE. Dans le but de détecter tous les oligonucléotides, il est possible de jouer sur le pH. Puisqu'un mauvais sens de migration est suspecté, le pH peut être diminué pour favoriser l'apparition d'une charge positive sur ces composés et ainsi changer le sens de migration. Néanmoins, le BGE utilisé est déjà acide (pH 2,2). Si le pH du BGE est encore diminué, cela risque de dégrader les digests mais aussi de fragiliser la silice du capillaire. Une autre possibilité est de changer la nature du BGE et notamment de passer à un BGE en milieu basique. Cela permettrait alors d'induire une charge négative aux oligonucléotides. En passant d'un BGE de pH acide à basique, il faudrait donc inverser la polarité de la tension appliquée pour que les composés migrent vers la sortie du capillaire. Il faudrait également veiller à ce que le BGE n'induisse pas d'effet Joule et s'assurer que le flux électroosmotique génère un débit suffisamment élevé pour former un spray ESI stable. Si ce flux électroosmotique n'était pas suffisant, il serait possible de le moduler en greffant la paroi interne du capillaire. Il existe différents types de greffage pouvant augmenter ou réduire le flux et également changer le sens [14]. Néanmoins, le changement de BGE pour améliorer la séparation n'a pas été évalué dans cette thèse.

Une autre solution a été envisagée pour détecter tous les digests. Dans le protocole, la RNase T<sub>1</sub> est une endonucléase qui clive les ARN au niveau des G, produisant des fragments terminant par un phosphate 3'. Ces phosphates terminaux sont à l'origine du ralentissement de la mobilité des oligonucléotides dans des conditions de séparation acide. En effet, dans ces conditions d'analyses, les phosphates 3' apportent des charges négatives aux oligonucléotides et ces charges diminuent alors la mobilité des digests. De plus, ces phosphates n'apportent aucune information pour le séquençage et peuvent réduire l'abondance d'ions qui fournissent des informations de séquence [15]. Ainsi, l'hydrolyse des phosphates terminaux a été évaluée pour améliorer la qualité de la séparation.

Le protocole de digestion des ARN a alors été adapté pour générer des oligonucléotides sans phosphates terminaux. Afin de déphosphoryler les oligonucléotides, la BAP a été utilisée. Pour comparer la séparation d'oligonucléotides avec et sans phosphates 3', un ARN de transfert purifié a été utilisé (ARNt<sup>Phe</sup> *S. Cerevisiae*). L'ARN a été digéré avec la RNase T<sub>1</sub> et la RNase T<sub>1</sub> + BAP puis les digests obtenus ont ensuite été analysés dans les mêmes conditions que les nucléosides et en appliquant une pression de 3 psi à partir de 100 min. La Figure 53 montre les séparations des digests de l'ARN issus de ces deux digestions.

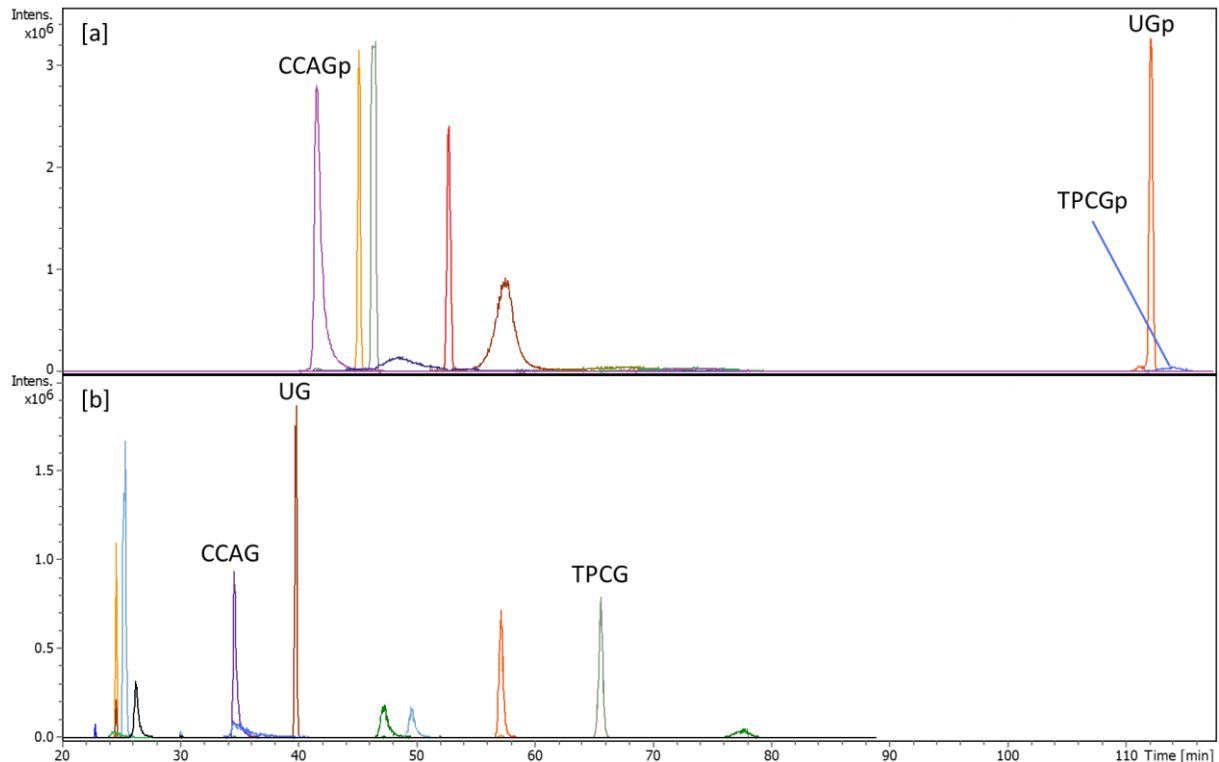


Figure 53 Électrophérogrammes, obtenus par CZE-MS/MS, de l'ARNt<sup>Phe</sup> digéré à la RNase T<sub>1</sub> avec [a] sans déphosphorylation et [b] déphosphorylé avec la BAP.

En comparant les électrophérogrammes, on observe que l'étape de déphosphorylation est grandement bénéfique pour la séparation des oligonucléotides car cela permet de réduire les temps de migration, d'améliorer l'efficacité des pics et la résolution mais aussi de détecter des oligonucléotides sans appliquer de pression.

En effet, sur l'analyse du digest non-déphosphorylé, le pic correspondant à Gp a été détecté à 52,8 min, alors que sur l'autre analyse, le pic G a été détecté bien plus tôt, à 22,9 min. L'application d'une pression, en plus de la tension, a été nécessaire pour détecter UGp (à 112,2 min) mais grâce à l'hydrolyse du phosphate 3', le pic UG a été observé à 39.8 min sans l'assistance d'une pression. Cela a également été remarqué pour le composé [m<sup>5</sup>U][ψ]CG qui a été observé, uniquement grâce au champ électrique, à 65,3 min alors que [m<sup>5</sup>U][ψ]CGp a été détecté à 114,1 min avec 3 psi. En plus de réduire les temps de migration, la qualité des pics obtenus pour les oligonucléotides en 3'-OH a également été améliorée. La figure ci-dessous présente un exemple de gain en efficacité de séparation.

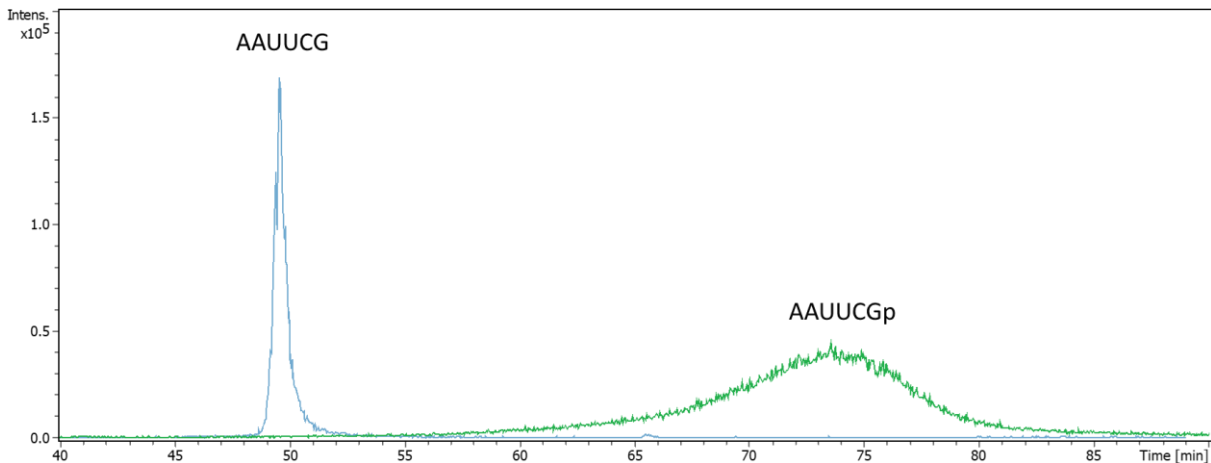


Figure 54 Électrophérogramme montrant le gain en efficacité de séparation obtenu grâce à l'hydrolyse des phosphates 3' des digests. En bleu, le digest déphosphorylé AAUUCG (930,15 m/z) et en vert AAUUCGp (970,13 m/z), digest produit par la digestion l'ARNt<sup>Phe</sup> à la RNase T<sub>1</sub>.

Cette figure montre non seulement une augmentation de la mobilité des digests déphosphorylés, mais surtout une amélioration significative de l'efficacité de pic. La finesse des pics permet d'obtenir un signal plus intense en MS et donc une plus grande sensibilité de détection (sur l'exemple, l'intensité du signal est triplée). Cette augmentation de la sensibilité, en plus de permettre la détection de faibles quantités de digest, permet d'obtenir des spectres MS/MS de meilleures qualités et donc un séquençage facilité.

Néanmoins, malgré tous les avantages apportés par la déphosphorylation sur la qualité des séparations, les temps d'analyses sont restés relativement long (80 min). De plus, les oligonucléotides les plus longs ou trop acide n'ont pas été détectés ou bien l'efficacité de pic reste médiocre malgré l'hydrolyse des phosphates 3' comme cela est illustré sur la Figure 55.

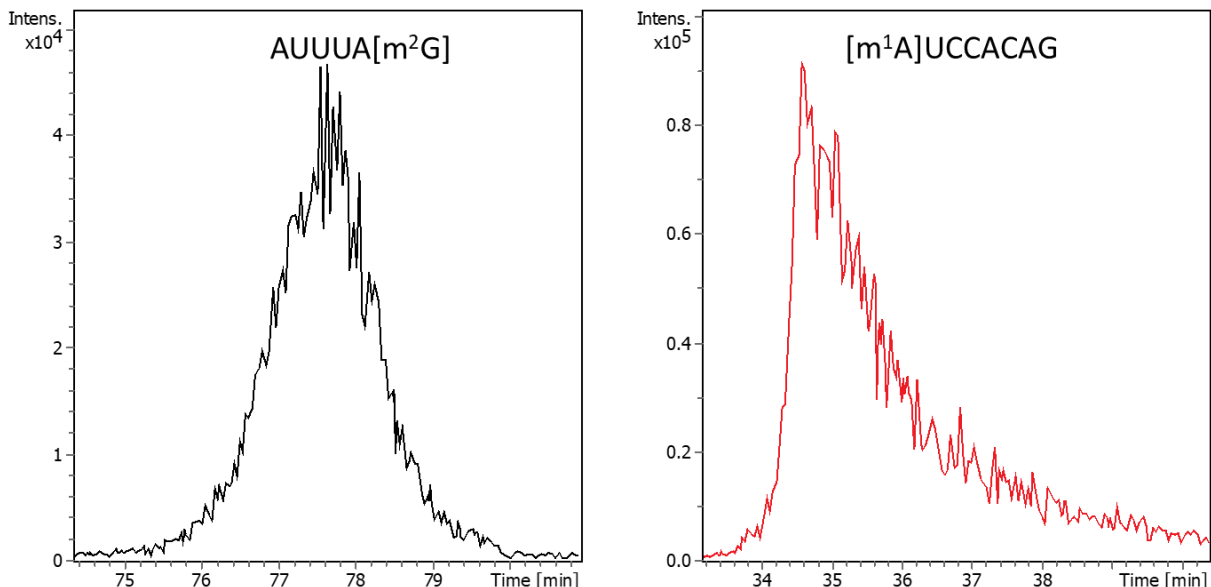


Figure 55 Pics correspondant à l'oligonucléotide AUUUA[m<sup>2</sup>G] (à gauche) et à [m<sup>1</sup>A]UCCACAG (à droite), montrant une efficacité de pic moyenne malgré l'étape d'hydrolyse des phosphates terminaux.

Bien qu'une amélioration de l'efficacité de pic ait été observée pour ces oligonucléotides, l'hydrolyse du phosphate terminal ne permet pas d'obtenir des pics aussi



efficaces que les oligonucléotides courts. Ceci peut s'expliquer simplement par le fait que plus l'oligonucléotide est long, plus il possède de groupements phosphodiester. Par conséquent, l'hydrolyse du seul phosphate terminal ne permet pas d'améliorer autant l'efficacité de pic que pour les petits oligonucléotides. La mauvaise détection de ces composés est un inconvénient pour notre étude car si des fragments ne sont pas détectés, cela diminue les recouvrements de séquence mais surtout la localisation de certains nucléotides modifiés ne pourrait être réalisée.

Par conséquent, dans le but d'améliorer le recouvrement de séquence, une autre stratégie a été mise en place. Une digestion des ARN est également effectuée à l'aide d'une autre enzyme spécifique : la RNase A. Cette endonucléase permet d'hydrolyser spécifiquement les pyrimidines et permet ainsi d'obtenir des oligonucléotides dont les séquences peuvent partiellement se recouvrir avec ceux de la RNase T<sub>1</sub>. Avec la RNase A, les oligonucléotides formées contiennent également une terminaison 3' phosphate. Ainsi, de la même façon qu'avec la RNase T<sub>1</sub>, les digestions à la RNase A ont été réalisées avec la BAP afin de retirer les phosphates terminaux et d'améliorer la séparation en CE. La RNase A présentent des avantages similaires à la RNase T<sub>1</sub> avec une grande spécificité et aussi l'absence de « miscleavage ». La grande différence réside dans la taille des oligonucléotides générés. En effet, cette enzyme est spécifique à deux nucléotides : C et U. Ainsi, les fragments formés sont en général plus petits. Cela présente donc un avantage pour la séparation électrophorétique comme cela a été discuté auparavant. Des qualités séparatives équivalentes ont été observées en analysant des digests de RNase A. Cependant, la probabilité que des oligonucléotides plus petits soient spécifiques à une séquence est beaucoup plus faible. De plus, la formation d'oligonucléotides de séquence redondante est plus fréquente. Toutefois, la double digestion individuelle d'un ARNt à l'aide de ces deux enzymes permet d'améliorer significativement le recouvrement de séquence comme cela est décrit plus tard dans le chapitre 7.

Ainsi, le développement de la méthode d'analyse CZE-MS/MS a été effectuée pour l'analyse des modifications post-transcriptionnelles d'ARN à deux niveaux. Les paramètres tels que la nature du BGE, la valeur de la tension de séparation et le volume d'injection ont été ajustés. Les conditions mises en place ont permis d'apporter les informations désirées telles que l'identification de nucléosides et le séquençage d'oligonucléotides. Les séparations électrophorétiques des espèces déphosphorylés ont montré d'excellentes performances en termes d'efficacité et de résolution malgré un temps d'analyse relativement long pour les oligonucléotides. Quant à la détection par MS et MS/MS, elle a permis la détection précise des analytes. De plus, la méthode mise en place a permis de réaliser les analyses des deux niveaux dans des conditions identiques. Il s'agit d'un réel avantage car classiquement, pour effectuer l'analyse des nucléosides et des oligonucléotides, il est nécessaire d'adapter les méthodes chromatographiques. Ainsi, les analyses des modifications d'ARN peuvent être effectuées à l'aide du même système.

Par la suite, il a été nécessaire de se pencher sur l'interprétation des données obtenues par CZE-MS/MS. Cette étape permet d'extraire les informations des spectres afin de pouvoir identifier les nucléotides modifiés et localiser ces modifications sur les structures primaires des ARN.

## 5.5 Références bibliographiques

1. Crain, P.F., *Preparation and Enzymatic-Hydrolysis of DNA and Rna for Mass-Spectrometry*. Methods in Enzymology, 1990. **193**: p. 782-790.
2. Neusüss, C., et al., *Coupling of Capillary Electromigration Techniques to Mass Spectrometry*. 2018. p. 293-311.
3. Hocker, O., C. Montealegre, and C. Neususs, *Characterization of a nanoflow sheath liquid interface and comparison to a sheath liquid and a sheathless porous-tip interface for CE-ESI-MS in positive and negative ionization*. Analytical and Bioanalytical Chemistry, 2018. **410**(21): p. 5265-5275.
4. Maxwell, E.J. and D.D.Y. Chen, *Twenty years of interface development for capillary electrophoresis-electrospray ionization-mass spectrometry*. Analytica Chimica Acta, 2008. **627**(1): p. 25-33.
5. Smith, R.D., et al., *Capillary zone electrophoresis mass-spectrometry using an electrospray ionization interface*. Analytical Chemistry, 1988. **60**(5): p. 436-441.
6. Moini, M., S.M. Demars, and H.L. Huang, *Analysis of carbonic anhydrase in human red blood cells using capillary electrophoresis/electrospray ionization-mass spectrometry*. Analytical Chemistry, 2002. **74**(15): p. 3772-3776.
7. Schmitt-Kopplin, P. and M. Frommberger, *Capillary electrophoresis - mass spectrometry: 15 years of developments and applications*. Electrophoresis, 2003. **24**(22-23): p. 3837-3867.
8. Thüring, K., et al., *Analysis of RNA modifications by liquid chromatography–tandem mass spectrometry*. Methods, 2016. **107**: p. 48-56.
9. Esmans, E.L., et al., *Liquid chromatography mass spectrometry in nucleoside, nucleotide and modified nucleotide characterization*. Journal of Chromatography A, 1998. **794**(1-2): p. 109-127.
10. Pomerantz, S.C. and J.A. McCloskey, *Analysis of Rna Hydrolyzates by Liquid-Chromatography Mass-Spectrometry*. Methods in Enzymology, 1990. **193**: p. 796-824.
11. Su, D., et al., *Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry*. Nature Protocols, 2014. **9**(4): p. 828-841.
12. Thomas, B. and A.V. Akoulitchev, *Mass spectrometry of RNA*. Trends in Biochemical Sciences, 2006. **31**(3): p. 173-181.
13. Huber, C.G. and H. Oberacher, *Analysis of nucleic acids by on-line liquid chromatography-mass spectrometry*. Mass Spectrometry Reviews, 2001. **20**(5): p. 310-343.
14. Horvath, J. and V. Dolnik, *Polymer wall coatings for capillary electrophoresis*. Electrophoresis, 2001. **22**(4): p. 644-655.
15. Krivos, K.L., B. Addepalli, and P.A. Limbach, *Removal of 3'-phosphate group by bacterial alkaline phosphatase improves oligonucleotide sequence coverage of RNase digestion products analyzed by collision-induced dissociation mass spectrometry*. Rapid Communications in Mass Spectrometry, 2011. **25**(23): p. 3609-3616.

# CHAPITRE 6. TRAITEMENT DES DONNEES

## 6.1 Introduction

L'interprétation des données permet de « convertir » les spectres obtenus en informations utiles pour caractériser les modifications. Étant donné que le protocole établi pour étudier les modifications d'ARN est sur deux niveaux (nucléoside et oligonucléotide), le traitement des données à ces deux niveaux est différent.

Dans les cas d'analyses des nucléosides, il s'agit d'identifier quels sont les nucléotides modifiés présents dans un ARN. Ainsi, la détection de l'ion parent et de l'ion fils spécifique au nucléoside permet, dans la grande majorité des cas, d'affirmer la présence du nucléoside. L'exploitation des données pour interpréter les analyses d'oligonucléotides est plus complexe car il s'agit d'obtenir des informations de séquençage. Des outils de bioinformatique ainsi que des moteurs de recherche, tels que RNAModMapper, Pytheas, RoboOligo, Ariadne ou plus récemment NucleicAcidSearchEngine [1-5], ont été développés pour épauler les utilisateurs en exploitant de façon automatique les analyses LC-MS/MS d'oligonucléotides. Ces outils ont été créés dans le but d'automatiser les exploitations de spectres MS et MS/MS afin d'éviter les interprétations manuelles fastidieuses et chronophages. De plus, ces logiciels permettent non seulement la reconstruction des séquences des ARN mais également la localisation des nucléotides modifiés sur la structure primaire. Toutefois, ces outils ont été développés uniquement pour des analyses réalisées en détection négative. Or, la méthode d'analyse d'oligonucléotides par CZE-MS/MS a été développée avec une détection positive. Ainsi, l'utilisation de ces logiciels est alors impossible dans le cadre de notre méthode. De plus, bien que ces logiciels aient démontré leur utilité dans le traitement des spectres, il est recommandé de toujours vérifier les résultats fournis par ces logiciels [1]. Par conséquent, durant cette thèse, toutes les données d'analyses ont été interprétées manuellement.

Dans ce chapitre, l'interprétation pour l'analyse des nucléosides puis des oligonucléotides sont décrites. Ici, c'est une recherche ciblée des informations qui a été réalisée. Les séquences d'ARN et leurs modifications post-transcriptionnelles (type et position sur la séquence) ont été recherchées dans la banque de données Modomics <https://iimcb.genesilico.pl/modomics/> [6]. Ensuite, une simulation des digestions a été réalisée à l'aide du calculateur MongoOligo, développé par l'équipe du Prof. Limbach, disponible à l'adresse : <https://mods.rna.albany.edu/masspec/Mongo-Oligo>. Enfin, les digests, générés *in silico*, sont recherchés dans les spectres bruts puis les résultats des analyses sont confrontés afin de déterminer la nature et la position des nucléotides modifiés sur la structure primaire de l'ARN.

Une procédure expérimentale détaillée pour interpréter les données a été rédigée au laboratoire et est proposée en annexe (cf. annexe).

## 6.2 Identification des nucléosides

Le but de l'analyse des nucléosides est d'identifier les modifications présentes dans un ARN. Ainsi, dans un premier temps, les informations sur les modifications contenues dans l'ARN sont recherchées dans la base de données Modomics. Ensuite les résultats de l'analyse CZE-MS/MS sont comparés avec les informations de la banque de données.

### 6.2.1 Simulation de l'hydrolyse totale

Afin de simuler les masses des nucléosides portant des modifications, la séquence de l'ARN est d'abord recherchée dans la base de données Modomics. Pour cela, le type d'ARN (ARNt, ARNr, etc.), le sous-type (Cys, Lys, etc. pour l'ARNt) et l'organisme (*E. Coli*, *H. Sapiens*, etc.) sont renseignés. Ensuite, le moteur de recherche affiche les séquences connus des ARN correspondant aux critères de recherche (voir Figure 56).

RNA sequences with modifications

RNA type: tRNA Subtype: Trp Organism: Saccharomyces cerevisiae

Visualise as: Modomics symbols Display Draw Modification Profile

PDF CSV Print

Search:

Type	AA type	Anticodon	Organism	Organelle	accept	D-domain	anticodon domain	variable region
					10	20	30  40	50
tRNA	Trp	CCA	Saccharomyces cerevisiae	cytosolic	012345678901234567-890--	1234567890123456789012345	eeeeeeeeeeeeeeeeeeee	67890
tRNA	Trp	!CA	Saccharomyces cerevisiae	mitochondrial	-GAAGCGGUKLCUCAAD--	#GD--AGAGCPPCGABUBCAAAPCGAAGG	-----7DUGCAG	
					-AAGGAUAUAGUUUAAD--	GGD--AAACAGUUGAPU!	CAHAPCAUCA-----	-U-UAGGA

Showing 1 to 2 of 2 entries

Figure 56 Résultats de la recherche de séquence de l'ARNt<sup>Trp</sup><sub>CCA</sub> (*S. Cerevisiae*) sur la base de données Modomics.

Dans les séquences, les nucléotides modifiés sont représentés par le code à une lettre et sont en bleus. La banque de données fournit également des informations utiles pour l'exploitation MS. Les informations disponibles sur les nucléosides modifiés dans Modomics sont :

- Formule brute.
- Masse monoisotopique.
- Masse moyenne.
- Masse de l'espèce protonée  $[M+H]^+$ .
- Masse d'un fragment spécifique issu du précurseur  $[M+H]^+$ .
- Temps d'élution normalisé (HPLC) [7].
- Ordre d'élution/caractéristiques (HPLC) [7].

Parmi les informations disponibles, celles qui ont été utilisées dans ces travaux pour la recherche et l'identification des nucléosides ont été le rapport  $m/z$  du nucléoside chargé 1+ et celui de l'ion fils caractéristique. En général, la masse du fragment obtenu par CID correspond à la masse du ribose ou de la base nucléique plus la masse de la modification, suivant la position de la modification sur le nucléotide.

