



**HAL**  
open science

# Investigation of RNA kissing complexes by native electrospray mass spectrometry : magnesium binding and ion mobility

Clemence Rabin

► **To cite this version:**

Clemence Rabin. Investigation of RNA kissing complexes by native electrospray mass spectrometry : magnesium binding and ion mobility. Agricultural sciences. Université de Bordeaux, 2017. English. NNT : 2017BORD0892 . tel-01717579

**HAL Id: tel-01717579**

**<https://theses.hal.science/tel-01717579>**

Submitted on 26 Feb 2018

**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.

THÈSE PRÉSENTÉE  
POUR OBTENIR LE GRADE DE  
**DOCTEUR DE**  
**L'UNIVERSITÉ DE BORDEAUX**

ÉCOLE DOCTORALE SCIENCES DE LA VIE ET DE LA SANTÉ

SPÉCIALITÉ INTERFACE CHIMIE-BIOLOGIE

Par Clémence RABIN

Ingénieur diplômée de l'ENSTBB

**Investigation of RNA kissing complexes by *native* electrospray  
mass spectrometry:  
magnesium binding and ion mobility**

Sous la direction de : Dr. Valérie GABELICA

Soutenue le 19 décembre 2017

Membres du jury :

Dr. LEIZE-WAGNER, Emmanuelle  
Dr. ENNIFAR, Eric  
Pr. FABRIS, Daniele  
Dr. DI PRIMO, Carmelo

Directeur de recherche, CNRS, LSMIS  
Directeur de recherche, CNRS, IBMC, Strasbourg  
Professeur à l'Université d'Albany, RNA institute  
Chargé de recherche, INSERM, IECB

Présidente  
Rapporteur  
Rapporteur  
Examineur



## REMERCIEMENTS

Après avoir écrit ce manuscrit et présenté ces travaux, il est temps de clôturer avec l'écriture de mes remerciements. Une thèse, ce n'est pas que soi-même face à son sujet, c'est une multitude de moments de joie, de doutes, de tristesse et de rires partagés avec de nombreuses personnes.

En premier lieu, je remercie Emmanuelle Leize-Wagner, Eric Ennifar, Daniele Fabris et Carmelo Di Primo pour avoir accepté d'évaluer mon travail. Je souhaiterais aussi remercier Jean-Louis, directeur de l'IECB, de l'unité U1212 et aussi chef de l'équipe Mergny, pour son accueil dans l'établissement.

Je souhaite ensuite remercier Valérie, notre « *Big Boss* ». Merci de m'avoir fait confiance durant ces 3 dernières années. Tu as vite compris que la chimie, ce n'était pas mon fort, pourtant tu m'as encouragée et suivie tout au long de ces années. Merci pour tes conseils, ta rigueur et ton exigence qui, même si parfois c'était dur, m'ont fait évoluer et avancer. J'ai pu découvrir grâce à toi le monde de la recherche en spectrométrie de masse et je ne le regrette absolument pas ! Merci.

Je tiens aussi à remercier les membres de la plateforme de spectrométrie de masse de l'IECB : Fred et Loïc. Vous faites partie intégrante de notre équipe. Fred, merci pour ton aide en instrumentation et surtout pour l'IMS, parfois capricieuse. Loïc, merci pour nos discussions, ton aide et ta patience (il en faut souvent beaucoup !).

Faire de la recherche c'est bien, mais en équipe c'est mieux ! Et quelle équipe... ! Ces 3 années m'ont permis de rencontrer de multiples personnalités, venant de tout horizon (Italie, Russie, Espagne, Inde, Canada, Japon, etc.) et que je ne risque pas d'oublier. Tout d'abord, je vais remercier les nouveaux arrivants : Anirban, Steven, Eric et Jorge. *We did not spent so much time together, but I wish you the best for the future. I hope you will enjoy working at the IECB as I liked it!* Ensuite, je souhaite remercier Adrien, Max, et Joséphine, collègues du « bureau au bout du couloir ». Adrien, comme je te l'ai déjà dit, je te remercie pour ton aide en début de thèse. Merci d'avoir été patient et de m'avoir transmis ton savoir ! Max... je vais te dire seulement une chose et tu comprendras « 😊 ». Joséphine, merci pour tout, autant professionnel que personnel. Même si tu as un petit côté dictateur, j'ai adoré travailler avec toi ! Aussi, tu as une petite famille au top, prenez bien soin de vous !

Ensuite, nous avons Nina. Nina, tu es une très belle rencontre ! Cette année a été pour toi aussi une grande année et je suis très heureuse d'avoir pu partager avec toi tous ces moments. Je me souviendrai toujours de cette partie de bowling endiablée pour ton EVJF ! Comme on dit chez toi : SPACIBA (si j'ai bien retenu un mot, se sera celui là ^^). Passons maintenant à mes collègues de bureau. Sandrine, ton départ n'a pas été une grande joie pour nous mais je suis contente que tu aies trouvé un poste qui te plaît. Merci pour ces bons moments et surtout pour les derniers mots que tu m'as écrits, ça m'a beaucoup touchée. Tu es la « force tranquille » incarnée, surtout ne change pas ! Ensuite, je tiens particulièrement à remercier Stefano. Tu es quelqu'un d'exceptionnel, toujours calme et posé.

J'aimerais tellement être comme toi pour ça ^^ . Tu es devenu un ami très cher. Je sais que ce n'est pas toujours facile d'allier professionnel et personnel mais tout va très bien se passer, tu verras. Et puis n'oublie pas : « *Quando il gatto non c'è, i topi ballano* » !

J'ai aussi eu la chance de pouvoir encadrer une stagiaire au top : Louise. Merci pour ton travail qui m'a beaucoup avancé. Merci pour ton humour et ta gentillesse ☺. J'en profite aussi pour remercier Marine, autre stagiaire, avec qui on a beaucoup ri et que j'apprécie particulièrement.

J'ai pu également faire la rencontre d'autres thésardes, devenues amies maintenant. Nassima (Chouchou), avec toi j'ai ri, j'ai pleuré, j'ai râlé, on s'est consolées mutuellement, et on n'a rien lâché ! Léonie, ma belle rencontre des JJC, avec toi aussi j'ai ri et surtout beaucoup râlé ! Ne lâche pas, si tu en es arrivée là c'est que tu as toutes les capacités pour ! Merci d'être vous les filles.

Je voudrais remercier aussi deux autres personnes de l'IECB. Elles restent souvent dans l'ombre alors qu'elles mériteraient d'être plus mises en lumière. Aurore, je pense que sans toi, l'équipe Mergny ne fonctionnerait pas aussi bien. Tu es un atout alors ne l'oublie pas. Mariline, notre hôtesse d'accueil. L'IECB serait bien triste sans toi. Merci pour ce que tu as fait pour moi ces derniers temps. Surtout ne changez pas !

Je voudrais aussi en profiter pour remercier tous les autres avec qui j'ai pu discuter, rire, et échanger durant ces 3 années : Laura, Ly, Clément, Solenne, Mona, Gilmar, Samir, Carmelo., et bien d'autres encore.

Je remercie aussi mes amis, notamment ceux qui sont là malgré la distance qui nous sépare : Pitch, Emy, Arnaud, Marie & Charlotte, ma petite Marie, Hélène.

Enfin mes derniers mots seront pour ma famille et amis proches. Je commencerai par deux personnes que j'ai l'impression de connaître depuis des années alors que ce n'est pas le cas. Mandy, tu es quelqu'un de formidable. Merci pour tout ce que tu fais pour nous. Ton amitié nous est précieuse et pour rien au monde je n'aimerais l'échanger. Tu sais trouver les bons mots que ce soit pour nous pousser ou nous rassurer. Ensuite, il y a Julien. Mon petit Ju, j'ai pu te découvrir tout au long de cette dernière année (et demi), et en toi j'ai trouvé un véritable ami. Merci d'être toi tout simplement. Ces derniers mois, tu m'as montrée que je pourrai toujours compter sur toi (et inversement). Ton amitié est inestimable. Merci... Ensuite, Emeline, dite Luna, tu es ma meilleure amie. Même si nous sommes loin l'une de l'autre, tu es toujours là pour moi. Tu as toujours cru en moi et je t'en remercie du fond du cœur. Je suis très fière d'être la marraine de votre petit roi ! Je tiens ensuite à remercier ma famille. Ma belle-famille qui est formidable et à laquelle je tiens beaucoup. Merci de votre soutien. Merci à Mathias pour sa gentillesse et son aide photographique. Merci à mon grand-père pour tout ce qu'il

m'apporte et d'être là tout simplement. Elise, ma sœur, comme tu me l'as si bien dit, tu as assisté aux turpitudes de mon parcours et à toutes mes angoisses. Ça n'a pas toujours été facile, mais tu vois, j'y suis arrivée ! Merci d'être là Sis. Je tiens particulièrement à remercier du plus profond de mon cœur mes parents. Vous me soutenez depuis 27 ans. C'est grâce à vous que j'en suis là aujourd'hui. Vous ne m'avez jamais lâchée et je sais que vous ne le ferez jamais. Je vous aime.

Mes (presque) derniers mots seront pour l'homme de ma vie, Morgan. Cela fait bientôt 5 ans que tu partages ma vie, 5 ans de bonheur avec toi. Merci de me soutenir tous les jours. Je serai, moi aussi, toujours là pour toi. Je t'aime.

Ma dernière pensée sera pour une personne qui nous a malheureusement quittés il y a quelques années mais que je n'oublierai jamais. J'espère qu'elle est fière de moi. Tu me manques...

*"If you can dream it, you can do it."*

*"All our dreams can come true if we have the courage to pursue them."*

**- Walt Disney -**





# Table of Contents

<b>ABSTRACT</b> .....	5
<b>RESUME - APERÇU EN FRANÇAIS</b> .....	7
<b>ABBREVIATIONS</b> .....	13
<b>1. INTRODUCTION</b> .....	15
I.    Nucleic Acids.....	15
I.1. Bases and primary structure.....	15
I.2. Secondary and tertiary structures .....	17
II.   An RNA story.....	20
II.1. The biological roles of RNA.....	20
II.2. RNA structures: from RNA motifs to more complex structures .....	25
II.3. RNA structure and Cations .....	32
II.4. Why and How to study RNA-cations structures and equilibria? .....	36
III.  Mass Spectrometry as a tool to study RNA .....	41
III.1. <i>Native</i> mass spectrometry.....	41
III.2. Ion mobility spectrometry (IMS).....	44
III.3. Tandem mass spectrometry .....	47
IV.   Scope of the thesis .....	49
IV.1. TAR-R06 kissing complex.....	49
IV.2. RNAlI-RNAlI kissing complex.....	52
References .....	54
<b>2. MATERIALS &amp; METHODS</b> .....	67
V.    Sample preparation .....	67
V.1. Kissing complexes .....	67
V.2. Duplexes.....	68

V.3. Sequences used .....	68
VI. Characterization in solution by thermal denaturation.....	71
VII. Mass Spectrometry experiments.....	72
VII.1. Sample preparation for mass spectrometry analysis .....	72
VII.2. Mass spectrometers used .....	72
VII.3. MS data analysis .....	78
VIII. Ion Mobility Spectrometry: how to obtain CCS values?.....	84
VIII.1. Determination of the CCS of the ATD peak center .....	84
VIII.2. CCS distribution reconstruction.....	85
IX. How to analyze MS/MS data .....	87
References .....	88
<b>3. RESULTS AND DISCUSSION.....</b>	<b>89</b>
<b>Chapter 1. Characterization of the interaction between Mg<sup>2+</sup> and kissing complexes by <i>Native</i> mass spectrometry .....</b>	<b>90</b>
X. <i>Native</i> MS from Magnesium solutions: optimization of the method.....	90
X.1. Introduction.....	90
X.2. Results.....	91
X.3. Discussion .....	98
XI. Determination of specific cations number .....	100
XI.1. Introduction.....	100
XI.2. Results .....	100
XI.3. Discussion .....	111
XII. The extent of kissing complex stabilization by Mg <sup>2+</sup> is correlated to the cation affinity, due to a displacement of equilibria .....	113
XII.1. Introduction.....	113
XII.2. Results and discussion.....	114

XII.3. Conclusion and perspectives .....	127
XIII. Mg <sup>2+</sup> localization .....	128
XIII.1. Introduction.....	128
XIII.2. Results .....	128
XIII.3. Discussion .....	140
<b>Chapter 2. Mass spectrometry to screen kissing complex affinities.....</b>	<b>145</b>
XIV. Introduction.....	145
XV. Kissing complexes used .....	145
XVI. Results .....	146
XVI.1. Mass spectrometry experiments.....	146
XVI.2. UV melting experiments.....	148
XVI.3. Surface Plasmon Resonance Experiments .....	149
XVII. Discussion and conclusion.....	152
References .....	153
<b>4. RESULTS AND DISCUSSION.....</b>	<b>157</b>
<b>Chapter 3. DNA and RNA duplexes structures in the gas phase .....</b>	<b>158</b>
XVIII. Introduction.....	158
XIX. Results and discussion.....	159
XIX.1. ESI-MS spectra .....	159
XIX.2. IM-MS results.....	163
XX. Conclusion and perspectives .....	176
XX.1. Perspectives: the effect of the charge state on <sup>DT</sup> CCS <sub>He</sub> .....	176
<b>Chapter 4. Kissing complexes structure in the gas phase .....</b>	<b>179</b>
XXI. Introduction.....	179
XXII. Results and discussion.....	179
XXII.1. Kissing complexes in the gas phase .....	179

XXII.2. Comparison RNA kissing complexes with RNA duplexes .....	189
XXIII. Conclusion and perspectives .....	190
References .....	191
<b>4. GENERAL DISCUSSION AND CONCLUSIONS</b> .....	<b>195</b>
XXIV. Discussion about mass spectrometry results (PART I) .....	196
XXIV.1. Mass spectrometry compatible with Magnesium salts .....	196
XXIV.2. Analysis of Mg <sup>2+</sup> binding to RNA by <i>native</i> mass spectrometry .....	196
XXIV.3. Mass spectrometry as a screening tool .....	199
XXV. Discussion about the ion mobility-mass spectrometry results (PART II) .....	200
XXV.1. Initial compaction in the gas phase .....	200
XXV.2. Collision-induced compaction .....	201
<b>5. APPENDICES</b> .....	<b>203</b>

## ABSTRACT

Besides being the molecular intermediate between DNA and proteins, RNA can have many other functions such as gene regulation (riboswitches), gene expression (mRNA and tRNA) or catalysis (ribozymes). RNA function is linked to its structure and its folding dynamics. Cations such as magnesium bind to RNA and are in some instances essential for proper folding and for stability. The need of structural and thermodynamic details about  $Mg^{2+}$  interactions is then of utmost importance in the study of the structure-function relationships. The first part of our work consists in characterizing the binding equilibria between magnesium and RNA model motifs, called kissing complexes, using *native* mass spectrometry (MS). MS makes it possible to distinguish individual binding stoichiometries, and the present work consisted in developing a method to quantify each species, taking into account the contribution of nonspecific adducts. We also explored how tandem mass spectrometry (MS/MS) could further help localizing magnesium ions. Further, we explored the structures of RNA complexes in the gas phase using ion mobility mass spectrometry (IM-MS), with the aim to detect shape changes upon cation or ligand binding. But in contrast with anticipations, we found that DNA and RNA duplexes as well as RNA kissing complexes undergo a significant compaction at charge states naturally produced by *native* ESI-MS, which may hide the effect of cations. Our work showcases how mass spectrometry can bring novel information on RNA-cation binding stoichiometries and affinities, but also discusses some limitations of a gas-phase method to probe solution structures.

**Keywords:** Mass spectrometry, ion mobility, RNA complexes, magnesium, gas-phase ion structure



# RESUME - APERÇU EN FRANÇAIS

## Introduction

En plus d'être l'intermédiaire entre l'ADN et les protéines, l'ARN est impliqué dans plusieurs processus biologiques : régulation et expression des gènes (riboswitches, ARNm et ARNt) ou encore catalyse (ribozymes). La fonction de chaque ARN est liée à sa structure et à sa dynamique de repliement. L'ARN peut adopter différents motifs structuraux secondaires tels que le simple brin, les duplexes ou encore les hairpins, mais aussi des motifs tertiaires comme les pseudoknots ou encore les kissing complexes. Des études ont montré l'importance des cations (tels que  $K^+$ ,  $Na^+$ ,  $Mg^{2+}$ ) dans la stabilité et le bon repliement des structures d'ARN. Les ARN sont entourés par ce que l'on appelle « l'atmosphère ionique ». La présence de ces ions va permettre de réduire la répulsion électrostatique entre les charges négatives présentes sur l'ARN, et va donc être essentielle pour l'obtention de structures fonctionnelles. Un des principaux cations divalents impliqué dans la stabilisation des complexes d'ARN est le magnésium divalent.

Un des défis actuels est de comprendre comment le magnésium interagit et stabilise les structures d'ARN. Actuellement plusieurs techniques biophysiques telles que la cristallographie aux rayons X, la RMN (Résonance Magnétique Nucléaire), la SPR (Surface Plasmon Resonance) ou la spectroscopie UV sont utilisées pour obtenir des caractéristiques structurales et/ou thermodynamiques. Cependant, ces techniques ne permettent pas toutes d'avoir des données directes sur les cations spécifiques (notamment constantes d'équilibre spécifiques), ou présentent des désavantages non négligeables.

La spectrométrie de masse (SM) *native* permet d'obtenir des données structurales mais aussi thermodynamiques, ce qui en fait une technique de choix. La spectrométrie de masse est dite *native* quand elle préserve, de par des conditions d'ionisation douces, les structures non-covalentes de la solution à la phase gazeuse. Son utilisation permet d'identifier les espèces présentes en solution et donne accès aux stœchiométries des complexes étudiés. La spectrométrie de mobilité ionique (SMI), lorsqu'elle est couplée à la SM, apporte par ailleurs une nouvelle dimension qui est la séparation des molécules en fonction de leur conformation. La mobilité ionique nous permet d'obtenir une mesure de la surface exposée aux collisions, la Section Efficace de Collision (CCS). Cette mesure est reliée à la conformation des espèces présentes en phase gazeuse et nous permet d'établir des corrélations avec les structures grâce à l'utilisation complémentaire de la modélisation moléculaire.

Ce travail est donc divisé en deux parties. La première concerne le développement d'une méthode de caractérisation des équilibres de liaison entre le magnésium et des motifs d'ARN modèles. La seconde

partie est centrée sur l'analyse de complexes d'ARN par spectrométrie de mobilité ionique avec pour objectif de détecter des changements conformationnels dus à la liaison de ligands ou cations. L'intérêt ici est d'évaluer le potentiel de la SMI, technique en phase gazeuse, pour l'étude de structure en solution. Les structures d'ARN pouvant être très complexes et de tailles diverses, l'étude s'est portée sur l'analyse du motif appelé « kissing complexe ». Les kissing complexes sont formés par deux hairpins d'ARN, liées entre elles par liaisons Watson-Crick grâce à la complémentarité des bases de leurs boucles. Deux modèles bien caractérisés dans la littérature ont été utilisés : le système TAR-R06 et le système RNAIi-RNAIi.

## Résultats

### **Partie 1 : Nouveaux développements en spectrométrie de masse, nouveaux aspects sur la structures de kissing complexes**

Dans un premier temps, nous avons développé une méthode afin de pouvoir obtenir des spectres reflétant directement les espèces présentes en solution et uniquement celles-ci. La force ionique est donc assurée par de l'acétate d'ammonium, classiquement utilisé en SM *native*. Nous avons pu doper ces solutions jusqu'à 1 mM en acétate de magnésium,  $Mg(OAc)_2$ . Nos résultats ont aussi montré qu'une optimisation des réglages instrumentaux par rapport à ceux habituellement utilisée par notre équipe était nécessaire. En effet, les conditions étaient trop douces et de nombreux adduits de  $NH_4^+$  étaient retenus. Pour notre étude, une réduction des adduits non-spécifiques d'ammonium est nécessaire, d'où l'utilisation de conditions plus dures. Il est à noter que les conditions optimisées activent les ions tout en conservant les liaisons non covalentes des complexes étudiés.

Dans un second temps, grâce au développement d'une méthode permettant de soustraire la contribution des adduits de magnésium non-spécifiques, les différentes espèces présentes ont pu être quantifiées. Les constantes d'équilibre individuelles de liaison de chaque magnésium spécifique ainsi que des pistes sur la localisation de ceux-ci ont pu être mis en évidence.

La première étape de notre méthode consiste en un traitement mathématique, utilisant les intensités mesurées sur une séquence de référence, qui permet de déterminer la part d'adduits spécifiques (liés au motif « kissing loop » avec une forte affinité) et de non-spécifiques (sites de liaison de faible affinité). Pour être une bonne référence, la séquence contrôle doit être de même type que la séquence d'intérêt (i.e. ADN ou ARN), de même taille et de même composition, mais ne pas posséder le motif auquel nous voulons attribuer les sites spécifiques. Par rapport à des duplex d'ARN choisis comme contrôles, nous avons

déterminé qu'il y a deux magnésiums spécifiques au kissing complexe TAR-R06, et seulement un pour le complexe RNAIi-RNAIIi.

La deuxième étape est la détermination des constantes apparentes d'équilibre de dissociation entre les deux hairpins de chaque complexe en fonction de la concentration en magnésium ( $K_{dKC,app}$ ). Afin de pouvoir corrélérer l'intensité des pics détectés aux concentrations en solution, une correction utilisant les facteurs de réponse est utilisée. La détermination des constantes apparentes d'équilibre a permis de confirmer l'effet stabilisateur de  $Mg^{2+}$  puisque son ajout augmente l'affinité entre les deux hairpins du kissing complexe, et ce quel que soit le kissing complexe étudié.

En combinant les données obtenues sur la stœchiométrie en adduits spécifiques et celles obtenues lors de la détermination de  $K_{dKC,app}$ , il est possible de calculer les constantes d'équilibre de dissociation de chaque  $Mg^{2+}$  spécifique au kissing complexe. Les constantes déterminées sont liées à la stœchiométrie en cation et non à leur site de liaison. Nous avons montré que même si plusieurs cations sont spécifiques à un motif kissing complexe, comme c'est le cas pour TAR-R06, ces cations ne présentent pas nécessairement la même affinité. Cette étude est innovante puisque c'est la première fois que ce niveau de détail est atteint dans l'analyse des équilibres liés à l'interaction  $Mg^{2+}$ -ARN.

Enfin, nous avons utilisé la spectrométrie de masse en tandem (SM/SM) afin d'obtenir des informations sur la localisation du magnésium le plus affin. Nos résultats montrent qu'il est possible de dissocier le kissing complexe en ces deux monomères. Lorsque du magnésium est ajouté, un  $Mg^{2+}$  est retrouvé sur chacune des deux hairpins. Ces résultats ont permis d'émettre l'hypothèse d'une coopérativité négative entre deux sites de liaison du magnésium. Pour RNAIi-RNAIIi la coopérativité négative serait telle que les deux sites de liaison ne peuvent être peuplés simultanément. Pour TAR-R06, la liaison du premier  $Mg^{2+}$  à l'un des deux sites de liaison, réduirait l'affinité du second  $Mg^{2+}$  pour le second site.

Dans un projet connexe, la spectrométrie de masse *native* est utilisée comme outil de criblage rapide pour classer des kissing complexes, différents par seulement une paire de base, en fonction de leur stabilité relative. L'avantage de la spectrométrie de masse est l'identification rapide des différentes espèces présentes en solution notamment la détection de complexes « inconnus » (i.e. formation d'homodimère par exemple). L'intensité relative des pics a été directement reliée à l'abondance en solution grâce à l'utilisation d'un standard interne et de simplifications.

## **Partie 2 : l'étude de complexes d'ARN et ADN par spectrométrie de masse couplée à la mobilité ionique révèle une compaction en phase gazeuse**

En spectrométrie de masse les molécules sont désolvatées, ionisées et analysées en phase gazeuse. Il est souvent admis pour les protéines que celles-ci conservent leur forme globulaire de la solution à la phase gazeuse. Les acides nucléiques, quant à eux, ne sont pas globulaires. Plusieurs questions peuvent alors se poser : qu'advient-il des structures des acides nucléiques en phase gazeuse ? Est-ce qu'une technique en phase gazeuse peut sonder les structures en solution ? La spectrométrie de masse seule ne peut répondre à ces questions. En revanche, la SM couplée à la spectrométrie de mobilité ionique apporte une nouvelle dimension qui est la séparation des molécules en fonction de leur forme. Grâce à cette technique, la Section Efficace de Collision (CCS) des molécules peut être obtenue. La CCS est une grandeur physique liée aux frictions entre les ions et le gaz tampon présent dans la chambre de mobilité. La comparaison entre les valeurs de CCS expérimentales et les CCS théoriques obtenues par modélisation permet d'avoir une idée sur la structure des ions en phase gazeuse et d'établir un modèle. L'un des autres buts de cette partie est aussi de détecter des changements de conformation dus à la liaison de ligands et/ou cations sur des complexes d'intérêt. Pour répondre aux questions posées précédemment, nous avons analysé en premier lieu des duplexes d'ADN ou d'ARN, puis des kissing complexes.