Les informations sur les nucléosides ont été utilisées pour rechercher les nucléosides modifiés sur les analyses CZE-MS/MS.

## 6.2.2 Exploitation des données d'analyse de nucléosides

Une fois l'analyse CZE-MS/MS des produits d'hydrolyses de l'ARN effectuée, les nucléosides sont recherchés grâce aux données spectrales. Pour cela, les masses des ions parents sont extraites avec  $\pm 0,02 m/z$ .

Lors d'une recherche de nucléosides modifiés, il est courant d'obtenir plusieurs pics pour une masse extraite. En effet, un nucléotide peut porter une même modification mais à une position différente. Ainsi, il est très important de vérifier la présence des ions fils afin d'attribuer correctement le pic à une modification. Cependant, certains nucléosides, portant un même type de modification mais à des positions différentes, ne peuvent être discriminés par MS/MS. À défaut de pouvoir réaliser des expériences MS<sup>n</sup>, ou bien de la mobilité ionique, il est possible d'assigner ces pics avec leurs temps de migration. Toutefois, cette solution n'est possible que si le standard est disponible et il est également nécessaire de s'assurer que les temps de migration soient reproductibles entre plusieurs analyses.

Ainsi, des tests de reproductibilités ont été réalisés afin de déterminer l'écart-type des temps de migration. Pour ce faire, les temps de migration des nucléosides canoniques ont été étudiés. Plusieurs échantillons de différents ARNt totalement hydrolysés ont été testés sur plusieurs jours par analyse CZE-MS/MS. L'évaluation a été réalisée avec  $n=8$  pour chaque nucléosides. Les résultats sont affichés sur la Figure 57.

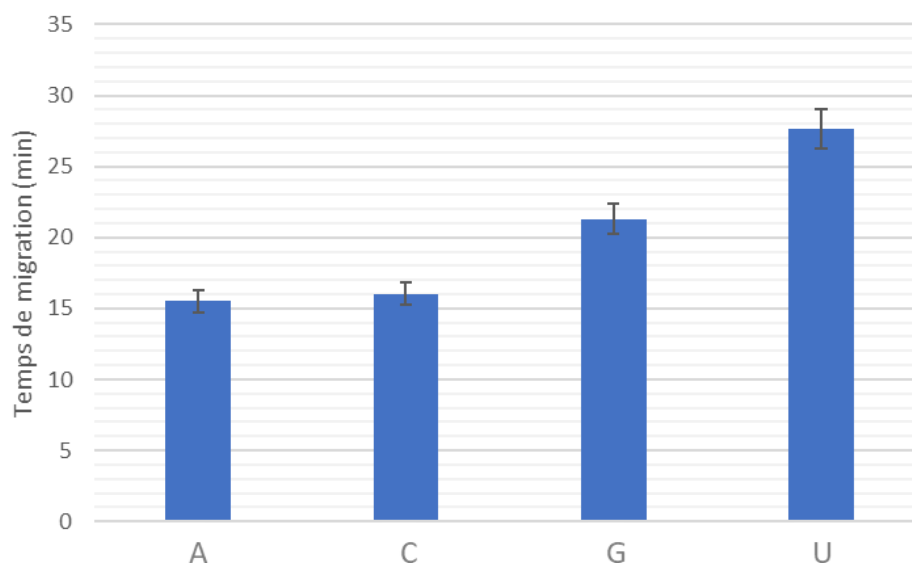


Figure 57 Histogramme montrant les temps de migration moyens des quatre nucléosides canoniques et leur écart-type.

L'écart-type minimum observé pour les temps de migration des nucléosides non-modifiés était de 0,8 min pour A et le maximum était de 2,1 min pour U. Cela correspond à un écart-type relatif minimum de 4,9 et un maximum de 7,6 %. Il a été observé que plus le temps de migration du composé est long, plus l'écart-type du temps de migration était grand. La reproductibilité, surtout en termes de temps de migration, est un des points faibles de la CE [8]. Dans cette thèse, les écarts observés étaient acceptables au vu de l'identification des pics correspondant à une même masse. Cependant, afin d'améliorer la reproductibilité, il aurait

été possible d'analyser simultanément des standards dont les mobilités électrophorétiques sont connues. Ensuite, les temps de migrations des nucléosides sont simplement convertis en mobilités électrophorétiques effectives. De cette façon, les écarts des temps de migration dus aux fluctuations du flux électroosmotique sont corrigés, permettant d'obtenir des résultats plus reproductibles [9].

Ainsi, les modifications des ARN sont identifiées principalement à l'aide des ions parents et des ions fils. Toutefois, dans certains cas, les temps de migrations sont nécessaires afin d'identifier certaines modifications. Par la suite, l'interprétation des données issus des analyses d'oligonucléotides a été développée.

### 6.3 Séquençage d'oligonucléotides et localisation de modifications

L'analyse des oligonucléotides permet le séquençage des ARN mais aussi la localisation des nucléotides modifiés par une stratégie dite « bottom-up ». Pour cela, les séquences des ARN ainsi que les positions des nucléotides modifiés sont d'abord recherchées dans la même base de données (Modomics). Par la suite, les séquences des ARN sont digérées *in silico*, à l'aide du calculateur MongoOligo, afin d'obtenir les séquences ainsi que les masses des oligonucléotides produits par les digestions spécifiques. Les données théoriques sont ensuite comparées avec les résultats des analyses des digests spécifiques par CZE-MS/MS. L'exploitation des analyses est identique pour les oligonucléotides obtenus par la RNase T<sub>1</sub> et la RNase A. L'interprétation des spectres MS/MS bruts a été effectuée manuellement afin d'éviter les erreurs et pour obtenir une localisation sans ambiguïté des modifications. L'interprétation de ces données a permis de séquencer les ARN et, dans le même temps, de déterminer les emplacements des nucléotides modifiés sur la séquence des ARN.

#### 6.3.1 Simulation de la digestion

Afin de simuler les masses des oligonucléotides obtenus par digestions spécifiques, la séquence de l'ARN est d'abord recherchée dans la base de données Modomics de la même manière que pour l'étude des nucléosides. La digestion de la séquence est ensuite simulée à l'aide du calculateur de masse MongoOligo. MongoOligo prend en compte les nucléotides modifiés et utilise également le code à une lettre pour les modifications. Ce simulateur fournit diverses informations utiles pour l'exploitation des données spectrales d'oligonucléotides telles que :

- Séquences (Digestion RNase T<sub>1</sub> et RNase A).
- Masses moléculaires.
- Rapports  $m/z$  ESI (mode positif ou négatif et adduits).

Il faut noter que les séquences déterminées par le simulateur pour les digestions aux RNases se terminent en 3'p. Ainsi, puisque dans le protocole développé dans cette thèse les oligonucléotides sont déphosphorylés à la BAP, les masses sont recalculées en tenant compte de la soustraction de la masse du phosphate terminal.

Il faut également noter que ce calculateur simule seulement l'hydrolyse des ARN au niveau des nucléotides non-modifiés, soit dans notre cas G et C/U. Or, les endonucléases, utilisées pour accomplir les digestions spécifiques, hydrolysent également les liaisons phosphodiester au niveau de certains nucléotides modifiés. Le tableau ci-dessous liste certains nucléotides modifiés qui sont reconnus par ces enzymes.

RNase T <sub>1</sub>	RNase A
I	D
m <sup>2</sup> G	ψ
	m <sup>5</sup> U
	m <sup>5</sup> C
	s <sup>4</sup> U

Tableau 11 Nucléotides modifiés reconnus par la RNase T<sub>1</sub> et la RNase A. D'après les observations réalisées durant ces travaux de thèse.

Il est très important de vérifier les séquences des oligonucléotides simulées par MongoOligo car, en recherchant uniquement les masses des oligonucléotides digérés au niveau des nucléotides non-modifiés, le recouvrement de séquence risque d'être plus faible. De plus, la localisation des modifications en question ne peut être réalisée si cette vérification est omise.

Afin d'attribuer un pic à un oligonucléotide, malgré la bonne précision de mesure du spectromètre, le seul rapport  $m/z$  est insuffisant. Pour attribuer correctement un pic, il est nécessaire de vérifier sa séquence en exploitant les spectres de fragmentation. Pour cela, il est possible d'interpréter les spectres par séquençage *de Novo* ou bien d'utiliser le simulateur MongoOligo. En effet, ce calculateur permet également de simuler des fragmentations CID d'oligonucléotides. MongoOligo calcul les rapports  $m/z$  des ions  $y$  et  $c$  mais également  $a-B$  et  $w$  en mode positif ou négatif, permettant de reconstruire les séquences. Ce simulateur fournit également d'autres informations pour des expériences CID notamment en simulant les fragments internes et aussi la perte de base. L'oligonucléotide intact peut perdre une base azotée sous forme d'anion ( $B^-$ ) ou une base neutre ( $B+H$ ). Quant aux fragments internes, ils résultent d'un double clivage du squelette. Ces deux dernières informations ne sont pas utiles dans le séquençage mais permettent d'apporter une preuve supplémentaire de la présence d'un oligonucléotide.

Une fois les digestions simulées, l'exploitation des analyses CZE-MS/MS des digests de RNase T<sub>1</sub> et RNase A est réalisée. Les résultats théoriques sont confrontés avec les données expérimentales.

### 6.3.2 Exploitation des données d'analyse d'oligonucléotides

Les échantillons d'ARN, obtenus par digestions spécifiques, sont analysés par CZE-MS/MS et le séquençage ainsi que la localisation des modifications sont effectués à l'aide des spectres. Dans ce but, les masses des ions parents, correspondant aux oligonucléotides simulés par le calculateur, sont extraites des électrophérogrammes avec une tolérance de  $\pm 0,02 m/z$ .

Après extraction des pics, une vérification des rapports  $m/z$  de chaque oligonucléotides est effectuée. En effet, lors de l'extraction, plusieurs pics correspondant à une masse peuvent apparaître. Par exemple, lorsque le pic correspondant à UAG chargé 2+ est extrait, il est possible d'obtenir deux pics : celui correspondant à UAG (460,09  $m/z$  chargé 2+) et celui correspondant au deuxième isotope de CAG (460,10  $m/z$  chargé 2+). Il est donc impératif de vérifier, dans un premier temps, l'état de charge ainsi que la masse monoisotopique de chaque pics de l'électrophérogramme.

Plus important encore, il est nécessaire de vérifier la séquence de l'oligonucléotide car une masse détectée peut correspondre à un grand nombre de séquences. Les ions fils de l'oligonucléotide, simulés précédemment, sont alors recherchés dans le spectre MS/MS brut avec également une tolérance de 0,02  $m/z$ . Dans la très grande majorité du temps, le séquençage en suivant les séries d'ions  $y$  et  $c$  est suffisant pour reconstruire la structure primaire de l'oligonucléotide. Néanmoins, dans certains cas, il est nécessaire de suivre les séries d'ions  $a$ - $B$  et  $w$  afin de confirmer la séquence. Dans le cas de dinucléotides, le séquençage n'a pas lieu d'être réalisé. En effet, pour les dinucléotides, la spécificité des endonucléases permet de générer des fragments qui se terminent forcément par un nucléotide hydrolysable suivant l'enzyme utilisée (G ou C/U). Ainsi, la masse du dinucléotide est suffisante pour affirmer sa séquence.

La localisation des modifications dans les séquences est réalisée simultanément avec le même jeu de données. Lorsqu'un nucléotide est modifié, cela génère un changement de sa masse (excepté pour l'isomérisation de U en  $\psi$ ). Ce changement de masse permet alors de repérer facilement la position du nucléotide modifié dans la séquence sur le spectre MS/MS.

### 6.3.3 Détermination du recouvrement de séquence

Une fois le travail d'interprétation des données terminé, il s'agit de déterminer le recouvrement de séquence de l'ARN. Pour cela, les résultats des interprétations des données obtenus pour les digests de RNase T<sub>1</sub> et RNase A sont combinés. L'utilisation de deux endonucléases de spécificité différentes augmente automatiquement le recouvrement car cela génère des oligonucléotides de séquences différentes.

Dans ces travaux, uniquement les oligonucléotides dont les séquences ont pu être totalement confirmées par MS/MS ont été considérées pour le séquençage. Il faut également noter que les RNases produisent des oligonucléotides (notamment des mono et des dinucléotides) qui possèdent des séquences redondantes et donc non spécifiques aux séquences des ARN. Ces digests non-spécifiques diminuent alors les recouvrements de séquence. Ainsi, seulement les oligonucléotides dont la séquence est unique ont été considérés pour le recouvrement de séquence.

L'interprétation des données, réalisée dans ces travaux de thèse, a été décrite dans ce chapitre. Bien que des logiciels existes pour interpréter informatiquement les spectres, aucun de ces logiciels n'est adaptés pour des analyses réalisées en polarité positive. Ainsi, les données obtenues par analyses CZE-MS/MS ont été interprétées manuellement. Ici, la



méthode d'interprétation décrite est utilisée dans le cadre de l'étude d'un ARN dont la séquence et les nucléotides modifiés sont connus.

L'interprétation manuelle décrite ici présente plusieurs défauts :

- Chronophage.
- Adapté uniquement pour des ARN dont la séquence et les modifications sont connues.

La méthode décrite ici est extrêmement chronophage car il s'agit de simuler les digestions pour les trois enzymes puis de rechercher manuellement toutes les informations dans les spectres. Cette méthode n'est donc pas adaptée dans le cadre de l'interprétation d'analyse d'ARN possédant un grand nombre de nucléotides.

De plus, la méthode permet uniquement la recherche ciblée des informations. Ainsi, elle n'est pas adaptée pour interpréter des données avec un ARN de séquence inconnue ou bien lorsque l'ARN contient un ou plusieurs nucléotides modifiés inconnus. Néanmoins, dans ce cas, il est possible de séquencer manuellement les oligonucléotides par *de Novo* puis de vérifier les résultats obtenus en utilisant le simulateur MongoOligo.

Dans cette partie du manuscrit, le développement de la méthode d'analyse par CZE-MS/MS ainsi que l'interprétation des données ont été détaillées. La méthode d'analyse ainsi que l'interprétation des données ont été adaptées pour caractériser les modifications post-transcriptionnelles des ARN sur les deux niveaux d'analyse. La CZE-MS/MS a permis de caractériser les digests d'ARN grâce aux excellentes séparations fournies par la CE ainsi que la précision et la possibilité d'obtenir des informations structurales par MS et MS/MS. Dans la partie suivante est décrite l'application de cette méthode développée pour l'analyse des modifications post-transcriptionnelles des ARN.

## 6.4 Références bibliographiques

1. 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**: p. 128-138.
2. D'Ascenzo, L., A. Popova, and J.R. Williamson, *Pytheas: A Software to Map RNA Modifications via Tandem Mass Spectrometry*. *Biophysical Journal*, 2019. **116**(3, Supplement 1): p. 287a.
3. 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 Research*, 2015. **43**(10).
4. Nakayama, H., et al., *Ariadne: a database search engine for identification and chemical analysis of RNA using tandem mass spectrometry data*. *Nucleic Acids Research*, 2009. **37**(6).
5. Wein, S., et al., *A computational platform for high-throughput analysis of RNA sequences and modifications by mass spectrometry*. *Nature Communications*, 2020. **11**(1): p. 926.
6. Boccaletto, P., et al., *MODOMICS: a database of RNA modification pathways. 2017 update*. *Nucleic Acids Research*, 2018. **46**(D1): p. D303-D307.

7. Kellner, S., et al., *Profiling of RNA modifications by multiplexed stable isotope labelling*. Chemical Communications, 2014. **50**(26): p. 3516-3518.
8. Buscher, J.M., et al., *Cross-Platform Comparison of Methods for Quantitative Metabolomics of Primary Metabolism*. Analytical Chemistry, 2009. **81**(6): p. 2135-2143.
9. Gonzalez-Ruiz, V., et al., *ROMANCE: A new software tool to improve data robustness and feature identification in CE-MS metabolomics*. Electrophoresis, 2018. **39**(9-10): p. 1222-1232.



## **PARTIE III : APPLICATIONS ET RESULTATS**

Dans cette dernière partie, la méthode précédemment développée a été appliquée pour caractériser des modifications post-transcriptionnelles d'ARN afin de démontrer les bénéfices du couplage CE-MS/MS. Dans un premier temps, la méthode a été appliquée à l'analyse d'ARN dont les séquences et les modifications sont connues afin d'avoir un point de comparaison. Ces travaux ont permis de vérifier la capacité de la méthode mise en place à fournir les informations utiles pour caractériser les modifications. Ensuite, la méthode a été appliquée pour l'exploration des modifications avec l'analyse de différents ARN issus d'un extrait provenant d'*E. coli*.

## CHAPITRE 7. PREUVE DE CONCEPT

### 7.1 Introduction

Dans la partie précédente, un protocole d'analyse complet pour la caractérisation des modifications d'ARN a été établi. La méthode d'analyse, basée sur le couplage CE-MS/MS, a été développée dans le but de fournir des informations sur les modifications grâce aux deux niveaux d'analyses. Enfin, l'exploitation des résultats a également été mise en place dans le but d'extraire les informations utiles à la caractérisation des modifications.

Dans les travaux détaillés de cette partie, l'objectif a été de prouver la possibilité de caractériser les modifications avec la méthode développée précédemment. Afin de démontrer l'efficacité de la méthode CE-MS/MS pour étudier les modifications post-transcriptionnelles, quatre ARN de transfert modèles, issus de *S. Cerevisiae*, ont été analysés. Les séquences ainsi que les informations concernant les modifications de ces ARNt sont totalement décrites dans la base de données Modomics. Ainsi, ces ARN ont été utilisés en tant que témoins afin de confirmer la pertinence de la méthode développée pour explorer les modifications des ARN.

Les protocoles de préparation des échantillons ont été adaptés pour l'analyse CE-MS/MS. Dans un premier temps, les ARNt ont été totalement hydrolysés à l'aide de la nucléase P<sub>1</sub> puis les nucléotides ont été hydrolysés en nucléosides grâce à la BAP. Les hydrolysats ont ensuite été analysés en moins de 30 min par la méthode CE-MS/MS. Puis l'exploitation des résultats a été réalisée manuellement en recherchant les nucléosides modifiés théoriquement présents.

Ensuite, les ARNt ont été individuellement digérés à la RNase T<sub>1</sub> et la RNase A afin de produire des oligonucléotides. Le phosphate terminal des oligonucléotides a également été hydrolysé avec la BAP dans le but d'obtenir une bonne qualité de séparation des digests. Une fois les digestions achevées, les digests ont été analysés par CE-MS/MS dans les mêmes conditions de séparation que les nucléosides. Les analyses ont ensuite été interprétées manuellement afin de séquencer les oligonucléotides et de déterminer les positions des nucléotides modifiés. Les résultats issus des deux digestions ont été combinés pour chaque ARN et ont permis d'obtenir de très bons recouvrements de séquence. De plus, plusieurs modifications ont pu être localisées sur les oligonucléotides.

Les résultats expérimentaux ont ainsi été comparés avec la banque de données. La comparaison a montré la cohérence des résultats obtenus et ainsi démontré que le couplage CE-MS/MS est une méthode viable pour étudier les modifications.

Les résultats présentés dans cette partie ont fait l'objet d'un article dans le journal scientifique *Analytical Chemistry* (ACS publications)

<https://doi.org/10.1021/acs.analchem.0c01345>

## 7.2 Publication

### **Characterization of post-transcriptional RNA modifications by sheathless capillary electrophoresis – high resolution mass spectrometry**

Antony Lechner<sup>1\*</sup>, Philippe Wolff<sup>2,3\*</sup>, Emmanuelle Leize-Wagner<sup>1</sup>, Yannis-Nicolas François<sup>1</sup>

<sup>1</sup> Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS) UMR 7140 (Unistra-CNRS), Université de Strasbourg, France.

<sup>2</sup> Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR 9002, F-67000 Strasbourg, France.

<sup>3</sup> Plateforme protéomique Strasbourg Esplanade, CNRS, FRC 1589, F-67000 Strasbourg, France.

\* These authors contributed equally to this work

#### **ABSTRACT:**

Since last decade there is a growing interest of RNA modifications analysis. High performance liquid chromatography-tandem mass spectrometry coupling (HPLC-MS/MS) is classically used to characterize post-transcriptional modifications of ribonucleic acids (RNAs). Here we proposed a novel and simple workflow based on capillary zone electrophoresis-tandem mass spectrometry (CE-MS/MS), in positive mode, to characterize RNA modifications at nucleoside and oligonucleotide levels. A first total digestion of purified RNA, prior CE-MS/MS analysis, enables to identify nucleoside modifications. Then, using a bottom-up approach, sequencing of RNAs and mapping of modifications were performed. Sequence coverages from 68% to 97% were obtained for four tRNAs. Furthermore, unambiguous identification and mapping of several modifications were achieved.

## Introduction

All the classes of RNAs are post-transcriptionally modified. RNA modifications increase the structural and functional diversity of RNA and as a consequence influence RNA structural stability, RNA base pairing and protein recognition. There is poor information on RNA modifications. To increase the knowledge of the community, it is important to use sensitive approaches. Several analytical methods were developed to study RNA modifications. Among them, RNA-seq techniques are the most common approach to locate RNA modifications. However, this method can be limited for the characterization of certain modifications due to blockage mechanism [1, 2]. In parallel, mass spectrometry (MS) has increasingly become a valuable technique to investigate modifications because it provides specificity, sensitivity and may also perform structural elucidations [3]. In the last decades, strategies using enzymatic digestions, prior MS detection, were established to study RNA [4]. However, despite the impressive last improvements of MS in terms of resolution and sensitivity, MS often needs to be hyphenated to separation methods in order to perform deeper characterization of biomolecules. Since the report of electrospray ionization (ESI), HPLC-ESI-MS/MS has become rapidly an important and popular tool for the investigation of nucleic acids modifications. Indeed, this tool can be used not only to characterize modifications but also to quantify them [1, 5-8]. It permitted the discovery of many unknown RNA modifications which, at this day, represent more than 160 natural chemical modifications reported in the literature [9]. Capillary zone electrophoresis (CZE) is a complementary method to liquid chromatography. In CZE, analytes are separated under an electrical field which offers different selectivity. In addition, CZE has major advantages as low sample consumption, fast and effective separations. CE equipped with optical detection were abundantly reported to study RNA at different levels [10-13]. However, no structural information can be extracted from this type of detector and/or standards should be used to identify modifications [14-16]. Hyphenation of CE-MS was a long time challenging due to the necessity to maintain two independent voltages (separation voltage and electrospray voltage). Since its first report in the late 80's by Olivares *et al.* [17], CE coupled with MS became of high interest for the analysis of biologics [18]. Concerning RNA analysis, Feng *et al* developed, in 2007, the pressure assisted electrokinetic injection technique to measure oligonucleotides and their adducts [19]. In 2016, Khan *et al*



reported the use of CE-MS to directly characterize modifications of microRNAs in cancer serum [20]. CZE coupled to MS was also used to estimate rate and equilibrium constant of the conformational dynamics of DNA G-Quadruplex [21]. More recently, Demellenne et al. performed CZE-MS analysis in order to characterize therapeutic oligonucleotides [22]. Different geometries of interface were developed since the first CE-MS hyphenation [23]. In 1997, Moini *et al* developed a sheathless interface [24]. The design of this interface has the ability to generate a nanoESI and provides an excellent sensitivity [25]. Indeed, several articles reported the use of this sheathless interface in many fields of application as metabolomic [26], proteomic [27] or glycomic [28, 29]. RNA modifications are chemical changes that generally lead to a mass shift and/or charge shift. Recently, Yu *et al* reported the first study about the quantification of eight RNA modified nucleotides by multiple reaction monitoring (MRM) using CZE-ESI-MS/MS [30]. In this work, a simple workflow based on CZE-ESI-MS/MS was developed in order to characterize RNA post-transcriptional modifications at two different levels using the same CZE experimental conditions. Although RNA are classically detected in negative ion mode, in the present study, detection of analytes were achieved in positive ion mode. In a first approach, we performed a total hydrolysis of RNA to screen modifications at the nucleoside level, then, a bottom-up approach was developed with the aim of sequencing RNA and mapping the modifications at the oligonucleotide level. The implemented strategy permitted to highlight the modifications present on the RNAs but also to locate them on the primary structure. For this study, four purified and desalted tRNA from *S. cerevisiae* (tRNA<sup>AspGUC</sup>; tRNA<sup>AlaAGC</sup>; tRNA<sup>TrpCCA</sup> and tRNA<sup>PheGAA</sup>) were selected as model molecules. These tRNAs contain about 75 nucleic acids and their sequences and modifications (type and location) are well-known and already described in the literature [9, 31].