Les résultats montrent qu'un ensemble de conformations est adopté par les duplexes d'ADN, d'ARN, et par les kissing complexes. Les structures étudiées étant flexibles, plusieurs conformations proches peuvent être adoptées, formant ainsi des distributions de CCS larges. De plus, les comparaisons entre duplexes d'ADN, d'ARN et kissing complexes de même taille et composition, ont révélé que ces structures couvrent la même gamme de CCS. Ces résultats montrent qu'une hélice B, une hélice A et kissing complexes ont des compacités très proches en phase gazeuse alors que les structures sont différentes en solution.

L'un des principaux résultats de notre étude est qu'il y a une compaction de la structure des acides nucléiques étudiés, i.e. duplexes et kissing complexes, du moins aux états de charge typiquement obtenus par spectrométrie de masse *native*, c'est-à-dire à force ionique physiologique. Notre hypothèse est qu'aux bas états de charge produits par SM *native*, les biomolécules ont tendance à devenir globulaires à cause de la formation de nouvelles liaisons hydrogène entre les phosphates. Cette compaction, d'environ 20% en termes de CCS, est de nature à masquer l'éventuel effet de cations ou ligands sur la conformation de l'acide nucléique en solution. Cette étude révèle donc une limitation quant à l'utilisation de la spectrométrie de mobilité ionique pour sonder des structures initialement présentes en solution.

Néanmoins, nous avons aussi vu que l'augmentation de l'état de charge grâce à l'utilisation d'un agent super-chargeant pourrait prévenir l'effet de compaction observé. Cette piste est donc à explorer.

Plusieurs études ont été réalisées sur l'effet de l'activation des ions sur leur structure en phase gazeuse. Pour les protéines, l'effet dépend du type d'activation et de l'énergie donnée aux ions. Afin d'en savoir davantage sur le comportement des acides nucléiques en phase gazeuse, nous avons réalisé ce même type d'expériences, i.e. activer plus les ions en jouant sur divers paramètres instrumentaux. Les résultats montrent qu'une plus grande compaction encore est observée lors de l'activation des longues séquences analysées. D'après ces résultats, la compaction s'effectue jusqu'à un seuil où la structure ne peut pas se compacter plus. La taille de la séquence, son état de charge ainsi que la distribution microscopique des charges (influencée par les cations restant liés) vont influencer cette étape.

## Conclusions et perspectives

En conclusion, la spectrométrie de masse *native* ainsi que la spectrométrie de mobilité ionique sont deux outils très polyvalents. Elles peuvent être utilisées pour cribler différents motifs car l'identification des différentes espèces présentes en solution, leur stœchiométrie ainsi qu'une analyse quantitative peuvent être effectuées. Notre méthode apporte de nouvelles connaissances sur l'interaction  $Mg^{2+}$ -cations. Elle peut être adaptée à différents motifs ou encore différents cations ou ligands. Nous avons aussi montré que la spectrométrie de mobilité ionique ne peut être utilisée dans l'analyse de structures d'ARN qu'en tenant compte de la forte compaction qui se produit en phase gazeuse, aux états de charge obtenus par *SM native*.

Nos résultats ont aussi montré l'intérêt d'analyser des acides nucléiques en présence de sels. Une autre étape intéressante serait de trouver des additifs afin d'influencer le processus électrospray et de jouer sur la proportion en adduits.

Notre étude est portée uniquement sur de petits motifs tertiaires d'ARN. De nouvelles connaissances ont donc été amenées sur le comportement des ARN en phase gazeuse et permettent de faire un pas en avant vers l'étude de structures beaucoup plus complexes telles que les riboswitches.



## ABBREVIATIONS

A	Adenine/Adenosine
ATD	Arrival Time Distribution
bp	Base pair
C	Cytosine/Cytidine
CCS	Collision Cross Section
CCSD	Collision Cross Section Distribution
CID	Collision induced dissociation
DNA	Deoxyribonucleic acid
ESI	Electrospray ionization
G	Guanine/Guanosine
HP	Hairpin
IM-MS	Ion mobility-mass spectrometry
IMS	Ion mobility spectrometry
KC	Kissing complex
L	Ligands
m/z	Mass-to-charge ratio
mRNA	Messenger RNA
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NH <sub>4</sub> OAc	Ammonium acetate
NMR	Nuclear magnetic resonance
PDB ID	Protein data bank identification number
Q	Quadrupole
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
SPR	Surface Plasmon resonance
T	Thymine/Thymidine
TAR	Trans Activating responsive RNA element
T <sub>m</sub>	Melting temperature

TOF

Time-of-flight

tRNA

Transfer RNA

UV

Ultraviolet

# 1. INTRODUCTION

## I. Nucleic Acids

Nucleic acids, and more particularly Deoxyribonucleic Acid (DNA), can be considered as the unit of life. Indeed, DNA contains all the coding genetic information of each component of the cell and is transmitted through generations.<sup>1</sup> Cells are using this genetic code to synthesize proteins needed for their good functioning. As DNA is present in the nucleus of eukaryotic cells and proteins are present in the cytoplasm, an intermediary is needed to go from the first one to the other. This is where Ribonucleic Acid (RNA) plays an important role. DNA will be transcribed into messenger RNA (mRNA), which is itself translated into proteins thanks to other RNAs. *The central dogma* of molecular biology is based on this simple principle.<sup>2</sup>

As for proteins, nucleic acids are biopolymers. They are coded by a specific alphabet and, depending on this alphabet arrangement and the biological environment, nucleic acids can adopt different structures. General information on DNA and RNA structures are discussed in the following paragraphs. RNA is discussed in more detail afterwards.

### I.1. Bases and primary structure

Nucleic acids are composed of a chain of nucleotides.<sup>1</sup> Nucleotides are formed by three different parts: the phosphate group, the sugar and the base (Figure 1, A). The first difference between DNA and RNA comes from the nature of the sugar that is a deoxyribose for DNA and a ribose for RNA. The five natural bases are: Adenine (A), Guanine (G), Cytosine (C), Thymine (T) and Uracil (U). The second difference between DNA and RNA is that thymine can only be found in DNA and uracil only in RNA. The purine bases (A and G) composed of two aromatic cycles and the pyrimidine bases (C, T and U) containing a single aromatic cycle (Figure 1, B).

The primary structure is made by the succession of nucleotides. The phosphate group is making an ester bond between the 3'-OH of the first nucleotide with the 5'-OH of the second one (Figure 1, C). The alphabet composing the genetic code is then composed of just 4 letters. One can write the sequence of DNA or RNA as the succession of A, C, G and T/U. The successive phosphate-sugar-phosphate moiety is called the backbone of the nucleic acid.

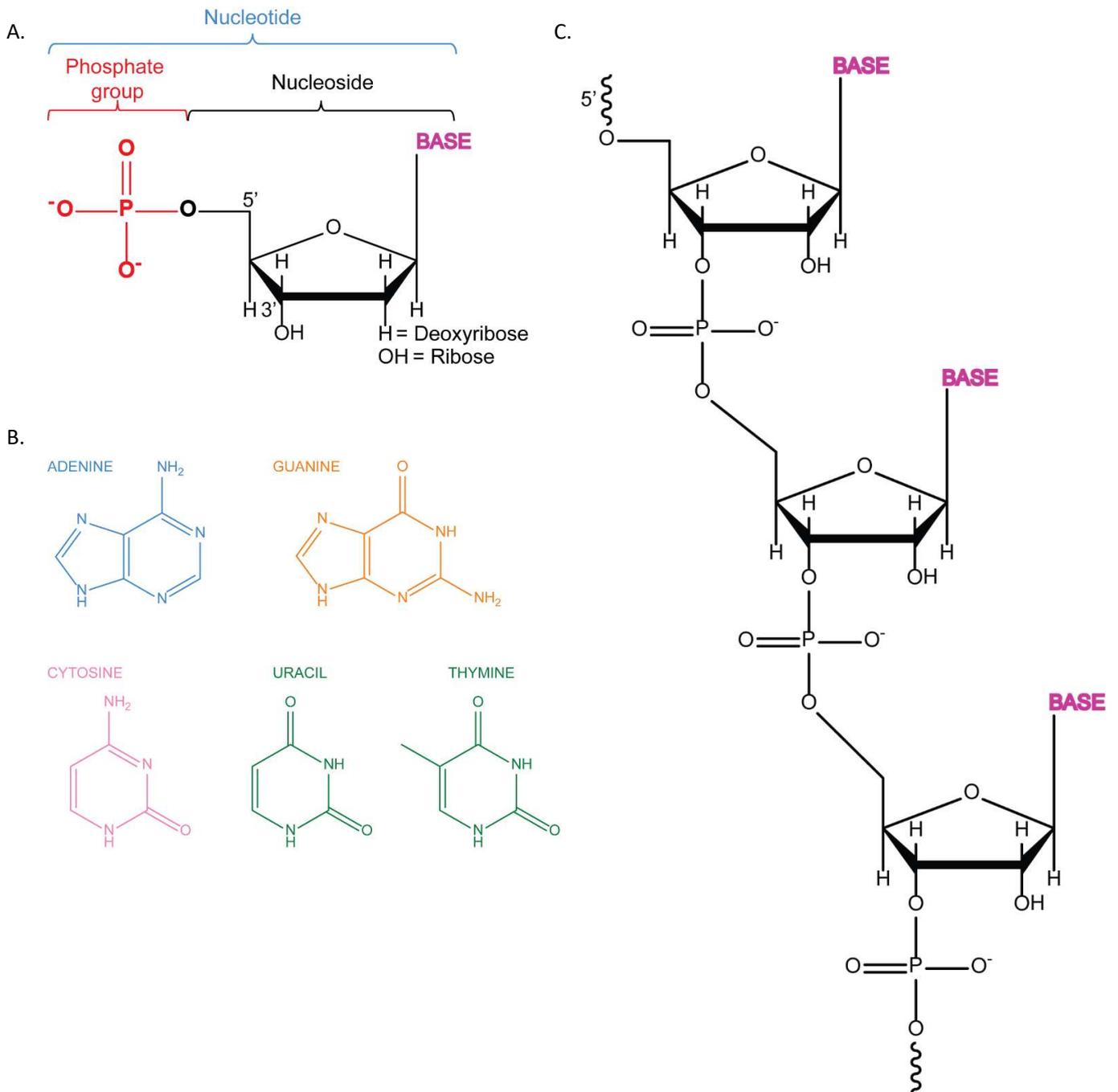


Figure 1: Bases and nucleic acids primary structure. A. Scheme of a nucleotide. B. Most abundant bases in DNA and RNA. C. Example of the primary structure of RNA (for DNA, the sugar is replaced by a Deoxyribose).

## 1.2. Secondary and tertiary structures

The secondary structure of DNA and RNA implies the formation of hydrogen bonds between two bases. The Watson-Crick base pairing is the most well-known hydrogen bonding. Adenine is binding to Thymine and Uracil via two hydrogen bonds, and Cytosine is binding to Guanine via three hydrogen bonds (Figure 2). Due to the number of H-bonds formed, a G-C base pair is more stable than an A-T or A-U base pair. In some cases, non-Watson-Crick pairing, or non-canonical base pair, can occur. The tertiary structure of nucleic acids derives from their secondary structure.

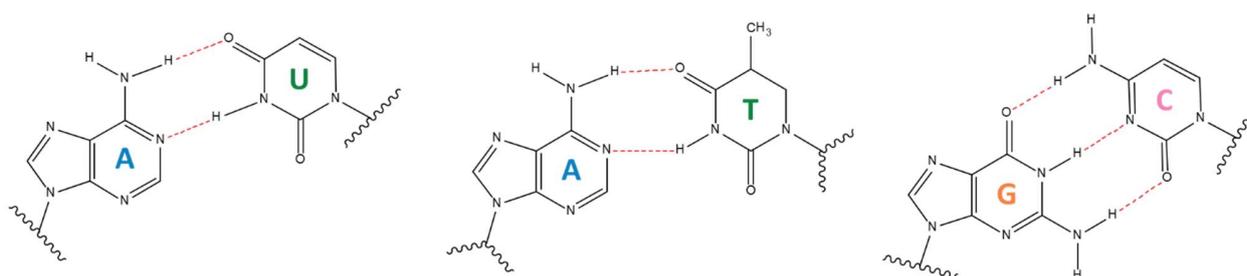


Figure 2: Watson-Crick base pairing between A-U, A-T and G-C (from left to right).

### 1.2.1. Most abundant three-dimensional structures based on Watson-Crick base pairs

#### B- and A-form duplex

The most famous structure is the double helix adopted by double stranded DNA and RNA. It was first described by J. Watson and F. Crick in 1953.<sup>3</sup> The B-form duplex, which is mostly adopted by DNA, is a right-handed double helix which consists of the binding of two antiparallel complementary strands held together by H-bonds between base pairs (Figure 3, left). The B-helix is defined by a turn composed of 10 residues and one can define a major and a minor groove.

The other well-known structure is the A-Helix, which is mainly adopted by RNA and can be adopted by DNA depending on the hydration conditions<sup>4</sup> (Figure 3, right). The A-helix is also a right-handed helix. The main difference between the A and B form comes from the placement of the base pair along the axis of the helix. In A-helix, the bases are more compressed and deviate from the vertical axis. This results in a difference of the major and minor grooves between A and B-form. In the A-form, the major groove is deeper which leads to a shallower minor groove.<sup>5</sup> Also in the A-form there are 11 residues per turn.

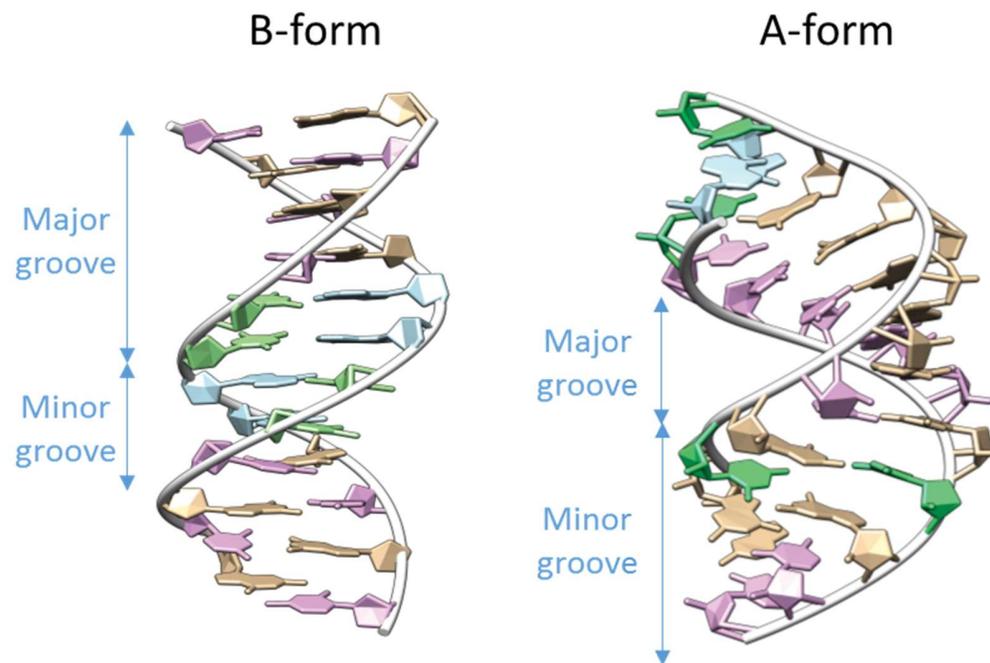


Figure 3: Example of a B-Helix on the left (PDB ID 1BNA<sup>6</sup>) and an A-Helix on the right (PDB ID 353D<sup>7</sup>).

### 1.2.2. Examples of structures based on non-Watson-Crick base pairs

#### Triplexes

A triplex DNA can be formed when a third strand is binding to the major groove of a duplex DNA via Hoogsteen bonding. This triplex can be intramolecular or an intermolecular triplex when the third strand comes from another DNA molecule. Triplexes can be found naturally but can also have potential therapeutic effects, for example as a molecular target for a specific duplex (Figure 4, A).<sup>8</sup>

#### G-quadruplexes

G-quadruplexes are formed by two or more quartets of Guanines (Figure 4, B). A quartet is formed by four guanines that are binding together via Hoogsteen H-bonds. G-quadruplexes are found in G-rich sequences and evidence has been brought that they are present in gene promoters or even in telomers.<sup>9</sup>

#### I-motifs

I-motifs can form in C-rich sequences and consist of the binding of two parallel stranded duplexes via C<sup>+</sup>-C base pairs (Figure 4, C). The Cytosines involved are exchanging a proton. I-motifs are stabilized in acidic conditions but may be formed in *in vivo* conditions.<sup>10</sup>

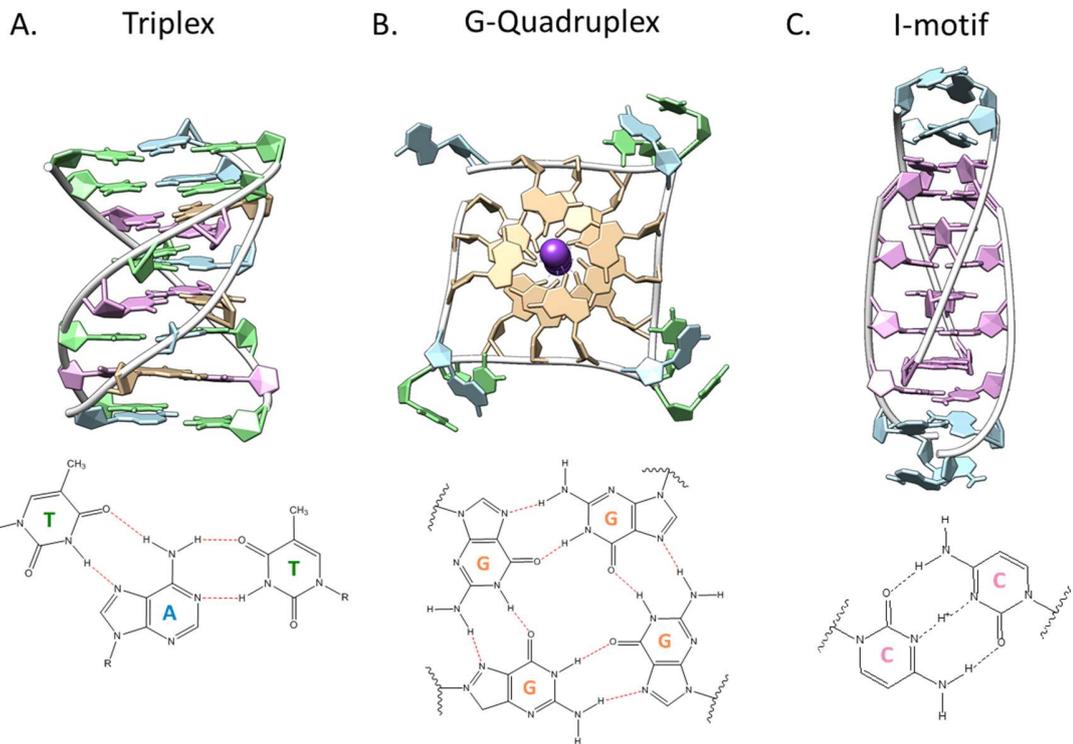


Figure 4: Examples of tertiary structures. A) Triplex DNA and the example of the Hydrogen bonds for a T-A-T triplet (PDB:1D3X<sup>11</sup>). B) Example of a G-quadruplex and the hydrogen network between Guanines (PDB ID 1KF1<sup>12</sup>). C) Example of an i-motif and the C<sup>+</sup>-C pair formed (PDB ID 1YBL<sup>13</sup>).

The structures and folding state of DNA and RNA are important to assure their functions in the cells. The next part is focusing on RNA and its structure-function relationships.

## II. An RNA story

In the *central dogma of molecular biology* exposed by F. Crick in 1958, the RNA is only considered as an intermediate between DNA and protein.<sup>2</sup> However, in 1970, F. Crick himself refines his own words about the transfer of the information between DNA, RNA and proteins.<sup>14</sup> He classified the transfer of information in three classes: general transfers, special transfers and unknown transfers (Figure 5). General transfers are the one previously described and admitted by the *Central dogma*. The special transfers may somehow happen depending on the cell type, and the unknown transfers may occur, but no proof was existing.

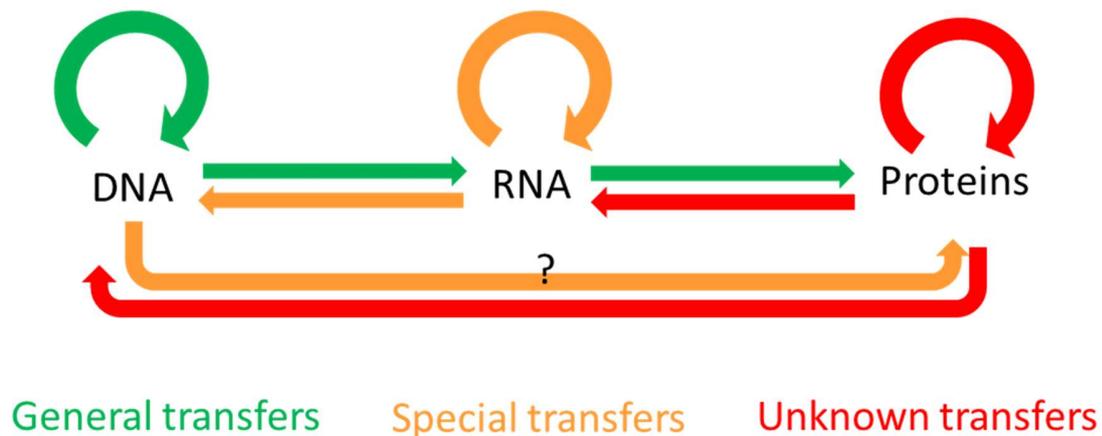


Figure 5: Refinement of the Central Dogma of Molecular Biology (adapted from Crick F.<sup>14</sup>)

From this fact arises the question of the origins of life, “who came first?” between DNA, RNA and proteins. This problem has become a chicken-and-egg problem. Several review papers are explaining the term of “RNA world” and why RNA should be considered first.<sup>15–18</sup> The term “RNA world” was used for the first time by Gilbert W. in 1986.<sup>15</sup> RNA was shown to be not only the intermediate between DNA and proteins, but has many diverse roles. RNA has been discovered to present catalytic abilities in the form of ribozymes and is involved in genetic regulation through riboswitches. All these capabilities of RNA suggest that an RNA world has existed before proteins and now all the different forms of RNAs are remnants of this ancient time.

### II.1. The biological roles of RNA

Here we present a list of the main biological roles of RNA and that constitute the vestiges of this so-called “RNA world”.

### **Main actors in protein synthesis: mRNA, tRNA and rRNA**

In 1962, when J. Watson and F. Crick obtained the Nobel prize of medicine for the discovery of the DNA structure, three types of RNA had been discovered and identified as key actors in protein synthesis: messenger RNA, transfer RNA and ribosomal RNA.<sup>1</sup>

Proteins are oligomers formed by the succession of amino acids. Proteins are then based on their own alphabet which is the amino acids succession. DNA contains all the genetic information coded under the succession of nucleotides. The questions asked were “How to go from DNA to proteins? How to go from nucleotides to amino acids?”. Figure 6 is showing a schematic view of the transcription-translation process. As DNA cannot go outside the nucleus of eukaryotic cells, DNA is transcribed into **messenger RNA** (mRNA) thanks to several proteins including the RNA polymerase II. mRNA is then translated in the cytoplasm to proteins thanks to ribosomes and transfer RNAs. The mRNA is “read” 3 nucleotides by 3 nucleotides following a specific code. Each triplet, called codon, corresponds to a specific amino acid. The synthesis of the protein starts at a starting codon, usually the AUG start codon, corresponding to Methionine in Eukaryotes and N-methylmethionine in Prokaryotes. **Transfer RNAs** (tRNAs) recognize the correct codon on the mRNA and ribosomes is the machinery allowing the interaction between mRNA and tRNA and the formation of the polypeptide chain. Each tRNA is unique because it is binding to only one specific amino acid and corresponds to one specific codon. The attachment of the proper amino acid to its corresponding tRNA is performed by the aminoacyl-tRNA synthetase. When the tRNA binds to its amino acid it is directed to the A-site inside the ribosome. tRNA and mRNA are interacting together inside the ribosome. Peptide bond and elongation of the polypeptide chain happen also in ribosomes. Ribosomes are composed of two subunits. In prokaryotes the large subunit is made of 34 proteins and two rRNAs (23S and 5S rRNA) and the small subunit comprises 21 proteins and only one rRNA (16S).<sup>19</sup> The small subunit mediates the interaction between mRNA and tRNA. The large subunit contains all the machinery to form peptide bonds. tRNAs pass through 3 different sites inside the ribosome: the A-site to interact with the corresponding codon, the P-site containing the elongated protein and the E site where tRNAs are released.