## Materials and Methods

**Chemicals and Reagents.** Chemicals used were of analytical grade or high purity. Water used to prepare samples solutions or buffers was obtained using a Direct-Q®3 UV-R water purification system (Merck Millipore, Amsterdam, The Netherlands). Bacterial Alkaline Phosphatase (BAP) (150 U/μL), RNase T1 and RNase A were purchased from ThermoFisher

Scientific (Marietta, OH, USA). RNase P1 from *Penicillium citrinum* was purchased from Sigma–Aldrich (Saint Louis, MO, USA)

**Sample preparation.** Purified tRNAs from *S. cerevisiae* (tRNA<sup>AspGUC</sup>; tRNA<sup>AlaAGC</sup>, tRNA<sup>TrpCCA</sup> and tRNA<sup>PheGAA</sup>) are homemade prepared and desalted by three successive precipitations with ammonium acetate. For nucleoside analysis, tRNAs were diluted to a concentration of 2 µg/µL in H<sub>2</sub>O. Digestion was carried out in this order: 7 µL H<sub>2</sub>O; 1 µL buffer P1 10X (2mM ZnCl<sub>2</sub>, 250mM NH<sub>4</sub>OAc, pH5.0); 1 µL of tRNA and 1 µL of P1 (0.5 U/µL). Mixture was incubated at 37°C for 2 hours. After P1 digestion, 10 µL of BAP (1.5 U/µL in 100 mM NH<sub>4</sub>OAc) was added to the mixture and incubated at 37°C for 4 hours. The mixture was diluted 2 times in NH<sub>4</sub>OAc 200 mM pH 4 and analyzed by CE-MS/MS. For oligonucleotides analysis, tRNAs were diluted to a concentration of 0.75 µg/µL. For RNase T1 digestion, a solution of 100 µL of RNase T1 (1/1000 in NH<sub>4</sub>OAc 100 mM) + 0.5 µL of BAP (150 U/µL) was made. 10 µL of this solution + 1 µL of tRNA were mixed and samples were incubated for overnight at 37°C. Then, the mixture was directly analyzed by CE-MS/MS. For RNase A digestion, a solution of 1000 µL of RNase A (1/1000 in NH<sub>4</sub>OAc 100 mM) + 0.5 µL of BAP (150 U/µL) was made. 10 µL of this solution + 1 µL of tRNA were mixed and samples were incubated for overnight at 37°C. Then, the mixture was directly analyzed by CE-MS/MS.

**Capillary electrophoresis.** All of the CE experiments have been done on a CESI8000 capillary electrophoresis system from Sciex Separation (Brea, CA, USA). The 32 Karat™ (Sciex separation) software was used for instrument control and data acquisition. Samples were stored at 7 °C in sample tray before injection. Bare fused-silica capillary (total length 100 cm; 30 µm i.d.) with porous tip on its final end, a second capillary filled during experiments with BGE (10% acetic acid) allows electrical contact. Conditioning of a new capillaries was performed by applying 10 min 50 psi of NaOH 1 M, NaOH 0.1 M and H<sub>2</sub>O. For CE experiments, hydrodynamic injection (4 psi for 9.1 sec) corresponding to a total volume of 5 nL of sample injected was used (for total and partial digestions). Separations were achieved applying +30 kV. After each run, capillary was rinsed by applying 1 min 100 psi of MeOH, H<sub>2</sub>O, NaOH 0.1 M, H<sub>2</sub>O, HCl 0.1 M, H<sub>2</sub>O and 2 min 100 psi of BGE.

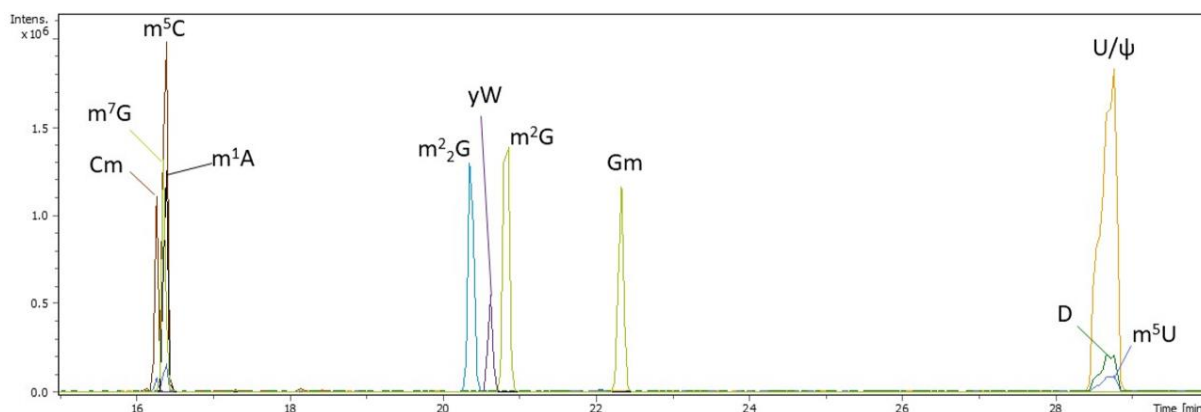
**Mass spectrometry.** For nucleoside analysis, the CE system was coupled to a maXis 4G mass spectrometer (Bruker, Bremen, Germany) by the sheathless interface. The maXis MS is equipped with a hybrid analyzer composed of quadrupoles followed by time-of-flight (TOF). The Otof control 3.4 software allowed to pilot the nanoESI source. Source settings were the following ones: Ion polarity: positive, nanoESI voltage: +1500 V, Source heating temperature: 130 °C, Dry gas: 1.2 L/min, Mass/charge ( $m/z$ ) range was 80-800 for MS and MS/MS detection. Gas-phase fragmentations were performed using  $N_2$ . Collision energies were set at 20 eV for CID experiments. For oligonucleotide analysis, the same source parameters were used except for  $m/z$  range which was 230-2300 for MS and MS/MS detection. Gas-phase fragmentations were performed using  $N_2$ . Collision energies for CID experiments were ranged from 6 to 48 eV according to  $m/z$  ratio of precursors.

**Data analysis.** Data obtained from CE-MS/MS experiments were analyzed using DataAnalysis software (Bruker, Bremen, Germany). All data were manually interpreted. Sequences of tRNA as well as locations of structures of modified nucleotides were found in Modomics database (<http://modomics.genesilico.pl/sequences/list/tRNA/>). This database is constantly verified and regularly updated.  $m/z$  ratios of theoretical digest products were simulated using MongoOligo mass calculator v2.08 (<https://mods.rna.albany.edu/masspec/Mongo-Oligo>). The  $m/z$  ratios of fragment ions generated by CID experiments were also calculated by MongoOligo. It is important to note that RNase T1 and RNase A cleave RNA after specific nucleotides, but could also cleave after certain modified nucleotides. For example, RNase T1 cleaves after m2G and RNase A cleave after D and m5U. MongoOligo simulator (v2.08) does not cleave modifications and checking of simulated fragments from this calculator should be done.

For the analysis of total hydrolysis, theoretical  $m/z$  of unmodified and modified compound were extracted with a tolerance of  $\pm 0.02$   $m/z$ . Identification of compounds were done using migration times, precursor ions and product ions.

For the analysis of oligonucleotides, theoretical  $m/z$  of compounds were extracted with a tolerance of  $\pm 0.02$   $m/z$ . Identifications of digested RNA fragments were done from CID spectra by reconstructing the sequence following y and/or c series. If not possible, other series

can be used. Modifications are localized with the same procedure. Then experimental results were confronted with theoretical.



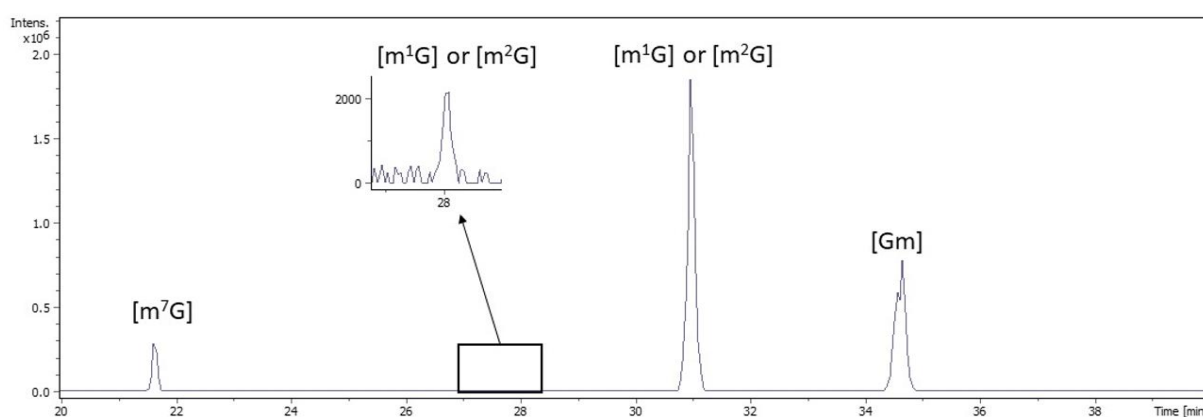
**Figure 1.** Multiple extracted ion electropherograms for the analysis of RNase P<sub>1</sub> BAP digested tRNA<sup>Phe</sup><sub>GAA</sub> from *S. cerevisiae* by CE-MS in positive ion mode, only modified bases are extracted. Injected volume 5 nL (10 fmol)

## Results and discussion

### Characterization at the nucleoside level

To study RNAs modifications, classical strategies based on different ribonucleases (RNases) were assessed [5, 7]. The first level of characterization consisted to perform the total hydrolyses of the sample with the aim of identifying modified nucleotides present in the RNAs. This step was performed using ribonuclease P1 (RNase P1) which cleaves 3' phosphate of any nucleotide. Remaining 5' phosphates were removed using bacterial alkaline phosphatase (BAP) to obtain a mixture of nucleosides. A 30 minutes separation of ribonucleosides were achieved by applying +30 kV with a background electrolyte constituted of 10% acetic acid (pH 2.2) and a sample volume of 5 nL (corresponding to 10 fmol injected) [30]. While RNA are classically detected in negative ion mode due to their acidic property, exploring the positive mode can provide another facet of RNA analysis and thus could show complementarity. To screen modified nucleosides, the analytes were then on-line detected by CZE-ESI-MS/MS in positive ion mode. Extracted Ion Electropherograms of nucleosides from tRNA<sup>Phe</sup><sub>GAA</sub> are shown in Figure 1 (for other tRNA, see Supporting Information Figure S1 to S3). As expected, the four canonical nucleosides and modified nucleosides were separated and characterized with  $\pm 0.02$  m/z. To increase identification confidence, MS/MS of nucleosides was performed. For peaks

assignments, identifications were confirmed when specific transition were detected [32]. For example, in the case of uridine and pseudouridine, these species were identified using MS/MS spectra because these isomers have different fragmentation patterns. However, these methods cannot distinguish nucleosides having the same fragmentation pattern. In that case, standards need to be used to characterize modification based on their migration times. For tRNA<sup>PheGAA</sup>, all the known modifications described in the Modomics database were detected and compiled in Table 1 (for other tRNA, see Supporting Information Table S1 to S3). The most important and common modified nucleosides were identified as dihydrouridine (D) and methylated nucleosides, however, in the same set of data, other more complex modifications as wybutosine (yW) were also characterized. Interestingly, in the case of tRNA<sup>TrpCCA</sup> which contains four methylated guanosines (1-methylguanosine m<sup>1</sup>G; N2-methylguanosine m<sup>2</sup>G; 2'-O-methylguanosine Gm and 7-methylguanosine m<sup>7</sup>G), separation of these isomers were achieved using CZE-ESI-MS/MS as shown in Figure 2. The excellent separation provided by CZE enables the best selectivity for these isomers. While Gm was characterized with its different fragmentation pattern and m<sup>7</sup>G with the analysis of P1 digested tRNA Val from *S. cerevisiae* standard (data not shown), m<sup>1</sup>G and m<sup>2</sup>G were not assigned because the standards were not available. Thus, we demonstrated the ability of CZE-ESI-MS/MS to easily screen and identify all modifications contained in the four tRNA. However, this level of characterization does not allow to locate modifications on RNA sequences.



**Figure 2.** Extracted ion electropherograms of 298.12 m/z for the analysis of RNase P1 BAP digested tRNA<sup>TrpCCA</sup> by CE-MS/MS in positive ion mode. Injected volume 5 nL (10 fmol)

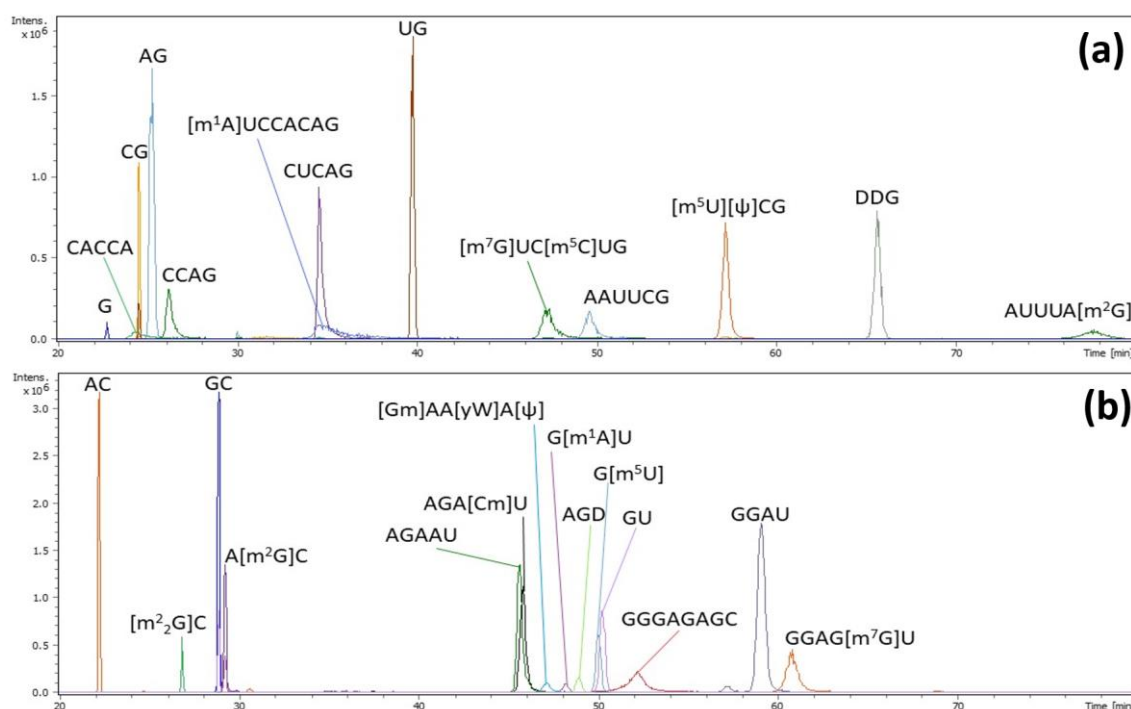
**Table 1. List of all digested products experimentally observed for tRNA<sup>PheGAA</sup>, after digestion with RNase P1 BAP and CE-MS/MS analysis.**

Precursor ion (m/z)	Product ion (m/z)	Molecular mass (Da)	Migration time (min)	Assigned nucleoside
244.09	112.05	243.09	16.0	C
245.08	113.04	244.07	28.7	U
268.10	136.06	267.10	16.5	A
284.10	152.06	283.09	21.9	G
298.12	166.07	297.11	20.9	m <sup>2</sup> G
247.09	115.05	246.08	28.7	D
312.13	180.09	311.12	20.4	m <sup>2</sup> <sub>2</sub> G
258.11	112.05	257.10	16.3	Cm
298.12	152.06	297.11	22.3	Gm
509.19	377.16	508.19	20.6	yW
245.08	125.04	244.07	28.7	ψ
258.11	126.07	257.10	16.4	m <sup>5</sup> C
298.12	166.07	297.11	16.4	m <sup>7</sup> G
259.09	127.05	258.08	28.7	m <sup>5</sup> U
282.12	150.08	281.11	16.4	m <sup>1</sup> A

### Characterization at the oligonucleotide level

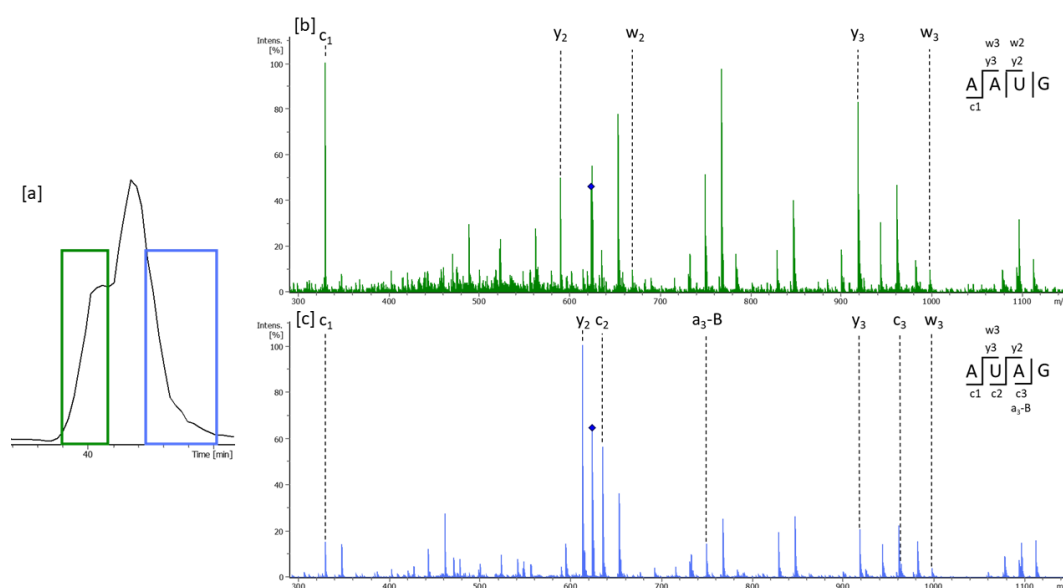
Hence, a strategy was implemented to map modified nucleotides on the RNA sequence. In addition to the location of previously identified modifications, primary structure characterization was simultaneously assessed using the same set of data. For this purpose, a bottom-up approach was developed using exactly the same CZE-ESI-MS/MS conditions used for ribonucleosides characterization. Based on a proteomic-like strategy, RNase digestions have been performed to obtain a pool of short oligonucleotides. To increase sequence coverages, a workflow based on two specific digestions was realized. Prior CZE-ESI-MS/MS analysis, samples were individually digested by RNase T1 and RNase A [33, 34]. RNase T1 cleaves specifically at 3' end of guanosine while RNase A cleaves specifically at 3' end of pyrimidines. Major advantages of these RNases are their strong specificity and their excellent reactivity involving the production of no miscleavages. In addition, dephosphorylation of terminal phosphates using BAP was also performed simultaneously to each digestion to improve separation (data not shown) and facilitate MS/MS spectra interpretation [35]. Following digestion step, oligonucleotides were analyzed in 60 to 80 minutes by CZE-ESI-MS/MS in positive polarity using the same experimental conditions described for the ribonucleosides study. The similarity of experimental conditions is a very important advantage in terms of analysis time because it allows to perform the two levels of characterization

without any instrumental and experimental modification. In comparison with LC-MS workflows, analysis of nucleosides and oligonucleotides have different experimental conditions (column, nature of mobile phase, etc.), which considerably complicates the implementation of RNA characterization at different levels and consequently increases the analysis time. Following the CZE-ESI-MS detection, fragmentation of oligonucleotides by MS/MS experiments were achieved using collision-induced dissociation (CID). Nowadays, softwares for data treatment of MS and MS/MS spectra of RNA were developed (RNAModMapper [36], RoboOligo [37] or Ariadne [38]), however, they are not yet available in positive mode. Then, we carried out the manual interpretation of MS/MS raw data in order to avoid errors for unambiguous mapping of modifications. Excellent separations of RNase T1 and RNase A oligonucleotides were achieved by CZE-ESI-MS/MS for each tRNA, facilitating significantly MS spectra interpretation. As an example, extracted ion electropherogram of oligonucleotides are shown in Figure 3 for tRNA<sup>PheGAA</sup> (For other tRNA, see Supporting Information Figure S4 to S9).



**Figure 3.** Multiple extracted ion electropherograms for the analysis of (a) RNase T1 BAP and (b) RNase A BAP digested tRNA<sup>PheGAA</sup> by CE-MS in positive ion mode. Injected volume 5 nL (10 fmol)

MS/MS spectra were also manually interpreted. Primary structures were reconstructed following the y and/or c series [39]. a-B and w ions can also be useful if the sequencing is difficult or to confirm the sequence. To validate oligonucleotides, we have considered only the ones whose were detected in MS and fully sequenced by MS/MS. For the larger ones, we often get MS spectra but sometimes they were not abundant enough to be fragmented. Therefore it drops the sequence coverage. It is also important to note that specific RNases produce mononucleotides or dinucleotides that are generally not sequence specific, hence diminishing the sequence recoveries. In our study, only unique mono- and dinucleotides were considered for sequence recoveries. Isobaric species also represent a common issue for RNA sequencing. For example, in the case of RNase T1 hydrolysis of tRNA<sup>AspGUC</sup>, fragments AUAG and AAUG were product. The performances of CE separation combined with the MS/MS spectral quality allow to characterize without any ambiguity both isobaric species (Figure 4).

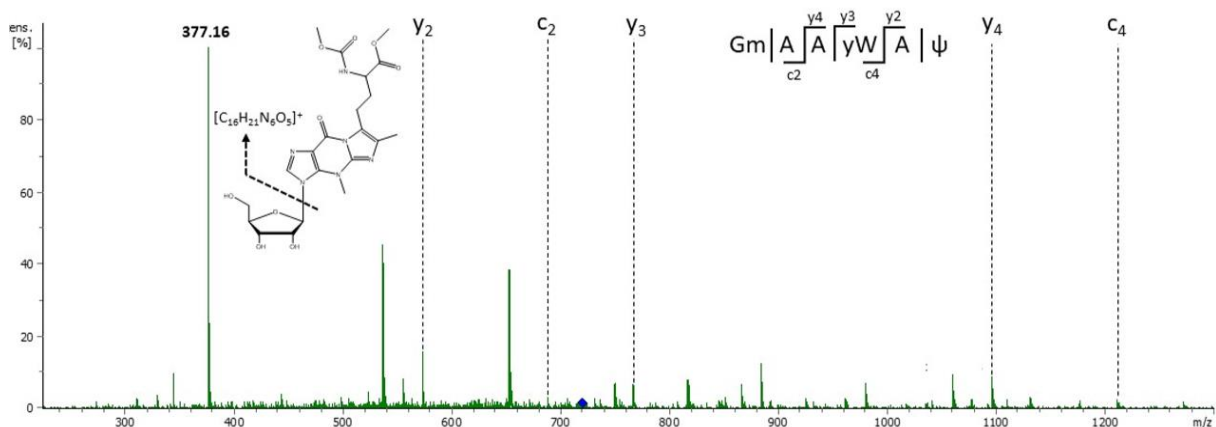


**Figure 4.** [a] Extracted ion electropherograms of m/z 624.2 corresponding to the tRNA<sup>AspGUC</sup> RNase T1 digestion products AAUG and AUAG [b] CID mass spectrum of m/z 624.21 @ 39.9 min and [c] CID mass spectrum of m/z 624.21 @ 40.4 min

Finally, compiling the two digestions, sufficient number of sequence specific oligonucleotides were detected by MS and MS/MS to obtain the sequence recoveries from 68 % to 97 % for studied tRNAs. Concerning RNA modifications, same manual data treatment has been realized to locate modifications on RNA sequence. Calculated m/z ratio corresponding



to modified oligonucleotides have been extracted using MS spectra and sequenced using MS/MS spectra. The Figure 5 shows an example of MS/MS sequencing for [Gm]AA[yW]A[ψ] oligonucleotide. Here, the fragment of 377.15 m/z ( $z=1$ ) correspond to the fragmentation of this modified nucleoside [yW]. Thus, loss of the modified base must be considered to sequence this oligonucleotide. Deduction of 376.15 Da was done for the fragments ions whose contains the modified base and permitted the successful sequencing of this modified oligonucleotide. MS/MS spectral quality allowed precise location of the modified nucleotides on sequences. However, if CZE-ESI-MS/MS analysis of RNA digests enables to determine the precise position onto RNA sequence, in the case of methylation it does not allow to locate the modification on the ribose and/or base. Furthermore, the list of identified fragments by the CZE-ESI-MS/MS analysis were summarized in Table 2 for tRNA<sup>PheGAA</sup> (for other tRNA, see Supporting Information Table S4 to S6). We were able to locate a large number of modifications previously characterized at the nucleoside level, including the wybutosine modification on tRNA<sup>PheGAA</sup>.



**Figure 5** CID mass spectrum of RNase A product [Gm] [13] [13] [13]AA[yW]A[ψ] (m/z 721.15) from *S. cerevisiae* tRNA<sup>PheGAA</sup> with fragmentation assignments. Peak of m/z 377.16 correspond to the loss of the modified base yW.

**Table 2. List of all digested products experimentally observed for tRNA<sup>PheGAA</sup>, after individual digestion with RNase T1 and RNase A. Identified sequences were underlined in black.**<sup>1</sup>MS/MS identification was done considering loss of m/z 376 corresponding to loss of modified base yW

Theoretical m/z	Experimental m/z	z	Molecular mass (Da)	Migration time (min)	Oligonucleotide
<b>RNase T1</b>					
<b>589.14</b>	589.14	+1	588.14	24.6	CG
<b>937.65</b>	937.64	+2	1873.29	77.6	AUUUA[m <sup>2</sup> G]
<b>765.13</b>	765.13	+2	1528.25	34.6	CUCAG
<b>900.18</b>	900.17	+1	899.18	65.6	DDG
<b>612.12</b>	612.12	+2	1222.23	26.3	CCAG
<b>940.16</b>	940.15	+2	1878.30	47.3	[m <sup>7</sup> G]UC[m <sup>5</sup> C]UG
<b>590.12</b>	590.12	+1	589.12	39.8	UG
<b>608.11</b>	608.10	+2	1214.20	57.2	[m <sup>5</sup> U][ψ]CG
<b>836.15</b>	836.14	+3	2505.40	35.3	[m <sup>1</sup> A]UCCACAG
<b>930.15</b>	930.14	+2	1858.29	49.6	AAUUCG
<b>756.64</b>	756.64	+2	1511.28	24.3	CACCA
<b>RNase A</b>					
<b>632.62</b>	632.61	+2	1263.22	59.1	GGAU
<b>932.21</b>	932.20	+1	931.21	29.3	A[m <sup>2</sup> G]C
<b>461.10</b>	461.10	+2	920.19	49.0	AGD
<b>876.48</b>	876.48	+3	2626.43	52.2	GGGAGAGC
<b>617.17</b>	617.17	+1	616.17	23.6	[m <sup>2</sup> G]C
<b>784.15</b>	784.14	+2	1566.28	45.9	AGA[Cm]U
<b>721.15</b>	721.15	+3	2160.44	47.1	[Gm]AA[yW]A[ψ] <sup>1</sup>
<b>984.67</b>	984.67	+2	1967.33	60.9	GGAG[m <sup>7</sup> G]U
<b>590.12</b>	590.12	+1	589.12	50.3	GU
<b>604.14</b>	604.14	+1	603.13	50.0	G[m <sup>5</sup> U]
<b>933.19</b>	933.19	+1	932.19	48.2	G[m <sup>1</sup> A]U
<b>789.14</b>	789.14	+2	1576.27	45.7	AGAAU
<b>268.10</b>	268.11	+1	267.10	18.5	A
<b>Total Sequence</b>					
<b><u>GCGGAUUUA[m<sup>2</sup>G]CUCAGDDGGGAGAGC[m<sup>2</sup>G]CCAGA[Cm]U[Gm]AA[yW]A[ψ]<sup>1</sup>[m<sup>5</sup>C]UGGAG[m<sup>7</sup>G]UC[m<sup>5</sup>C]UGUG[m<sup>5</sup>U][ψ]CG[m<sup>1</sup>A]UCCACAGAAUUCGCACCA</u></b>					

## Conclusion

To summarize, we developed a simple CZE-ESI-MS/MS workflow for the successful characterization of post-transcriptional RNA modifications. Based on strategies using three different ribonucleases, RNA sequencing and modification characterization have been performed at nucleoside and oligonucleotide levels. Using the same platform with the same experimental conditions, we managed to obtain excellent sequence recoveries and were able to identify and map modifications onto primary sequences of tRNAs. CE-MS/MS could become a powerful tool for unambiguous chemical characterization and sequence placement of modifications onto tRNAs sequences (including mitochondrial tRNAs) with software support. Moreover, CE-MS/MS could be applied to other modified non-coding RNA, especially ribosomal RNAs.

## References

1. Carell, T., et al., *Structure and Function of Noncanonical Nucleobases*. Angewandte Chemie International Edition, 2012. **51**(29): p. 7110-7131.
2. Motorin, Y. and M. Helm, *Methods for RNA Modification Mapping Using Deep Sequencing: Established and New Emerging Technologies*. Genes, 2019. **10**(1): p. 35.
3. Hunt, D.F., C.E. Hignite, and K. Biemann, *Structure elucidation of dinucleotides by mass spectrometry*. Biochemical and Biophysical Research Communications, 1968. **33**(3): p. 378-383.
4. Crain, P.F., *Preparation and Enzymatic-Hydrolysis of DNA and Rna for Mass-Spectrometry*. Methods in Enzymology, 1990. **193**: p. 782-790.
5. Kowalak, J.A., et al., *A Novel Method for the Determination of Posttranscriptional Modification in Rna by Mass-Spectrometry*. Nucleic Acids Research, 1993. **21**(19): p. 4577-4585.
6. Cai, W.M., et al., *A Platform for Discovery and Quantification of Modified Ribonucleosides in RNA: Application to Stress-Induced Reprogramming of tRNA Modifications*. Rna Modification, 2015. **560**: p. 29-71.
7. Ross, R.L., X.Y. Cao, and P.A. Limbach, *Mapping Post-Transcriptional Modifications onto Transfer Ribonucleic Acid Sequences by Liquid Chromatography Tandem Mass Spectrometry*. Biomolecules, 2017. **7**(1).
8. Thomas, B. and A.V. Akoulitchev, *Mass spectrometry of RNA*. Trends in Biochemical Sciences, 2006. **31**(3): p. 173-181.
9. Boccaletto, P., et al., *MODOMICS: a database of RNA modification pathways. 2017 update*. Nucleic Acids Research, 2018. **46**(D1): p. D303-D307.
10. Jiang, X.M., et al., *Direct UV detection of underivatized amino acids using capillary electrophoresis with online sweeping enrichment*. Journal of Separation Science, 2009. **32**(11): p. 1927-1933.
11. Li, N., et al., *Separation of miRNA and its methylation products by capillary electrophoresis*. Journal of Chromatography A, 2008. **1202**(2): p. 220-223.
12. Zabzdyr, J.L. and S.J. Lillard, *UV- and visible-excited fluorescence of nucleic acids separated by capillary electrophoresis*. Journal of Chromatography A, 2001. **911**(2): p. 269-276.
13. Saevels, J., A. Van Schepdael, and J. Hoogmartens, *Capillary electrophoresis of RNA oligonucleotides: Catalytic activity of a hammerhead ribozyme*. Analytical Biochemistry, 1999. **266**(1): p. 93-101.
14. Skeidsvoll, J. and P.M. Ueland, *Analysis of RNA by capillary electrophoresis*. Electrophoresis, 1996. **17**(9): p. 1512-1517.
15. Liebich, H.M., et al., *Capillary electrophoresis of urinary normal and modified nucleosides of cancer patients*. Journal of Chromatography A, 1998. **793**(2): p. 341-347.
16. Cornelius, M.G. and H.H. Schmeiser, *RNA analysis by MEKC with LIF detection*. Electrophoresis, 2007. **28**(21): p. 3901-3907.
17. Olivares, J.A., et al., *Online Mass-Spectrometric Detection for Capillary Zone Electrophoresis*. Analytical Chemistry, 1987. **59**(8): p. 1230-1232.
18. Kleparnik, K., *Recent advances in combination of capillary electrophoresis with mass spectrometry: Methodology and theory*. Electrophoresis, 2015. **36**(1): p. 159-178.

19. Feng, Y.-L., H. Lian, and J. Zhu, *Application of pressure assisted electrokinetic injection technique in the measurements of DNA oligonucleotides and their adducts using capillary electrophoresis–mass spectrometry*. *Journal of Chromatography A*, 2007. **1148**(2): p. 244-249.
20. Khan, N., G. Mironov, and M.V. Berezovski, *Direct detection of endogenous MicroRNAs and their post-transcriptional modifications in cancer serum by capillary electrophoresis-mass spectrometry*. *Analytical and Bioanalytical Chemistry*, 2016. **408**(11): p. 2891-2899.
21. Mironov, G.G., et al., *Conformational Dynamics of DNA G-Quadruplex in Solution Studied by Kinetic Capillary Electrophoresis Coupled On-line with Mass Spectrometry*. *ChemistryOpen*, 2014. **3**(2): p. 58-64.
22. 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*. *Journal of Chromatography A*, 2020. **1614**: p. 460716.
23. Neusüss, C., et al., *Coupling of Capillary Electromigration Techniques to Mass Spectrometry*. 2018. p. 293-311.
24. Cao, P. and M. Moini, *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): p. 561-564.
25. 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*. *Analytical and Bioanalytical Chemistry*, 2014. **406**(4): p. 1029-1038.
26. Sánchez-López, E., et al., *Sheathless CE-MS based metabolic profiling of kidney tissue section samples from a mouse model of Polycystic Kidney Disease*. *Scientific Reports*, 2019. **9**(1): p. 806.
27. Han, X., et al., *Sheathless Capillary Electrophoresis-Tandem Mass Spectrometry for Top-Down Characterization of Pyrococcus furiosus Proteins on a Proteome Scale*. *Analytical Chemistry*, 2014. **86**(22): p. 11006-11012.
28. Kammeijer, G.S.M., et al., *Sialic acid linkage differentiation of glycopeptides using capillary electrophoresis – electrospray ionization – mass spectrometry*. *Scientific Reports*, 2017. **7**(1): p. 3733.
29. Giorgetti, J., et al., *Monoclonal antibody N-glycosylation profiling using capillary electrophoresis - Mass spectrometry: Assessment and method validation*. *Talanta*, 2018. **178**: p. 530-537.
30. Yu, Y., et al., *Ultrasensitive and simultaneous determination of RNA modified nucleotides by sheathless interfaced capillary electrophoresis-tandem mass spectrometry*. *Chemical Communications*, 2019. **55**(53): p. 7595-7598.
31. Kellner, S., et al., *Profiling of RNA modifications by multiplexed stable isotope labelling*. *Chemical Communications*, 2014. **50**(26): p. 3516-3518.
32. Su, D., et al., *Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry*. *Nature Protocols*, 2014. **9**(4): p. 828-841.
33. Antoine, L., et al., *Mapping post-transcriptional modifications in Staphylococcus aureus tRNAs by nanoLC/MSMS*. *Biochimie*, 2019. **164**: p. 60-69.

34. Antoine, L. and P. Wolff, *Mapping of Posttranscriptional tRNA Modifications by Two-Dimensional Gel Electrophoresis Mass Spectrometry*, in *RNA Spectroscopy: Methods and Protocols*, V. Arluison and F. Wien, Editors. 2020, Springer US: New York, NY. p. 101-110.
35. Krivos, K.L., B. Addepalli, and P.A. Limbach, *Removal of 3'-phosphate group by bacterial alkaline phosphatase improves oligonucleotide sequence coverage of RNase digestion products analyzed by collision-induced dissociation mass spectrometry*. *Rapid Communications in Mass Spectrometry*, 2011. **25**(23): p. 3609-3616.
36. 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**: p. 128-138.
37. 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 Research*, 2015. **43**(10).
38. Nakayama, H., et al., *Ariadne: a database search engine for identification and chemical analysis of RNA using tandem mass spectrometry data*. *Nucleic Acids Research*, 2009. **37**(6).
39. Mcluckey, S.A. and S. Habibigoudarzi, *Ion-Trap Tandem Mass-Spectrometry Applied to Small Multiply-Charged Oligonucleotides with a Modified Base*. *Journal of the American Society for Mass Spectrometry*, 1994. **5**(8): p. 740-747.

## Supporting Information

### Characterization of post-transcriptional RNA modifications by sheathless capillary electrophoresis – high resolution mass spectrometry

Antony Lechner<sup>1\*</sup>, Philippe Wolff<sup>2,3\*</sup>, Emmanuelle Leize-Wagner<sup>1</sup>, Yannis-Nicolas François<sup>1</sup>

<sup>1</sup> Laboratoire de Spectrométrie de Masse des Interactions et des Systèmes (LSMIS) UMR 7140 (Unistra-CNRS), Université de Strasbourg, France.

<sup>2</sup> Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR 9002, F-67000 Strasbourg, France.

<sup>3</sup> Plateforme protéomique Strasbourg Esplanade, CNRS, FRC 1589, F-67000 Strasbourg, France.

\* These authors contributed equally to this work

### Table of content

**Figure S1** Multiple extracted ion electropherograms for the analysis of RNase P1 BAP digested tRNA<sup>Ala</sup> by CE-MS in positive ion mode, only modified bases are extracted. **Page 3**

**Figure S2** Multiple extracted ion electropherograms for the analysis of RNase P1 BAP digested tRNA<sup>Asp</sup> by CE-MS in positive ion mode, only modified bases are extracted. **Page 3**

**Figure S3** Multiple extracted ion electropherograms for the analysis of RNase P1 BAP digested tRNA<sup>Trp</sup> by CE-MS in positive ion mode, only modified bases are extracted. **Page 3**

**Figure S4** Multiple extracted ion electropherograms for the analysis of RNase T1 BAP digested tRNA<sup>Ala</sup> by CE-MS in positive ion mode. **Page 4**

**Figure S5** Multiple extracted ion electropherograms for the analysis of RNase A BAP digested tRNA<sup>Ala</sup> by CE-MS in positive ion mode. **Page 4**

**Figure S6** Multiple extracted ion electropherograms for the analysis of RNase T1 BAP digested tRNA<sup>Asp</sup> by CE-MS in positive ion mode. **Page 4**

**Figure S7** Multiple extracted ion electropherograms for the analysis of RNase A BAP digested tRNA<sup>Asp</sup> by CE-MS in positive ion mode. **Page 5**

**Figure S858** Multiple extracted ion electropherograms for the analysis of RNase T1 BAP digested tRNA<sup>Trp</sup> by CE-MS in positive ion mode. **Page 5**

**Figure S9** Multiple extracted ion electropherograms for the analysis of RNase A BAP digested tRNA<sup>Trp</sup> by CE-MS in positive ion mode. **Page 5**

**Table S1.** List of all digested products experimentally observed for tRNA<sup>Ala</sup> GUC, after digestion with RNase P1 BAP and CE-MS/MS analysis. **Page 6**

**Table S2.** List of all digested products experimentally observed for tRNA<sup>Asp</sup> GUC, after digestion with RNase P1 BAP and CE-MS/MS analysis. **Page 7**

**Table S3.** List of all digested products experimentally observed for tRNA<sup>Trp</sup> CCA, after digestion with RNase P1 BAP and CE-MS/MS analysis. **Page 8**

**Table S4.** List of all digested products experimentally observed for tRNA<sup>Ala</sup> AGC, after individual digestion with RNases T1 and A. Identified sequences were underlined in black. **Page 9**

**Table S5.** List of all digested products experimentally observed for tRNA<sup>Asp</sup> GUC, after individual digestion with RNases T1 and A. Identified sequences were underlined in black. **Page 10**

**Table S6.** List of all digested products experimentally observed for tRNA<sup>Trp</sup> CCA, after individual digestion with RNases T1 and A. Identified sequences were underlined in black. **Page 11**

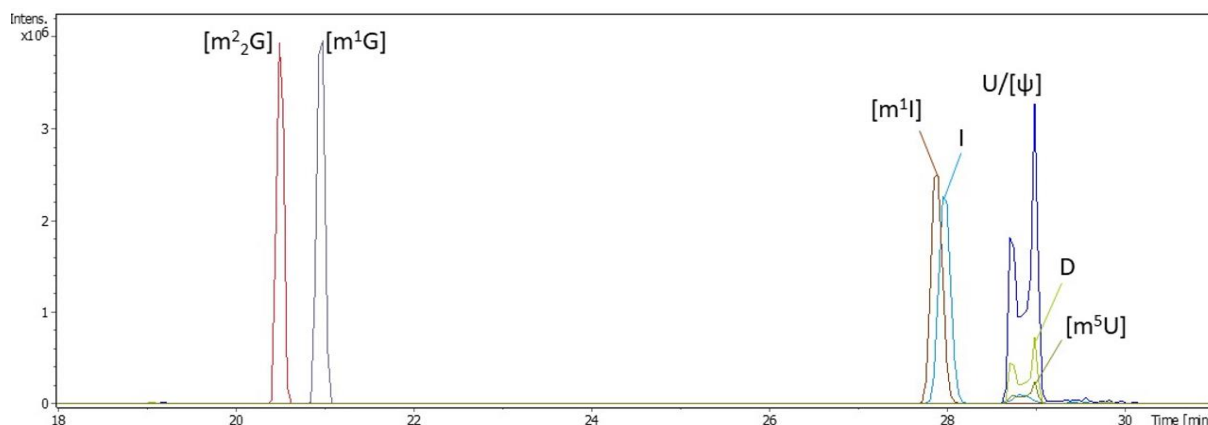


Figure S1 Multiple extracted ion electropherograms for the analysis of RNase P1 BAP digested tRNA<sup>Ala</sup> by CE-MS in positive ion mode, only modified bases are extracted.

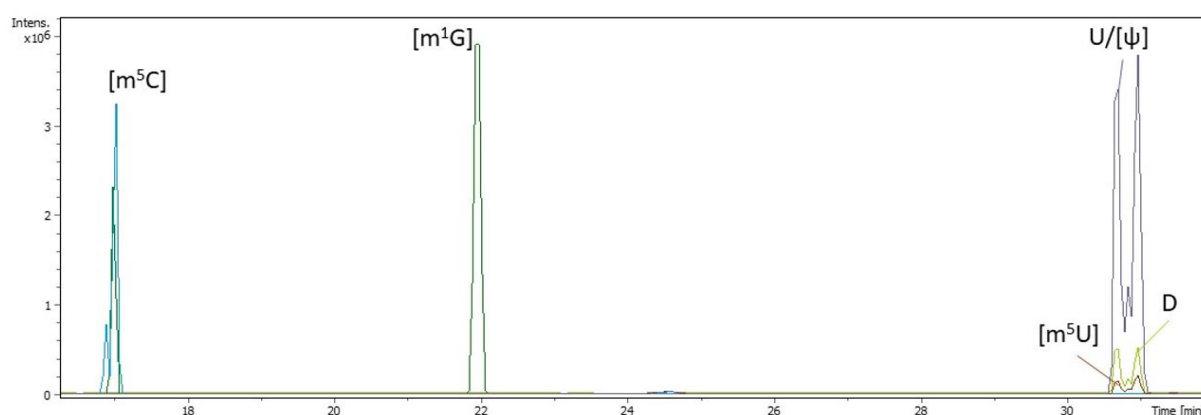


Figure S2 Multiple extracted ion electropherograms for the analysis of RNase P1 BAP digested tRNA<sup>Asp</sup> by CE-MS in positive ion mode, only modified bases are extracted.

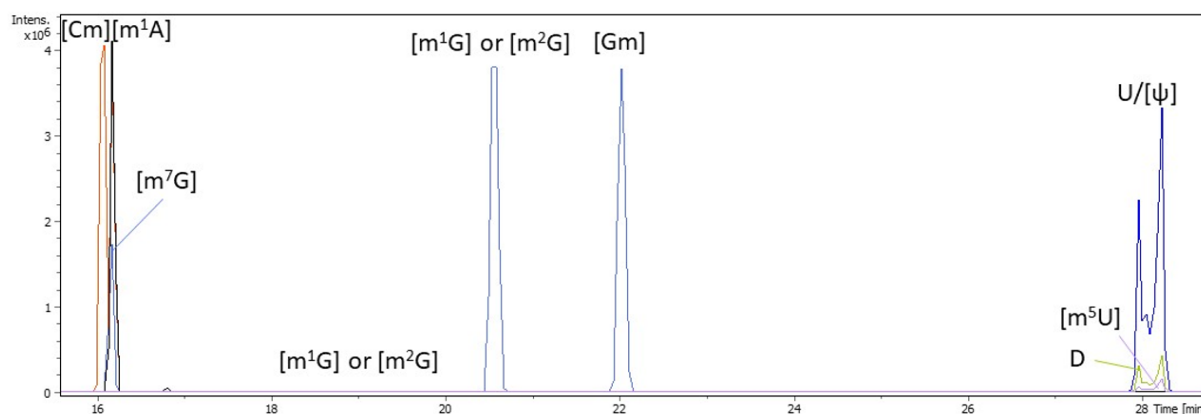


Figure S3 Multiple extracted ion electropherograms for the analysis of RNase P1 BAP digested tRNA<sup>Trp</sup> by CE-MS in positive ion mode, only modified bases are extracted.



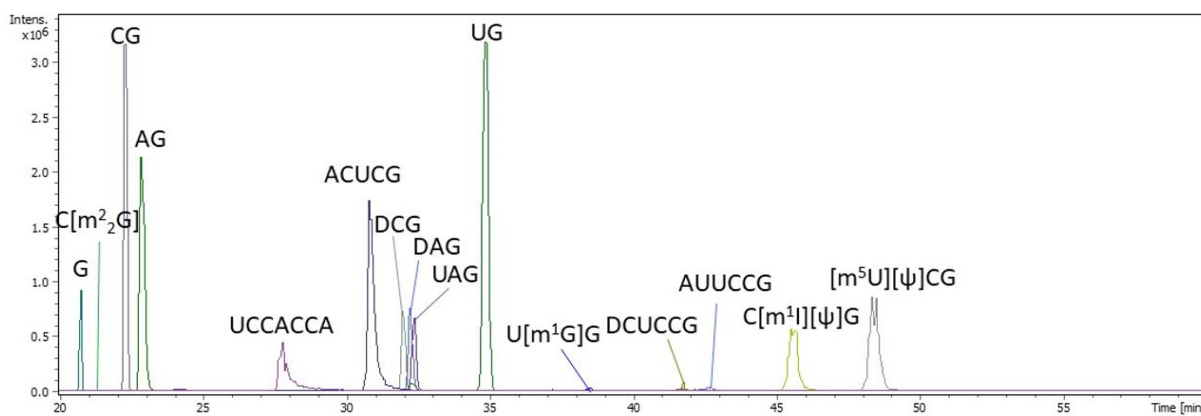


Figure S4 Multiple extracted ion electropherograms for the analysis of RNase T1 BAP digested tRNA<sup>Ala</sup> by CE-MS in positive ion mode.

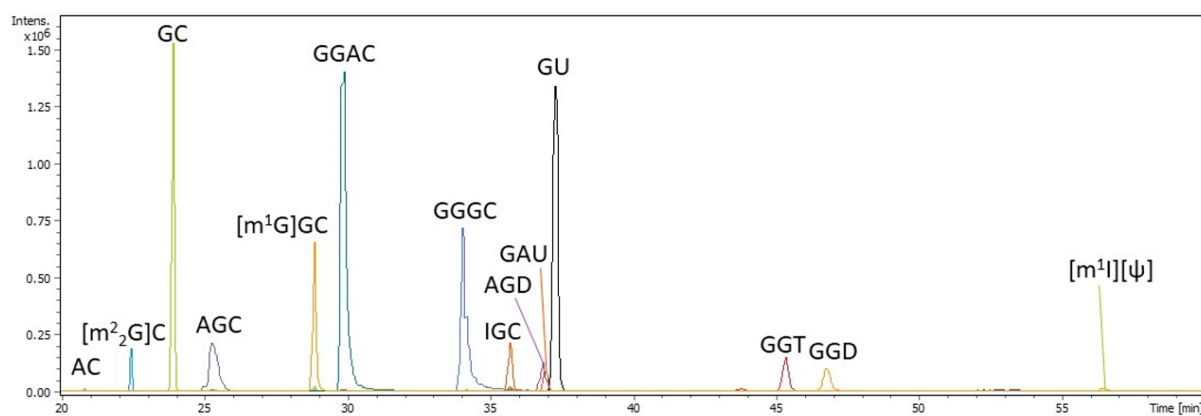


Figure S5 Multiple extracted ion electropherograms for the analysis of RNase A BAP digested tRNA<sup>Ala</sup> by CE-MS in positive ion mode.

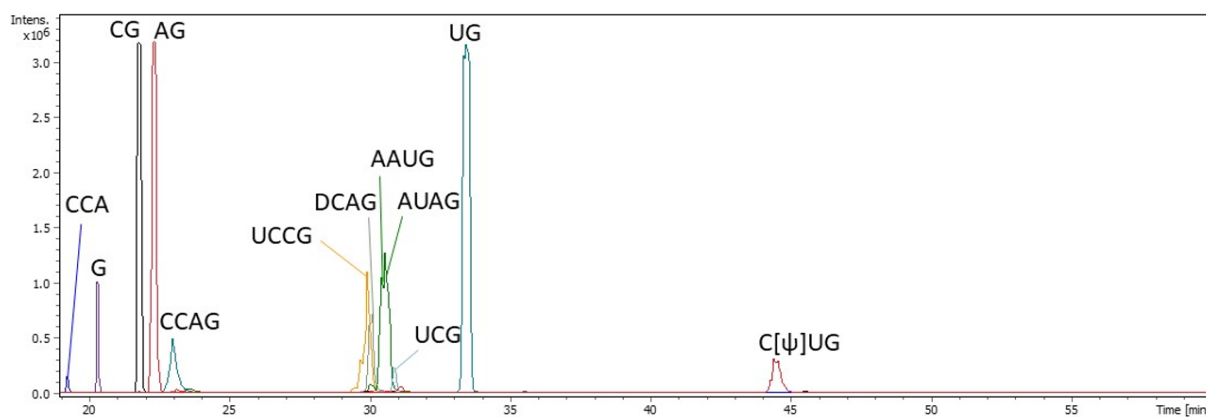


Figure S6 Multiple extracted ion electropherograms for the analysis of RNase T1 BAP digested tRNA<sup>Asp</sup> by CE-MS in positive ion mode.