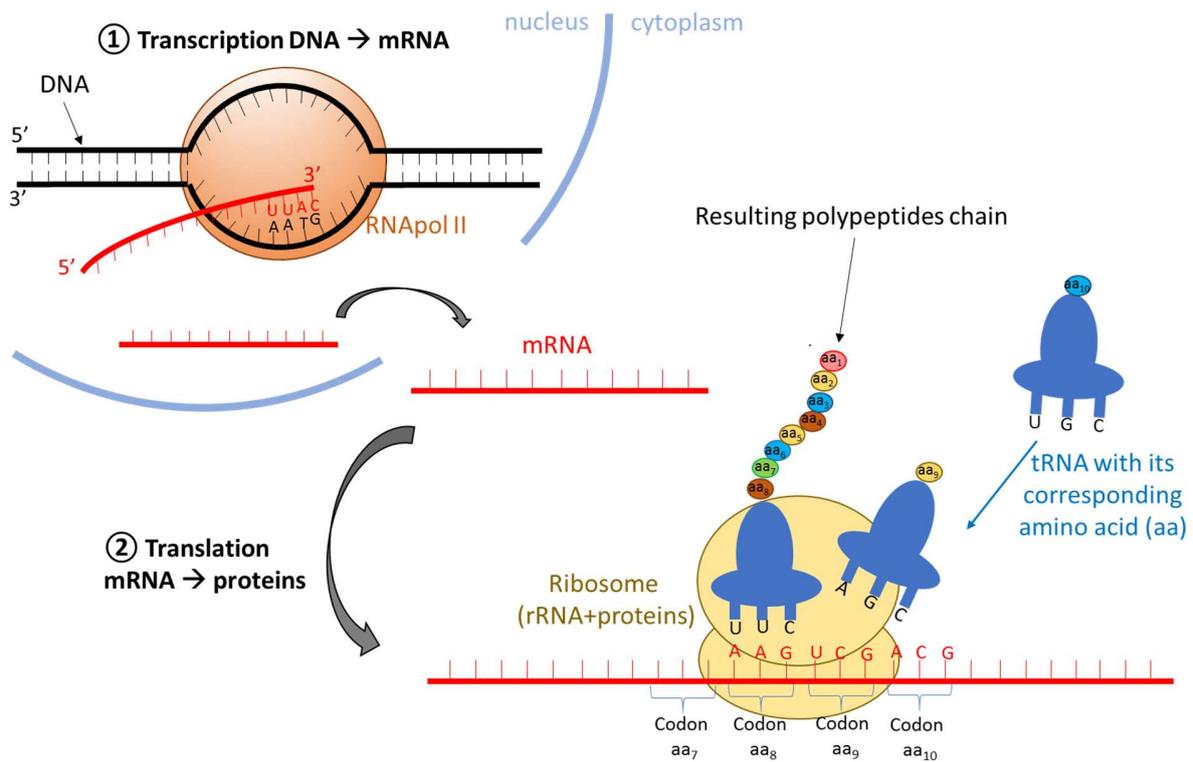


Figure 6: Scheme of the transcription and translation step to illustrate the three fundamental roles of RNA in protein synthesis (adapted from Lodish H. et al<sup>1</sup>).

Crystal structures of tRNAs and part of ribosomes were obtained and have revealed that both RNAs are structured in a specific manner to ensure their function.<sup>19,20</sup> Structural details will be given in part II.2 and II.3.

### RNA as catalyst: discovery of ribozymes

For a long time people thought that catalysis was only performed by enzymes. However, in the early 80's, Sidney Altman and Thomas Cech had discovered independently that RNA can be involved in catalysis reactions. For their respective works they received the Nobel Prize in chemistry in 1989. They discovered respectively that Ribonuclease P (RNase P)<sup>21</sup> and the *Tetrahymena* Group I intron<sup>22</sup> possess enzyme properties. RNase P processes tRNA precursor (pre-tRNA) by catalyzing the hydrolysis of pre-tRNA. RNase P is present in prokaryotes and eukaryotes.<sup>23</sup> The *Tetrahymena* Group I intron on rRNA precursor was shown to cleave and ligate itself without the intervention of proteins, and has thus autocatalytic properties. Since then, Group I intron were found in various organisms, from bacteria to eukaryotic cells. They catalyze their own excision from RNA through successive transesterification reactions (Figure 7, A).

Group I intron are considered as self-splicing RNAs as they do not need enzymes to be removed from RNA during mRNA splicing.<sup>23</sup>

The name Ribozymes for “**Ribonucleic Enzymes**” was then given to RNA molecules that present catalytic properties. Since the discovery of the two first ribozymes (RNase P and Group I intron), several other types of ribozymes emerged: the Hammerhead ribozyme which became the most studied and characterized ribozyme, the Hepatitis Delta Virus (HDV) ribozyme, hairpin ribozyme, Varkud Satellite (VD) ribozyme, the *glmS* ribozyme which is also a riboswitch and the rRNA in ribosomes which was found to catalyze the formation of the peptide bound.<sup>23–26</sup> Ribozymes catalyze three main types of reactions: transesterification, hydrolysis and peptidyl transfer. They are involved in many biological processes which may be important in the life of cells (RNA translation, RNA splicing, etc). As these RNA molecules are chemical catalysts and can act without the need of proteins, they may be a kind of proof of what is left from the RNA world where all the reactions were done without proteins.

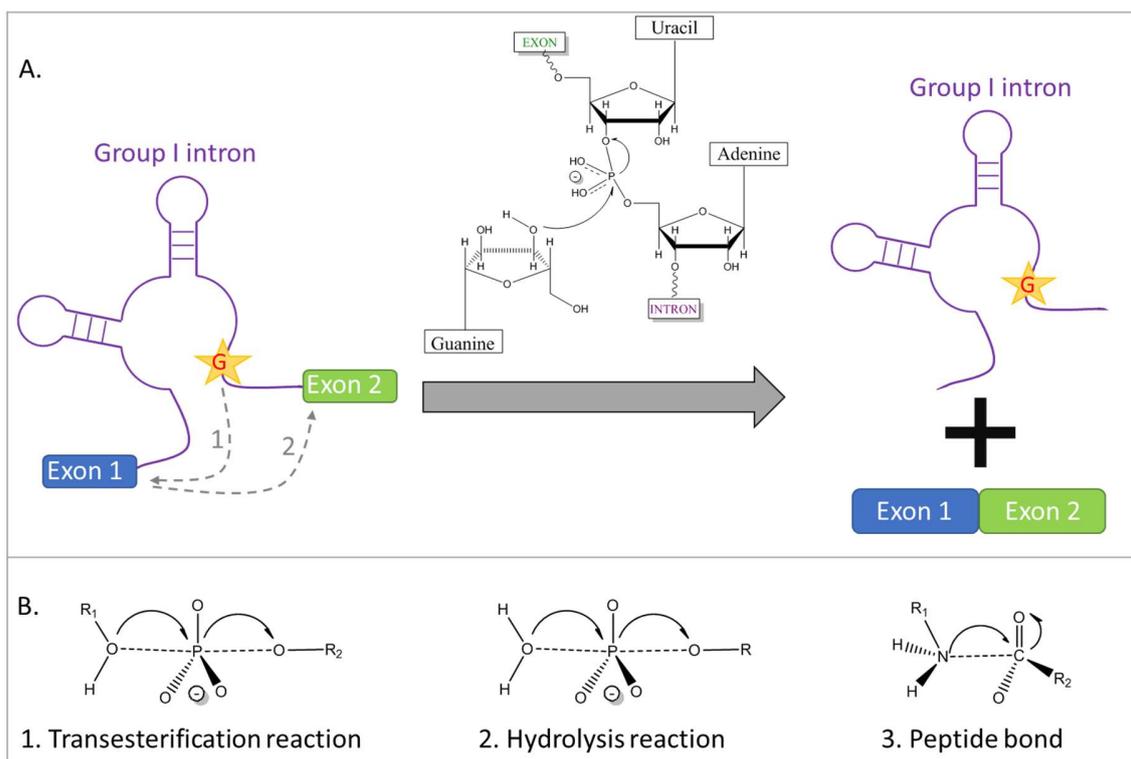


Figure 7: A) Schematization of Group I intron cleavage by two successive transesterifications autocatalyzed by the ribozyme (adapted from Ramesh A. et al<sup>23</sup>). B) Three main reactions catalyzed by ribozymes: transesterification like in Group I intron, hydrolysis like with RNase P and peptidyl transfer for rRNA in ribosomes.

It has been shown that ribozymes adopt specific structures and that they are composed of key motifs like helical parts or pseudoknots. In some cases, metal ions like magnesium are also binding and help for the

catalytic properties. These different structural features will be discussed in part II.2 and II.3 of this introduction.

### **RNA as gene regulator: discovery of riboswitches**

In 2002, researchers proved that vitamin derivatives like thiamine pyrophosphate (TPP, derivative of vitamin B<sub>1</sub>), Flavin mononucleotide (FMN, derivative of vitamin B<sub>2</sub>) and the coenzyme B<sub>12</sub> (also called AdoCbl, derivative of vitamin B<sub>12</sub>) interact with the mRNA controlling the expression of the corresponding vitamin.<sup>27,28</sup> They showed that each vitamin derivative was binding to a specific domain of the mRNA and that, above a certain threshold, it conduces to the regulation of the relative gene. The name riboswitch was given by Ronald R. Breaker to the part of mRNA that regulate gene expression through the binding of a specific metabolite. In the last fifteen years, riboswitches have become a trendy subject. More than 20 classes of riboswitches were discovered in many different organisms. Many researches and reviews have been published since their discovery in 2002.<sup>29–35</sup>

Riboswitches are found at the 5' untranslated region (UTR) of mRNA in eubacteria. However, a riboswitch binding to Thiamine Pyrophosphate (TPP) was found in plants, fungi and some other eukaryotic cells.<sup>36–38</sup> Riboswitches are composed of two domains: the aptamer part where the specific ligand is binding and the expression platform, directly downstream the aptamer, which transduces the ligand binding event into a gene control response (Figure 8). The binding of the ligand to the aptamer part induces a conformational change in all the riboswitch and more particularly in the expression platform, which leads either to the repression or to the expression of the gene. Riboswitches regulate genes through various mechanisms. The most common ones are the transcription termination, the translation initiation (Figure 8) and splicing control. Depending on the configuration of the riboswitch, the regulation can be an activation of the gene or a repression.<sup>30,34,35,39</sup>

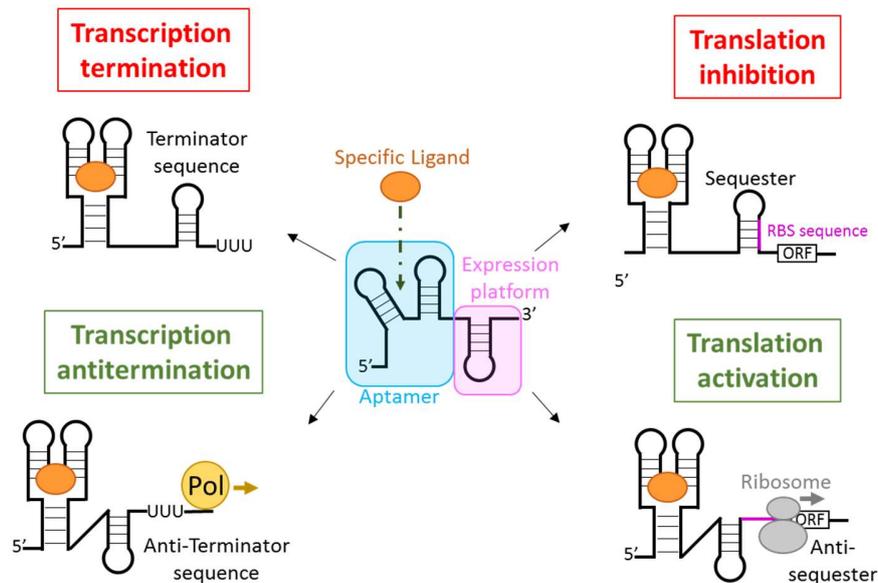


Figure 8: Mechanisms of gene regulation, examples with the transcription and translation (adapted from Serganov et al<sup>35</sup>). In red are the repression mechanisms and in green the activation mechanism. Transcription termination: no binding of the polymerase due to the termination forming sequence. Transcription anti-termination: polymerase can bind and transcribe the gene. Translation inhibition: the binding of the ligand induce a conformational change that hide the ribosome binding site (RBS). Translation activation: the RBS is accessible to the ribosome so translation can happen.

Riboswitches sense a large variety of ligands: coenzyme and vitamin derivatives as seen previously (TPP, FMN, coenzyme B<sub>12</sub>)<sup>27,28,40</sup>, purines and derivatives<sup>41,42</sup>, amino acids<sup>43,44</sup> or even cations like magnesium<sup>45,46</sup>.

As for ribozymes, the structure of riboswitches is well organized and is of great importance for their function as a conformational change is involved. We will discuss about these structural features in the following parts, II.2 and II.3.

## II.2. RNA structures: from RNA motifs to more complex structures

As seen previously, RNA ensures many functions and is involved in various biological processes. To do so it must adopt specific secondary and tertiary structures. Several reviews are detailing the different types of building blocks from secondary to tertiary structure motifs.<sup>47-50</sup> In this section, we will present the main secondary motifs and tertiary motifs that lead to more complex three-dimensional structures.

### II.2.1. Secondary structure motifs

The secondary structure of an RNA is defined by the interactions between nucleotides, which can bind through canonical or non-canonical base pairing, and by the succession of nucleotides. RNA adopts several

secondary structure motifs that are considered as building blocks for ribozymes or riboswitches structures.<sup>47,48</sup> We present here the main building blocks found in RNA. Another important part in RNA structure, which will not be described here, is base stacking.<sup>47</sup>

### **Single stranded RNA**

In 3D RNA structures, some parts remain as a single strand (Figure 9, A).

### **Duplex**

Duplex parts correspond to double stranded RNA (Figure 9, B). Usually they form a right-handed A-helix. The A-helix has 11 residues per turn and the base pairs are tilted from the helical axis of about 18° so the bases are displaced of 4 Å from the axis. Those constraints make the major groove deeper and narrower and on the contrary the minor groove will be shallower and wider compare to a B-helix.

### **Bulges and internal loops**

A bulge occurs when unpaired nucleotides are present on one of the two strands forming a double helix (Figure 9, C). The smallest bulge is composed of only one nucleotide but their size can go to several ones. The formation of bulges can bend the general A-form of the double helix, however the type of nucleotides, the number of nucleotides in the bulge, the nucleotide surrounding the bulge and the presence of cations can influence the percentage of bending of the helix. Bulges can be a preferred binding site for proteins, ligands or cations.<sup>51</sup> One can quote the example of the HIV-1 TAR RNA-Tat protein system where the TAR RNA sequence is forming a bulge and a hairpin and it was demonstrated that the bulge is involved in Tat recognition.<sup>52</sup>

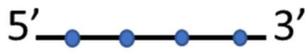
Contrary to bulges, internal loops are formed when unpaired nucleotides are present on both strands of the double helix (Figure 9, D). When one or two nucleotides are involved it is called a mismatch, if more than three nucleotides are involved it is an internal loop. The nucleotides of an internal loop or mismatch form non Watson-Crick base pairs. One of the most widespread mismatches is the G•U wobble base pair. It often constitutes a recognition site for proteins, ligands or cations.<sup>53,54</sup> Mismatches create a small distortion of the A-helix but do not bend the helix.<sup>47</sup> As internal loop geometry depends on the number and type of nucleotides involved, it can have a non-negligible effect on the bending on the A-helix.

## Hairpins

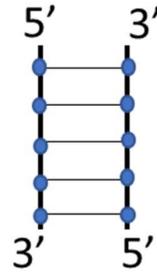
Hairpins are formed by a duplex part, the stem, and end by a loop of unpaired nucleotides (Figure 9, E). The size of the loop varies from two to several nucleotides. Hairpins are flexible depending on their composition, the size of the stem and loop. The hairpin is a motif of choice in RNA interactions and is important for RNA folding. For example, it is one of the main motifs involved in tRNA structure. Indeed, three hairpins are present on the structure and give this familiar structure of tRNA.<sup>20,55–57</sup> The most widespread loop is the tetraloop and more particularly the GNRA tetraloop (where N can be all nucleotides and R a purine). It is a motif naturally present in a variety of RNAs like rRNAs, group I introns or ribozymes like the Hammerhead.<sup>49,58,59</sup> The formation of this tetraloop involves a network of hydrogen bonds between the different bases, which is important for RNA structure stability.

## Junctions between helical parts

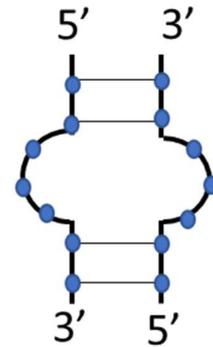
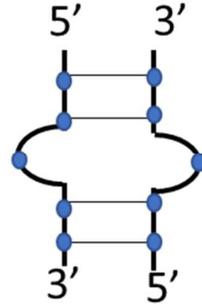
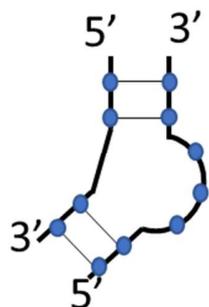
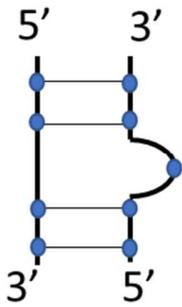
As said in their name, the junctions between helical domains are the regions where at least three different helical parts are connected (Figure 9, F). These regions are important for RNA folding as they determine the number of branches and so the global structure. For example, the three-way junction in the TPP-riboswitch<sup>60,61</sup> or in the Hammerhead ribozyme<sup>59</sup> gives the Y-shape of both structures. tRNAs are defined by a four-way junction.<sup>62</sup> In junctions, regions of unpaired nucleotides stabilize the overall structure. They can constitute the site of base triples or binding pockets. For example, in the Hammerhead ribozyme, the cleavage site is located on the unpaired bases in the junction.<sup>47,59</sup>



A. Single strand

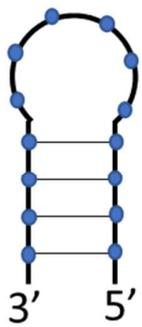


B. Duplex

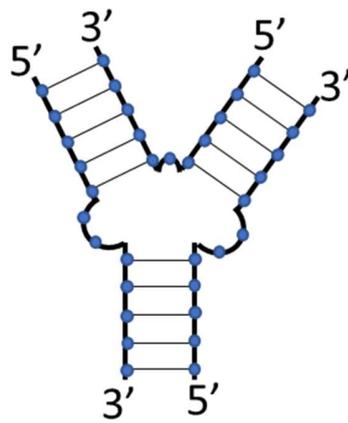


C. Bulges

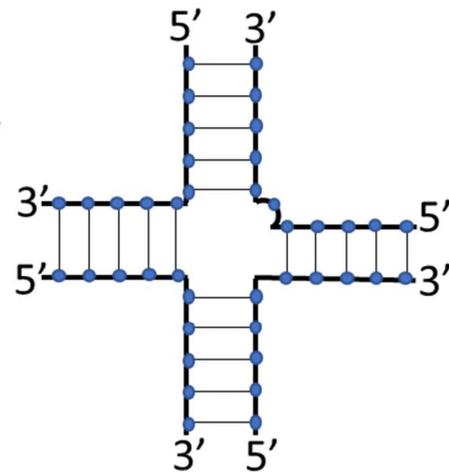
D. Internal loop



E. Hairpin



Three-way junction



Four-way junction

F. Junctions

Figure 9: Secondary structure motifs of RNA.

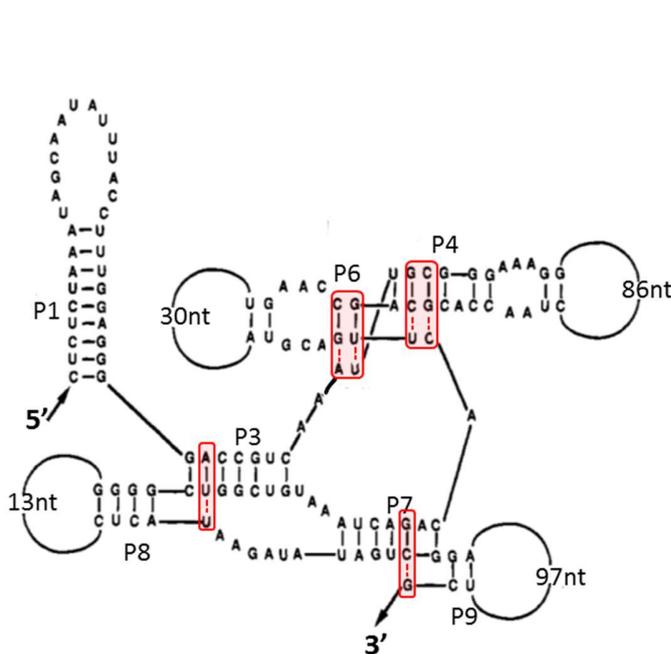
### II.2.2. Tertiary structure motifs

The secondary structure motifs described above associate and interact to form tertiary structure motifs, important for the global three-dimensional structure of RNA.

#### Base Triple interactions

A base triple happens when an unpaired nucleotide is making hydrogen bonds with a Watson-Crick base pair present in a duplex. It is a common motif found in most of complex structures of RNA because the formation of new hydrogen bonds helps to maintain and stabilize the RNA conformation. The known triples include (UA)U, (CG)C, (CG)G, A(GC), (UA)A, (CG)A and (AU)G.<sup>48</sup> Base triples were shown to be present in various RNA structures like tRNA<sup>Phe 55–57</sup>, group I intron ribozymes<sup>63</sup> or even riboswitches like the PreQ1-riboswitch<sup>64</sup>. Figure 10 presents two examples of structures involving base triples: Group I intron and the aptamer part of the PreQ1-riboswitch in *T. Tengcongensis*.

#### A. Group I intron



#### B. PreQ1 riboswitch aptamer

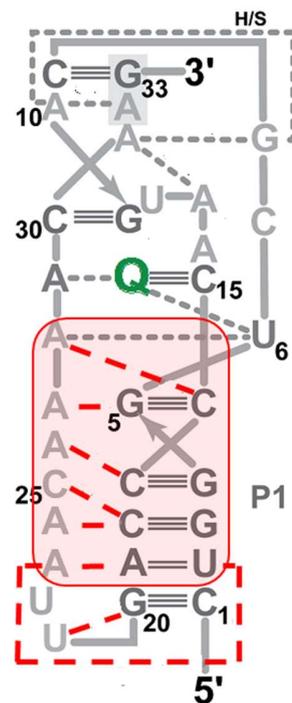


Figure 10: Example of base triple interactions involved in RNA structure. A. Group I intron RNA structure showing 6 base triples (adapted from Chastain et al<sup>63</sup>). B. Aptamer part of the PreQ1-riboswitch from *T. Tengcongensis* (adapted from Jenkins et al<sup>64</sup>). Hydrogen bonds and bases involved in base triples are shown in red.



## Pseudoknots

A pseudoknot is formed when unpaired nucleotides in a loop make hydrogen bonds with nucleotides of a single stranded region. One can consider this motif as a fusion of two hairpins. It is one of the most widespread tertiary motifs present in RNA structures. When the region between the two stems is at least one unpaired nucleotide, it bends the overall structure of the pseudoknot. If there are no bases between the two stems, the stems are coaxially stacked and form a continuous helix. This bending on the pseudoknots can have an influence on the RNA function, for example it has an effect on the efficiency of frameshifting during retroviral RNA translation.<sup>68</sup> The pseudoknot motif is easily found in riboswitches<sup>66,69,70</sup> or ribozymes<sup>65</sup>. Figure 11 presents some examples of RNA structures based on the pseudoknot element.

## Kissing complexes

Kissing complexes are formed when the nucleotides in the loops of two RNA hairpins are complementary. The smallest loop-loop interaction is formed by only two nucleotides in each loop<sup>71</sup>, and can go to several nucleotides. If the loop is too long, the probability to form other structures, like bulges or internal loops, is higher. Studies on the flanking bases at the interface between the loop and the stem show that this base pair is important for the kissing complex's stability. For example, in 2000, Ducongé *et al.* have shown that a G-A closing base pair makes the kissing complex more stable than any other base pair.<sup>72</sup> Study of loop-loop interactions were extensively done on the biological RNAI-RNAII kissing complex involved in the replication of the ColE1 plasmid in *Escherichia Coli* and on the TAR RNA sequence from the mRNA of the HIV-1 virus. These two kissing complexes will be described in more detail in the scope of the thesis. Except these two kissing complexes, this motif is present in other biologically relevant systems: the dimerization initiation site (DIS) of the HIV-1 mRNA which induces the dimerization of the viral genome by forming a kissing complex<sup>73,74</sup>, a kissing complex formed in the *Neurospora Varkud* Satellite ribozyme involved in the catalytic domain<sup>75,76</sup>, a kissing loop that improves the binding of the ligand and help for proper folding in the *btuB* riboswitch in *E.Coli*<sup>77</sup> or a loop-loop interaction formed in the *add* A-riboswitch after ligand binding<sup>41</sup>. Based on the study of several kissing complexes, the structure of this type of motif is bent at the junction between the two loops, and the formation of phosphate clusters in the major groove of the loop-loop interaction can possibly be metal binding sites.<sup>78,79</sup> Figure 12 presents some examples of RNA structures based on the kissing complex element.