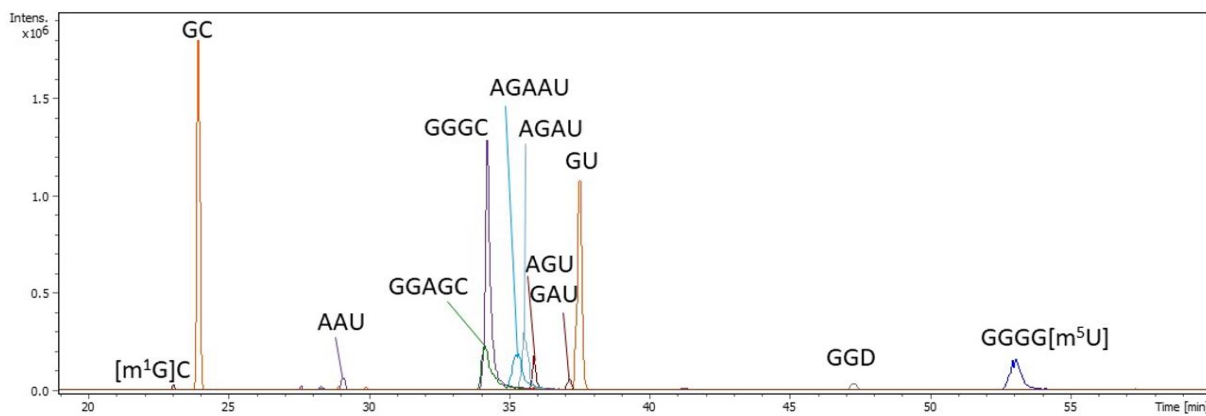


Figure S7 Multiple extracted ion electropherograms for the analysis of RNase A BAP digested tRNA<sup>Asp</sup> by CE-MS in positive ion mode.

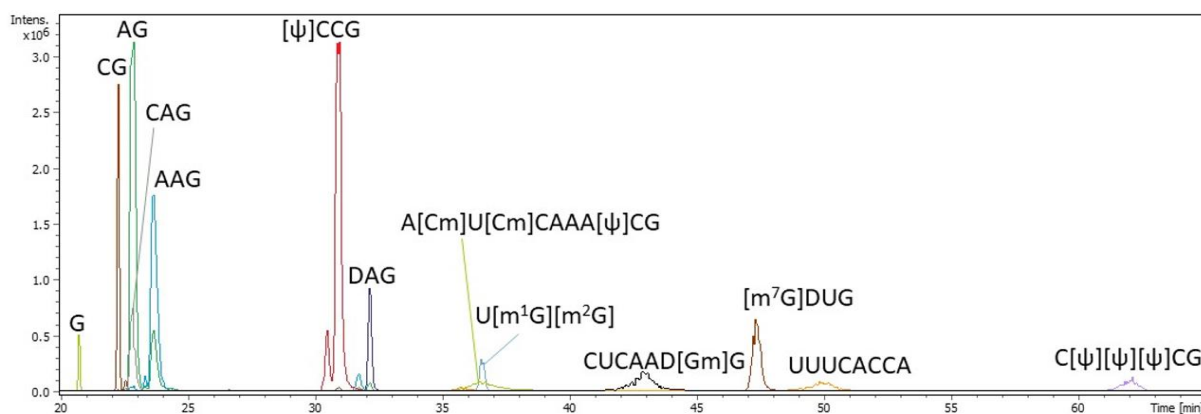


Figure S859 Multiple extracted ion electropherograms for the analysis of RNase T1 BAP digested tRNA<sup>Trp</sup> by CE-MS in positive ion mode.

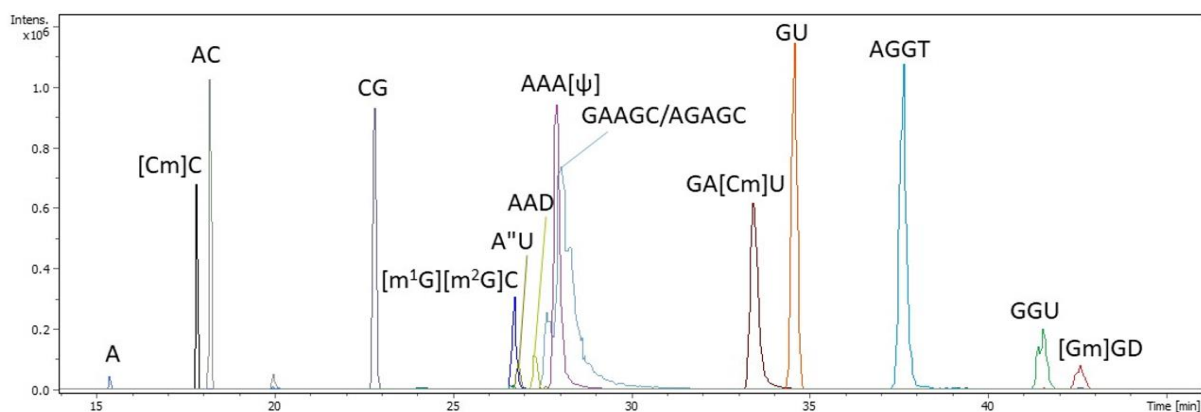


Figure S9 Multiple extracted ion electropherograms for the analysis of RNase A BAP digested tRNA<sup>Trp</sup> by CE-MS in positive ion mode.

Table S1. List of all digested products experimentally observed for tRNA<sup>Ala</sup> GUC, after digestion with RNase P1 BAP and CE-MS/MS analysis.

Precursor ion (m/z)	Product ion (m/z)	Molecular mass (Da)	Migration time (min)	Assigned nucleoside
244.09	112.05	243.09	16.1	C
245.08	113.04	244.07	28.8	U
268.10	136.06	267.10	16.6	A
284.10	152.06	283.09	22.0	G
298.12	166.07	297.11	21.0	m <sup>1</sup> G
247.09	115.05	246.08	28.8	D
312.13	180.09	311.12	20.5	m <sup>2</sup> G
269.09	137.05	268.08	28.0	I
283.10	151.06	282.10	27.9	m <sup>1</sup> I
245.08	125.04	244.07	28.8	ψ
259.09	127.05	258.08	28.8	m <sup>5</sup> U

Table S2. List of all digested products experimentally observed for tRNA<sup>Asp</sup> GUC, after digestion with RNase P1 BAP and CE-MS/MS analysis.

Precursor ion (m/z)	Product ion (m/z)	Molecular mass (Da)	Migration time (min)	Assigned nucleoside
244.09	112.05	243.09	16.6	C
245.08	113.04	244.07	30.8	U
268.10	136.06	267.10	17.2	A
284.10	152.06	283.09	23.2	G
245.08	125.04	244.07	30.8	ψ
247.09	115.05	246.08	30.8	D
298.12	166.07	297.11	22.0	m <sup>1</sup> G
258.11	126.07	257.10	17.1	m <sup>5</sup> C
259.09	127.05	258.08	30.8	m <sup>5</sup> U

Table S3. List of all digested products experimentally observed for tRNA<sup>Trp</sup> CCA, after digestion with RNase P1 BAP and CE-MS/MS analysis.

Precursor ion (m/z)	Product ion (m/z)	Molecular mass (Da)	Migration time (min)	Assigned nucleoside
244.09	112.05	243.09	15.8	C
245.08	113.04	244.07	28.1	U
268.10	136.06	267.10	16.3	A
284.10	152.06	283.09	21.6	G
298.12	166.07	297.11	19.3 or 20.6	m <sup>1</sup> G
298.12	166.07	297.11	19.3 or 20.6	m <sup>2</sup> G
247.09	115.05	246.08	28.1	D
298.12	152.06	297.11	22.0	Gm
245.08	125.04	244.07	28.1	ψ
258.11	112.05	257.10	16.1	Cm
298.12	166.07	297.11	16.2	m <sup>7</sup> G
259.09	127.05	258.08	28.1	m <sup>5</sup> U
282.12	150.08	281.11	16.2	m <sup>1</sup> A

Table S4. List of all digested products experimentally observed for tRNA<sup>Ala</sup> AGC, after individual digestion with RNases T1 and A. Identified sequences were underlined in black.

Theoretical m/z	Experimental m/z	z	Molecular mass (Da)	Migration time (min)	Oligonucleotide
<b>RNase T1</b>					
<b>590.12</b>	590.12	+1	589.12	34.9	UG
<b>475.10</b>	475.10	+2	948.18	37.2	U[m <sup>1</sup> G]G
<b>919.18</b>	919.18	+1	918.17	32.4	UAG
<b>897.18</b>	897.18	+1	896.17	32.1	DCG
<b>921.19</b>	921.19	+1	920.18	32.3	DAG
<b>620.11</b>	620.12	+2	1238.21	45.5	C[m <sup>1</sup> ][ψ]PG
<b>608.11</b>	608.11	+2	1214.20	48.5	[m <sup>5</sup> U][ψ]CG
<b>765.13</b>	765.14	+2	1528.25	30.9	ACUCG
<b>1062.18</b>	1062.18	+2	2122.34	27.8	UCCACCA
<b>RNase A</b>					
<b>640.12</b>	640.12	+2	1278.23	34.1	GGGC
<b>948.20</b>	948.21	+1	947.20	28.9	[m <sup>1</sup> G]GC
<b>469.10</b>	649.10	+2	936.18	36.9	GGD
<b>459.60</b>	459.60	+1	917.18	25.4	AGC
<b>589.14</b>	589.14	+1	588.13	24.0	GC
<b>617.17</b>	617.17	+1	616.16	22.5	[m <sup>2</sup> ₂G]C
<b>919.18</b>	919.18	+1	918.17	35.7	IGC
<b>475.10</b>	475.10	+2	948.18	45.4	GG[m <sup>5</sup> U]
<b>919.18</b>	919.18	+1	918.18	37.0	GAU
<b>632.12</b>	632.13	+2	1262.23	29.9	GGAC
<b>573.15</b>	573.15	+1	572.14	19	AC
<b>268.10</b>	268.11	+1	267.10	15.9	A
<b>Total Sequence</b>					
<b><u>GGGCGUGU[m<sup>1</sup>G]GCGUAGDCGGDAGCGC[m<sup>2</sup>₂G]CUCUUIGC[m<sup>1</sup>][ψ]GGGAGAGGDCUCCGG[m<sup>5</sup>U]</u></b>					
<b><u>[ψ]CGAUUCCGGACUCGUCCACCA</u></b>					

Table S5. List of all digested products experimentally observed for tRNA<sup>Asp</sup> GUC, after individual digestion with RNases T1 and A. Identified sequences were underlined in black.

Theoretical m/z	Experimental m/z	z	Molecular mass (Da)	Migration time (min)	Oligonucleotide
<b>RNase T1</b>					
<b>600.61</b>	600.61	+2	1199.20	30.0	UCCG
<b>624.62</b>	624.62	+2	1247.22	30.6	AUAG
<b>613.62</b>	613.62	+2	1225.23	30.1	DCAG
<b>624.62</b>	624.62	+2	1247.22	30.1	AAUG
<b>601.10</b>	601.10	+2	1200.19	44.6	C[ψ]UG
<b>612.12</b>	612.13	+2	1222.23	23.1	CCAG
<b>895.17</b>	895.17	+1	894.16	30.9	UCG
<b>613.15</b>	613.16	+1	612.14	22.4	AG
<b>439.60</b>	439.60	+2	877.18	19.3	CCA
<b>RNase A</b>					
<b>919.18</b>	919.18	+1	918.17	37.2	GAU
<b>919.18</b>	919.18	+1	918.17	35.9	AGU
<b>469.10</b>	469.10	+2	936.18	47.3	GGD
<b>789.14</b>	789.15	+2	1576.27	35.3	AGAAU
<b>640.12</b>	640.12	+2	1278.23	34.3	GGGC
<b>624.62</b>	624.62	+2	1247.22	35.6	AGAU
<b>820.14</b>	820.15	+2	1638.27	52.9	GGGG[m <sup>5</sup> U]

<b>452.09</b>	452.10	+2	902.17	29.2	AAU
<b>804.65</b>	804.65	+2	1607.28	34.2	GGAGC
<b>268.10</b>	268.11	+1	267.10	15.9	A
<b>Total Sequence</b>					
<u>UCCGUGAUAGUU[ψ]AADGGDCAGAAUGGGCGC[ψ]UGUC[m<sup>1</sup>G]CGUGCCAGAU[m<sup>5</sup>C]GGGG[m<sup>5</sup>U][ψ]CAAUU</u> <u>CCCCGUCGCGGAGCCA</u>					

Table S6. List of all digested products experimentally observed for tRNA<sup>Trp</sup> CCA, after individual digestion with RNases T1 and A. Identified sequences were underlined in black.

Theoretical m/z	Experimental m/z	z	Molecular mass (Da)	Migration time (min)	Oligonucleotide
<b>RNase T1</b>					
<b>942.20</b>	942.21	+1	941.20	23.8	AAG
<b>589.14</b>	589.14	+1	588.13	22.4	CG
<b>963.20</b>	963.20	+1	962.19	36.6	U[m <sup>1</sup> G][m <sup>2</sup> G]
<b>1263.21</b>	1263.21	+2	2524.41	43.0	CUCAAD[Gm]G
<b>921.19</b>	921.20	+1	920.18	32.2	DAG
<b>613.15</b>	613.15	+1	612.14	23.0	AG
<b>906.63</b>	906.64	+2	1811.25	62.1	C[ψ][ψ][ψ]CG
<b>629.12</b>	629.12	+2	1256.22	47.3	[m <sup>7</sup> G]DUG
<b>918.19</b>	918.20	+1	917.18	22.9	CAG
<b>600.61</b>	600.61	+2	1199.20	30.9	[ψ]CCG
<b>1215.68</b>	1215.68	+2	2429.35	49.9	UUUCACCA
<b>RNase A</b>					
<b>796.65</b>	796.65	+2	1591.28	28.1	GAAGC
<b>468.09</b>	468.09	+2	934.16	41.6	GGU
<b>962.22</b>	962.22	+1	961.21	26.7	[m <sup>1</sup> G][m <sup>2</sup> G]C
<b>453.10</b>	453.11	+2	904.19	27.3	AAD
<b>476.11</b>	476.11	+2	950.19	42.6	[Gm]GD
<b>796.65</b>	796.65	+2	1591.28	27.7	AGAGC
<b>619.62</b>	619.62	+2	1237.22	33.4	GA[Cm]U
<b>563.15</b>	563.15	+1	562.14	17.9	[Cm]C
<b>616.62</b>	616.62	+2	1231.23	27.9	AAA[ψ]
<b>639.62</b>	639.63	+2	1277.23	37.6	AGG[m <sup>5</sup> U]
<b>459.10</b>	459.11	+2	916.19	26.9	A[m <sup>1</sup> A]U
<b>573.15</b>	573.15	+1	572.14	18.2	AC
<b>268.10</b>	268.10	+1	267.10	15.4	A
<b>Total Sequence</b>					
<u>GAAGCGGU[m<sup>1</sup>G][m<sup>2</sup>G]CUCAAD[Gm]GDAGAGC[ψ][ψ][ψ]CGA[Cm]U[Cm]CAAA[ψ]CGAAGG[m<sup>7</sup>G]DUGCAG</u> <u>G[m<sup>5</sup>U][ψ]CA[m<sup>1</sup>A]UCCUG[ψ]CCGUUUCACCA</u>					

### 7.3 Conclusion

Les travaux présentés dans ce chapitre ont concerné la caractérisation des modifications d'ARN par le couplage de la CE à la MS. Ces travaux avaient pour objectif de démontrer l'intérêt du couplage CE-MS/MS pour la caractérisation des modifications post-transcriptionnelles des ARN. Dans ce but, quatre ARNt modèles provenant de *S. cerevisiae* et possédant une variété de modifications situées à plusieurs positions sur les séquences ont été choisis. La caractérisation des modifications a été réalisée en utilisant les stratégies de préparation d'échantillons classiquement utilisées en LC-MS.

Grâce à cette stratégie de préparation d'échantillon basée sur l'utilisation de trois nucléases, il a été possible d'identifier et de localiser sans ambiguïté les nucléotides modifiés présents dans les ARNt. Pour cela, les ARN ont été individuellement digérés à la nucléase P<sub>1</sub> (hydrolyse totale) mais aussi à la RNase T<sub>1</sub> et la RNase A (digestions partielles). Les phosphates terminaux des digests ont été hydrolysés avec la phosphatase alcaline bactérienne dans le but d'améliorer la détection des analytes. Une fois les digestions accomplies, les échantillons ont été analysés avec le système CE-MS/MS en mode positif. De plus, dans ces travaux, les analyses des échantillons ont été effectuées avec la même plateforme dans des conditions d'analyses identiques.

Les excellentes séparations ainsi que la sensibilité fournies par le système d'électrophorèse ont permis de faciliter grandement l'interprétation des spectres. Ainsi, l'utilisation de différents ARNt de référence a permis de confirmer que, grâce au couplage CZE-MS/MS, il est possible non seulement d'identifier mais aussi de localiser avec certitude les nucléotides modifiés dans les ARNt. Les analyses des digests partielles par CZE-MS/MS ont permis de fournir des recouvrements de séquence de 67% jusqu'à 98% grâce à la stratégie de la double digestion. Le protocole simple a permis d'identifier et de localiser des modifications importantes comme la wybutosine (508,19 Da) située sur la position 37 de l'anticodon de l'ARNt<sup>Phe</sup>.

La comparaison des résultats obtenus aux résultats théoriques a mis en évidence l'efficacité de l'utilisation de la CE en tant que méthode séparative en amont de la MS dans le cadre de la caractérisation des modifications post-transcriptionnelles des ARN. Le couplage apparaît donc comme une méthode viable au service de la caractérisation des modifications d'ARN. Par conséquent, le système CZE-MS/MS pourrait devenir un nouvel outil puissant pour l'exploration des modifications d'ARN.

Les résultats de ce chapitre ont servi de preuve de concept pour l'analyse des modifications d'ARN. La suite de ces travaux a été de mettre en œuvre cette méthode afin d'explorer et de caractériser des modifications d'ARN provenant d'échantillons biologiques. Cette étude a pour objectif de démontrer l'utilité de cette méthode dans la caractérisation de modifications mais également dans l'éventuelle découverte de modifications inconnues.

## CHAPITRE 8. MISE EN APPLICATION DE LA METHODE

### 8.1 Introduction

Le chapitre précédent a mis en lumière l'intérêt du couplage CZE-MS/MS pour l'étude des modifications des ARN de transfert. Des ARNt, provenant de levure et préalablement purifiés, ont été utilisés dans le but de prouver l'efficacité du couplage pour la caractérisation des modifications. Dans le dernier chapitre de ce manuscrit, des travaux préliminaires ont été réalisés afin de mettre en place un protocole de préparation d'échantillons permettant l'étude de différents types d'ARN provenant d'autres organismes et directement issus de matrices biologiques. L'objectif est de cartographier tous les ARN présents et d'étudier les modifications de chaque ARN.

Ici, afin de démontrer la possibilité d'étudier d'autres types d'ARN par CZE-MS/MS, c'est un extrait total d'ARN d'*Escherichia Coli* qui a été analysé. Les ARN ont, dans un premier temps, été extraits et purifiés. Les ARN ont ensuite été isolés par électrophorèse sur gel afin de permettre l'étude individuelle de chaque ARN. Ici, c'est une électrophorèse sur gel à deux dimensions qui a été réalisée selon les conditions décrites par Antoine *et al* [1]. Afin d'étudier les modifications, les ARN isolés ont été digérés individuellement *in gel* avec plusieurs enzymes puis dessalés. Les différents digests ont ensuite été analysés par CZE-MS/MS dans les conditions décrites dans le chapitre précédent. L'identification des ARN dans les bandes du gel ainsi que la caractérisation des modifications ont été effectuées grâce à l'interprétation des données spectrales.

Concernant l'interprétation des données, ce n'est pas une recherche ciblée qui a été réalisée mais une recherche sans *a priori*. Dans le cadre de la stratégie d'identification des ARN, les pics obtenus par CZE-MS/MS ont été séquencés manuellement grâce aux spectres MS/MS par *de Novo*. Une fois les oligonucléotides séquencés, les séquences sont soumises à un moteur de recherche qui a permis de proposer des ARN. Ce moteur de recherche classe, avec un score, les ARN qui peuvent correspondre aux oligonucléotides détectés. Par la suite, la localisation des modifications a été réalisée avec le même jeu de données. Lorsque l'interprétation et l'identification ont été accomplies, les résultats obtenus sont comparés avec la base de données Modomics afin de vérifier la pertinence de ces résultats.

Néanmoins, à la suite des circonstances exceptionnelles de l'année 2020 concernant la pandémie de coronavirus, toutes les bandes du gel n'ont pas pu être analysées. De plus, l'analyse au premier niveau qui permet d'identifier les nucléotides modifiés n'a pas été réalisée. Toutefois, les résultats préliminaires obtenus sur les bandes analysées dans ce chapitre sont très encourageants puisque l'identification de plusieurs ARN, dont l'identification d'un ARNr, a été accomplie avec succès. En outre, la localisation de diverses modifications a également été démontrée par la méthode établie.

Le protocole a permis la caractérisation des modifications d'ARN sur différents niveaux (hydrolyse total et stratégie « bottom-up »). Ce travail a permis non seulement d'identifier des

ARN mais également de caractériser les modifications post-transcriptionnelles de ces ARN grâce à la stratégie d'analyse et au couplage CZE-MS/MS. Les résultats ont permis de démontrer la possibilité de caractériser les modifications d'autres types d'ARN provenant d'échantillons biologiques avec le protocole de préparation d'échantillon mis en place. De plus, l'identification des bandes d'ARN a été réalisée montrant la puissance du couplage CZE-MS/MS pour la caractérisation des ARN.

## 8.2 Préparation des ARN

Le protocole mis en place a permis d'isoler les ARN afin de les étudier individuellement par CZE-MS/MS. Un extrait d'ARN total (*E. coli*) a d'abord été purifié par électrophorèse sur gel de polyacrylamide en deux dimensions (2D PAGE). Les deux dimensions permettent une meilleure séparation des ARN. Les bandes ont ensuite été découpées puis les ARN ont été digérés individuellement *in gel*, avec la RNase T<sub>1</sub> et la RNase A. Les phosphates terminaux ont également été hydrolysés à l'aide de la BAP. Une fois les digestions achevées, les échantillons, préalablement dessalés, ont été analysés par CZE-MS/MS en mode positif.

### 8.2.1 Isolation des ARN par PAGE 2D

Afin de réaliser les digestions individuelles des ARN, l'extrait d'ARN (*E. coli*) a été divisé en deux pour effectuer deux séparations en parallèle. Ensuite, les extraits d'ARN total sont soumis à une première séparation sur un gel d'électrophorèse. La première dimension de cette séparation a été effectuée dans des conditions dénaturantes, avec un gel de polyacrylamide à 12,5% (acrylamide:bisacrylamide 19 :1 ; 8 M urée ; tampon 1 x Tris, Borate EDTA). Avant de réaliser la séparation, le gel a été soumis à une puissance constante de 30 W pendant 30 min. L'extrait d'ARN total a été coloré en bleu à l'aide d'une solution de bleu de bromophénol et de xylène cyanol afin de suivre le bon déroulement de la migration des composés sur le gel. Par la suite, les ARN ont été chargés sur le gel (20 µg) et la séparation a été réalisée pendant 7 heures à une puissance constante de 30 W. Dans ces conditions de séparation, tous les ponts hydrogènes présents dans les ARN sont rompus. Grâce à la forte concentration d'urée et la température élevée du gel (70°C), les ARN sont totalement dénaturés et sont séparés exclusivement en fonction de leur taille dans le gel d'électrophorèse. Une fois la 1<sup>ère</sup> dimension de la séparation accomplie, la partie du gel contenant les ARN a été coupée afin de réaliser la seconde dimension de la séparation.

Dans le but de réaliser la seconde dimension de la séparation, le gel coupé de la 1<sup>ère</sup> dimension a été placé sur le haut d'un second gel d'électrophorèse. La seconde dimension de la séparation a été réalisée en condition semi-dénaturante, avec un gel de polyacrylamide à 20% (acrylamide:bisacrylamide 19 :1 ; 4 M urée ; tampon 1 x Tris, Borate EDTA). Ici, la séparation a été effectuée à une puissance de 6W pendant 24 heures. Avec les conditions semi-dénaturantes (4M d'urée et gel à température ambiante), les liaisons hydrogènes entre les bases des ARN sont partiellement reformées, permettant la séparation des ARN dans une seconde dimension suivant la conformation secondaire partielle. Une fois la séparation terminée, le gel a été plongé dans une solution de bromure d'éthidium afin d'observer les spots d'ARN sous lumière UV. La Figure 60 montre la séparation 2D PAGE obtenue après



révélation UV (302 nm). Sur cette figure, on observe plusieurs bandes d'ARN dont l'identification de leur contenu a été réalisée par CZE-MS/MS. Pour cela, les différents spots de gel contenant des ARN ont été découpés et les bouts de gel ont été séchés sous vide. Une fois séchés, il est possible de conserver les bouts de gel à -20°C ou bien de réaliser directement les digestions enzymatiques. Dans la partie suivante, le protocole de digestion *in gel* des ARN est décrit ainsi que le dessalage des échantillons avant analyse CE-MS/MS.

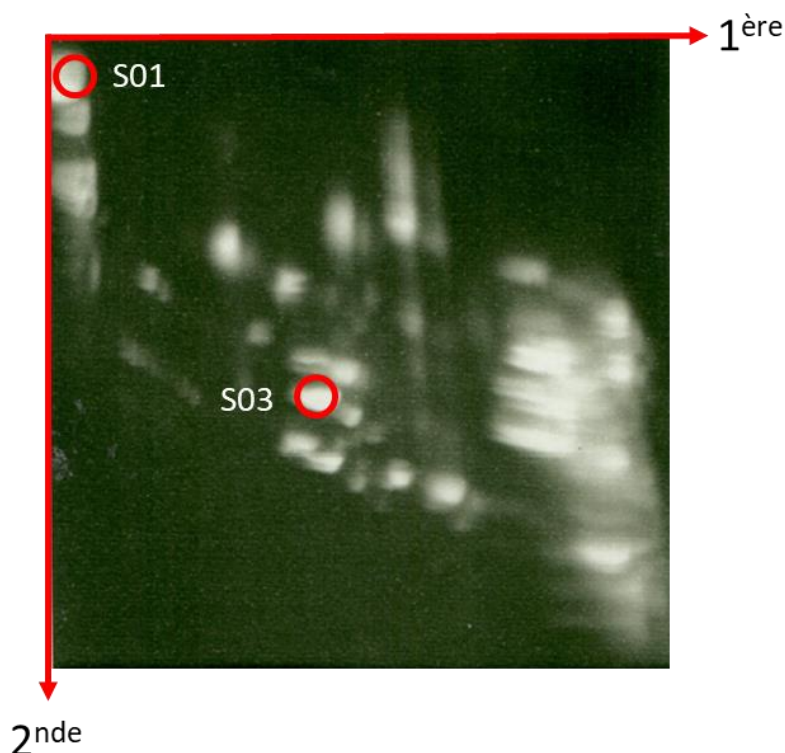


Figure 60 Isolation sur gel de polyacrylamide de l'extrait total d'*E. coli* (20 µg). Les ARN ont été révélés par lumière UV grâce à la coloration au bromure d'éthidium. Les flèches rouges indiquent le sens de migration des composés. Les cercles rouges indiquent les morceaux de gel qui ont été analysés dans ces travaux.

### 8.2.2 Digestions enzymatiques et dessalage des échantillons

Les digestions des ARN ont été précédemment réalisées en solution. Ici, ce sont les mêmes digestions qui ont été effectuées *in gel* [2], avec la RNase T<sub>1</sub> et la RNase A accompagnées de BAP. Les morceaux de gel, qui ont été auparavant séchés, ont été réhydratés dans les solutions contenant les enzymes. Cette réhydratation permet d'agrandir les mailles du gel d'acrylamide et de libérer les ARN piégés. Les ARN ont ensuite été digérés par l'action des enzymes pour obtenir les hydrolysats à analyser. Toutefois, avant l'analyse CE-MS/MS, il a été nécessaire de réaliser une étape de dessalage de l'échantillon afin de retirer les ions non-volatils qui sont incompatibles avec l'ionisation ESI.

Dans le but de retirer les sels non-volatils, c'est un dessalage par extraction en phase solide qui a été effectué. En effet, le dessalage en utilisant des pointes de pipette (10 µL) avec un lit de résine C<sub>18</sub> est tout à fait adapté pour les petits volumes de digestion. Pour réaliser ce type de dessalage, la résine a été dans un premier temps activée avec un mélange acétonitrile/eau 50/50 (v/v) suivi d'une solution d'AcONH<sub>4</sub> 200 mM. Ensuite, les digests d'ARN ont été immobilisés sur la phase solide, en veillant à ne pas prélever des morceaux de gel.

Après, les digests d'ARN ont été dessalés en échangeant les cations alcalins avec des cations ammoniums qui sont compatibles avec l'ionisation ESI. Le dessalage a été réalisé avec la solution d'AcONH<sub>4</sub> 200 mM (au moins 10 fois). Les hydrolysats, ainsi dessalés, ont été élués de la phase stationnaire avec le mélange acétonitrile/eau 50/50 (v/v) puis l'échantillon a été évaporé à sec. Enfin, l'échantillon a été resolubilisé dans le tampon d'analyse (AcONH<sub>4</sub> 200 mM pH4) et a été directement analysé par CZE-MS/MS.

### 8.3 Analyse et interprétation des données

Afin d'identifier les ARN, les digests déphosphorylés de RNase T<sub>1</sub> et RNase A ont été analysés par CZE-MS/MS dans les conditions décrites précédemment (cf. 5.4). Les pics de l'analyse ont été séquencés manuellement grâce aux spectres MS/MS. Les oligonucléotides non-modifiés séquencés avec les analyses du morceau de gel S03 sont disponibles dans le Tableau 12.

RNase T <sub>1</sub>	RNase A
CG	GAAGGU
AAG	GGC
UG	AGAC
CUAG	AGC
ACG	AGU
CUUAG	GGAC
UUAG*	AAGU
CACCA	

Tableau 12 Tableau regroupant les oligonucléotides non-modifiés séquencés du spot S03 grâce aux analyses CZE-MS/MS. À gauche, issus de la digestion à la RNase T<sub>1</sub> et à droite, issus de la digestion à la RNase A. \*Il s'est avéré plus tard que cet oligonucléotide était PUAG.

Une fois les oligonucléotides séquencés, les séquences obtenues ont été soumises à un moteur de recherche : MassSpec-Toolkit for RNAs, un outil développé par l'IBMC. Parmi d'autres fonctionnalités, cet outil a été conçu pour rechercher à quels ARN peuvent appartenir les oligonucléotides séquencés. Pour cela, il est possible de faire une recherche avec les masses des ions parents ou bien il est aussi possible d'indiquer les séquences des oligonucléotides. Il est aussi possible de renseigner le type d'ARN, le domaine, l'espèce et l'enzyme utilisée. Une fois les paramètres indiqués, l'outil effectue la recherche et classe les ARN par pertinence en attribuant un score, 1.0 étant le score le plus élevé. Un score de 1.0 indique que tous les oligonucléotides renseignés sont présents dans l'ARN. Bien qu'il soit possible qu'un spot de gel contienne plusieurs ARN, l'identification d'un ARN est vérifiée grâce à l'analyse des oligonucléotides des deux digestions. Néanmoins, cet outil ne prend pas en charge les nucléotides modifiés et, à ce jour, uniquement les séquences d'ARN de 6 espèces sont disponibles (dont *E. coli*). Les résultats de la recherche obtenus pour le spot S03 digéré par la RNase T<sub>1</sub> est disponible sur la Figure 61.

Submitted Data			
Data type	subsequences		
Values	AAG;ACG;CG;CUUCAG;CACCA;UG;CUAG;UUAG		
RNA type	tRNA		
Include seqs w/ mods	at position 34: no AND at position 37: no		
Domain	All		
Species	All		
Enzyme	RNaseT1		
Mode	positive		

=> 496 matching RNA sequences:

Score	RNA	Sequence	Species	Domain
1.0	Leu (CAG)	GCGAAGGUGGCGGAAUUGGUAGACGCGCUAGCUUCAGGUGUUAGUGUCCUACGGACGUGGGGUUCAAGUCCCCCCCCUCGCACCA	Escherichia coli str. K-12 substr. MG1655	Bacteria
1.0	Leu (CAG)	GCGAAGGUGGCGGAAUUGGUAGACGCGCUAGCUUCAGGUGUUAGUGUCCUACGGACGUGGGGUUCAAGUCCCCCCCCUCGCACCA	Escherichia coli str. K-12 substr. MG1655	Bacteria
0.75	Arg (CCU)	GUCCUCUAGUUAAUUGGAAUUAACGAGCCCCUUAAGGGCUAAUUGCAGGUUCGAUUCUCCAGGGGACACCA	Escherichia coli str. K-12 substr. MG1655	Bacteria
0.75	Gly (GCC)	GCGGUGGUAGUCUAGCCUGGUCUAGGACGCCACCCUGCCAAGGUGGAGACCCGGGUUCAAAUCCCGGCCACCCACCA	Pyrococcus furiosus COM1	Archaea
0.75	Val (GAC)	GGGCCCGUGGUCUAGACUGGUUAUGACGCCACCCUGACAAGGUGGAGGUCCGGGUUCAAAUCCCGGCCACCCACCA	Pyrococcus furiosus COM1	Archaea

Figure 61 Résultats de la recherche d'ARN avec les séquences d'oligonucléotides déterminées par CZE-MS/MS sur l'analyse de l'échantillon du morceau de gel S03 digéré par la RNase T<sub>1</sub> et la BAP.

Ainsi, dans le spot S03, c'est l'ARNt<sup>Leu</sup><sub>CAG</sub> (*E. coli*) qui a été identifié avec le score le plus élevé. La même recherche a été réalisée avec les oligonucléotides issus de la RNase A et les résultats ont permis de confirmer, également avec un score de 1.0, qu'il s'agissait bien de l'ARNt<sup>Leu</sup><sub>CAG</sub> (*E. coli*).

Concernant les modifications, elles ont également pu être localisées avec précision grâce aux spectres MS/MS. En revanche, la caractérisation par MS/MS ne permet pas de discerner des nucléotides modifiés présentant la même masse (par exemple U et ψ). L'identification préalable de l'ARN a permis de rechercher les types de modifications dans la banque de données Modomics, afin d'assigner correctement les types de modifications. Les localisations de ces modifications ont également été comparées et vérifiées avec la base de données. Ainsi, en compilant les résultats des analyses CZE-MS/MS des digests de RNase T<sub>1</sub> et de RNase A, un recouvrement de séquence de 65 % a été obtenu (les séquences identifiées sont soulignées).

GCGAAGGUGGCGGAA[D][D][Gm]G[D]AGACGCGCUAGCUUCAG[m<sup>1</sup>G][ψ]G[ψ]UAGUGUCCUU  
ACGGACGUGGGG[m<sup>5</sup>U][ψ]CAAGUCCCCCCCCUGACCA

Ce recouvrement de séquence, relativement faible, s'explique par l'enchaînement de nucléotides C et U (surligné en rouge dans la séquence). En effet, avec une digestion à la RNase T<sub>1</sub>, cela produit des oligonucléotides à chaîne longue (par exemple UCCCCCCCCUCG) [3]. Or, les oligonucléotides longs sont difficilement détectés par la méthode CZE-MS/MS. D'un autre côté, cette répétition de nucléotides dans la séquence génère, avec la RNase A, des nucléotides qui ne sont d'aucune utilité dans le séquençage [4]. Toutefois, dans ce cas, les digestions ont produit suffisamment d'oligonucléotides spécifiques pour permettre l'identification de l'ARN et les nucléotides modifiés ont été localisés avec succès sur la structure primaire de l'ARN.

Par la suite, le spot S01 a également été préparé de la même manière et analysé par CZE-MS/MS. L'exploitation des données a permis de séquencer les oligonucléotides regroupés dans le Tableau 13 :

RNase T <sub>1</sub>	RNase A
UG	GC
CCUG	GGC
CCG	GU
UAG	AGC
AAG	GGU
AAACG	AC
AUG	GAC
AG	AU
AACUG	GAAC
CCAG	GAAAC
CAU	GAU
	AGU
	AGGC

Tableau 13 Tableau listant les oligonucléotides séquencés du spot S01 grâce aux analyses CZE-MS/MS. À gauche, issus de la digestion à la RNase T<sub>1</sub> et à droite, issus de la digestion à la RNase A.

Dans l'analyse des digests de la RNase T<sub>1</sub>, l'oligonucléotide CAU a été détecté et séquencé. Il s'agit de l'oligonucléotide situé à l'extrémité 3' de l'ARN puisque sa séquence ne se termine pas par un G (normalement produit par une digestion à la RNase T<sub>1</sub>). Or, les séquences d'ARNt se terminent, dans la très grande majorité des cas, par CCA. De plus, sur les oligonucléotides détectés, aucune modification n'a été localisée alors que les ARNt présentent systématiquement des modifications. Ainsi, l'ARN présent dans ce morceau de gel est nécessairement un ARN d'un autre type.

Les résultats de la recherche ont indiqué que l'ARN contenu dans ce morceau de gel était l'ARNr 5S (*E. coli*). C'est un ARN composé de 120 nucléotides et qui ne porte pas de modification [5]. Ainsi, le recouvrement de séquence suivant a été obtenu en compilant les résultats des analyses des deux digests (les séquences identifiées sont soulignées) :

UGCCUGGCGGCCGUAGCGCGGUGGCCACCUGACCCCAUGCCGAACUCAGAAGUGAAACGCCG  
UAGCGCCGAUGGUAGUGUGGGGUCUCCCAUGCGAGAGUAGGGAACUGCCAGGCAU

Ici, un faible recouvrement de séquence (31%) a été obtenu. En effet, plusieurs oligonucléotides possédant une courte chaîne sont retrouvés à plusieurs reprises dans la séquence et n'ont pas été pris en compte puisqu'ils ne sont pas spécifiques. Cet ARN ne porte pas de modification, il est alors plus probable de générer des oligonucléotides de même séquence. De plus, des oligonucléotides longs, dont le séquençage n'a pas pu être réalisé sans ambiguïté par MS/MS, n'ont pas été comptabilisés malgré leur détection en MS. Cependant, un nombre suffisant d'oligonucléotides spécifiques ont été détectés pour identifier l'ARN avec certitude.

Ainsi, les résultats préliminaires, obtenus avec la méthode mise en place pour étudier les ARN issus de matrices biologiques, ont démontré l'utilité de la CZE-MS/MS pour l'identification et la caractérisation des modifications. Néanmoins, il est essentiel d'évaluer la méthode sur des morceaux de gel contenant plusieurs ARN afin de déterminer la possibilité d'identifier plusieurs ARN et également d'étudier les modifications de chacun des ARN. De

plus, il est également nécessaire de tester l'analyse des morceaux de gel au premier niveau. Pour cela, la digestion de l'échantillon sera également accomplie *in gel* puis le dessalage de l'échantillon sera réalisé avant l'analyse CZE-MS/MS. Enfin, l'interprétation des données sera réalisée pour identifier les nucléosides en analysant les pics obtenus avec les masses des ions parents et des ions fils.

#### 8.4 Références bibliographiques

1. Antoine, L., et al., *Mapping post-transcriptional modifications in Staphylococcus aureus tRNAs by nanoLC/MSMS*. Biochimie, 2019. **164**: p. 60-69.
2. Taoka, M., et al., *In-Gel Digestion for Mass Spectrometric Characterization of RNA from Fluorescently Stained Polyacrylamide Gels*. Analytical Chemistry, 2010. **82**(18): p. 7795-7803.
3. Wagner, T.M., et al., *A novel method for sequence placement of modified nucleotides in mixtures of transfer RNA*. Nucleic Acids Symposium Series, 2004. **48**(1): p. 263-264.
4. Solivio, B., et al., *Improving RNA modification mapping sequence coverage by LC-MS through a nonspecific RNase U2-E49A mutant*. Analytica Chimica Acta, 2018. **1036**: p. 73-79.
5. Correll, C.C., et al., *Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain*. Cell, 1997. **91**(5): p. 705-712.

# CONCLUSION GENERALE ET PERSPECTIVES

Les ARN sont des espèces essentielles dans une cellule puisqu'ils assurent plusieurs rôles comme la transmission et la traduction de l'information génétique. L'ARN est une copie d'ADN qui permet de produire une protéine fonctionnelle. Afin de rendre l'ARN opérationnel, il est courant que celui-ci porte des modifications chimiques couramment appelées modifications post-transcriptionnelles. L'introduction de ces modifications sur un ARN met en jeu des processus enzymatiques complexes. Ces modifications naturelles apportent les propriétés physico-chimiques nécessaires pour garantir le bon fonctionnement de l'ARN. De nombreuses études ont démontré que la nature mais aussi la position des modifications sur l'ARN sont primordiales. La caractérisation des modifications est donc nécessaire pour comprendre l'influence de celles-ci sur l'ARN. Plusieurs outils analytiques ont alors été développés pour étudier les modifications avec notamment des méthodes biologiques (par exemple PCR) et spectrales (par exemple LC-MS). L'étude des modifications d'ARN reste un challenge pour les sciences analytiques et le développement de méthodes orthogonales est nécessaire afin de valider des informations et éventuellement pour les compléter.

L'objectif principal de cette thèse était alors de développer une méthode d'analyse simple permettant la caractérisation des modifications post-transcriptionnelles des ARN à l'aide du couplage CE-MS.

La première partie de cette thèse s'est focalisée sur l'étude bibliographique des ARN et des modifications post-transcriptionnelles. Cette partie permet de placer le contexte de cette thèse. Premièrement, elle fournit les notions de base permettant la compréhension de la chimie et de la biologie des ARN. Elle permet notamment de décrire et de comprendre les propriétés chimiques des ARN qui ont été exploitées pour justifier l'analyse par CE-MS mais aussi pour développer les séparations électrophorétiques. Par la suite, les rôles des principaux ARN ainsi que leurs mécanismes de productions sont décrits. Une autre partie discute des modifications post-transcriptionnelles des ARN. Les intérêts des modifications et leurs conséquences dans les propriétés des ARN sont expliqués en trois parties et illustrés avec plusieurs exemples. Enfin, un état de l'art, sur les méthodes existantes basées sur la MS et les stratégies actuelles utilisées pour étudier les modifications, est dressé. Aujourd'hui, la technique de référence pour étudier les modifications est le couplage LC-MS. Néanmoins, la CE-MS apparaît comme une méthode tout à fait adaptée pour l'analyse des modifications. En effet, cette méthode a déjà prouvé son efficacité dans la séparation et la détection de biomolécules. Par conséquent, une méthode d'analyse des modifications d'ARN basée sur le couplage CE-MS a été proposée.

Par la suite, la deuxième partie du manuscrit était consacrée au développement de la méthode CE-MS pour caractériser les modifications d'ARN. La stratégie d'analyse qui a été choisie repose sur des approches déjà établies pour l'analyse LC-MS avec l'utilisation de plusieurs enzymes (approche sur deux niveaux). Ces deux niveaux d'analyses permettent d'obtenir des informations complémentaires sur les modifications telles que leur nature et

aussi leur position dans la séquence de l'ARN. Le premier niveau consistait à effectuer l'hydrolyse totale de l'ARN. L'analyse de l'hydrolysate a permis d'identifier les modifications portées par l'ARN. Dans le second niveau, il s'agissait de réaliser la digestion partielle de l'ARN. Cette approche « bottom-up » permet de produire des oligonucléotides. Les informations générées par l'analyse CE-MS/MS des oligonucléotides permettent de reconstruire les séquences mais également de localiser les nucléotides modifiés sur ces séquences. Ici, c'est une double digestion individuelle des ARN qui a été réalisée afin d'obtenir un plus grand recouvrement de séquence et ainsi permettre la localisation d'un plus grand nombre de modifications. La séparation électrophorétique et la détection MS ainsi que la fragmentation MS/MS ont été mises en œuvre pour analyser les digests sur ces deux niveaux et en mode positif. Pour cela, plusieurs paramètres ont été ajustés dans le but d'obtenir une séparation efficace et résolutive des digests d'ARN permettant de simplifier l'interprétation des spectres de masse. Lors de l'analyse des digests partiels, des pics peu résolutifs correspondant aux oligonucléotides à chaîne longue ont été obtenus. Ainsi, dans le but d'améliorer la migration de ces composés, l'hydrolyse enzymatique des phosphates 3' des oligonucléotides a été mise en place. Cette déphosphorylation des digests a permis d'améliorer significativement la qualité de séparation. Un avantage de la méthode est que les analyses des échantillons du premier et second niveau sont accomplies dans les mêmes conditions de séparation et en utilisant une seule plateforme d'analyse. Ainsi, aucune adaptation de la méthode de séparation n'est requise pour l'analyse des deux niveaux contrairement à l'analyse par LC-MS. Ensuite, un protocole d'interprétation des résultats a été mis en place. En effet, les logiciels n'étant pas adaptés pour exploiter les résultats dans nos conditions d'analyses, les spectres bruts ont été interprétés manuellement. Les premiers résultats ont montré la capacité de la CE à séparer de façon efficace les composés d'intérêt et donc la possibilité d'obtenir des informations sur les nucléotides modifiés d'un ARN.

Enfin, dans la dernière partie du manuscrit, l'objectif était de démontrer que la méthode développée était adaptée pour caractériser les modifications d'ARN. Afin de mettre en évidence la pertinence de la méthode CE-MS pour étudier les modifications, quatre ARN modèles ont été analysés. Les analyses ont été accomplies sur les deux niveaux puis les données ont été interprétées manuellement. Finalement, grâce à la méthode CE-MS, plusieurs modifications ont été identifiées et localisées sur les séquences d'ARN. Des recouvrements de séquences allant jusqu'à 97% ont été obtenus. De plus, les résultats expérimentaux étaient cohérents avec les informations issues d'une banque de données. Ces résultats ont donc démontré la possibilité de caractériser les nucléotides modifiés des ARN par CE-MS/MS. Ainsi, la suite des travaux a été d'évaluer cette méthode sur des ARN issus d'échantillons biologiques. Pour cela, c'est un extrait total d'ARN provenant d'*E. coli* qui a été utilisé. Le but a été de cartographier les ARN présents dans l'échantillon et aussi de caractériser individuellement les modifications de chaque ARN. Dans un premier temps, les ARN de cet extrait ont été isolés à l'aide de la PAGE 2D. Par la suite, les ARN isolés ont été digérés *in gel* en adaptant le protocole de préparation d'échantillon. Les digests ont ensuite été analysés par la méthode CZE-MS/MS après avoir été dessalés. Les résultats préliminaires ont montré la possibilité d'identifier les ARN isolés et également d'étudier avec précision les nucléotides modifiés des ARN.

La méthode décrite dans cette thèse a permis la caractérisation de plusieurs modifications d'ARN. Toutefois, l'analyse des oligonucléotides par cette méthode présente certains défauts comme un temps d'analyse long mais aussi la difficulté de détecter les oligonucléotides longs ou trop acides. Ces composés, non-détectés, diminuent alors le recouvrement de séquence et de ce fait certaines modifications risquent de ne pas être localisées. Par conséquent, des perspectives sont envisagées dans le but d'améliorer l'analyse des oligonucléotides. En effet, dans cette thèse, certains paramètres n'ont pas été évalués avec par exemple l'utilisation d'un BGE basique. Avec une analyse réalisée en condition basique, la déphosphorylation n'est pas nécessaire puisque les phosphates vont contribuer à augmenter la mobilité des oligonucléotides. Une autre possibilité est de mettre en place une technique d'isotachophorèse transitoire. Cette technique permet la préconcentration en ligne des analytes. Cela permet d'améliorer significativement l'efficacité des pics et donc l'intensité de détection. Il est également possible d'évaluer l'apport du greffage des parois internes du capillaire. Il existe plusieurs types de greffage tels que des greffages positifs, neutres ou super hydrophobes qui pourraient être évalués pour la séparation d'oligonucléotides. Dans ces travaux, uniquement l'électrophorèse capillaire de zone a été utilisée. Il pourrait être intéressant d'évaluer les autres modes de séparation compatibles ESI-MS. Enfin, il s'agira également de comparer cette méthode CE-MS avec les méthodes LC-MS existantes.

Par la suite, l'analyse des modifications d'autres familles d'ARN telles que des ARN ribosomiques est envisagée. Étant donné que ces types d'ARN sont composés d'un nombre important de nucléotides, le développement d'un logiciel d'exploitation permettrait d'interpréter automatiquement les analyses CE-MS/MS et de rendre cette étape beaucoup moins chronophage et fastidieuse.