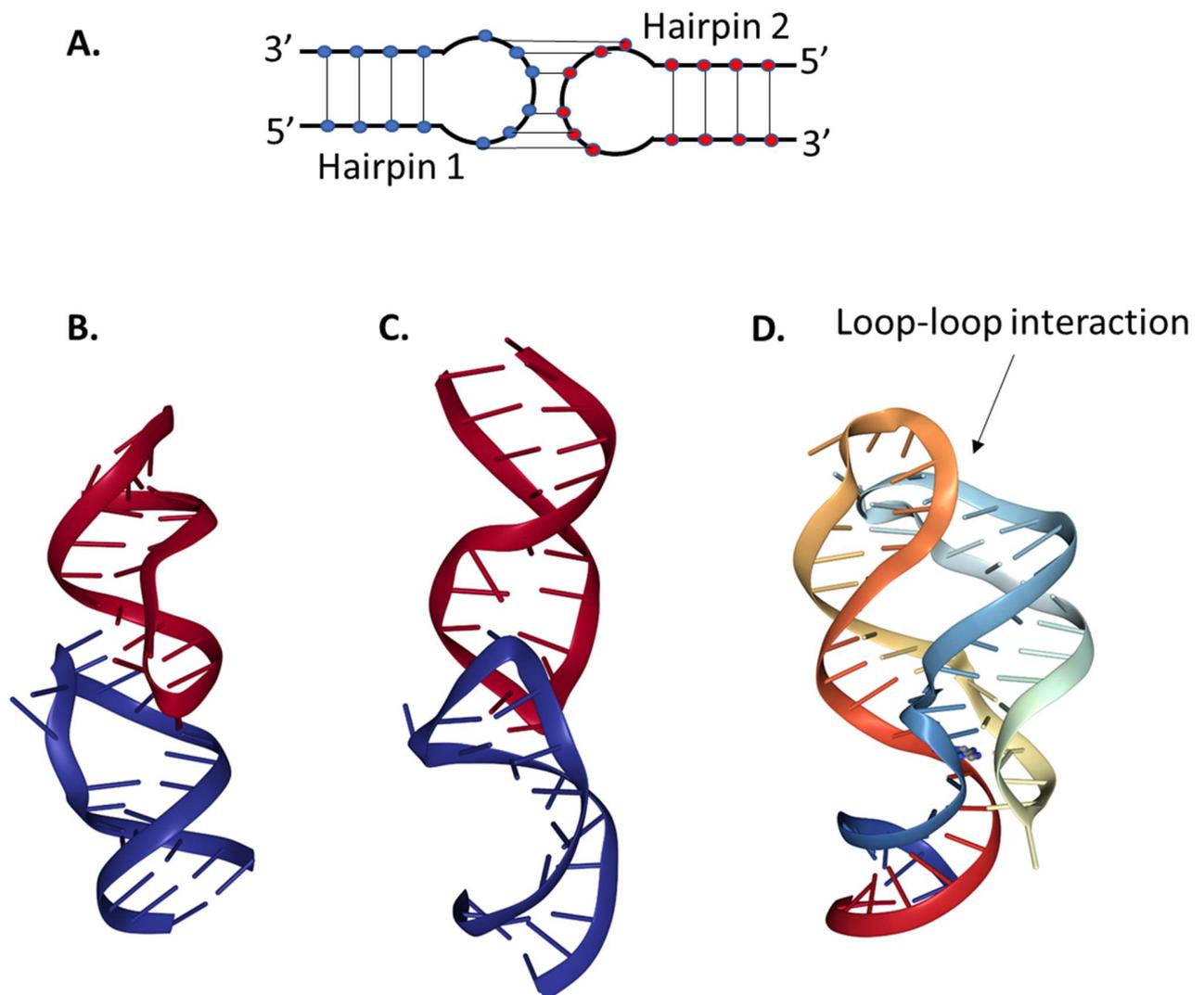


Figure 12: Examples of kissing loop structures. (A) Schematization of the kissing complex fold. (B) X-ray structure of the HIV-1 DIS Lai kissing complex (PDB: 2B8R). (C) X-ray structure of the Neurospora Varkud Satellite kissing complex region (PDB: 2M10). (D) X-ray structure of the add A-riboswitch from *Vibrio Vulnificus* (PDB: 1Y26)

### II.3. RNA structure and Cations

Most of the tertiary motifs and three-dimensional final structures of RNA bind to cations.<sup>50,80</sup> Several reviews show the importance of cations, and more particularly magnesium, in the folding and stability of tertiary structures.<sup>80-85</sup> In the following part, we will discuss about the importance of ions surrounding nucleic acids and we will focus more particularly on the role of magnesium cations.

### 11.3.1. The “ion atmosphere”

DNA and RNA are negatively charged polyelectrolytes: each phosphate is carrying a negative charge. In the cell, DNA and RNA are surrounded by ions that form the “ion atmosphere”. This “ion atmosphere” involves many monovalent and divalent cations depending on the ionic composition. This ionic environment reduces the electrostatic repulsion between all the negative charges present on nucleic acids. For example, without surrounding ions, the electrostatic repulsion in the folding of the 400 nucleotide-long *Tetrahymena* intron would correspond to  $\approx 600$  kcal/mol.<sup>86–88</sup> The ion atmosphere is thus of utmost importance in RNA folding and stabilization. Indeed, monovalent and divalent cations will help to maintain RNA in their functional structures by overcoming the electrostatic repulsion.<sup>83,84</sup>

The effect of the “ion atmosphere” on RNA folding structure will depend on its composition in monovalent and divalent cations. The main monovalent cations are  $K^+$  and  $Na^+$ . It has been shown that without  $Mg^{2+}$  and with a high concentration of monovalent cations, some RNA structures can still fold. Also the size of the monovalent ion has an impact on RNA folding. For example the A-riboswitch is more stable with  $Li^+$  than  $Cs^+$  (ionic radii:  $Li^+ < Cs^+$ ).<sup>89,90</sup> Some tertiary structures form mostly in the presence of monovalent cations; like G-quadruplexes DNA or RNA in which  $K^+$ ,  $Na^+$  or  $NH_4^+$  bind in-between each G-quartets to fold G-quadruplexes.<sup>91</sup>

Several theoretical models are used to describe the interactions between the ion atmosphere and nucleic acids. They take into account electrostatic interactions between charged element. However, we will not describe the different models in this manuscript. A review written by Lipfert et al<sup>88</sup> is explaining very well the basics of the most popular models like the Counterions Condensation (based on ion condensation around nucleic acids until a certain charge critical value is reached) or the Poisson-Boltzmann theory (based on an average description of electrostatic interactions). Modelling is challenging because one needs to take into account a lot of charges and interactions between them.

The ion atmosphere is mostly invisible by X-ray crystallography because this technique does not consider all the layers of ions that are surrounding RNA or DNA in a non-periodical way. X-ray crystallography instead shows direct contacts between nucleic acids and ions. One of the main experimental techniques allowing to “count” the ion pertaining to the “ion atmosphere” is the buffer equilibration-atomic emission spectroscopy (BE-AES). Briefly, this technique “counts” how many ions are present in a solution by comparing the sample containing the nucleic acid and a control composed of only buffer. It is then possible to count the number of ions which are in excess in the “ion atmosphere”. The quantity of nucleic acid is determined by the quantification of the phosphorus atoms.<sup>87,88</sup>

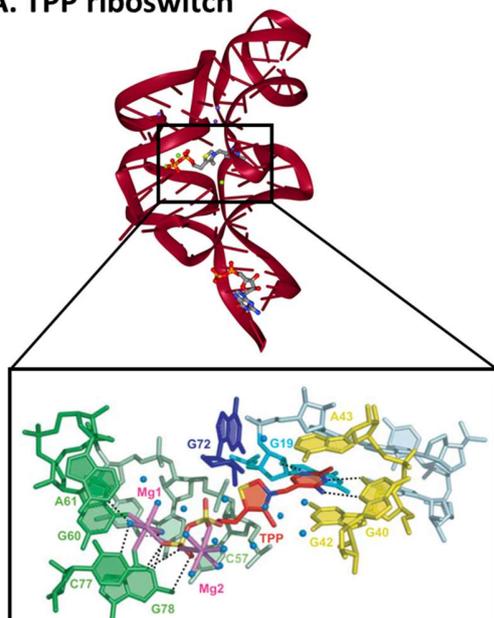
### *II.3.2. Role of Magnesium ions on RNA folding and stability*

It is now widely admitted that RNA folding and stability depend on the cation environment, and that among all ions the main cation involved is  $Mg^{2+}$ .<sup>80,82–85</sup> But one of the main challenges is to understand how magnesium is stabilizing RNA structures.

In the cell, the total concentration of Magnesium is about 17 to 20 mM. However, magnesium is involved in a lot of biological processes, involving proteins or nucleic acids.<sup>92</sup> Consequently, the free concentration of magnesium is only around 0.25 to 1 mM, because most of the magnesium is used somewhere in the cell.<sup>93</sup>  $Mg^{2+}$  has a small ionic radius (0.72 Å<sup>94</sup>) and in pure water it is coordinated to 6 water molecules. The dehydration of  $Mg^{2+}$  is energetically more costly than that of  $Na^+$  for example.<sup>85</sup>

Nowadays a lot of X-ray structures of complex RNA have been elucidated and present new insights into magnesium binding. One of the first structures obtained with  $Mg^{2+}$  binding, in the early 80's, is that of the tRNA<sup>Phe</sup>. From these structures, the number and location of magnesium ions were determined. The number of  $Mg^{2+}$  binding sites depends on the type of crystal obtained: for the orthorhombic form 4  $Mg^{2+}$  are binding<sup>95,96</sup> to tRNA<sup>Phe</sup> whereas for the monoclinic form, 5  $Mg^{2+}$  were thought to bind but after refinement two more  $Mg^{2+}$  were found<sup>20</sup>. In both cases,  $Mg^{2+}$  are located in the D-stem and in the anticodon-stem. All  $Mg^{2+}$  contribute to the stabilization of the global tRNA shape<sup>82,96</sup>. In 2006, crystallization of the aptamer part of the TPP-riboswitch has revealed that two hexa-coordinated  $Mg^{2+}$  were interacting with both the RNA and TPP. They helped TPP to bind to the aptamer by stabilizing its interaction, and also stabilize the overall Y-shape of the activated aptamer.<sup>60,61</sup> More recently, in 2013, a study of SAM-I riboswitch and  $Mg^{2+}$  showed that the active structure of the aptamer is obtained only when SAM interacts with magnesium.<sup>97</sup> From all these studies, magnesium presents different roles from conformation stabilization, interaction with specific ligands (i.e. riboswitches) or is involved in the cleavage reaction of ribozymes<sup>98</sup>. Figure 13 present the examples of the TPP and SAM-I riboswitches.

### A. TPP riboswitch



### B. SAM-I riboswitch

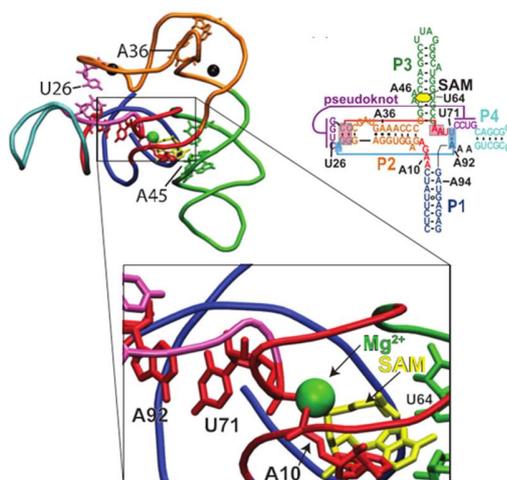
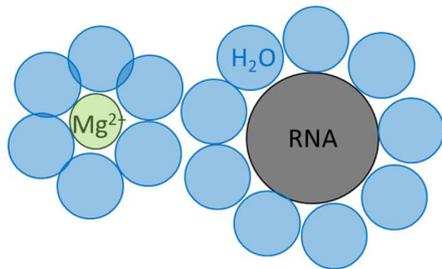


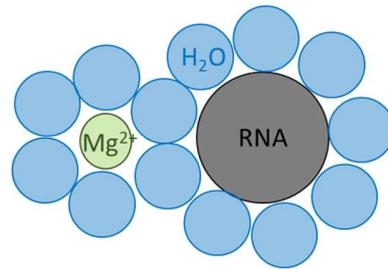
Figure 13: Examples of X-ray structures determined with  $Mg^{2+}$  ions. A) TPP riboswitch on top with a view of the location of specific  $Mg^{2+}$  (PDB: 2GDI).<sup>61</sup> B) SAM-I riboswitch extracted from Hennely et al<sup>97</sup>.

The acquisition of X-ray structures and the analysis of the interaction between  $Mg^{2+}$  and RNA have led to the classification of  $Mg^{2+}$  ions into two main groups: “diffuse ions” and “site-binding” ions.<sup>82,83,85</sup> Schematic representation of these two binding modes are shown on Figure 14. Diffuse ions bind to RNA through long-range electrostatic interactions. The first hydration shell of each molecules, i.e. ions and the RNA, are fully conserved and separated (Figure 14, A). Diffuse ions are moving around RNAs. On the contrary to diffuse ions, site-bound  $Mg^{2+}$  are closer to the surface of the RNA and are divided in two subgroups. The first group, defined by the term of “outer-sphere” complex, is represented by ions and RNA that share their first hydration shell. This type of ions can be trapped into electrostatic pockets from which they cannot move. The second group, composed by “inner-sphere” complexes, is defined by RNA-ions direct contact. In this case,  $Mg^{2+}$  is directly coordinated to the RNA without  $H_2O$  intermediate binding (Figure 14, B). Usually  $Mg^{2+}$  cations are binding to phosphates or oxygens. In some cases they are binding to nitrogen.<sup>85,99</sup> Experimentally it will be difficult to distinguish the different types of binding but several techniques can be used to obtain various information on stoichiometries and locations. Combining all the data can possibly give information on the type of binding. Biophysical techniques are described in the section II.4. Several theoretical models based on electrostatics and thermodynamics have been proposed to understand what is behind these different binding modes<sup>82,83,88</sup>, but they will not be discussed in this manuscript.

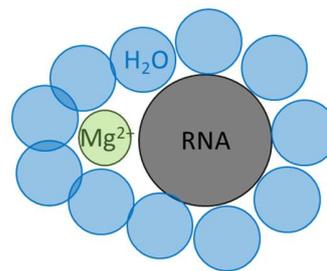
### A. Diffuse ions



### B. "Site-binding" ions



Outer-sphere



Inner-sphere

Figure 14: Representation of the different  $Mg^{2+}$  bindings to RNA. 1) Diffuse ions, 2) Site binding ions with the "outer sphere" complex and the "inner sphere" complex. Adapted from Misra et al<sup>82</sup>.

## II.4. Why and How to study RNA-cations structures and equilibria?

### II.4.1. WHY?

We have seen that the three-dimensional structure of RNA is important to assure its function. Also magnesium was shown to be involved in RNA folding and stabilization, like in riboswitches or ribozymes.<sup>82,100</sup> Understanding how ions are helping RNA to fold is of great importance, for example to understand the folding pathways involved in riboswitches or ribozymes. As an example, Figure 15 shows the hypothesis of the folding pathway of the A-riboswitch which may depends on cations and ligand binding.

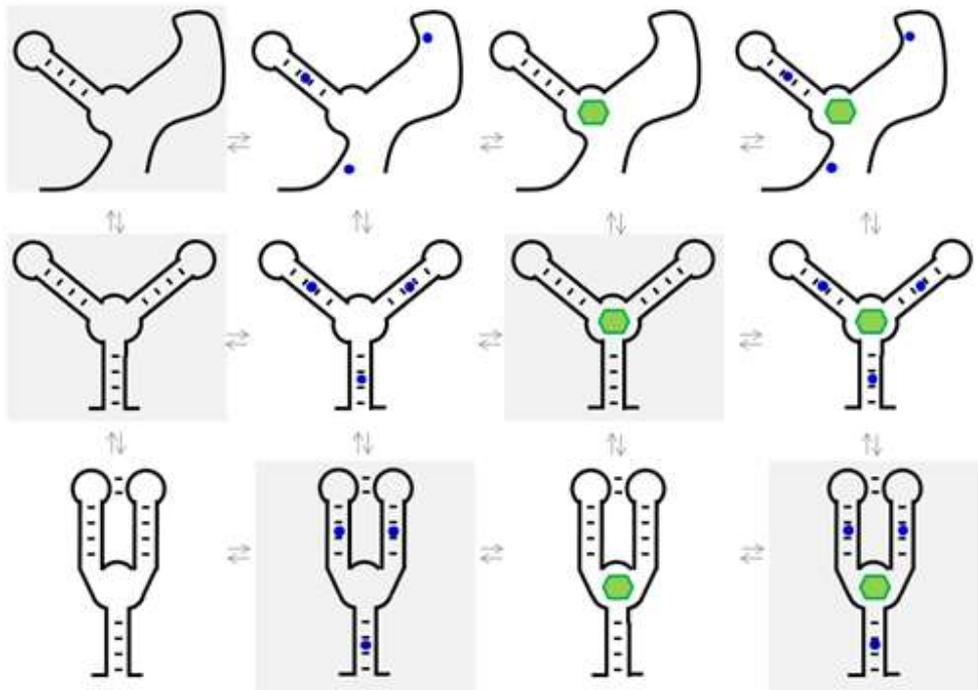


Figure 15: Hypothesis of folding pathways of an A-riboswitch aptamer. Several conformational states may exist and may depend on the binding of cations (blue dots) and/or ligands (green). The grey-shaded structures have been discussed in the literature.

Riboswitches and ribozymes are more and more studied nowadays and become interesting as drug targets or therapeutic tools.<sup>101</sup> As riboswitches are involved in gene regulation of mostly bacteria, they become a target of choice for new antimicrobial molecules.<sup>102</sup> For example, the TPP-riboswitch recognizes not only TPP as ligand but also Pyrithiamine pyrophosphate (PTPP) which is toxic for fungi and bacteria. PTPP blocks the binding of TPP and then induces gene repression.<sup>102</sup> Engineered hammerhead ribozymes have been created to cleave some specific parts of mRNA that are responsible of the expression of proteins important in the development of cancer cells.<sup>101</sup>

To engineer such structures, it is thus important to have as much information as possible on the folding and stabilization pathway. The fields of characterization of global structure of RNA complexes and RNA-Mg<sup>2+</sup> interactions are still challenging as it is still difficult to obtain structures of big RNA complexes, to understand clearly how magnesium is interacting with three-dimensional structures, where it is located and to quantify the effect of magnesium.

#### II.4.2. HOW?

To characterize RNA structures and interactions between RNA and Mg<sup>2+</sup> several biophysical techniques are used. Thanks to them we can have access to RNA stability, structural arrangement, Mg<sup>2+</sup> stoichiometry

or even  $Mg^{2+}$  location. However, all the existing techniques present advantages and disadvantages. In this section, we will describe some main techniques used to analyze RNA- $Mg^{2+}$  interaction and structures, their advantages and their limitations.

### **X-ray Crystallography**

X-ray crystallography is based on the analysis of the diffraction of X-rays by a crystal. By the acquisition of diffraction spectra, one can go back to electronic density map and then have access to the position of atoms and so to the structure and position of bound ions. Many structures of riboswitches<sup>61,66,103</sup>, tRNA<sup>20,55,56</sup> and ribozymes<sup>59,65</sup> were obtained by X-ray diffraction. The pros of this technique are its high resolution and the acquisition of atomistic details. Also, X-ray crystallography allows the identification of inner-sphere bound ions if resolution is good enough ( $\leq 3 \text{ \AA}$ ). So we can have access to magnesium stoichiometry.<sup>99</sup> However, in a recent study published in 2016, Auffinger and collaborators showed that mistakes are often made on  $Mg^{2+}$  assignment and one should be careful when using and analyzing the data.<sup>99</sup> One of the main drawback of X-ray diffraction is the need for crystals. Finding the good conditions can be time consuming, and not all conditions can be studied. In addition, by using X-ray diffraction one will have a rather static view of the structure, and no information about equilibria between RNA and cations is accessible.

### **Nuclear Magnetic Resonance spectroscopy (NMR)**

NMR is based on the resonance frequency of atoms (H, C, N mostly) in the presence of a magnetic field. As a function of their chemical environment, the resonance frequency of the atoms will shift. Distances between atoms are determined using 2D and 3D experiments. When the signal between two atoms is strong, it means that they are close to one another. The possible structure is reconstructed based on the defined distances. NMR is a solution technique compared to X-ray crystallography, so physiological conditions are easier to mimic. Like X-ray diffraction, NMR is a high-resolution technique that allows the determination of RNA structures. For example, structures of kissing complexes<sup>78,79</sup> and pseudoknots<sup>68</sup> were obtained by NMR spectroscopy. The limitations of this technique are first that  $Mg^{2+}$  is invisible to NMR, only assumptions on its position can be made. It is possible to do  $^{25}Mg$ -NMR<sup>104</sup> but this isotope has a low sensitivity so the experiments might be very long. Secondly, experiments can be long, costly, sample-consuming and the structure determination can be fastidious.

### Single-molecule Förster Resonance Energy Transfer (smFRET)

FRET is based on energy transfer between two chromophores, the acceptor and the donor. The donor will be excited. If the two dyes are close to one another, the donor will give energy to the acceptor. If the two dyes are far from each other, only the donor fluorescence will be higher and the acceptor will not fluoresce (Figure 16).

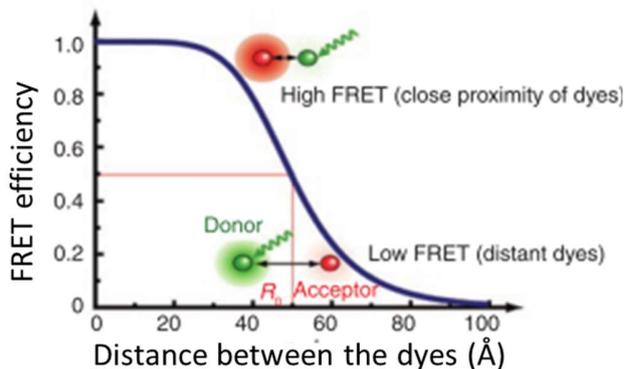


Figure 16: FRET efficiency as a function of the distance between the two dyes: principle of the smFRET experiments (adapted from Roy R. et al).<sup>105</sup>

The FRET efficiency is then calculated by the ratio of the fluorescence intensity of the acceptor over the intensity of the acceptor and donor. Then we can have access to the distance separating the two fluorophores as the FRET efficiency is inversely proportional to the distance. The smFRET is done at the single molecule level and observed under confocal microscopy. The goal of smFRET experiments is to obtain information on the distance between the two dyes. Information about folding equilibria and transitions can be extracted. The change in distances is correlated to the dynamics of the structure studied.<sup>105</sup> As an example, the folding of the TPP riboswitch with  $Mg^{2+}$  and/or ligands was studied using smFRET and revealed that the aptamer part is folding through a first step involving magnesium and then bind to TPP to form the activating conformation.<sup>106</sup> The advantage of this technique is that it can provide insights into possible folding pathway, kinetics and conformational changes can be observed. However, two of the drawbacks are that the molecule studied must be immobilized and it has to be tagged. These two steps may have an impact on the global folding of the molecule. Also, the effect of magnesium ions is measured only indirectly, via its impact on folding and dynamics.

### **Surface Plasmon Resonance (SPR)**

SPR allows one to determine binding constants between two partners. SPR is based on the measurement of refractive index changes. Briefly, one of the partner molecules (called the ligand) is immobilized on a sensorchip surface and the other one is injected over it (called the analyte). When there is binding between the analyte and the ligand, the change in mass will be detected using the total internal reflection of a laser light on the surface. The angle at which there is plasmon resonance will change due to the change in refractive index caused by the binding of the two partners. The main drawback here is that binding affinity of  $Mg^{2+}$  cannot be determined, because the change in mass will not be large enough. As a consequence, specific binding constants of  $Mg^{2+}$  cannot be obtained. The only data will be the determination of binding constant between partners as a function of  $Mg^{2+}$  concentration for example. It is a technique often used to characterize the affinity between two hairpins for examples.<sup>107,108</sup>

### **UV spectroscopy**

As nucleic acids are absorbing light at a maximum of 260 nm, one of the main technique used is UV spectroscopy to follow the concentration of the oligonucleotide. One of the most widespread application is to follow the absorbance as a function of temperature, to deduce the stability in solution of the nucleic acid sequence studied. Thermal denaturation experiments give access to the melting temperature ( $T_m$ ), the temperature at which 50% of the population is folded (or unfolded). It can be possible through this type of experiment to determine binding constants by calculating the reaction free energies ( $\Delta G^\circ$ ), and to extract reaction enthalpies ( $\Delta H$ ).