Ainsi, la méthode CE-MS/MS développée dans cette thèse a permis d'identifier et de localiser les modifications post-transcriptionnelles des ARN. Cette méthode vient donc s'ajouter comme nouvel outil pour l'exploration et l'étude des modifications des ARN. La CE-MS/MS permettrait éventuellement de découvrir des modifications encore inconnues à ce jour. Elle pourrait également caractériser les intermédiaires des modifications pour comprendre les enzymes impliquées dans la création des modifications.

Lors de cette thèse, d'autres projets ont été abordés avec la quantification absolue d'anticorps monoclonaux dans des sérums biologiques par CE-MS/MS et aussi la mise en place d'un système d'interface CE-MS dite « nanoflow sheath liquid ». Brièvement, le projet de quantification propose une méthode permettant de titrer de façon sélective des anticorps d'intérêts. Ces protéines sont extraites sélectivement à l'aide de billes magnétiques sur lesquels sont fixés des récepteurs. Ensuite, un standard interne marqué correspondant à l'anticorps est ajouté. La digestion enzymatique est réalisée puis les digests sont analysés par CE-MS/MS. La quantification est ensuite réalisée sur des peptides spécifiques à l'anticorps et grâce aux standards internes. Les premiers résultats obtenus avec des anticorps en solution ont montré une très bonne linéarité entre concentration et réponse du détecteur. La suite du projet sera d'évaluer la méthode sur des sérums sains dopés avec l'anticorps. L'autre projet consistait à mettre en œuvre une interface CE-MS nanoflow sheath liquid. En effet, le laboratoire ne possède que l'interface sheathless. Cette interface présente des inconvénients



comme la robustesse relativement faible et la difficulté de greffer les capillaires de faible diamètre interne. Ainsi, une interface nanoflow sheath liquid a été mise en place afin de réaliser des séparations CE-MS en utilisant des capillaires greffés tout en améliorant la robustesse. De plus, cette interface permet de réduire les effets de dilution comparé à l'interface « sheath liquid » classique et donc d'améliorer la sensibilité. Après ajustement des paramètres, les résultats obtenus sur un digest tryptique d'une protéine ont permis d'obtenir 91 % de recouvrement de séquence.



# LISTES DES COMMUNICATIONS SCIENTIFIQUES

## Publications avec comité de lecture

*Monoclonal Antibody N-Glycosylation Profiling using Capillary Electrophoresis – Mass Spectrometry: Assessment and Method Validation*, J. Giorgetti, V. D'Atri, J. Canonge, **A. Lechner**, D. Guillaume, O. Colas, E. Wagner-Rousset, A. Beck, E. Leize-Wagner, YN. François, *Talanta*, 2018, **178**, p. 530-537

DOI : 10.1016/j.talanta.2017.09.083

*Intact monoclonal antibodies separation and analysis by sheathless capillary electrophoresis-mass spectrometry*, J. Giorgetti, **A. Lechner**, E. Del Nero, A. Beck, YN. Francois, E. Leize-Wagner, *Eur. J. Mass Spectrom.*, 2018, **25**, p. 324-332

DOI : 10.1177/1469066718807798

*Insights from capillary electrophoresis approaches for characterization of monoclonal antibodies and antibody drug conjugates in the period 2016-2018*, **A. Lechner**, J. Giorgetti, R. Gahoual, A. Beck, E. Leize-Wagner, YN. Francois, *J. Chromatogr. B*, 2019, **1122**, p. 1-17

DOI : 10.1016/j.jchromb.2019.05.014

*Characterization of post-transcriptional RNA modifications by sheathless capillary electrophoresis-high resolution mass spectrometry*, **A. Lechner**, P. Wolff, E. Leize-Wagner, YN. François, *Anal. Chem.*, 2020, **92**, p. 7363-7370

DOI : 10.1021/acs.analchem.0c01345

## Communications orales

**Auteurs :** Antony Lechner, Philippe Wolff, Yannis-Nicolas François, Emmanuelle Leize-Wagner

**Titre :** « Mapping of post-transcriptional RNA modifications by capillary electrophoresis-tandem mass spectrometry »

**Congrès :** Rencontres franco-allemand Strasbourg (Présenté à Strasbourg le mardi 09 décembre 2019)

**Auteurs :** Antony Lechner, Philippe Wolff, Yannis-Nicolas François, Emmanuelle Leize-Wagner

**Titre :** « Structural characterization of RNA post-transcriptional modifications by CE-MSMS »

**Congrès :** Journée des doctorants Strasbourg (Présenté à Strasbourg le mardi 03 décembre 2019)

**Auteurs :** Antony Lechner, Emine Sakka, Yannis-Nicolas François, Emmanuelle Leize-Wagner

**Titre :** « Implementation and evaluation of a homemade nanoflow interface for capillary electrophoresis-mass spectrometry coupling »

**Congrès :** Congrès national de spectrométrie de masse – analyse protéomique SMAP 2019 Strasbourg (Présenté à Strasbourg le mardi 17 septembre 2019)

**Auteurs :** Jérémie Giorgetti, Antony Lechner, Elise Del Nero, Alain Beck, Emmanuelle Leize-Wagner, Yannis-Nicolas François

**Titre :** « Mise au point d'un greffage cationique en CE-MS pour l'analyse d'anticorps monoclonaux »

**Congrès :** journée scientifique de l'AfSep (Présenté à Illkirch-Graffenstaden le jeudi 7 juin 2018)

**Auteurs :** Antony Lechner, Philippe Wolff, Yannis-Nicolas François, Emmanuelle Leize-Wagner

**Titre :** « Development of a capillary electrophoresis-mass spectrometry method for RNA characterization.»

**Congrès :** journée des doctorants et master (Présenté à Strasbourg le mardi 17 avril 2018)

**Auteurs :** Jérémie Giorgetti, Antony Lechner, Elise Del Nero, Alain Beck, Yannis-Nicolas Francois, Emmanuelle Leize-Wagner,

**Titre :** « Development of a capillary electrophoresis-mass spectrometry method for monoclonal antibodies characterization »

**Congrès :** Rencontre du Club Jeune de Spectrométrie de Masse RCJSM 2018 (Présenté à Saint-Martin-de-Londres le mardi 27 mars 2018)

### **Communications par affiche**

**Auteurs :** Antony Lechner, Jérémie Giorgetti, Julie Canonge, Valentina D’Atri, Davy Guillarme, Olivier Colas, Elsa Wagner-Rousset, Alain Beck, Emmanuelle Leize-Wagner, Yannis-Nicolas François

**Titre :** « Monoclonal antibody Fc-glycosylation profiling using a capillary electrophoresis-mass spectrometry coupling »

**Congrès :** Congrès national de méthode SEParative SEP Paris 2019 (Présenté à Paris du 26 au 28 mars 2019)

**Auteurs :** Rabah Gahoual, Antony Lechner, Philippe Wolff, Yannis-Nicolas François, Emmanuelle Leize-Wagner

**Titre :** « Sheathless CE-MS interface as an original and powerful infusion platform for nanoESI study: from intact proteins to high molecular mass noncovalent complexes »,

**Congrès :** European Mass Spectrometry Convention EMSC Saarbruck 2018 (Présenté à Saarbruck le mardi 13 mars 2018)

### **Enseignements délivrés en tant que moniteur (Université de Strasbourg)**

Années universitaires 2017-2018 ; 2018-2019 et 2019-2020

Travaux dirigés module « Chimie pour les Sciences du Vivant/Atomistique » (niveau Licence 1 Sciences de la Vie)

Travaux pratiques de « chimie analytique » (niveau Licence 3 Chimie biologie)

### **Autre**

Elu représentant des doctorants du campus Esplanade à l’ED 222 de 2017 à 2020



### LC-MS Information

Sum formula	C <sub>10</sub> O <sub>5</sub> N <sub>3</sub> H <sub>15</sub>
Monoisotopic mass	257.1012
Average mass	257.2467
[M+H] <sup>+</sup>	258.1082
Product ions	112
Normalized LC elution time *	0.90 (Kellner 2014); 0.92 (Kellner 2014)
LC elution order/characteristics	between U and G (Kellner 2014, Kellner 2014)

Fichier Excel :

	A	B	C	D	E
31	Nom du fichier d'analyse				
32	Date d'analyse				
33	M+H	MS2	t(min)	HTML name	Short name
34	244.092	112.05		C	
35					
36	245.076	113.03		U	
37					
38	268.104	136		A	

## 1.2 Interprétation des data

### 1.2.1 Extraction des pics

#### Exemple : tRNA Trp CCA from *S. Cerevisiae* digéré avec RNase P<sub>1</sub> et BAP analysé en polarité MS +

Indiquer, sur la feuille Excel, le nom du fichier d'analyse et la date de l'analyse comme indiqué sur la figure ci-dessous :

Nom du fichier d'analyse					
Date d'analyse					
M+H	MS2	t(min)	HTML name	Short name	
244.092	112.05		C		

Nommer la feuille Excel comme indiqué sur l'image, suivant l'ARN analysé et l'enzyme utilisée → Clic droit sur l'onglet de la feuille puis renommer (Organisme, Sous-type, Anticodon, Enzyme).

16.2 "	m1A
LE Trp CCA P1BAP	

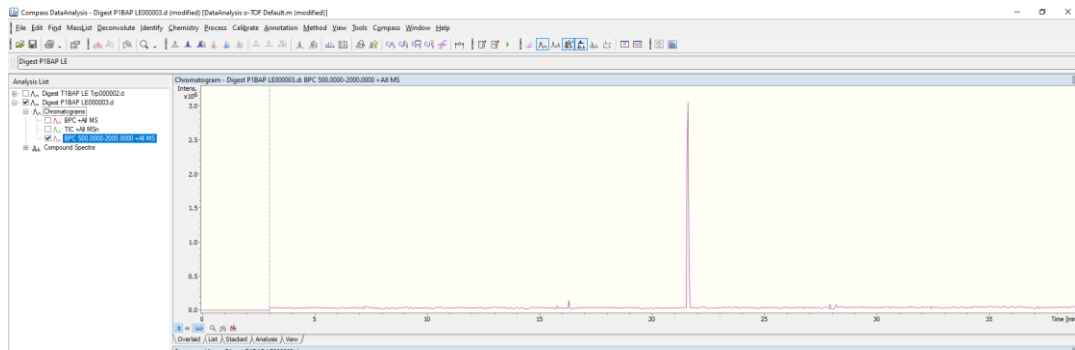
Ouvrir le logiciel DataAnalysis et ouvrir le fichier à analyser.

**Pour Calibration ext** (si nécessaire mais vivement recommandé).

Sur DataAnalysis :

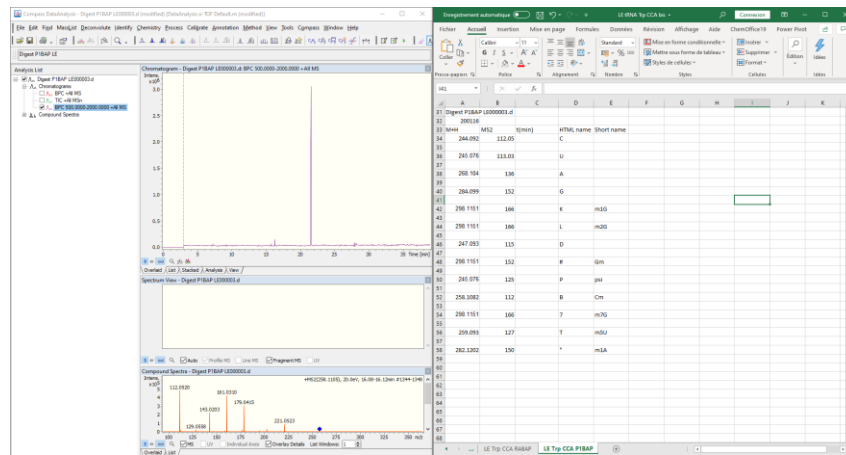
-Calibrate → External → Load et choisir un fichier de calibration → OK.

Puis (pour indiquer que le fichier est calibré) F7 → Type : BPC Filter : All MS Scan mode : All Masses : 500-2000 Width : +/- 0.02 → Cliquer sur Add et OK.



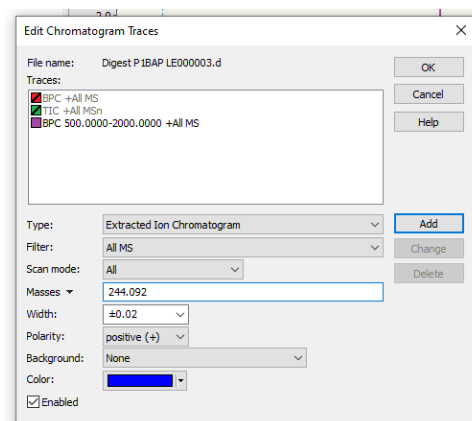
Ensuite extraire les pics précédemment simulés.

Pour cela afficher DataAnalysis et la feuille Excel sur l'écran voir figure ci-dessous :



Sur DataAnalysis → F7 et mettre les paramètres comme indiqué sur l'image.

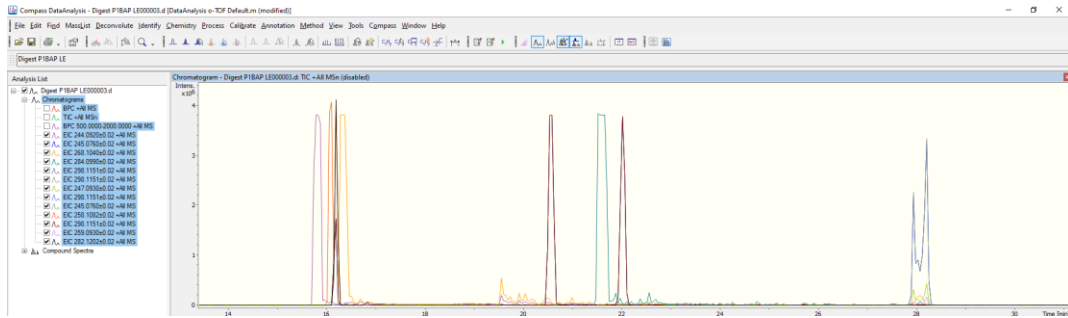
Copier et coller le m/z d'un nucléoside dans la case Masses :



Cliquer sur Add et répéter l'opération pour tous les m/z et tous les nucléosides.

Une fois toutes les masses rentrées cliquer sur OK et les pics correspondant aux nucléosides sont affichés.





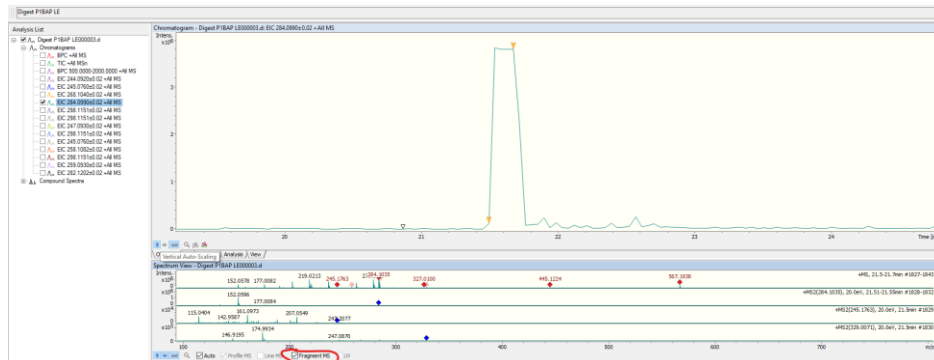
Vérifier la présence des m/z pour chaque nucléosides (surligner en vert si présent sinon en rouge) et noter les temps de migrations.

**!/ Attention il est possible que des nucléosides modifiés possèdent la même masse → pour attribuer les pics : MS/MS ou temps de migration d'un standard.**

Vérifier également la présence des ions fils spécifiques à chaque nucléosides.

Pour cela, sur DataAnalysis, dans la colonne Analysis List, cocher le pic à vérifier (ici G m/z 284 et ions fils : 152 m/z) et intégrer le pic en maintenant clic droit au début du pic et en relâchant à la fin du pic.

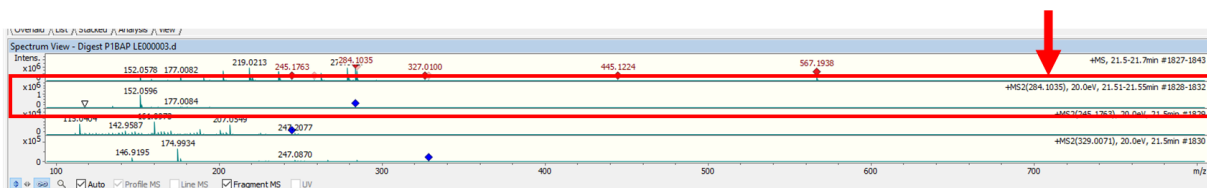
Dans la fenêtre Spectrum View il faut que Fragment MS soit coché pour afficher les spectres MS<sup>2</sup> :



Le premier spectre de cette fenêtre correspond au spectre MS du pic intégré (indiqué en haut à droite du spectre) et les carrés rouges au-dessus des pics indiquent que le spectro a effectué de la MS<sup>2</sup> sur ces masses.

Les spectres suivants sont les spectres MS<sup>2</sup> du pic intégré (ion précurseur indiqué en haut à droite du spectre).

Chercher le spectre MS<sup>2</sup> de l'ion à séquencer (ici 284.0990) :



Clic droit sur ce spectre et Copy to Compound Spectra → Le spectre MS<sup>2</sup> s'affiche dans la fenêtre Compound Spectra.



## Mongo Oligo Mass Calculator v2.08

Sequence (5'-->3')

GAAGCGGUKLCUCAA#GDAGAGCPPCGABUBCAAAPCGAAGG7D  
 UGCAGGTPCA"UUCUGPCCGUUUCACCA

DNA2RNA RNA2DNA CleanUp Clear sequence

Calculations: molecular mass, electrospray series, CID fragments, internal fragments, base losses, **RNase T1 digest**, RNase O2 digest, RNase A digest, 5' exonuclease digest, 3' exonuclease digest, modified bases list

Options:  average mass,  monoisotopic mass,  negative mode,  positive mode,  DNA,  RNA,  allow undercut

Terminals:  3'-OH,  5' phosphate,  3'-OH,  3' phosphate,  3' cyclic phosphate

Memories: GAAGCGGUKLCUCAA... copy paste clear all

User residue 'N': base anion mass, sugar add mass

Output:  close button,  output summary, sequence by  3  10  50,  extended output,  show messages,  sorted output

Calculate Clear Form Help

written by [Jef Rozenski](#) (1999)

Cliquer sur Calculate et une fenêtre apparaît avec les séquences et masses (3'p et 3>p) des oligonucléotides issus d'une digestion T<sub>1</sub> :

Sequence :

GAA GCG GUK LCU CAA D#G DAG AGC PPP CGA BUB CAA APC GAA GG7 DUG CAG  
 GTP CA" UUC CUG PCC GUU UCA CCA

---

5'OH - RNA [75mer] - 3'OH  
 C:16 U:10 T:1 A:16 G:16 #:1 B:2 ":1 7:1 D:3 K:1 L:1 P:6  
 monoisotopic mass, positive mode

RNASE T1 DIGEST

pos	mass	3'>p	sequence
G1:G1	363.057	345.047	Gp
A2:G4	1021.161	1003.151	AAp
C5:G6	668.098	650.088	Cp
G7:G7	363.057	345.047	Gp
U8:G18	3628.522	3610.512	U <sup>1</sup> LCUCAA#Gp
D19:G21	1000.150	982.140	Dp
A22:G23	692.109	674.099	Ap
C24:G29	1891.214	1873.204	CPPCGp
A30:G40	3539.511	3521.501	ABUBCAAAPCGp
A41:G43	1021.161	1003.151	AAp
G44:G44	363.057	345.047	Gp
745:G48	1336.185	1318.175	7DUGp
C49:G51	997.150	979.140	CAGp
G52:G52	363.057	345.047	Gp
T53:G63	3494.441	3476.431	TPCA"UUCUGp
F64:G67	1279.164	1261.154	FCCGp
U68:A75	2429.346		UUUCACCA

Close

**⚠ Attention les masses (« mass ») correspondent aux espèces 3'-phosphorylés. Or avec une digestion enzyme + BAP les oligonucléotides sont déphosphorylés.**

Recopier les séquences des oligonucléotides sur une feuille Excel :

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Date de l'analyse													
2	Nom du fichier													
3	Sequence	ch	M+H	t(min)	MS <sup>2</sup>	GAAGCGGUKLCUCAAD#GDAGAGCPPPGABUBCAAAPCGAAGG7DUGCAGGTPCA*UUCUGPCGUUUCACCA								
4	G					24162.29 Da								
5														
6	AAG													
7														
8														
9	CG													
10														
11														
12	UKLCUCAAD#G													
13														
14	UKL	CL												
15														
16														
17	CUCAAD#G	CL												
18														
19														
20														
21	DAG													
22														
23														
24	AG													
25														
26														
27	CPPPCG													
28														
29														
30														
31	ABUBCAAAPCG													
32														
33														
34														
35														
36	7DUG													
37														
38														

**⚠ ATTENTION : Il est possible que des modifications soient clivées par les enzymes.**

RNase T1		RNase A	
Cleave	Ne cleave pas	Cleave	Ne cleave pas
G	K	C	B
I	#	U	T
L	7	D	
(R)		P	
		T	
		?	
		4	

D'après mes observations au cours de la thèse

Par exemple, ici MongoOligo a simulé le digest UKLCUCAAD#G mais sur les analyses UKL et CUCAAD#G ont été observés.

Dans Excel, mettre l'oligonucléotide simulé par Mongo et mettre en dessous les digests clivés (CL).

Ensuite simuler les m/z de tous les oligonucléotides avec MongoOligo.

Exemple avec AAG : Sélectionner « molecular mass » et « electrospray series » et cocher les options indiquées sur l'image :

## Mongo Oligo Mass Calculator v2.08

Sequence (5'-->3')

AAG

DNA2RNA RNA2DNA CleanUp Clear sequence

**Calculation**

- molecular mass
- electrospray series**
- CID fragments
- internal fragments
- base losses
- RNase T1 digest
- RNase U2 digest
- RNase A digest
- 5' exonuclease digest
- 3' exonuclease digest
- modified bases list

**Options**

- average mass
- monoisotopic mass
- negative mode
- positive mode
- DNA
- RNA
- allow undercut

**Terminals**

- 5'-OH
- 5' phosphate
- 3'-OH
- 3' phosphate
- 3' cyclic phosphate

**Memories**

GAAGCGGUKLCUCAA...

copy paste clear all

**User residue 'N'**

base anion mass

sugar add mass

**Output**

- close button
- output summary
- sequence by  3  10  50
- extended output
- show messages
- sorted output

Calculate Clear Form Help

written by [Jef Rozenski](#) (1999)

Cliquer sur Calculate et une fenêtre apparaît avec les masses ESI de l'oligonucléotide :

Sequence :  
AAG

5'OH - RNA[3mer] - 3'OH  
A:2 G:1  
monoisotopic mass, positive mode

**MOLECULAR MASS**

M = 941.195 [M+H]<sup>+</sup> = 942.203 [M-H]<sup>-</sup> = 940.187

**ELECTROSPRAY SERIES**

ch	m/z	M+Na	M+K	M+2Na
1	942.203	964.185	980.159	986.167
2	471.605	482.596	490.583	493.587

Close

Recopier les masses m/z (1<sup>ère</sup> colonne « Electro spray series ») sur la feuille Excel.

En pratique, avec la CE-MS polarité +, les oligonucléotides sont généralement chargés de 1+ à 3+ et plus rarement 4+.

**⚠ Sur Mongo le symbole ? ne correspond pas à m5C (on peut remplacer ? par B).**

3	Sequence	ch	M+H	t(min)	MS <sup>2</sup>	24162.29 Da								
4	G		284.099											
5														
6	AAG		942.203											
7			471.605											
8														
9	CG		589.14											
10			295.07395											
11														
12	UKLUCUAA#G													
13														
14	UKL	CL	963.202											
15			482.105											
16														
17	CUCAAD#G	CL	2525.414											
18			1263.21											
19			842.476											
20														

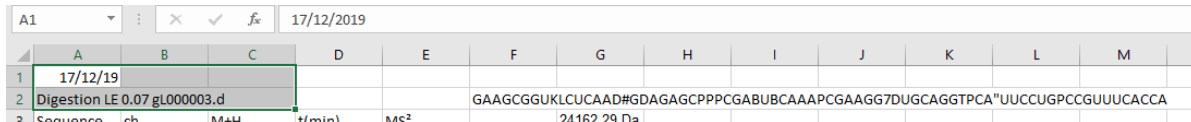
La digestion *in silico* est terminée.

## 2.2 Interprétation des data

### 2.2.1 Extraction de pics

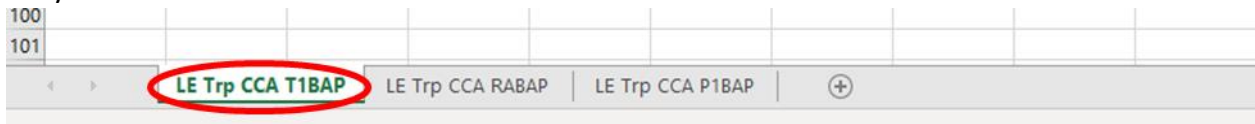
#### Exemple : tRNA Trp CCA from *S. Cerevisiae* digéré avec RNase T1 et BAP analysé en polarité MS +

Indiquer, sur la feuille Excel, le nom du fichier d'analyse et la date de l'analyse comme indiqué sur la figure ci-dessous :



A1														
A	B	C	D	E	F	G	H	I	J	K	L	M		
1	17/12/19													
2	Digestion LE 0.07 gL000003.d					GAAGCGGUKLCUCAAD#GDAGAGCPPPGABUBCAAAPCGAAGG7DUGCAGGTPCA"UUCUGPCCGUUACCA								
3	Sequence	rch	M+H	t(min)	MS <sup>2</sup>	24162 29 Da								

Nommer la feuille Excel comme indiqué sur l'image, suivant l'ARN analysé et l'enzyme utilisé → Clic droit sur l'onglet de la feuille puis renommer avec Organisme, Sous-type, Anticodon, Enzyme :



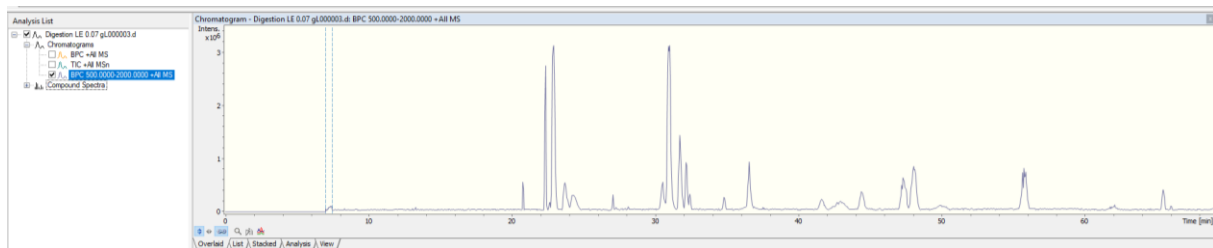
Ouvrir le logiciel DataAnalysis et ouvrir le fichier à analyser.

**Pour Calibration ext** (si nécessaire mais vivement recommandé)

Sur DataAnalysis :

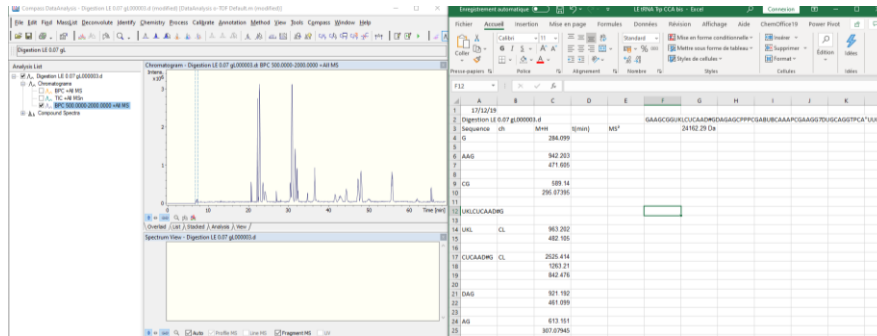
-Calibrate → External → Load et choisir un fichier de calibration → OK

Puis (pour indiquer que le fichier est calibré) F7 → Type : BPC Filter : All MS Scan mode : All Masses : 500-2000 Width : +- 0.02 → Cliquer sur Add et OK.



Ensuite extraire les pics précédemment simulés.

Pour cela afficher DataAnalysis et la feuille Excel sur l'écran. Voir figure ci-dessous :



Sur DataAnalysis → F7 et mettre les paramètres comme indiqué sur l'image et copier/coller un m/z de l'oligonucléotides dans la case Masses.

File name: Digestion LE 0.07 gl.000003.d

Traces:

- BPC +All MS
- TIC +All MSn
- BPC 500.0000-2000.0000 +All MS

Type: Extracted Ion Chromatogram

Filter: All MS

Scan mode: All

Masses: 284.099

Width: ±0.02

Polarity: positive (+)

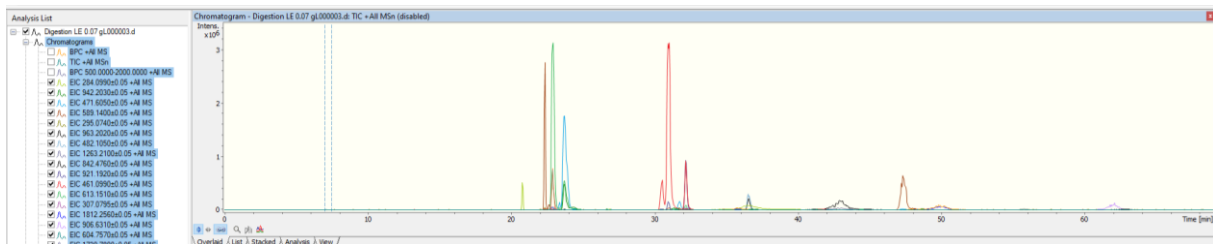
Background: None

Color:

Enabled

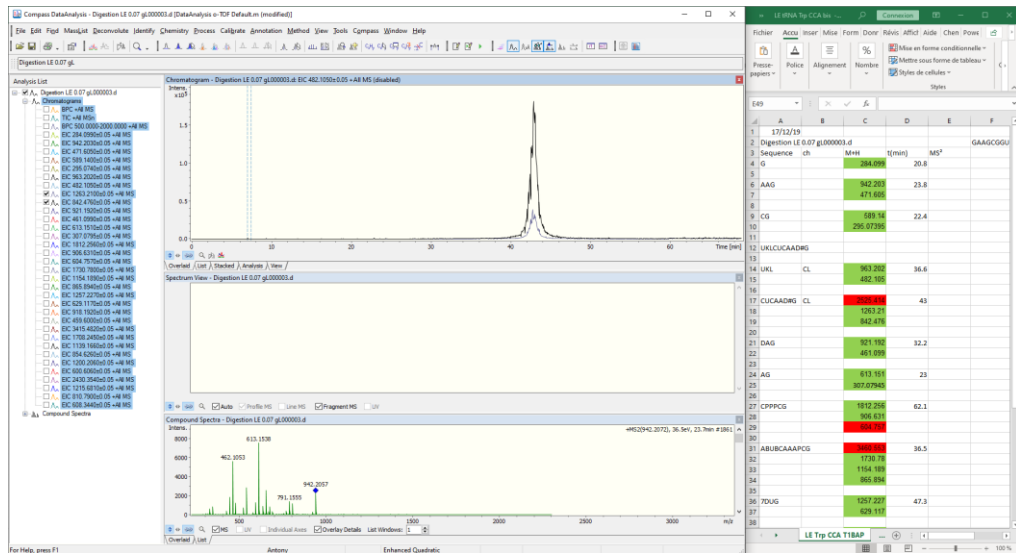
Cliquer sur Add et répéter l'opération pour tous les m/z et tous les oligonucléotides.

Une fois toutes les masses rentrées cliquer OK et le logiciel affiche les pics correspondant aux digests :



Vérifier la présence des m/z pour chaque oligonucléotide (surligné en vert si présent sinon en rouge) et noter les temps de migrations.

**⚠ ATTENTION :** Lorsqu'il y a plusieurs pics sur l'électrophérogramme pour 1 m/z, il est impératif de vérifier l'état de charge des pics et/ou le spectre MS/MS (voir dans la partie MS/MS).



Une fois l'extraction des pics terminée, il faut vérifier la séquence des oligonucléotides avec les spectres MS/MS.

## 2.2.2 Séquençage MS<sup>2</sup> des oligonucléotides

Remarques :

- 1- Pour les mono- et les dinucléotides, le séquençage est inutile (pour les dinucléotides on sait que le deuxième nucléotide est forcément un nucléotide clivable suivant l'enzyme (G ou C/U ou modif)).
- 2- Pour les trinucléotides si ce sont des oligonucléotides de type AAG le séquençage n'est pas forcément utile non plus.

Pour vérifier la séquence des oligonucléotides, il faut déjà simuler les fragments CID → Mettre la séquence de l'oligonucléotide dans Mongo, sélectionner et cocher les paramètres suivants :

Ici exemple avec 7DUG :

### Mongo Oligo Mass Calculator v2.08

Sequence (5'-->3')

7DUG

DNA2RNA RNA2DNA CleanUp Clear sequence

Memories  
GAAGCGGUKLCUCA...  
copy paste clear all

Calculation  
 molecular mass  
 electrospray series  
 CID fragments  
 internal fragments  
 base losses  
 RNase T1 digest  
 RNase U2 digest  
 RNase A digest  
 5' exonuclease digest  
 3' exonuclease digest  
 modified bases list

Options  
 average mass  
 monoisotopic mass  
 negative mode  
 positive mode  
 DNA  
 RNA  
 allow undercut

Terminals  
 5'-OH  
 5' phosphate  
 3'-OH  
 3' phosphate  
 3' cyclic phosphate

User residue 'N'  
 base anion mass  
 sugar add mass

Output  
 close button  
 output summary  
 sequence by  3  10  50  
 extended output  
 show messages  
 sorted output

Calculate Clear Form Help

Cliquer sur Calculate et une fenêtre est affichée :



Sequence :  
7DU G

5'OH - RNA[4mer] - 3'OH  
U:1 G:1 7:1 D:1  
monoisotopic mass, positive mode

MOLECULAR MASS  
M = 1256.219 [M+H]<sup>+</sup> = 1257.227 [M-H]<sup>-</sup> = 1255.211

ELECTROSPRAY SERIES

ch	m/z	M+Na	M+K	M+2Na
1	1257.227	1279.209	1295.183	1301.191
2	629.117	640.108	648.095	651.099
3	419.747	427.075	432.399	434.402

CID FRAGMENTS

n	ch	a-B	w	y	d-H2O
1	1		364.065	284.099	360.070
2	1	474.101	670.090	590.124	668.111
	2		335.548	295.565	334.559
3	1	782.142	978.130	898.164	974.136
	2	391.575	489.569	449.586	487.572
	3		326.715	300.060	325.383

INTERNAL FRAGMENTS

pos	1	2
D2:D2	503.045	
D2:U3	809.070	405.039
U3:U3	501.030	

BASE LOSSES

base	ch	M-B	M-BH
U	1	1146.208	1145.200
	2	573.607	573.104
	3	382.741	382.405
G	1	1107.185	1106.178
	2	554.096	553.592
	3	369.733	369.397
7	1	1093.170	1092.162
	2	547.089	546.585
	3	365.061	364.726
D	1	1144.192	1143.184
	2	572.600	572.096
	3	382.069	381.733

La partie CID FRAGMENTS est celle qui permet de vérifier la séquence des oligonucléotides : Ici, les ions fils (c et y mais aussi a-B et w) de l'oligonucléotides sont simulés.

→ Généralement, pour les ARN, ce sont les ions y et c («d-H2O» dans le tableau) qui vont être utiles.

Si le séquençage avec y et c est difficile on peut utiliser les ions a-B et w.

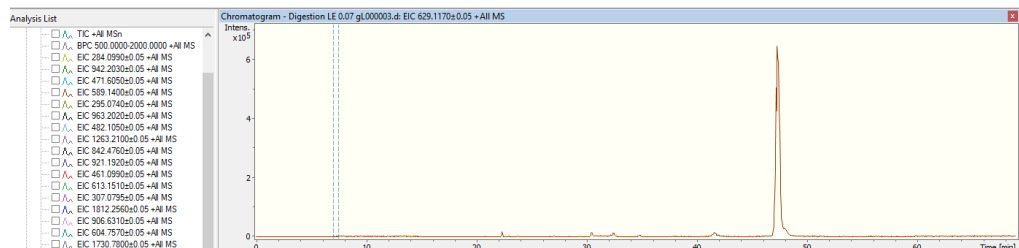
INTERNAL FRAGMENTS et BASE LOSSES permettent d'augmenter la confiance de la présence de l'oligonucléotide mais il n'est pas impératif de les observer.

Internal fragments : fragment issu d'un double clivage dans la chaîne.

Base lose : perte d'une base nucléique de l'oligonucléotide intact.

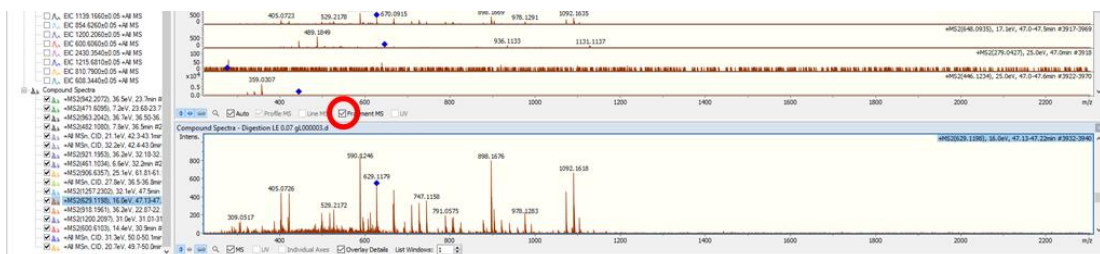
**⚠** Sur Mongo la modif ? ne correspond pas à m5C (on peut remplacer ? par B)

Sur DataAnalysis, dans la colonne Analysis List, cocher le pic à séquencer (ici 7DUG → 629.1170 (2+)).



Et intégrer le pic en maintenant clic droit au début du pic et en relâchant à la fin du pic.

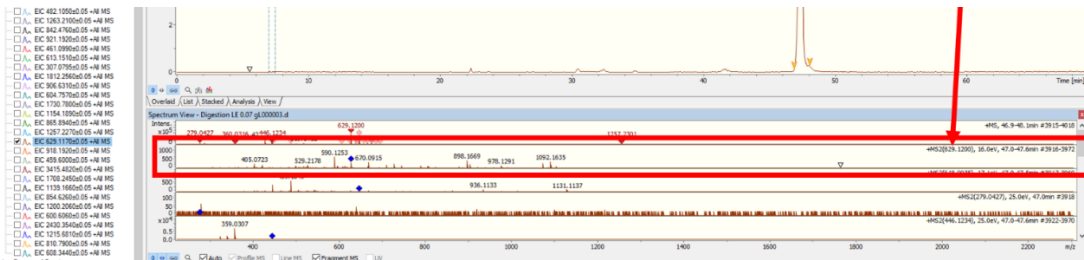
Dans la fenêtre Spectrum View, il faut que Fragment MS soit coché pour voir les spectres MS<sup>2</sup>.



Le premier spectre de cette fenêtre correspond au spectre MS du pic intégré (indiqué en haut à droite du spectre) et les carrés rouges au-dessus des pics indiquent que le spectro a fait de la MS<sup>2</sup> sur ces masses.

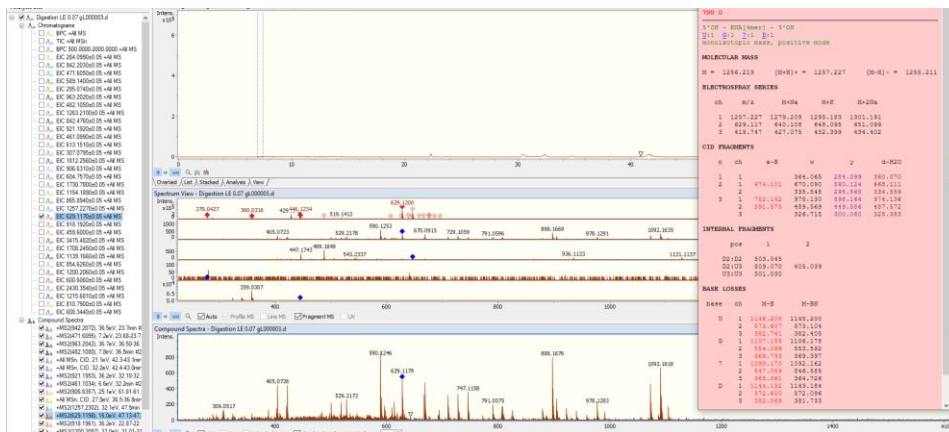
Les spectres suivants sont les spectres MS<sup>2</sup> du pic intégré (ion précurseur indiqué en haut à droite du spectre).

Chercher le spectre MS<sup>2</sup> de l'ion à séquencer (ici 629.1170).



Clic droit sur ce spectre et Copy to Compound Spectra. Le spectre MS<sup>2</sup> s'affiche dans la fenêtre Compound Spectra.

Puis chercher manuellement les pics simulés dans CID fragments.



(On peut noter tous les ions présents ou absents mais TRES laborieux).

	A	B	C	D	E
1	17/12/19				
2	Digestion LE 0.07 gL00003.d				GA
3	Sequence	ch	M+H	t(min)	MS <sup>2</sup>
4	G		284.099	20.8	
5					
6	AAG		942.203	23.8	
7			471.605		
8					
9	CG		589.14	22.4	
10			295.07395		
11					
12	UKLCUCAAD#G				
13					
14	UKL	CL	963.202	36.6	

Si le séquençage est excellent : remplir la case MS<sup>2</sup> en vert foncé.

Si ce n'est pas très intense mais que le séquençage est faisable : en vert clair.

Si le spectro a fait de la MS<sup>2</sup> mais que le spectre MS<sup>2</sup> ne permet pas du tout de séquencer : en jaune.

Si le spectro n'a pas fait de MS<sup>2</sup> : en rouge.

Répéter le séquençage pour chaque oligonucléotides détectés.

Les ions n=1 peuvent manquer → Commencer par n=2

CG/GC → 589.14 ; UG/GU → 590.124 ; AG → 613.151 ; AC → 573.145 ; AU → 574.129

On peut également faire du séquençage *de Novo* si le pic est inconnu. Correspondances des Δ de masse avec les nucléotides non-modifiés :

305 → C ; 306 → U ; 329 → A ; 345 → G

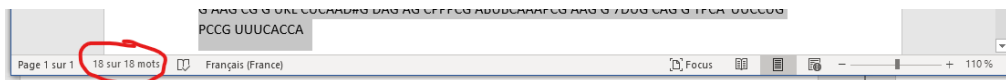
## 2.2.3 Recouvrement de séquence

Pour calculer le recouvrement de séquence RNase T<sub>1</sub> et RNase A :

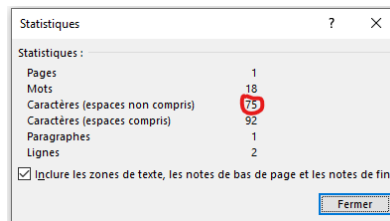
Recopier la séquence de l'ARN sur un fichier Word et ajouter des espaces où l'enzyme hydrolyse les liaisons (Ici RNase T<sub>1</sub> = hydrolyse G).

Exemple : G AAG CG G UKL CUCAAD#G DAG AG CPPPCG ABUBCAAAPCG AAG G 7DUG CAG G TPCA"UUCCUG PCCG UUUCACCA

Sélectionner la séquence et cliquer sur XXX sur XXX mots :



Une fenêtre apparaît :



Le nombre de nucléotide total de la séquence est affiché sur la ligne « Caractères ».

-Ensuite surligner en vert les oligonucléotides qui ont été détectés en MS et séquencés en MS<sup>2</sup>.

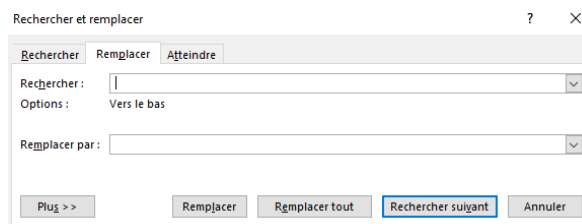
-Surligner en bleu les oligonucléotides qui ont été détectés en MS mais non séquencés en MS<sup>2</sup>.

-Ne pas surligner les oligonucléotides qui n'ont pas été détectés.

G **AAG** **CG** G UKL CUCAAD#G DAG AG CPPPCG ABUBCAAAPCG **AAG** G **7DUG** CAG G TPCA"UUCCUG **PCCG** UUUCACCA

Remarque : les oligonucléotides qui ne sont pas uniques ne sont pas pris en compte dans le séquençage (ici on a deux fois AAG (en gras) et 4 fois G → pas d'info sur la séquence).

Pour calculer le recouvrement de séquence : Sélectionner toute la séquence et appuyer sur Ctrl + H :



Mettre un espace dans la barre « Rechercher » et laisser vide dans la barre « Remplacer par » et cliquer sur « Remplacer tout » puis cliquer sur Oui.

GAAGCGGUKLCUCAAD#GDAGAGCPPPCGABUBCAAAPCGAAGG7DUGCAGGTPCA"UCCUGPC  
CGUUUCACCA

Sélectionner uniquement les oligos totalement séquencés (surligné vert) et cliquer sur XXX sur XXX mots pour que le logiciel compte les nucléotides. (Raccourcis pour sélection = Maintenir Maj et sélectionner avec la touche directionnelle « Droite »).

Additionner les séquences surlignées en vert = Recouvrement de séquence.

Ici : **MS/MS** 44/75 : 58 % et **MS** 65/75 : 65/75 : 86 %

Répéter le même protocole pour les résultats obtenus avec la RNase A et pour déterminer la séquence totale.

Mettre les deux séquences sans espaces l'un en dessous de l'autre et une troisième séquence sans surlignage et sans espace :

RNase T1 :

GAAGCGGUKLCUCAAD#GDAGAGCPPPCGABUBCAAAPCGAAGG7DUGCAGGTPCA"UCCUGPC  
CGUUUCACCA

RNase A :

GAAGCGGUKLCUCAAD#GDAGAGCPPPCGABUBCAAAPCGAAGG7DUGCAGGTPCA"UCCUGPC  
CGUUUCACCA

Total :

GAAGCGGUKLCUCAAD#GDAGAGCPPPCGABUBCAAAPCGAAGG7DUGCAGGTPCA"UCCUGPC  
CGUUUCACCA

Puis surligner manuellement les oligonucléotides qui se recouvrent :

RNase T1 :

GAAGCGGUKLCUCAAD#GDAGAGCPPPCGABUBCAAAPCGAAGG7DUGCAGGTPCA"UCCUGPC  
CGUUUCACCA

RNase A :

GAAGCGGUKLCUCAAD#GDAGAGCPPPCGABUBCAAAPCGAAGG7DUGCAGGTPCA"UCCUGPC  
CGUUUCACCA

Total :

GAAGCGGUKLCUCAAD#GDAGAGCPPPCGABUBCAAAPCGAAGG7DUGCAGGTPCA"UCCUGPC  
CGUUUCACCA

**TOTAL : MS 75/75 (100%) MS/MS 62/75 (83%)**

**Développement de l'électrophorèse capillaire  
couplée à la spectrométrie de masse pour  
l'étude des modifications post-  
transcriptionnelles des ARN.**

## Résumé

Les modifications post-transcriptionnelles sont fondamentales dans différents aspects pour assurer les rôles biologiques des acides ribonucléiques (ARN) dans le processus de traduction. La caractérisation chimique des modifications est par conséquent essentielle afin de permettre la compréhension des intérêts de ces modifications. Les travaux, présentés dans ce manuscrit de thèse, décrivent la mise en œuvre d'une nouvelle approche basée sur le couplage de l'électrophorèse capillaire à la spectrométrie de masse de haute résolution (CE-MS) au service de l'étude des modifications d'ARN. Ici, une stratégie sur deux niveaux a été développée. Le premier niveau d'analyse a permis de déterminer la nature des modifications présentes sur un ARN. Le second niveau, s'inspirant d'une stratégie « bottom-up », a permis de localiser les modifications sur les structures primaires des ARN. La méthodologie développée a permis de caractériser, sans ambiguïté, un grand nombre de modifications sur différents types d'ARN et notamment des ARN de transfert, ouvrant la voie à la CE-MS comme un nouvel outil d'analyse pour la caractérisation de modifications d'ARN.

Mots-clés : Spectrométrie de masse, Électrophorèse capillaire, Acide ribonucléique, Modification post-transcriptionnelle

## Summary

Post-transcriptional modifications are fundamental in different aspects to ensure the biological roles of ribonucleic acids (RNA) in the translation process. The chemical characterization of the modifications is therefore essential in order to understand their influences. The work presented in this thesis manuscript describes the implementation of a new approach based on the coupling of capillary electrophoresis to high resolution mass spectrometry (CE-MS) for the investigation of RNA modifications. Here, a two-level strategy has been developed. The first level of analysis enable to determine the nature of the modifications contained on an RNA. The second level of this strategy, inspired by the "bottom-up" approach, permit to map the modifications onto the primary structures of RNA. The methodology developed allows to characterize, without ambiguity, a large number of modifications on different types of RNA and particularly on transfer RNA, paving the way to CE-MS as a new analysis tool for the characterization of RNA modifications.

Keywords : Mass spectrometry, Capillary electrophoresis, Ribonucleic acids, Post-transcriptional modification