### III. Mass Spectrometry as a tool to study RNA

The biophysical techniques presented in the previous section are providing either details on structure or binding constants, but it is difficult to obtain precise data on cations and more particularly on the equilibria in which each magnesium ion is involved. Here we propose to use mass spectrometry as another tool to study RNA complexes. The goal of the following paragraphs is to describe why mass spectrometry and more particularly *Native* mass spectrometry can be used to study biomolecular complexes such as RNA complexes, and what kind of information can be obtained by mass spectrometry.

#### III.1. *Native* mass spectrometry

MS is based on the identification of the components of a solution by separating the molecules according to their mass to charge ( $m/z$ ) ratio. The principle of mass spectrometry is the following: 1) the molecules are ionized and desolvated. 2) They are brought to the analyzer where they are separated according to their  $m/z$  ratio. They can pass through the instrument thanks to electric fields which guide the ions. 3) Ions are detected. Using mass spectrometry many applications are possible, but the one which is of interest to us is mass spectrometry to study non-covalent complexes. RNA complexes are non-covalent complexes, so the study of their structure involved the preservation of all the interactions from the solution to the gas phase. *Native* MS is thus the technique of choice.

*Native* mass spectrometry is an approach based on electrospray ionization (ESI). A recent review paper written by Albert Heck and co-workers<sup>109</sup> gives an excellent definition of *Native* MS. One can say that the term “native” is undue because the molecules are analyzed in the gas phase. The name “native” comes from the fact that it uses gentle conditions to preserve molecules in their initial state (their stoichiometry, for mass detection, and even their fold, for gas-phase structural characterization), from the solution to the gas phase. Solvents used in *native* MS are non-denaturing, close to physiological conditions and have to be volatile.

Because non-covalent complexes are very important in biology, their study have grown and so is the use of *native* MS. *Native* MS has become a technique of choice in structural biology study.

##### III.1.1. *Beginning of Native MS: importance of the electrospray mechanism*

When mass spectrometry was invented, more than 100 years ago, the unique goal was to measure mass to charge ratio, and at the beginning only small molecules were analyzed.<sup>109</sup> As the interest for biologically relevant systems, like proteins, was growing, one also needed new ionization methods. The electrospray ionization (ESI) method, as we know it nowadays, was first described in 1984 by John B. Fenn.<sup>110</sup> He

received in 2002 the Nobel Prize in chemistry for his developments in this field. He used ESI to ionize macromolecules from nucleic acids (14-mer) to large proteins (bovine albumin dimer, 133000 Da).<sup>111</sup> ESI ionization is soft enough to allow the transmission of such biomolecules into the gas phase without disrupting them.

The ESI process occurs at atmospheric pressure. The mechanism can be divided in three main parts: 1) formation of charged droplets, 2) fission of those droplets and then 3) ion desolvation. These steps are happening in the microsecond time scale.<sup>112</sup>

#### 1) Formation of the charged droplets

The solution is injected inside a capillary, or needle. A voltage is applied between this capillary tip and the entrance of the mass spectrometer. If the instrument is set in negative mode, the counter electrode (entrance of the mass spectrometer) is more positive than the capillary and thus negative ions are attracted. Due to the electric field applied, the meniscus at the end of the tip is disturbed and then form what is called a "Taylor cone" (Figure 17, A. step 1), from which small droplets are emitted.

#### 2) Droplet fissions

Those parent droplets formed can then evaporate thanks to collisions with the ambient gas (often Nitrogen gas). The size of the parent droplets will decrease until a certain limit which is called the Rayleigh limit (when the Coulomb repulsion between the droplet charges becomes larger than the cohesive forces, i.e. the surface tension). Then offspring droplets are formed, and will in turn undergo evaporation and fission processes (Figure 17, A. step 2).

#### 3) Production of desolvated ions

Two main models are proposed to explain the formation of the final ions after the droplet fission: the ion evaporation model (IEM)<sup>113</sup> (Figure 17, B.) and the charged residue model (CRM)<sup>114</sup> (Figure 17, C). In the first model, which is mostly addressed for small molecules, when the droplets reach a certain size, ions will be directly emitted from these droplets. In the second model, the final droplets contain only one ion. This model is commonly accepted for large molecules, and we believe it applies to nucleic acids, which are hydrophilic.

Currently, several studies are still done to understand clearly all the processes that are involved in the electrospray ionization.<sup>115</sup>

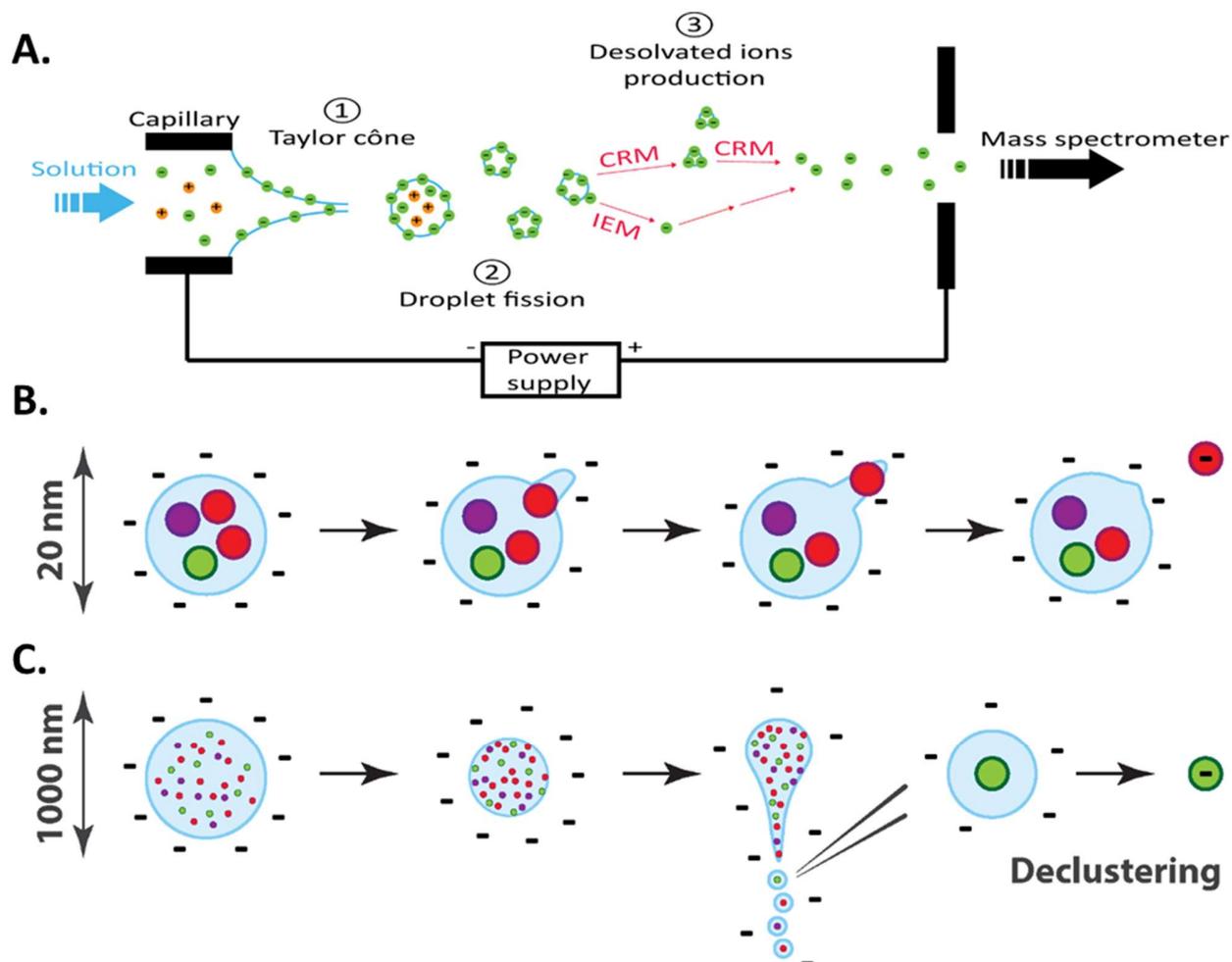


Figure 17: Principle of the electrospray process, example in negative ion mode.

As it is one of the softer way to ionize molecules and keep them intact, ESI-MS has become a powerful tool to study large biomolecules. Nowadays, ESI is widely used to study various non-covalent complexes like proteins or nucleic acids but also complexes between them.<sup>116</sup> Concerning nucleic acids, it has been possible to analyze a large range of complexes between DNA, RNA, proteins in complex together or with cations and ligands.<sup>109,117,118</sup> For example, it has been possible to detect G-quadruplexes DNA, DNA duplexes and triplexes binding to specific cations and ligands.<sup>119–123</sup> Concerning RNA, several studies have been done on the TAR RNA sequence and its interaction with the Tat protein<sup>124–126</sup>. Also the group of Daniele Fabris has focused several studies on the analysis of the HIV-1 dimerization sequence, forming a kissing complex, and its interaction with the nucleocapsid protein.<sup>127–129</sup>

### III.1.2. What can we learn from mass spectrometry experiments?

First, it allows the identification of the different molecules that form complexes in solution, when knowing the masses of each partner. As a consequence, it allows the determination of stoichiometries of complexes between several molecules (biomolecules with ions, ligands, cations). What will be important here is the declustering step during electrospray ionization. Indeed, the ionization has to be energetic enough to remove solvent and non-specific adducts ( $\text{NH}_4^+$  mostly) but without altering the stoichiometry or structure further, in order to measure interactions between the partners as they were in solution.

It is also possible to relate the intensities of each species detected in MS to their concentrations in the initial solution. To do so, we need to be aware that the ionization efficiency of each species could be different, and so the intensities of the different species may not reflect directly the concentration in solution. However, a correction method was developed in order to take into account the possible difference in ionization response of the different species.<sup>130</sup> This point will be described in more details in the material and methods section. By knowing the concentration of the different partners, equilibrium binding constants in solution can be determined because what is present in the gas phase is reflecting what was in solution.

Another advantage of MS is that it is a label free technique. As it does not involve the use of labels or immobilization, it avoids changes of conformation that may happen when derivatizing the system.

Mass spectrometry does not show only advantages. Even though the sample consumption is low, samples cannot be reused for other studies because MS is a destructive method. Also, it is a relatively expensive technique as it needs special instrumentation. Also, just by using mass spectrometry, structural details cannot be obtained. However, it is possible to add a new dimension to MS by coupling it to ion mobility spectrometry (IMS).

### III.2. Ion mobility spectrometry (IMS)

In ion mobility spectrometry (IMS), in addition to the separation by the  $m/z$  ratio, the molecules will also be separated according to their shape. Coupling IMS to MS (IM-MS) will allow one to obtain structural information, which cannot be accessible just by using MS. The field of IM-MS has grown significantly in the past 10 years, since commercial instruments became available.<sup>131</sup> Nowadays, *native* Ion Mobility coupled to Mass Spectrometry (IM-MS) is also an emerging tool in structural biology. It is a label free technique, small amount of samples is needed, and screening of molecules can thus be quick.<sup>132–134</sup> Native

IM-MS is used mainly to study protein complexes, but also nucleic acids, and complexes between various type of molecules.<sup>132,135–138</sup>

Three types of ion mobility instruments exist: temporally-dispersive (ions will arrive at different arrival time), spatially-dispersive(ions are separated based on the difference in mobility between different electric field) and confinement and selective release (ions are trapped and then ejected following their mobility).<sup>131</sup> Two main temporally-dispersive techniques are mostly used: Traveling Wave Ion Mobility Spectrometry (TWIMS) and Drift Tube Ion Mobility Spectrometry (DTIMS). As the instrument used for the study described in this manuscript is a DTIMS, only its functioning will be addressed. The principle is illustrated in Figure 18.

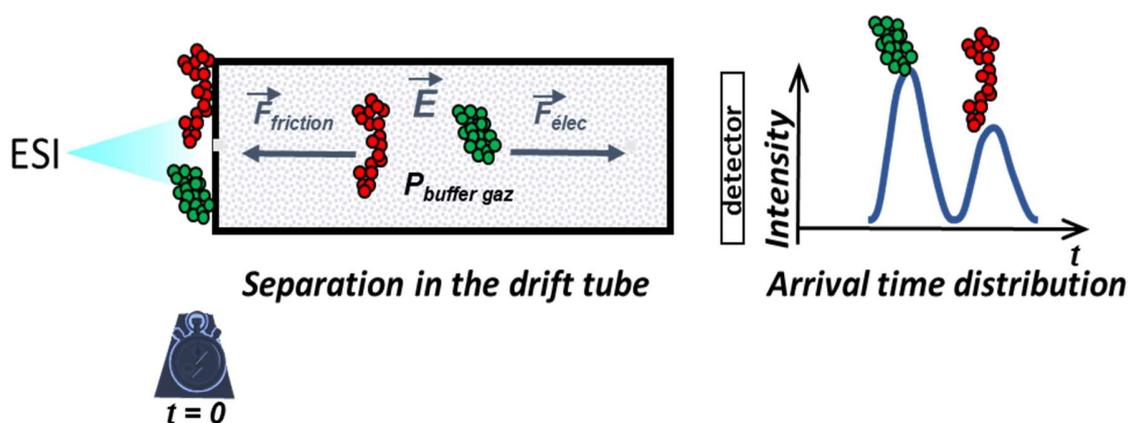


Figure 18: Schematization of the principle of Drift tube ion mobility spectrometry.

First, all ions are injected at the same time into a mobility cell which contains a neutral buffer gas. In the cell, the ions are accelerated thanks to an uniform and linear electric field but also slowed down by collisions with the buffer gas. The balance of forces results in a steady-state velocity. The ions are detected as a function of their arrival time. An extended molecule will arrive after a more compact one (for a same charge and  $m/z$  ratio). This arrival time and the mobility values are correlated to the collision cross section (CCS) which has the dimension of an area (unit:  $\text{\AA}^2$ ). The conversion from the arrival time distribution (ATD) to CCS distribution is described in the material and method section VIII. The CCS will be the value that will allow us to make a link with a three-dimensional structure. In *native* MS, theoretical calculations of CCS are done on structures obtained by molecular dynamics in the gas phase.<sup>135,139</sup> Comparison between theoretical CCS and experimental ones is the clue to find the most probable structure in the gas phase.

### III.2.1. What is the collision cross section of a molecule?

The CCS is related to the orientationally averaged momentum transfer collision integral.<sup>140,141</sup> This value will be related to the frictions between the ion and the buffer gas. Indeed, IMS is measuring the friction that are slowing down the different ions depending on their size, charge and shape. The CCS value will also depend on the buffer gas, usually Nitrogen or Helium, the temperature and pressure. Indeed, all these parameters can affect the velocity of the different ions and gas molecules and so affect their collisions.<sup>141</sup>

One of the biggest challenges in the determination of theoretical CCS of biomolecules comes from the fact that they have cavities and possess a specific topography, which is defining their particular structure. The friction will depend on this topography. It is difficult then to quantify the “real” size or volume of a biomolecule.<sup>140</sup> Several models have been developed to calculate the  $CCS_{theo}$  of a molecule. In a recent paper, D’Atri et al<sup>139</sup> are explaining the different models using a rigid DNA G-quadruplex,  $d(TG_4T)_4$ , as example. Briefly, three main models exist. The Projection Approximation model (PA) considers the atoms as hard spheres and the  $CCS_{theo}$  is defined as the average area of the “shadow” of the molecule obtained by projection, like the shadow on a wall due to the light. This method is rejected for our systems because it does not account for the effects of surface roughness. The second method, and the chosen one for nucleic acids, is the Exact Hard Sphere Scattering model (EHSS). Here also atoms are considered as hard spheres and the model measures the scattering angles of the buffer gas due to the collisions with the ion of interest. This model takes into account the cavities of the structure and the possibility of multiple collisions due to cavities. The last method is the trajectory method (TM) which takes into account the long- and short-range interactions. This method is the most precise, because it also accounts for the fact that atoms and molecules are not hard spheres, but it is also time consuming, and thus less practical for large molecules like proteins or nucleic acids.<sup>141</sup> The EHSS method gives reliable enough agreement with experiments ( $\approx 1.5\%$  deviation for  $d(TG_4T)_4$ )<sup>139</sup>.

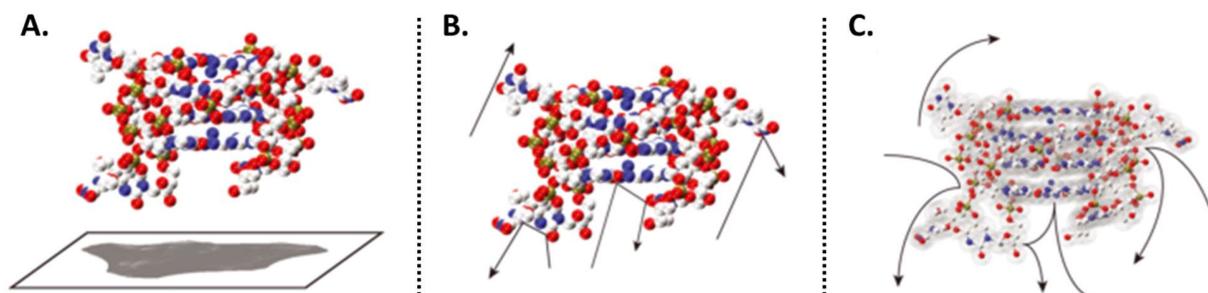


Figure 19: Schematic representation of the three different model to calculate  $CCS_{theo}$  (adapted from D’Atri et al).<sup>139</sup> A) Projection approximation model (PA). B) Exact Hard Sphere Scattering model (EHSS) and C) Trajectory method (TM).

### III.2.2. Pros and cons?

One of the main characteristics of native IM-MS is that the experiments are carried out in the gas phase. The advantage is that one can focus only on intrinsic properties of the molecules as we are in a solvent-free environment. Also, the mass resolving power allows to separate individual cation binding stoichiometries. But working in the gas phase constitutes also a disadvantage as we are not anymore in solution and that the environment surrounding biomolecules is biologically relevant. That is why it is important to answer the question “what happens to biomolecules structure in the gas phase?”. A lot of studies were focused on proteins and how they behave in the gas phase. It has been shown that proteins keep a long-time memory of their *native* state in the gas phase but also that they undergo rearrangement, collapsing and sometimes unfolding.<sup>142–144</sup> In this work, we have asked ourselves this question for nucleic acids, and more details will be given in the introduction of the part II of the results.

In our case, IMS could potentially be very useful to probe conformational changes due to the binding of cations and/or ligands. Also, it could help us to discriminate if several conformations are present for the same  $m/z$  ratio.

### III.3. Tandem mass spectrometry

Tandem mass spectrometry (MS/MS) is a technique based on ion fragmentation. It is used to obtain structural details on biomolecules, study their relative stability and it is also possible to study the dissociation of complexes in the gas phase. For example, stability in the gas phase of 16-mer DNA duplexes were assessed by MS/MS and the results suggested that some hydrogen bonding and base stacking interactions initially present in solution were conserved in the gas phase.<sup>145</sup> MS/MS was also used to probe binding modes of some ligands to nucleic acids.<sup>121</sup> Study of fragmentation pathways in the gas-phase were also done on small oligonucleotides from 4 to 8 bases.<sup>146</sup> More recently, MS/MS was used to study the binding between the TAR sequence and the protein Tat.<sup>126</sup> The authors show that MS/MS is also a powerful tool to obtain information on binding sites and stoichiometry. In our case, it is used to obtain more knowledge on the dissociation channels of the RNA complexes of interest and to potentially localize specific cations and/or ligands.

Several types of fragmentation methods exist, with variations on the way ions are activated: by electrons, by collisions or by photons. A recent review by J. Brodbelt is presenting the different techniques and their use for peptide and proteins analysis.<sup>147</sup> The most widely used method is Collision Induce Dissociation (CID). More details will be given to this latest method as it is the one used in our study.

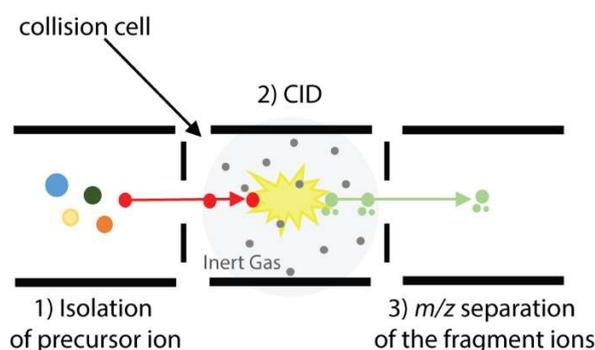


Figure 20: Principle of collision induced dissociation (CID). 1) The precursor ion is first isolated. 2) Then it is fragmented in the collision cell where it acquires a certain amount of energy which allow its fragmentation. 3) The fragment ions are analyzed through their  $m/z$  ratio.

A parent ion population is isolated according to its  $m/z$  ratio. This ensemble of ions will then undergo collisions with a neutral buffer gas in a collision cell, which conduces to its fragmentation (Figure 20). A potential difference,  $\Delta V$ , is applied on the collision cell in order to activate the ions. The ions will acquire a certain kinetic energy ( $E_{rel}$ ), which is related to the laboratory collision energy ( $E_{lab}$ ), the energy given by the electric field to the ion.

$$E_{lab} = q \times \Delta V \quad (1)$$

$$E_{rel} = \frac{m_n}{m_n + m_M} \times E_{lab} \quad (2)$$

Where  $m_n$  is the mass of the neutral gas and  $m_M$  is the mass of the ion. Part of  $E_{rel}$  will be converted into internal energy ( $E_{int}$ )<sup>148</sup>.  $E_{int}$  is redistributed as vibrational energy ( $E_{vib}$ ) that conduces to the ion fragmentation. Experimentally, we cannot have access to  $E_{int}$ , but we can plot the relative abundance of each species (parent ions and fragment ions) as a function of the experimental value that was used to change the internal energy (i.e  $\Delta V$ ), provided that we compare only sequences that have the same mass and the same charge state. We can then have information on the dissociation channels.

## IV. Scope of the thesis

In the introduction we have seen that RNA adopts a variety of specific structures that are involved in various biological processes. Furthermore, RNA structures are stabilized by addition of cations like magnesium. A lot of biophysical techniques have been used to characterize as much as possible RNA structures, folding pathways and RNA-Mg<sup>2+</sup> interactions. However, some structural and thermodynamic details are still missing.

In the present work we will demonstrate that *native* mass spectrometry and ion mobility mass spectrometry are powerful tools and that they can bring new knowledge in the field of RNA-cations interactions and induced conformational changes. The study will be then divided in two parts:

- PART I is focused on the characterization of binding equilibria between magnesium and RNA model motifs using *native* mass spectrometry. In this part, we will discuss about the development of a direct method to analyze the RNA-Mg<sup>2+</sup> interaction. The method enabled us to determine the specific stoichiometry of Mg<sup>2+</sup>, their relative binding affinities and their possible localization. We will show also in the last chapter that *native* MS can be used as a screening method.
- PART II explores the structures of RNA complexes in the gas phase using ion mobility-mass spectrometry, with the aim to detect shape changes upon cation or ligand binding. We will then discuss about the use of a gas phase technique to probe solution structures.

As RNA structures are very diverse in size, role, complexity, and because most of the tertiary structures needs cations to fold, different models have been chosen for our study. We have decided to focus our study on RNA-Mg<sup>2+</sup> interaction on one type of tertiary motif: the kissing complex motif. To do so, we also chose two well-characterized systems: the TAR-R06 kissing complex and the RNAlI-RNAlII kissing complex. The state of the art about these two models is discussed hereafter. For the IM-MS study, we have chosen to compare DNA and RNA duplexes and RNA kissing complexes. Both structures are important structural element in RNA as described in the introduction.

### IV.1. TAR-R06 kissing complex

The first model used is the TAR-R06 system. The TAR sequence, meaning Trans-Activation Responsive RNA element, is a part of the Human Immunodeficiency Virus (HIV) genomic RNA. The Human Immunodeficiency Virus type 1 (HIV-1) is a *lentivirus* which belongs to the *retrovirus* family, discovered in the 1980's.<sup>149</sup> It is responsible of the HIV infection and in the longer term leads to acquired immunodeficiency syndrome (AIDS). HIV virus infects mostly immune cells. The virus is composed of an

envelope containing a capsid in which two copies of positive single-stranded RNA are found. Consequently to infection, the two RNAs are released in the cytoplasm where they are reverse transcribed into double-stranded DNA by viral proteins. This double-stranded DNA then migrates into the nucleus and integrates into the host genome. Then the host machinery takes the lead to transcribe the newly integrated DNA. The integrated virus genome is composed of two regulatory regions (5' long terminal repeat (LTR) and 3'LTR) and 9 protein-coding genes that lead to the synthesis of proteins necessary for viral replication and for the creation of new viruses (Figure 21).<sup>150</sup>

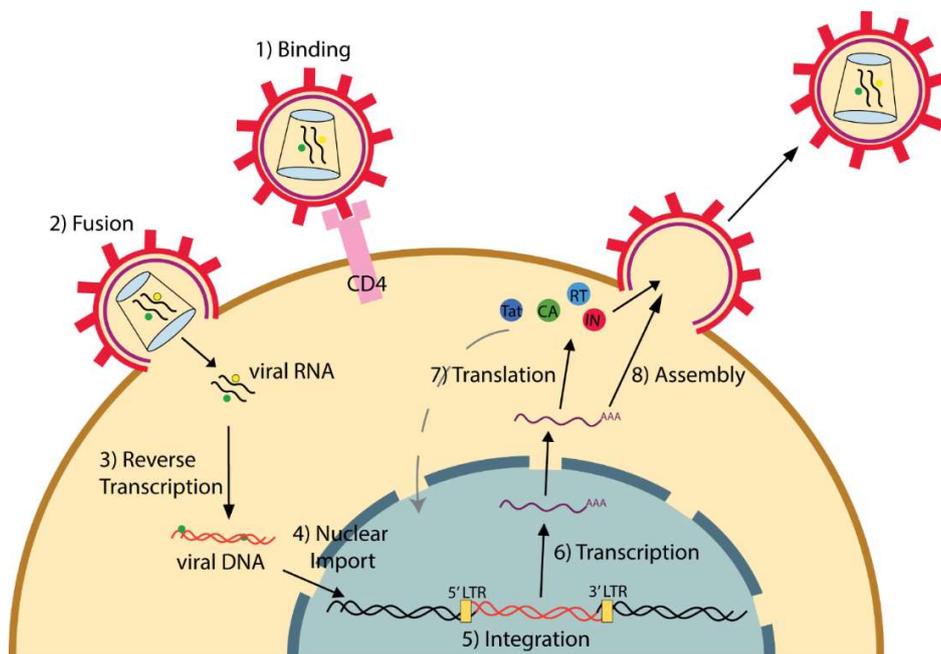


Figure 21: HIV replication cycle (adapted from Engelman et al<sup>150</sup>).

One of these essential proteins is the Tat protein. Tat, meaning *Trans*-activator protein, stimulates and induces a high-level transcription of the integrated viral mRNA. In other words, Tat is essential for RNA elongation.<sup>52</sup> To perform its role, Tat is binding to a small region of the 5' LTR called *Trans*-activation Responsive (TAR) element. The original TAR element consists of a 57-nt RNA sequence that forms a hairpin with a bulge in the stem region. However, studies have shown that not all the TAR sequence is essential for protein recognition, only the hairpin and the bulge are used for protein binding (~20-nt).<sup>52</sup> R06 is an RNA aptamer synthesized and selected to bind specifically to the hairpin TAR.<sup>151</sup> The kissing complex formed could prevent the transcription of the HIV RNA by stopping the binding of the different proteins to TAR and more particularly prevent the formation of the Tat-TAR complex.<sup>151</sup> One can consider the kissing complex formed between TAR and R06 as synthetic as R06 does not exist in nature.

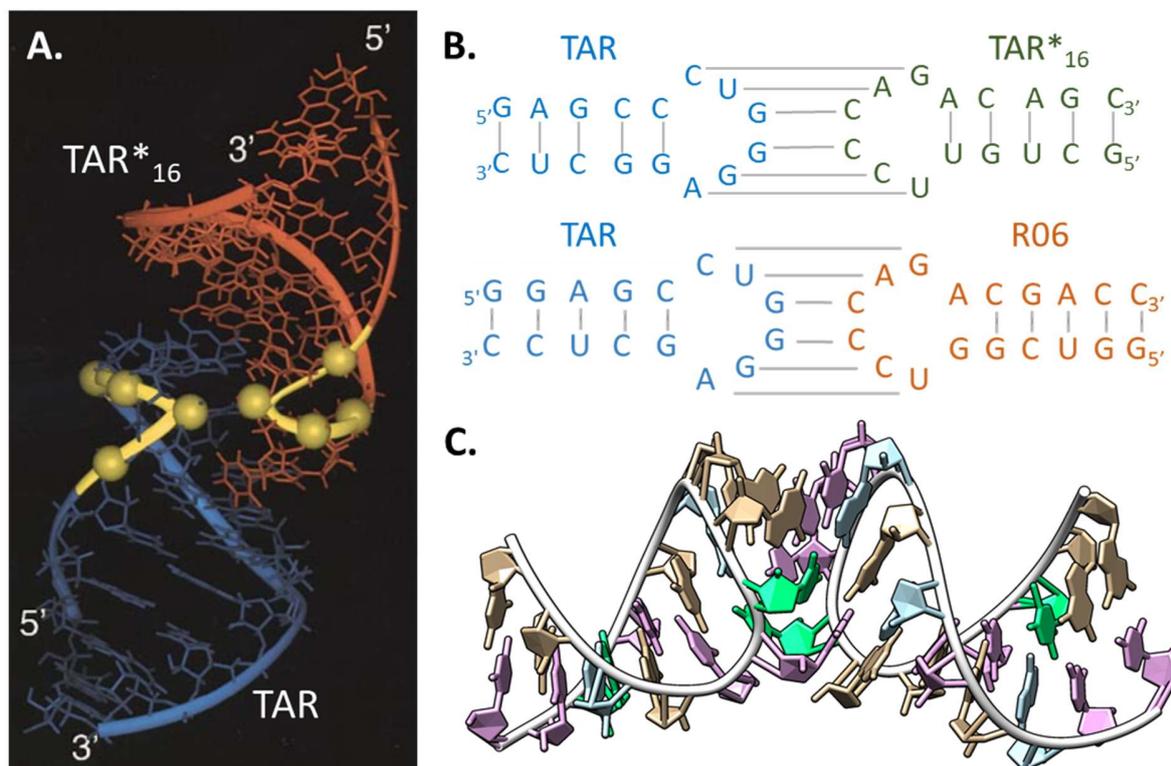


Figure 22: Structure obtained for kissing complexes involving the TAR sequence. A. NMR structure of TAR-TAR\*<sub>16</sub> (from Chang et al<sup>78</sup>) showing the two phosphate clusters in yellow. B) Sequences of the two kissing complexes: TAR-TAR\*<sub>16</sub> and TAR-R06. C) X-ray structure of TAR-R06 (PDB:2JLT).<sup>107</sup>

Figure 22 shows several structures of TAR forming a kissing complex with two complementary hairpins done by NMR<sup>78,152</sup> and by X-ray crystallography<sup>107</sup>. The TAR sequence used is a 16-nt truncated version of the original one. The X-ray structure of TAR-R06 shows that the global shape of the kissing complex is the A-helix but the helix is bent at the junction between the two hairpins.<sup>107</sup> Based on the NMR data obtained with another hairpin, TAR\*, two phosphate clusters flanking the major groove of the loop-loop helix can form binding pockets for magnesium.<sup>78</sup> Also magnesium was shown to stabilize the TAR-R06 kissing complex and dissociation constant between the two hairpins was determined (seen by thermal denaturation and SPR experiments).<sup>107</sup> Magnesium stoichiometries were determined indirectly by using the enthalpy of formation of the kissing complex, determined by thermal denaturation experiments. They found that  $\Delta\text{Mg}^{2+} = 1.7 \pm 0.1$  at 3 mM of  $\text{MgCl}_2$ , where  $\Delta\text{Mg}^{2+}$  is the number of  $\text{Mg}^{2+}$  uptake for the kissing complex formation.<sup>72</sup> However, the magnesium binding constants are not known.

## IV.2. RNAI-RNAII kissing complex

The second model is the RNAI-RNAII kissing complex. This kissing complex is involved in the replication of the plasmid colE1 in *E. Coli*. Study of this kissing complex has begun in the 80's. The plasmid ColE1 codes for several genes but especially the one of Colicin E1, which is a toxin for some strains of *E. Coli*.

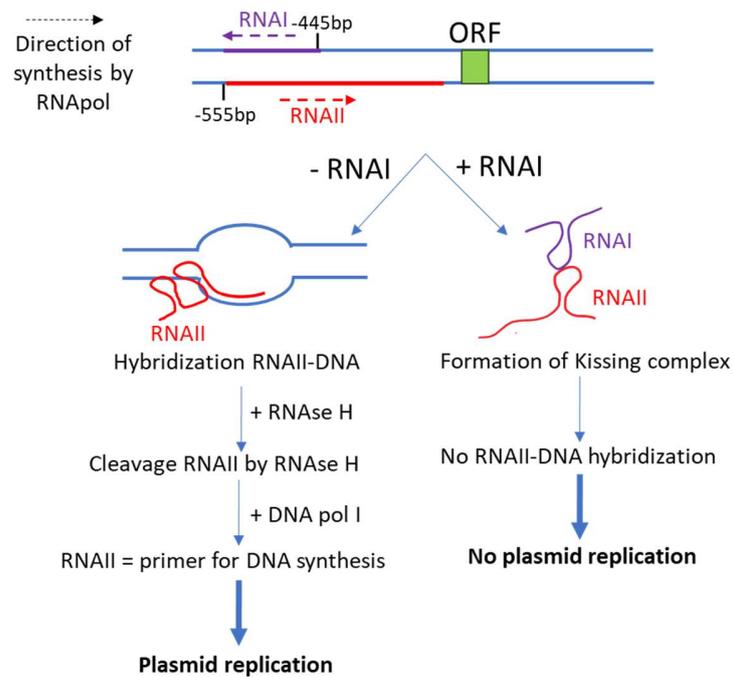


Figure 23: Scheme of the role of RNAII and RNAI in the replication of the plasmid ColE1.

Figure 23 illustrates the replication of ColE1 controlled by RNAI and RNAII. The replication of the ColE1 plasmid begins by the initiation of the synthesis of a primer transcript, called RNAII, by the RNA polymerase at 555 bp upstream of the origin of replication. This RNA is hybridized with the template DNA near the origin of replication. After its cleavage by the RNase H, RNAII will serve as a primer for the DNA synthesis by the DNA polymerase I. The replication of the plasmid is controlled by another RNA which is the RNAI. The synthesis of RNAI takes place 445 bp upstream the origin of replication and is synthesized in the opposite direction of RNAII synthesis, and terminates close to the initiation site of RNAII synthesis. RNAI will then interact with a preferred region of RNAII by forming a kissing complex. This newly formed structure will prevent the hybridization of RNAII to the DNA template by changing its conformation and so prevent the replication of the plasmid. Studies have shown that the rate of binding of RNAI to RNAII influence the control of the replication. Indeed, if the binding is too slow, RNAII can hybridize with the DNA template and replication happens. The binding rate is controlled by a protein called Rom.<sup>153-155</sup>

Several studies have shown that when the loop sequence of RNAi and RNAII are inverted from 5' to 3', the stability of the complex is 350-fold more stable than the original sequence.<sup>156</sup> That is why for biophysical studies the inverted sequence, RNAIi and RNAIIi, are preferentially used. The stability of the kissing complex was assessed by thermal denaturation with different cations and also by evaluating the effect of magnesium concentration on the stability. The results shown that the kissing complex is more stable with an increased concentration of Mg<sup>2+</sup> and that the stability with Co<sup>2+</sup> and Mn<sup>2+</sup> is equivalent to the one with Mg<sup>2+</sup>.<sup>156</sup> Like for TAR-R06 system, the Mg<sup>2+</sup> stoichiometry was determined indirectly, using the same method, and was estimated at 2.<sup>156</sup> Figure 24 shows the NMR structure of RNAIi-RNAIIi kissing complex<sup>79,157</sup> The kissing complex is bent at the junction between the two hairpins. As for TAR-TAR\*<sub>16</sub>, two phosphates clusters flanking the major groove of the loop-loop interaction are appearing. These clusters can constitute preferential binding sites for magnesium cations and also can constitute recognition site for proteins like Rom.<sup>79</sup>

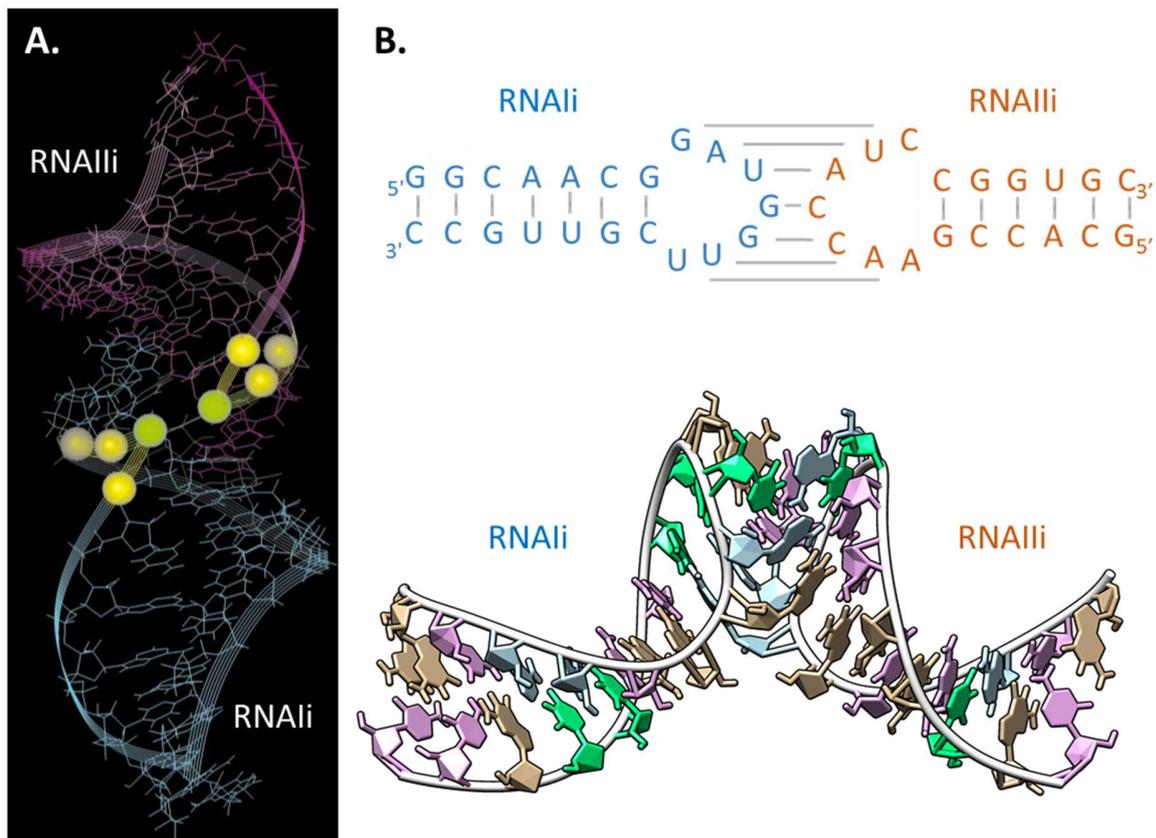


Figure 24: NMR structure obtained for RNAIi-RNAIIi (PDB: 2BL2).<sup>79</sup> The two phosphate clusters are shown in yellow on A. B is showing the sequence of the two hairpin and another view of the kissing complex.

## References

- (1) Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., and Darnell, J. (2000) *Molecular Cell Biology*. 4th edition. *New York W. H. Free*.
- (2) Crick, F. (1958) On Protein Synthesis. *Symp. Soc. Exp. Biol.* 138–166.
- (3) Watson, J. ., and Crick, F. H. . (1953) Molecular Structure of Nucleic Acids: a structure for deoxyribose nucleic acid. *Nature* 171, 737–738.
- (4) Ghosh, A., and Bansal, M. (2003) A glossary of DNA structures from A to Z. *Acta Crystallogr. - Sect. D Biol. Crystallogr.* 59, 620–626.
- (5) Dickerson, R. E. (1992) DNA structure from A to Z. *Methods Enzymol.* 211, 67–111.
- (6) Drew, H. R., Wingt, R. M., Takano, T., Broka, C., Tanakt, S., Itakura, K., and Dickerson, R. E. (1981) Structure of a B-DNA dodecamer : Conformation and dynamics. *Biochemistry* 78, 2179–2183.
- (7) Betzel, C., Lorenz, S., Fürste, J., Bald, A., Zhang, M., Schneider, T. R., Wilson, K. S., and Erdmann, V. . (1994) Crystal structure of domain A of *Thermus flavus* 5S rRNA and the contribution of water molecules to its structure. *FEBS Lett.* 351, 159–164.
- (8) Jain, A., Wang, G., and Vasquez, K. M. (2008) DNA Triple Helices: biological consequences and therapeutic potential. *Biochimie* 90, 1117–1130.
- (9) Lipps, H. J., and Rhodes, D. (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.* 19, 414–422.
- (10) Leroy, J. louis, Guéron, M., Mergny, J. louis, and Hélène, C. (1994) Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an I-motif. *Nucleic Acids Res.* 22, 1600–1606.
- (11) Tarköy, M., Phipps, A. K., Schultze, P., and Feigon, J. (1998) Solution structure of an intramolecular DNA triplex linked by hexakis(ethylene glycol) units: d(AGAGAGAA-(EG)<sub>6</sub>-TTCTCTCT-(EG)<sub>6</sub>-TCTCTCTT). *Biochemistry* 37, 5810–5819.
- (12) Parkinson, G. N., Lee, M. P. H., and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* 417, 876–880.
- (13) Esmaili, N., and Leroy, J. L. (2005) i-motif solution structure and dynamics of the d ( AACCCC ) and d ( CCCCAA ) tetrahymena telomeric repeats. *Nucleic Acids Res.* 33, 213–224.

- (14) Crick, F. H. C. (1970) Central Dogma of Molecular Biology. *Nature*.
- (15) Gilbert, W. (1986) The RNA world. *Nature*.
- (16) Bernhardt, H. S. (2012) The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others)(a). *Biol. Direct* 7, 23.
- (17) Pressman, A., Blanco, C., and Chen, I. A. (2015) The RNA world as a model system to study the origin of life. *Curr. Biol.* 25, R953–R963.
- (18) Nelson, J. W., and Breaker, R. R. (2017) The lost language of the RNA World. *Sci. Signal.* 10, 8812.
- (19) Moore, P. B., and Steitz, T. a. (2011) The roles of RNA in the synthesis of protein. *Cold Spring Harb. Perspect. Biol.* 3, a003780.
- (20) Shi, H., and Moore, P. B. (2000) The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: A classic structure revisited. *RNA* 6, S1355838200000364.
- (21) Koski, R. A., Bothwell, A. L. M., and Altman, S. (1976) Identification of a ribonuclease P-like activity from human KB cells. *Cell* 9, 101–116.
- (22) Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982) Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* 31, 147–157.
- (23) Ramesh, A., and Winkler, W. C. (2014) Metabolite-binding ribozymes. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1839, 989–994.
- (24) Pyle, a M. (1993) Ribozymes: a distinct class of metalloenzymes. *Science* 261, 709–714.
- (25) Scott, W. G. (2007) Ribozymes. *Curr. Opin. Struct. Biol.* 17, 280–286.
- (26) Lilley, D. M. J., and Eckstein, F. (2007) Ribozymes and RNA Catalysis.
- (27) Mironov, A. S., Gusarov, I., Rafikov, R., Lopez, L. E., Shatalin, K., Kreneva, R. A., Perumov, D. A., and Nudler, E. (2002) Sensing small molecules by nascent RNA: A mechanism to control transcription in bacteria. *Cell* 111, 747–756.
- (28) Nahvi, A., Sudarsan, N., Ebert, M. S., Zou, X., Brown, K. L., and Breaker, R. R. (2002) Genetic control by a metabolite binding mRNA. *Chem. Biol.* 9, 1043–1049.

- (29) Mandal, M., and Breaker, R. R. (2004) Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* 5, 451–63.
- (30) Barrick, J. E., and Breaker, R. R. (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol.* 8, R239.
- (31) Roth, A., and Breaker, R. R. (2009) The Structural and Functional Diversity of Metabolite-Binding Riboswitches. *Annu. Rev. Biochem.* 78, 305–334.
- (32) Serganov, A. (2009) The long and the short of riboswitches. *Curr. Opin. Struct. Biol.* 19, 251–259.
- (33) Breaker, R. R. (2011) Prospects for Riboswitch Discovery and Analysis. *Mol. Cell* 43, 867–879.
- (34) Breaker, R. R. (2012) Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol.* 4, 1–15.
- (35) Serganov, A., and Nudler, E. (2013) A decade of riboswitches. *Cell* 152, 17–24.
- (36) Li, S., and Breaker, R. R. (2013) Eukaryotic TPP riboswitch regulation of alternative splicing involving long-distance base pairing 41, 3022–3031.
- (37) Bocobza, S. E., and Aharoni, A. (2008) Switching the light on plant riboswitches.
- (38) Wachter, A., Tunc-ozdemir, M., Grove, B. C., Green, P. J., Shintani, D. K., and Breaker, R. R. (2007) Riboswitch Control of Gene Expression in Plants by Splicing and Alternative 3' End Processing of mRNAs. *Plant Cell* 19, 3437–3450.
- (39) Nudler, E., and Mironov, A. S. (2004) The riboswitch control of bacterial metabolism. *Trends Biochem. Sci.* 29, 11–7.
- (40) Winkler, W., Nahvi, A., and Breaker, R. R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* 419, 952–6.
- (41) Serganov, A., Yuan, Y.-R., Pikovskaya, O., and Al, E. (2004) Structural Basis for Discriminative Regulation of Gene Expression by Adenine- and Guanine-Sensing mRNAs. *Chem. Biol.* 11, 1729–1741.
- (42) Mandal, M., and Breaker, R. R. (2004) Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat. Struct. Mol. Biol.* 11, 29–35.
- (43) Serganov, A., Huang, L., and Patel, D. J. (2008) Structural insights into amino acid binding and gene control by a lysine riboswitch 7217, 1263–1267.

- (44) Serganov, A., and Patel, D. J. (2009) Amino acid recognition and gene regulation by riboswitches. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1789, 592–611.
- (45) Cromie, M. J., Shi, Y., Latifi, T., and Groisman, E. A. (2006) An RNA Sensor for Intracellular Mg<sup>2+</sup>. *Cell* 125, 71–84.
- (46) Dann, C. E., Wakeman, C. A., Sieling, C. L., Baker, S. C., Irnov, I., and Winkler, W. C. (2007) Structure and Mechanism of a Metal-Sensing Regulatory RNA. *Cell* 130, 878–892.
- (47) Chastain, M., and Tinoco, I. (1991) Structural elements in RNA, in *Progress in Nucleic Acid research and molecular biology*, pp 131–179.
- (48) Nowakowski, J., and Ignacio, T. (1997) RNA Structure and Stability. *Semin. Virol.* 8, 153–165.
- (49) Batey, R. T., Rambo, R. P., and Doudna, J. A. (1999) Tertiary motifs in RNA structure and folding. *Angew. Chemie - Int. Ed.* 38, 2326–2343.
- (50) Hermann, T., and Patel, D. J. (1999) Stitching together RNA tertiary architectures. *J. Mol. Biol.* 294, 829–49.
- (51) Hermann, T., and Patel, D. J. (2000) RNA bulges as architectural and recognition motifs. *Structure* 8, 47–54.
- (52) Karn, J. (2000) Tat, a novel regulator of HIV transcription and latency. *HIV Seq. Compend.* 2–18.
- (53) Allain, F. H. T., and Varani, G. (1995) Divalent metal ion binding to a conserved wobble pair defining the upstream site of cleavage of group I self-splicing introns. *Nucleic Acids Res.* 23, 341–350.
- (54) Varani, G., and McClain, W. H. (2000) The G·U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep.* 1, 18–23.
- (55) Ladner, J. E., Jack, a, Robertus, J. D., Brown, R. S., Rhodes, D., Clark, B. F., and Klug, a. (1975) Structure of yeast phenylalanine transfer RNA at 2.5 Å resolution. *Proc. Natl. Acad. Sci. U. S. A.* 72, 4414–4418.
- (56) Jack, A., Ladner, J. E., and Klug, A. (1976) Crystallographic refinement of yeast phenylalanine transfer RNA at 2.5Å resolution. *J. Mol. Biol.* 108, 619–649.
- (57) Holbrook, S. R., Sussman, J. L., Warrant, R. W., and Kim, S. H. (1978) Crystal structure of yeast phenylalanine transfer RNA. II. Structural features and functional implications. *J. Mol. Biol.* 123, 631–60.

- (58) Woese, C. R., Winker, S., and Gutell, R. R. (1990) Architecture of ribosomal RNA: constraints on the sequence of “tetra-loops”. *Proc. Natl. Acad. Sci.* *87*, 8467–8471.
- (59) Wedekind, J. E., and McKay, D. B. (1998) Crystallographic structures of the hammerhead ribozyme: Relationship to ribozyme folding and catalysis. *Annu. Rev. Biophys. Biomol. Struct.* *27*, 475–502.
- (60) Thore, S., Leibundgut, M., and Nenad, B. (2006) Structure of the Eukaryotic Thiamine. *Science* (80- ). *312*, 1208–1211.
- (61) Serganov, A., Polonskaia, A., Phan, A. T., Breaker, R. R., and Patel, D. J. (2006) Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature* *441*, 1167–71.
- (62) Westhof, E., and Auffinger, P. (2001) Transfer RNA Structure. *Encycl. Life Sci.* 1–11.
- (63) Chastain, M., and Tinoco, I. (1992) A base-triple structural domain in RNA. *Biochemistry* *31*, 12733–12741.
- (64) Jenkins, J. L., Krucinska, J., McCarty, R. M., Bandarian, V., and Wedekind, J. E. (2011) Comparison of a PreQ1 riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. *J. Biol. Chem.* *286*, 24626–24637.
- (65) Ferre-D’Amare, A. R., Zhou, K., and Doudna, J. A. (1998) Crystal structure of a hepatitis delta virus ribozyme. *Nature* *395*, 567–574.
- (66) Gilbert, S. D., Rambo, R. P., Van Tyne, D., and Batey, R. T. (2008) Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nat. Struct. Mol. Biol.* *15*, 177–82.
- (67) Kang, M., Eichhorn, C. D., and Feigon, J. (2014) Structural determinants for ligand capture by a class II preQ<sub>1</sub> riboswitch. *Proc. Natl. Acad. Sci.* *111*, E663–E671.
- (68) Chen, X., Kang, H., Shen, L. X., Chamorro, M., Varmus, H. E., and Tinoco Jr, I. (1996) A Characteristic Bent Conformation of RNA Pseudoknots Promotes –1 Frameshifting during Translation of Retroviral RNA. *J. Mol. Biol.* *260*, 479–483.
- (69) Rieder, U., Lang, K., Kreutz, C., Polacek, N., and Micura, R. (2009) Evidence for pseudoknot formation of class I preQ1 riboswitch aptamers. *ChemBioChem* *10*, 1141–1144.
- (70) Klein, D. J. (2006) Structural Basis of glmS Ribozyme Activation by Glucosamine-6-Phosphate. *Science* (80- ). *313*, 1752–1756.

- (71) Stephenson, W., Asare-okai, P. N., Chen, A. A., Keller, S., Santiago, R., Tenenbaum, S. A., Garcia, A. E., Fabris, D., and Li, P. T. X. (2013) The Essential Role of Stacking Adenines in a Two-Base-Pair RNA Kissing Complex. *J. Am. Chem. Soc.* *135*, 5602–5611.
- (72) Ducongé, F., Di Primo, C., and Toulmé, J. J. (2000) Is a closing “GA pair” a rule for stable loop-loop RNA complexes? *J. Biol. Chem.* *275*, 21287–21294.
- (73) Paillart, J. C., Westhof, E., Ehresmann, C., Ehresmann, B., and Marquet, R. (1997) Non-canonical interactions in a kissing loop complex: the dimerization initiation site of HIV-1 genomic RNA. *J. Mol. Biol.* *270*, 36–49.
- (74) Mundigala, H., Michaux, J. B., Feig, A. L., Ennifar, E., and Rueda, D. (2014) HIV-1 DIS stem loop forms an obligatory bent kissing intermediate in the dimerization pathway. *Nucleic Acids Res.* *42*, 7281–7289.
- (75) Bouchard, P., and Legault, P. (2014) A remarkably stable kissing-loop interaction defines substrate recognition by the *Neurospora Varkud* Satellite ribozyme. *RNA* *20*, 1451–64.
- (76) Bouchard, P., and Legault, P. (2014) Structural insights into substrate recognition by the *neurospora varkud* satellite ribozyme: Importance of u-turns at the kissing-loop junction. *Biochemistry* *53*, 258–269.
- (77) Lussier, A., Bastet, L., Chauvier, A., and Lafontaine, D. A. (2015) A kissing loop is important for *btuB* riboswitch ligand sensing and regulatory control. *J. Biol. Chem.* *290*, 26739–26751.
- (78) Chang, K. Y., and Tinoco, I. (1997) The structure of an RNA “kissing” hairpin complex of the HIV TAR hairpin loop and its complement. *J. Mol. Biol.* *269*, 52–66.
- (79) Lee, A. J., and Crothers, D. M. (1998) The solution structure of an RNA loop–loop complex: the ColE1 inverted loop sequence. *Structure* *6*, 993–1007.
- (80) Pyle, A. M. (2002) Metal ions in the structure and function of RNA. *J. Biol. Inorg. Chem.* *7*, 679–690.
- (81) Laing, L. G., Gluick, T. C., and Draper, D. E. (1994) Stabilization of RNA Structure by Mg Ions. *J. Mol. Biol.*
- (82) Misra, V. K., and Draper, D. E. (1998) On the Role of Magnesium Ions in RNA Stability. *Biopolymers* *48*, 113–135.
- (83) Draper, D. E. (2004) A guide to ions and RNA structure. *RNA* *10*, 335–343.
- (84) Draper, D. E. (2008) RNA Folding: Thermodynamic and Molecular Descriptions of the Roles of Ions.

*Biophys. J.* 95, 5489–5495.

(85) Bowman, J. C., Lenz, T. K., Hud, N. V., and Williams, L. D. (2012) Cations in charge: Magnesium ions in RNA folding and catalysis. *Curr. Opin. Struct. Biol.* 22, 262–272.

(86) Bai, Y., Das, R., Millett, I. S., Herschlag, D., and Doniach, S. (2005) Probing counterion modulated repulsion and attraction between nucleic acid duplexes in solution. *Proc. Natl. Acad. Sci.* 102, 1035–1040.

(87) Bai, Y., Greenfeld, M., Travers, K. J., Chu, V. B., Lipfert, J., Doniach, S., and Herschlag, D. (2007) Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *J. Am. Chem. Soc.* 129, 14981–8.

(88) Lipfert, J., Doniach, S., Das, R., and Herschlag, D. (2014) Understanding nucleic acid-ion interactions. *Annu. Rev. Biochem.*

(89) Shiman, R., and Draper, D. E. (2000) Stabilization of RNA tertiary structure by monovalent cations. *J. Mol. Biol.* 302, 79–91.

(90) Lambert, D., Leipply, D., Shiman, R., and Draper, D. E. (2009) The influence of monovalent cation size on the stability of RNA tertiary structures. *J. Mol. Biol.* 390, 791–804.

(91) Largy, E., Mergny, J.-L., and Gabelica, V. (2016) Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability, in *The Alkali Metal ions: Their Role for Life*, pp 203–258.

(92) Romani, A. M. P. (2011) Cellular magnesium homeostasis. *Arch. Biochem. Biophys.* 512, 1–23.

(93) Grubbs, R. D. (2002) Intracellular magnesium and magnesium buffering. *BioMetals* 15, 251–259.

(94) (1991) Ionic radii in crystals 14–16.

(95) Holbrook, S. R., Sussman, J. L., Warrant, R. W., Church, G. M., and Kim, S. H. (1977) RNA-ligand interactions. (I) Magnesium binding sites in yeast tRNAPhe. *Nucleic Acids Res.* 4, 2811–2820.

(96) Quigley, G. J., Teeter, M. M., and Rich, A. (1978) Structural analysis of spermine and magnesium ion binding to yeast phenylalanine transfer RNA. *Proc Natl Acad Sci U S A.* 75, 64–8.

(97) Hennelly, S. P., Novikova, I. V., and Sanbonmatsu, K. Y. (2013) The expression platform and the aptamer: Cooperativity between Mg<sup>2+</sup> and ligand in the SAM-I riboswitch. *Nucleic Acids Res.* 41, 1922–1935.

- (98) Schnabl, J., and Sigel, R. K. (2010) Controlling ribozyme activity by metal ions. *Curr. Opin. Chem. Biol.* *14*, 269–275.
- (99) Leonarski, F., D’Ascenzo, L., and Auffinger, P. (2016) Mg<sup>2+</sup> ions: do they bind to nucleobase nitrogens? *Nucleic Acids Res.* *45*, 987–1004.
- (100) Draper, D. E. (2008) RNA Folding: Thermodynamic and Molecular Descriptions of the Roles of Ions. *Biophys. J.* *95*, 5489–5495.
- (101) Mulhbachter, J., St-Pierre, P., and Lafontaine, D. A. (2010) Therapeutic applications of ribozymes and riboswitches. *Curr. Opin. Pharmacol.* *10*, 551–556.
- (102) Blount, K. F., and Breaker, R. R. (2006) Riboswitches as antibacterial drug targets. *Nat. Biotechnol.* *24*, 1558–1564.
- (103) Klein, D. J., Edwards, T. E., and Ferré-D’Amaré, A. R. (2009) Cocrystal structure of a class I preQ1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nat. Struct. Mol. Biol.* *16*, 343–4.
- (104) Rose, D. M., Blears, M. L., Record Jr., M. T., and Bryant, R. G. (1980) Mg NMR in DNA solutions: Dominance of site binding effects. *Proc. Natl. Acad. Sci. U. S. A.* *77*, 6289–92.
- (105) Roy, R., Hohng, S., and Ha, T. (2008) A practical guide to single-molecule FRET. *Nat. Methods* *5*, 507–516.
- (106) Haller, A., Altman, R. B., Soulière, M. F., Blanchard, S. C., and Micura, R. (2013) Folding and ligand recognition of the TPP riboswitch aptamer at single-molecule resolution. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 4188–93.
- (107) Lebars, I., Legrand, P., Aimé, A., Pinaud, N., Fribourg, S., and Di Primo, C. (2008) Exploring TAR-RNA aptamer loop-loop interaction by X-ray crystallography, UV spectroscopy and surface plasmon resonance. *Nucleic Acids Res.* *36*, 7146–7156.
- (108) Di Primo, C. (2008) Real time analysis of the RNAI-RNAII-Rop complex by surface plasmon resonance: From a decaying surface to a standard kinetic analysis. *J. Mol. Recognit.* *21*, 37–45.
- (109) Leney, A. C., and Heck, A. J. R. (2017) Native Mass Spectrometry: What is in the Name? *J. Am. Soc. Mass Spectrom.* *28*, 5–13.

- (110) Yamashita, M., and Fenn, J. B. (1984) Electrospray ion source. Another variation on the free-jet theme. *J Phys.Chem.* *88*, 4451–4459.
- (111) Fenn, J. B., Mann, M., Meng, C. K. A. I., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science (80- )*. *246*, 64–71.
- (112) Kebarle, P., and Verkerk, U. H. (2009) Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass Spectrom. Rev.* *28*, 898–917.
- (113) Iribarne, J. V. (1976) On the evaporation of small ions from charged droplets. *J. Chem. Phys.* *64*, 2287.
- (114) Dole, M., Mack, L. L., Hines, R. L., Mobley, R. C., Ferguson, L. D., and Alice, M. B. (1968) Molecular Beams of Macroions. *J. Chem. Phys.* *49*, 2240–2249.
- (115) Konermann, L., Ahadi, E., Rodriguez, A. D., and Vahidi, S. (2013) Unraveling the mechanism of electrospray ionization. *Anal. Chem.* *85*, 2–9.
- (116) Loo, J. A. (1997) STUDYING NONCOVALENT PROTEIN COMPLEXES BY ELECTROSPRAY IONIZATION MASS SPECTROMETRY. *Mass Spectrom. Rev.* *16*, 1–23.
- (117) Beck, J. L., Colgrave, M. L., Ralph, S. F., and Shei, M. M. (2001) Electrospray ionization mass spectrometry of oligonucleotide complexes with drugs, metals, and proteins. *Mass Spectrom. Rev.* *20*, 61–87.
- (118) Hofstadler, S. A., and Griffey, R. H. (2001) Analysis of noncovalent complexes of DNA and RNA by mass spectrometry. *Chem. Rev.* *101*, 377–390.
- (119) Gabelica, V., De Pauw, E., and Rosu, F. (1999) Interaction between antitumor drugs and a double-stranded oligonucleotide studied by electrospray ionization mass spectrometry. *J. Mass Spectrom.* *34*, 1328–1337.
- (120) Rosu, F., Nguyen, C. H., De Pauw, E., and Gabelica, V. (2007) Ligand Binding Mode to Duplex and Triplex DNA Assessed by Combining Electrospray Tandem Mass Spectrometry and Molecular Modeling. *J. Am. Soc. Mass Spectrom.* *18*, 1052–1062.
- (121) Rosu, F., De Pauw, E., and Gabelica, V. (2008) Electrospray mass spectrometry to study drug-nucleic acids interactions. *Biochimie* *90*, 1074–1087.

- (122) Marchand, A., and Gabelica, V. (2014) Native electrospray mass spectrometry of DNA G-quadruplexes in potassium solution. *J. Am. Soc. Mass Spectrom.* 25, 1146–1154.
- (123) Marchand, A., Granzhan, A., Iida, K., Tsushima, Y., Ma, Y., Nagasawa, K., Teulade-Fichou, M.-P., and Gabelica, V. (2015) Ligand-Induced Conformational Changes with Cation Ejection upon Binding to Human Telomeric DNA G-Quadruplexes. *J. Am. Chem. Soc.* 137, 750–756.
- (124) Sannes-Lowery, K. a, Hu, P., Mack, D. P., Mei, H. Y., and Loo, J. a. (1997) HIV-1 Tat peptide binding to TAR RNA by electrospray ionization mass spectrometry. *Anal. Chem.* 69, 5130–5135.
- (125) Mei, H., Cui, M., Heldsinger, A., Lemrow, S. M., Loo, J. A., Sannes-Lowery, K. A., Sharmeen, L., and Czarnik, A. W. (1998) Inhibitors of Protein-RNA Complexation That Target the RNA: Specific Recognition of Human Immunodeficiency Virus Type 1 TAR RNA by Small Organic Molecules. *Biochemistry* 37, 14204–14212.
- (126) Schneeberger, E. M., and Breuker, K. (2017) Native Top-Down Mass Spectrometry of TAR RNA in Complexes with a Wild-Type tat Peptide for Binding Site Mapping. *Angew. Chemie - Int. Ed.* 56, 1254–1258.
- (127) Hagan, N., and Fabris, D. (2003) Direct mass spectrometric determination of the stoichiometry and binding affinity of the complexes between nucleocapsid protein and RNA stem-loop hairpins of the HIV-1 TAR-recognition element. *Biochemistry* 42, 10736–10745.
- (128) Turner, K. B., Hagan, N. A., and Fabris, D. (2007) Understanding the isomerization of the HIV-1 dimerization initiation domain by the nucleocapsid protein. *J. Mol. Biol.* 369, 812–28.
- (129) Hagan, N. A., and Fabris, D. (2007) Dissecting the protein-RNA and RNA-RNA interactions in the nucleocapsid-mediated dimerization and isomerization of HIV-1 stemloop 1. *J. Mol. Biol.* 365, 396–410.
- (130) Gabelica, V., Rosu, F., and De Pauw, E. (2009) A simple method to determine electrospray response factors of noncovalent complexes. *Anal. Chem.* 81, 6708–15.
- (131) May, J. C., and Mclean, J. A. (2015) Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation. *Anal. Chem.* 87, 1422–1436.
- (132) Zhong, Y., Hyung, S. J., and Ruotolo, B. T. (2009) Ion mobility–mass spectrometry for structural proteomics. *Expert Rev Proteomics* 9, 47–58.
- (133) Konijnenberg, A., Butterer, A., and Sobott, F. (2013) Native ion mobility-mass spectrometry and

related methods in structural biology. *Biochim. Biophys. Acta - Proteins Proteomics* 1834, 1239–1256.

(134) Lanucara, F., Holman, S. W., Gray, C. J., and Eyers, C. E. (2014) The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat. Chem.* 6, 281–294.

(135) Uetrecht, C., Rose, R. J., van Duijn, E., Lorenzen, K., and Heck, A. J. R. (2010) Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* 39, 1633–1655.

(136) Arcella, A., Portella, G., Ruiz, M. L., Eritja, R., Vilaseca, M., Gabelica, V., and Orozco, M. (2012) Structure of triplex DNA in the Gas phase. *J. Am. Chem. Soc.* 134, 6596–6606.

(137) Zheng, X., Liu, D., Klärner, F. G., Schrader, T., Bitan, G., and Bowers, M. T. (2015) Amyloid  $\beta$ -Protein Assembly: The Effect of Molecular Tweezers CLR01 and CLR03. *J. Phys. Chem. B* 119, 4831–4841.

(138) Soper-Hopper, M. T., Eschweiler, J. D., and Ruotolo, B. T. (2017) Ion Mobility-Mass Spectrometry Reveals a Dipeptide That Acts as a Molecular Chaperone for Amyloid  $\beta$ . *ACS Chem. Biol.* 12, 1113–1120.

(139) D'Atri, V., Porrini, M., Rosu, F., and Gabelica, V. (2015) Linking molecular models with ion mobility experiments. Illustration with a rigid nucleic acid structure. *J. Mass Spectrom.* 50, 711–726.

(140) Wyttenbach, T., Bleiholder, C., and Bowers, M. T. (2013) Factors contributing to the collision cross section of polyatomic ions in the kilodalton to gigadalton range: Application to ion mobility measurements. *Anal. Chem.* 85, 2191–2199.

(141) Gabelica, V., Marklund, E., Acides, L., and Régulations, N. (2018) Fundamentals of ion mobility spectrometry. *Curr. Opin. Chem. Biol.*

(142) Breuker, K., and McLafferty, F. W. (2008) Stepwise evolution of protein native structure with electrospray into the gas phase, 10  $\mu$ s to 10<sup>2</sup> s. *Proc. Natl. Acad. Sci. U. S. A.* 105, 18145–18152.

(143) Meyer, T., Gabelica, V., Grubmüller, H., and Orozco, M. (2013) Proteins in the gas phase. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 3, 408–425.

(144) Michaelievski, I., Eisenstein, M., and Sharon, M. (2010) Gas-phase compaction and unfolding of protein structures. *Anal. Chem.* 82, 9484–9491.

(145) Gabelica, V., and De Pauw, E. (2002) Collision-induced dissociation of 16-mer DNA duplexes with various sequences: Evidence for conservation of the double helix conformation in the gas phase. *Int. J.*

*Mass Spectrom.* 219, 151–159.

(146) Mcluckey, S. A., Berkel, J. Van, and Glish, L. (1992) Tandem Mass Spectrometry of Small Multiply Charged Oligonucleotides. *J. Am. Soc. Mass Spectrom.* 3, 60–70.

(147) Brodbelt, J. S. (2016) Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* 88, 30–51.

(148) Mcluckey, S. A. (1992) Principles of Collisional Activation in Analytical Mass Spectrometry. *Anal. Chem.* 3, 599–614.

(149) Barré-Sinoussi, F., Ross, A. L., and Delfraissy, J.-F. (2013) Past, present and future: 30 years of HIV research. *Nat. Rev. Microbiol.* 11, 877–883.

(150) Engelman, A., and Cherepanov, P. (2012) The structural biology of HIV-1: mechanistic and therapeutic insights. *Nat. Rev. Microbiol.* 10, 279–290.

(151) Duconge, F., and Toulme, J. J. (1999) In vitro selection identifies key determinants for loop-loop interactions : RNA aptamers selective for the TAR RNA element of HIV-1. *RNA* 5, 1605–1614.

(152) Chang, K. Y., and Tinoco, I. (1994) Characterization of a “kissing” hairpin complex derived from the human immunodeficiency virus genome. *Proc. Natl. Acad. Sci. U. S. A.* 91, 8705–9.

(153) Tomizawa, J. (1984) Control of ColE1 Plasmid Replication: The Process of Binding of RNA I to the primer transcript. *Cell* 38, 861–870.

(154) Tomizawa, J. ichi. (1986) Control of colE1 plasmid replication: Binding of RNA I to RNA II and inhibition of primer formation. *Cell* 47, 89–97.

(155) Tomizawa, J. (1990) Control of ColE1 Plasmid Replication Interaction of Rom Protein with an Unstable Complex Formed by RNA I and RNA II. *J Mol Biol* 212, 695–708.

(156) Gregorian, R. S., and Crothers, D. M. (1995) Determinants of RNA hairpin loop-loop complex stability. *J. Mol. Biol.* 248, 968–984.

(157) Marino, J., Gregorian, R., Csankovszki, G., and Crothers, D. (1995) Bent helix formation between RNA hairpins with complementary loops. *Science.* 268, 1448–1454.



## 2. MATERIALS & METHODS

### V. Sample preparation

DNA and RNA sequences used in this study are purchased lyophilized from Integrated DNA Technology (IDT, Leuven, Belgium) with a standard desalting. All sequences are first dissolved in RNase free water (Ambion, Fisher Scientific, Illkirch, France) and stored at -20°C. Then, the preparation depends on the nature of the structure (kissing complexes or duplexes) and the type of experiments.

#### V.1. Kissing complexes

A stock at ~ 100 µM of each hairpin separately is prepared in the appropriate buffer. For MS experiments, the buffer should be MS-compatible, which means that it has to be volatile. Also, solutions conditions have to be close to biological conditions. To ensure the ionic strength, the electrolyte used is Ammonium Acetate (NH<sub>4</sub>OAc, Sigma-Aldrich, Saint-Quentin Fallavier, France) at 150 mM unless otherwise specified. The concentration is such as the ionic strength is high enough to mimic the monovalent cations concentration in the cell. The ionic strength decreases the repulsion between charges of the same sign.

Hairpin stocks are then desalted using Amicon Ultra-0.5 3K ultrafiltration devices (Merck Millipore, Saint-Quentin en Yvelines, France). This procedure reduces as much as possible the Na<sup>+</sup> traces. The filter cut-off should be lower than the mass of the sequence of interest.

The concentration of stock solutions after desalting is measured using UV absorption spectroscopy at 260 nm (absorbance wavelength of the bases). Using the Beer-Lambert law, the concentration can be determined.

$$A_{\lambda} = \varepsilon \times l \times C \quad (1)$$

Where  $A_{\lambda}$  is the absorbance at wavelength  $\lambda$ ,  $l$  is path length of the beam (1 cm here),  $\varepsilon$  is the extinction coefficient (in L.mol<sup>-1</sup>.cm<sup>-1</sup>) given by the oligo manufacturer following the calculations described in the literature<sup>1</sup> and  $C$  is the concentration (in mol.L<sup>-1</sup>).

Then, the hairpin stocks are heated at 90°C for 90 sec and then placed on ice and allowed to cool for 10 min. The samples are then prepared, using these stock solutions, by mixing the two complementary hairpins at the desired concentration.

## V.2. Duplexes

As for kissing complexes, intermediate stock solutions of DNA or RNA duplexes are prepared in NH<sub>4</sub>OAc. This solution consists in mixing the two corresponding single-strands and anneal the solution at 85°C for several minutes, let the samples cool down at room temperature and then store them at 4°C overnight. The procedure allows the good formation of duplexes. Desalting is performed only if needed.

## V.3. Sequences used

The list of the sequences used is presented on Table 1.

	SEQUENCES	Mass (Da)
<b>KISSING COMPLEXES</b>		
TAR	5'- GGA GCC UGG GAG CUC C -3'	5151.2
R06	5'- GGU CGG UCC CAG ACG ACC -3'	5745.5
RNAIi	5'- GGC AAC GGA UGG UUC GUU GCC -3'	6744.1
RNAIii	5'- GCA CCG AAC CAU CCG GUG C -3'	6034.7
K1	5'- UGC UCG GCC CCG CGA GCA -3'	5721.5
K1AU	5'- GC UCG GCC CCG CGA GC -3'	5086.1
K'1UC	5'- ACG AGC UGG GGC GCU CGU -3'	5802.6
K'1CC	5'- ACG AGC CGG GGC GCU CGU -3'	5801.6
K'1CU	5'- ACG AGC CGG GGU GCU CGU -3'	5802.6
K'1UU	5'- ACG AGC UGG GGU GCU CGU -3'	5803.5
<b>DUPLEXES</b>		
<b>RNA</b>		
17R1	5'- GGA GCU CCC AGA CGA CC -3'	5423.4
17R2	5'- GGU CGU CUG GGA GCU CC -3'	5434.3

17R3	5'- GAG CCG CAA AUG CCC CG -3'	5423.4
17R4	5'- CGG GGC AUU UGC GGC UC -3'	5434.3
20R1	5'- GUG AGC UCC CAG ACG ACC UG -3'	6380.9
20R2	5'- CAG GUC GUC UGG GAG CUC AC - 3'	6397.9
20R3	5'- GGA GCC GCA AAU GCC CCG UG -3'	6380.9
20R4	5'- CAC GGG GCA UUU GCG GCU CC -3'	6397.9
r <sub>66</sub>	5'-CGCGAAUUCGCG -3'	3810.3
r <sub>100</sub>	5'- CGCGGGCCCGCG -3'	3840.4
r <sub>66+66</sub>	5'-CGCGAAUUCGCGCGCGAAUUCGCG -3'	7682.7
r <sub>100+100</sub>	5'- CGCGGGCCCGCGCGCGGGCCCGCG -3'	7742.7
r <sub>100+100+100</sub>	5'- CGCGGGCCCGCGCGCGGGCCCGCGCGCGGGCCCGCG -3'	11645.1
r <sub>14a</sub>	5'- UAAUACGACUUAAC -3'	4230.9
r <sub>14b</sub>	5'- GUUAAGUCGUUAUA -3'	4292.9
<b>DNA</b>		
d <sub>17R1</sub>	5'- GGA GCT CCC AGA CGA CC -3'	5165.4
d <sub>17R2</sub>	5'- GGT CGT CTG GGA GCT CC -3'	5218.4
d <sub>20R1</sub>	5'- GTG AGC TCC CAG ACG ACC TG -3'	6103.02
d <sub>20R2</sub>	5'- CAG GTC GTC TGG GAG CTC AC - 3'	6134.03
d <sub>66</sub>	5'-CGCGAATTCGCG -3'	3646.4
d <sub>100</sub>	5'- CGCGGGCCCGCG -3'	3648.4
d <sub>66+66</sub>	5'-CGCGAATTCGCGCGCGAATTCGCG -3'	7354.8
d <sub>100+100</sub>	5'- CGCGGGCCCGCGCGCGGGCCCGCG -3'	7358.7

d <sub>100+100+100</sub>	5'- CGCGGGCCCCGCGCGCGGGCCCCGCGCGCGGGCCCCGCG -3'	11069.1
d <sub>14a</sub>	5'- TAATACGACTTAAC -3'	4230.9
d <sub>14b</sub>	5'- GTTAAGTCGTATTA -3'	4292.9
18a	5'- CCTACTCGTTACCTTCTT -3'	5351.5
18b	5'- AAGAAGGTAACGAGTAGG -3'	5645.8
22a	5'- CCTACTCGTTACCTTCTTCTGA -3'	6587.3
22b	5'- TCAGAAGAAGGTAACGAGTAGG -3'	6881.6
24a	5'- CCTACTCGTTACCTTCTTCTGATA -3'	7204.7
24b	5'- TATCAGAAGAAGGTAACGAGTAGG -3'	7499
32a	5'- CCTACTCGTTACCTTCTTCTGATACTGTTAA -3'	9666.4
32b	5' - TTAAACAGTATCAGAAGAAGGTAACGAGTAGG -3'	9978.6
36a	5'- CCTACTCGTTACCTTCTTCTGACTTCCCTCTTCTT -3'	10756
36b	5'- AAGAAAGAGGGAAGTCAGAAGAAGGTAACGAGTAGG -3'	11362.5

Table 1: List of the sequences used. The average molecular weight is indicated.

## VI. Characterization in solution by thermal denaturation

To evaluate the stability of a complex in function of the temperature, we used thermal denaturation: increasing progressively the temperature will unfold the structure present in solution. The most common way to follow this unfolding is to use UV absorbance at 260 nm for nucleic acids.

Using this method, we will have  $A_{260}=f(T)$  where  $A_{260}$  is the absorbance at 260 nm and  $T$  the temperature. In the curve, if there is a transition of the species conformation, the curve will be a sigmoid (Figure 25). The melting temperature ( $T_m$ ) is defined as the temperature at which 50% of the species in solution is folded. This temperature corresponds to the inflexion point of the curve. To determine  $T_m$ , the median between the high and low baseline is drawn. An alternative to this method is the determination of the first derivative of the curve. Generally the maximum of the derivative presents a deviation in comparison the “real”  $T_m$  value.<sup>2</sup> However, when the baseline cannot be determined and for simplification, we will take the maximum of the first derivative as an approximation for the  $T_m$  value.

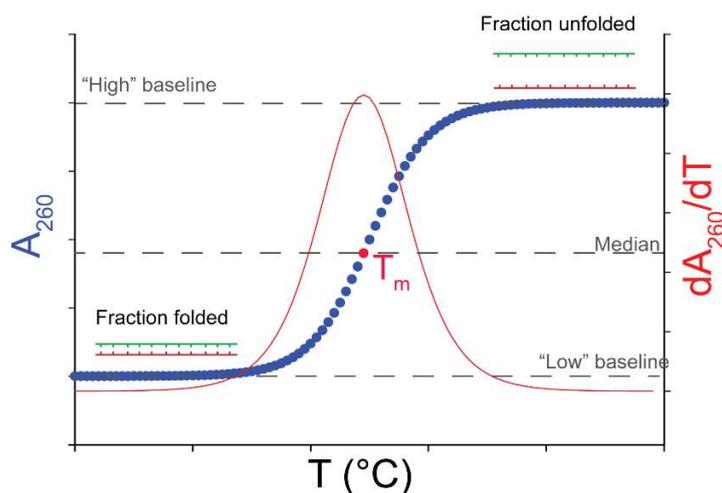


Figure 25: Schematic representation of UV melting experiment. In blue:  $A_{260}$  and in red:  $dA_{260}/dT$ . In practice, one needs to draw both base lines and the median of those curves. The crossing point between the median and the data will be the  $T_m$ . It is also the maximum of the first derivative curve.

The samples were prepared at 1  $\mu\text{M}$  of each hairpin for all kissing complexes and 1  $\mu\text{M}$  of each strand for a duplex. We used 580- $\mu\text{L}$  quartz cuvette (Hellma, Müllheim, Germany). The temperature ramp was set at 0.2°C/min from 4°C to 90°C.

## VII. Mass Spectrometry experiments

Mass spectrometry (MS) is used first to identify the different species present in solution. But MS can also be used in a more elaborate way to determine cation binding stoichiometries and binding constants. The section below details how the experiments are done and how the data are treated.

### VII.1. Sample preparation for mass spectrometry analysis

For kissing complexes, the two hairpins are mixed at 10  $\mu\text{M}$  each, unless otherwise mentioned. For most experiments, the electrolyte is 150 mM  $\text{NH}_4\text{OAc}$ , but other buffers have been tested as well (see results and discussion). Different cations salts are also tested, they are all provided by Sigma Aldrich (purity > 99%, Sigma-Aldrich, Saint-Quentin Fallavier, France).

In all MS experiments, 2  $\mu\text{M}$  of  $\text{dT}_6$  is added to the solution.  $\text{dT}_6$  is a 6-nucleotides long DNA which serves as an internal standard. As an internal reference, it does not have any interaction with the other species present in solution.  $\text{dT}_6$  is used to do quantification. Also it serves as an internal reference for CCS values when ion mobility is performed ( $^{\text{DT}}\text{CCS}_{\text{He}}$  of  $\text{dT}_6^{2-} = 306 \pm 1 \text{ \AA}^2$ , according to measurements done by V. Gabelica on the Bowers 5-cm drift tube<sup>3</sup>).

### VII.2. Mass spectrometers used

Three main instruments were used, depending on their characteristics or availability: the Agilent DTIMS-Q-TOF 6560, the LCT premier (Waters) and the Thermo LCQ Fleet.

#### VII.2.1 Agilent DTIMS-Q-TOF 6560

The DTIMS-Q-TOF 6560 (Agilent Technologies, Santa Clara, CA, USA) is coupling mass spectrometry and drift tube ion mobility spectrometry: ions are separated by their  $m/z$  ratio but also by their shape. It is also possible to perform tandem mass spectrometry. Most of the experiments are done on this instrument owing to its triple function.

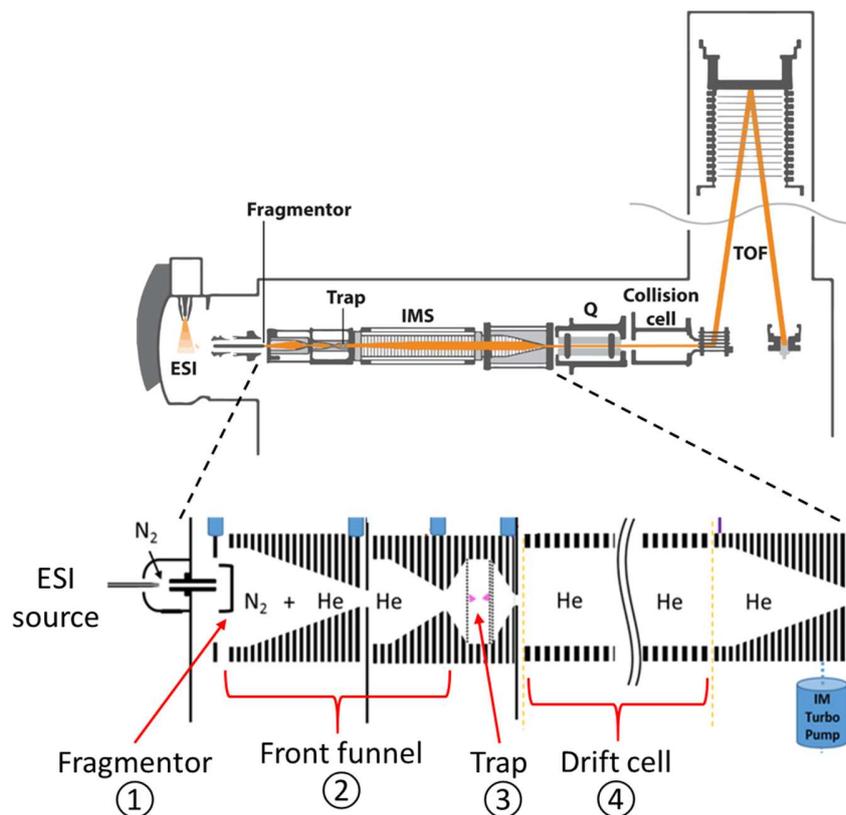


Figure 26: Schematic representation of the Agilent IMS Q-TOF 6560 mass spectrometer.

#### VIII.2.1.1. Pre-IMS device

Molecules are ionized using an ESI source. The source temperature, i.e. cone temperature, is set at 200-205°C. The source is at atmospheric pressure, whereas all the instrument is under  $P_{atm}$ . Then the ions are guided through the mass spectrometer using ion funnels where a radiofrequency is applied. Before entering the mobility cell, the ions are trapped in order to be accumulated, then sent as a bunch into the mobility cell. In this pre-IMS section, it is possible to activate more or less the ions, by playing on different parameters. The first one is called the fragmentor voltage, which helps to obtain a higher desolvation of the ions. Then it is possible to play on some voltages inside the trapping region. The following paragraphs will describe the different parameters. Some parameters can be changed and their relative effects are described in the results and discussion part.

- **Fragmentor**

The fragmentor is located between the source and the first ion funnel which conducts the ions towards the mass spectrometer (Figure 26, ①). Changing its value impacts on the desolvation and declustering. It

can help to activate the ions. Tests on the effect of the fragmentor voltage are presented in the results and discussion part.

- **Front funnel voltages**

The trap is preceded by two funnels to guide the ions and focus them (Figure 26, ②). The voltages are not changed in these parts. In the front funnel, the High Pressure Funnel Delta is set at 150 V, the High pressure Funnel RF at 200 V, Trap Funnel Delta at 140 V, Trap Funnel RF at 210 V and Trap funnel exit at 10 V. These parameters ensure a good transmission of the ions prior the trap.

- **Trapping voltages**

First, one can change the voltages of the trap which is just upstream to the ion mobility cell (Figure 26, ③). The trap is composed of three trapping grids: one at the entrance and two at the exit (grid 1 and 2) (Figure 27). All voltages can be controlled. A tune method contains the necessary parameters for two ways of functioning of the trap: trapping mode and release mode. For each grid (1 entrance grid and 2 exit grids), there is a “low” value and a “delta” value. “low”+“delta” define the “high” value. In Figure 3B, the voltage set used during trapping are shown in red, and those used while releasing the ions to the IMS are shown in blue. The trap fill time is fixed at 1 ms, and the trap release time at 100  $\mu$ s.

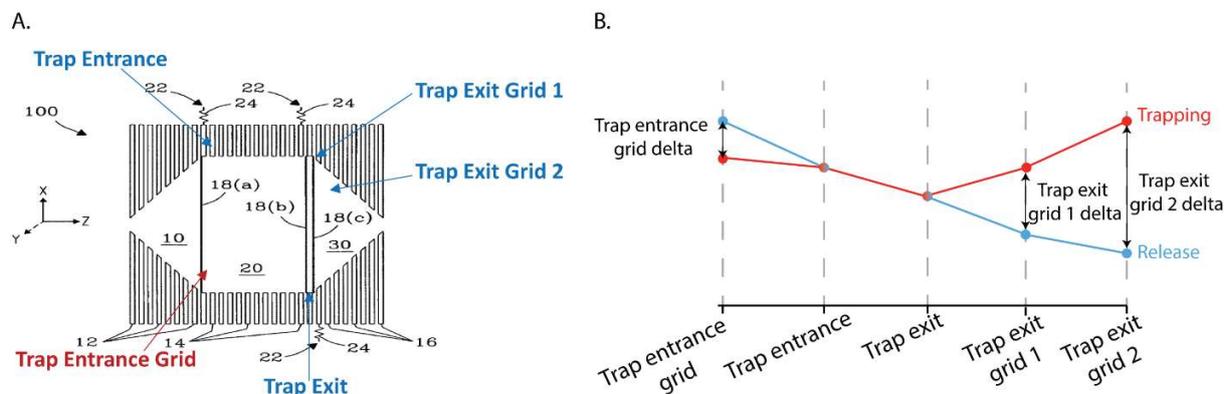


Figure 27: A) Scheme of the trapping region upstream the ion mobility cell<sup>4</sup>. B) Scheme of the trapping parameters voltages.

We have defined two sets of parameters tested during our experiments: SOFT and HARD parameters, the latter corresponding to harsher ion activation conditions. The different parameters applied for each tune file are given in Table 2. All voltages depend on each other to maintain transmission of ions.

	<b>SOFT parameters</b>	<b>HARD parameters</b>
Trap Ent. Grid low	102 V	105 V
Trap Entrance	101 V	101 V
Trap exit	99 V	99 V
Trap Ex. Grid Low 1	95 V	97 V
Trap Ex. Grid $\Delta 1$	6 V	6 V
Trap Ex. Grid Low 2	94 V	96 V
Trap Ex. Grid $\Delta 2$	10 V	10 V

Table 2: Voltages of the two sets of trapping parameters used in this work: SOFT and HARD parameters.

- **Trap Entrance Grid Delta**

The trap entrance grid delta is the potential difference between the trap entrance grid voltage in the release mode (higher value) and the same voltage in the trapping mode (lower value). This voltage can have a dramatic impact on the ion activation. Changing this parameter has an impact on declustering by favoring ammonia loss. We will test different TEGD voltages in the results and discussion section.

#### VII.2.1.2. Ion mobility device

This instrument includes a drift tube mobility cell (DTIMS) where ion mobility can be performed (Figure 26, ④). The principle of IM-MS is described in the introduction. Briefly, ions will collide with a buffer gas that will slow them down depending on their shape. The drift cell is filled with Helium at 3.89 Torr. A uniform and linear electric field is applied in the drift cell to accelerate the ions inside the cell. Usually for IMS experiments, 5 different  $\Delta V$  are applied in consecutive segments of the mass spectrum: 390 V, 490V, 590V, 690 V and 790V. Each segment is acquired during 1 min. This step-field method will allow the determination of the collision cross section.

A control of the pressure inside the drift cell is important to prevent contamination by  $N_2$  present in the parts after the mobility cell (i.e collision cell). For all measurements, the helium pressure in the drift tube was  $3.89 \pm 0.01$  Torr. The arrival time distribution is recorded over 60 ms, then the next bunch of ions is injected.

### VII.2.1.3. Quadrupole and collision cell for MS/MS analysis

Another functionality of this instrument is the MS/MS option. When performing a simple IMS-MS experiments, the ions are passing through the quadrupole and collision cell, where minimal voltages are applied to ensure ion transmission but avoid further activation. In the case of MS/MS, it is possible to select a parent ion using the quadrupole and to fragment it using collision-induced dissociation (CID) inside the collision cell. The collision cell is filled with  $N_2$  gas. The principle of CID is explained in the introduction. Energy is given to the ions for fragmentation by increasing the voltages inside the cell. By increasing the voltages, ions are increasingly accelerated and by the collisions with  $N_2$  molecules, they will acquire a higher amount of internal energy. This energy will conduce to the fragmentation of the ion. The time spent in the collision cell is  $\leq 1$  ms and this value cannot be changed.

### VII.2.1.4. Mass analyzer

The mass analyzer here is a Time-of-flight (TOF) analyzer. Briefly, ions are accelerated by an electric field ( $\Delta V$ ) which gives them a certain kinetic energy. They are pushed at the same time inside the analyzer. The TOF measures the time spent by the ions to go from the pusher to the detector and this time depends on the mass to charge ratio of the ion. Ions with a high  $m/z$  ratio will arrive more slowly to the detector than smaller ions. The resolution will depend on the length of the TOF. The resolution will increase will the length of the TOF.

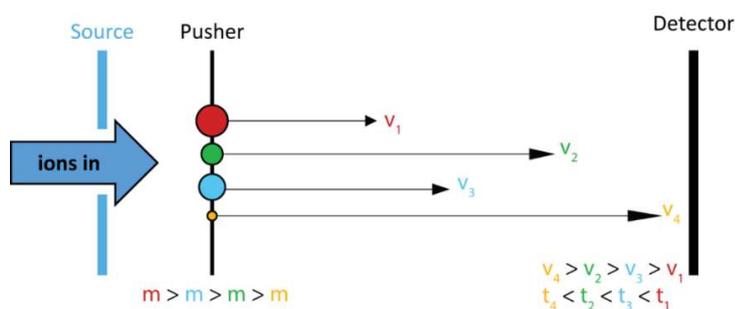


Figure 28: Time Of Flight (TOF) principle.

Experimentally, a calibration is needed to calibrate the TOF. To do so, Agilent is providing a solution of several calibrants (tunemix) for which the  $m/z$  values are known. The calibration will allow to obtain accurate value of  $m/z$ . The duty cycle of the TOF is set at  $160 \mu s$  (for the 3000  $m/z$  range). As the ions drift for 60 ms in the drift cell, there will be 375 mass spectra between consecutive drift time points.

#### VII.2.1.5. Software

Two different softwares, both part of the Agilent Mass Hunter suite, are used to visualize and analyze the data: IM-MS Browser B.07.01 and Qualitative Analysis B.07.00 (Agilent Technologies, Santa Clara, CA, USA). The first one is used to analyze IMS data ( $m/z$ , drift time and intensity) and the other one to have access to the MS spectra.

#### VII.2.2. Waters LCT premier

The LCT premier is a commercially available instrument from Waters. Its mass analyzer is a TOF as the AGILENT instrument. However, the length of the TOF tube is much lower than the one on the AGILENT. The resolution will be then lower. The extraction of the data is done using the MassLynx 4.1 software provided with the instrument.

#### VII.2.3. Thermo LCQ Fleet

The LCQ Fleet (Thermo Fisher Scientific, Waltham, MA, USA) is based on a quadrupolar ion trap mass analyzer. Complete and detailed explanations about the functioning of ion trap are given in the literature<sup>5,6,7</sup>. Briefly, the ion trap is composed of 3 electrodes: a ring electrode, and 2 endcaps one at the entry of the trap and the last one at the exit (Figure 29). The trap is filled with Helium and the pressure is around  $10^{-3}$  Torr. Thanks to electric fields and collisions with He, the ions are confined in the center of the trap. The instrument allows the ejection of ions and so the selection of a parent ion by changing the applied V (amplitude of the RF of the ring electrode). Changing the voltage between the two endcaps will activate the parent ion and allow its fragmentation. With the LCQ Fleet, it is possible to vary this activation voltage and to control the activation time. MS<sup>n</sup> can be performed by re-selecting a fragment ion, activating it, and analyzing its fragments.

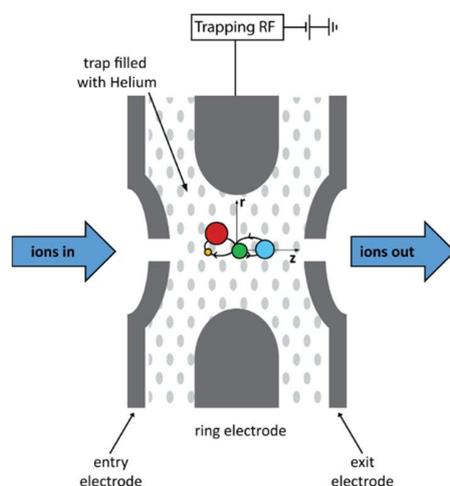


Figure 29: Schematic representation of the ion trap (A) and the Mathieu stability diagram (B).

The Thermo Xcalibur Qual Browser 2.2 (Thermo Fisher Scientific, Waltham, MA, USA) software is used to visualize and analyze the data obtained.

### VII.3. MS data analysis

#### VII.3.1. Peak integration

The first step, before any quantification analysis, is to extract the areas of the peaks of interest. Peaks are integrated through a defined  $m/z$  range that comprises the desired peak. Then background, which correspond to the integration of the signal prior to the peak of interest, over a similar  $m/z$  span, is subtracted.

#### VII.3.2. Adduct cleaning applied to kissing complexes

When solutions are doped with  $Mg^{2+}$  cations, adducts due to this cation are appearing. From this distribution, we would like to distinguish non-specific adducts from specific ones. First, it is important to define “non-specific” and “specific”.  $Mg^{2+}$  ions are non-specific if they are already present in solution but not at a particular location.  $Mg^{2+}$  ions are specific if they are binding to a specific site on the structure, or belong to a particular motif in comparison to a reference.

The method applied to discriminate between specific and non-specific adducts was first developed by Klassen and co-workers.<sup>8</sup> The aim of the method is to distinguish the proportion of specific and non-specific adducts under the distribution of cation adducts. The method is based on a mathematical treatment involving the subtraction of the distribution of non-specific adducts from the total distribution of adducts. The non-specific adducts distribution is defined as the distribution of adducts on a reference

sequence which does not form a kissing complex. The sequences chosen are RNA duplexes that have the same number of bases and composition than the sequence of interest.

1) The first step is to do the titration of the kissing complex and reference duplex by  $Mg^{2+}$  ions. The different points of the titration correspond each to one sample containing a different concentration of  $Mg^{2+}$ . In each spectrum, the integrals of the peaks, corresponding to the kissing complex or duplex, with and without adducts are determined.

2) We define ratio corresponding to the proportion of non-specific adducts (equation (2) and Figure 30,A). Here 4 non-specific adducts are taken into account.

$$x_1 = \frac{A_{Ref,1Mg}}{A_{Ref,0Mg}} \quad x_2 = \frac{A_{Ref,2Mg}}{A_{Ref,0Mg}} \quad x_3 = \frac{A_{Ref,3Mg}}{A_{Ref,0Mg}} \quad x_4 = \frac{A_{Ref,4Mg}}{A_{Ref,0Mg}} \quad (2)$$

$A_{Ref,iMg}$  is the area of the peak corresponding to the  $i^{th}$   $Mg^{2+}$  adduct.

3) Then, by applying the equations below, we can have access to the distribution of specific  $Mg^{2+}$  (Figure 30B).

$$I_{0Mg} = i0Mg * (1 + x_1 + x_2 + x_3 + x_4)$$

$$I_{1Mg} = (i1Mg - i0Mg * x_1) * (1 + x_1 + x_2 + x_3 + x_4)$$

$$I_{2Mg} = (i2Mg - i0Mg * x_2 - (i1Mg - i0Mg * x_1) * x_1) * (1 + x_1 + x_2 + x_3 + x_4)$$

$$I_{3Mg} = (i3Mg - i0Mg * x_3 - (i1Mg - i0Mg * x_1) * x_2 - (i2Mg - i0Mg * x_2 - (i1Mg - i0Mg * x_1) * x_1) * x_1) * (1 + x_1 + x_2 + x_3 + x_4)$$

$$I_{4Mg} = (i4Mg - i0Mg * x_4 - (i1Mg - i0Mg * x_1) * x_3 - (i2Mg - i0Mg * x_2 - (i1Mg - i0Mg * x_1) * x_1) * x_2 - (i3Mg - i0Mg * x_3 - (i1Mg - i0Mg * x_1) * x_2 - (i2Mg - i0Mg * x_2 - (i1Mg - i0Mg * x_1) * x_1) * x_1) * x_1) * (1 + x_1 + x_2 + x_3 + x_4)$$

Where  $I_{0Mg}$ ,  $I_{1Mg}$ ,  $I_{2Mg}$ ,  $I_{3Mg}$  and  $I_{4Mg}$  reflect the specific  $Mg^{2+}$  intensities,  $i0Mg$ ,  $i1Mg$ ,  $i2Mg$ ,  $i3Mg$  and  $i4Mg$  are the integrals of the  $i^{th}$  adduct of  $Mg^{2+}$  on the kissing complex peak distribution.

Experimentally, all the experiments are done at least 3 times in order to have assess the repeatability and calculate the standard error on the average value. For all titrations, samples contain 10  $\mu$ M of each hairpin, 2  $\mu$ M of  $dT_6$  (internal standard), 150 mM of  $NH_4OAc$  and a varying concentration of  $Mg^{2+}$ .

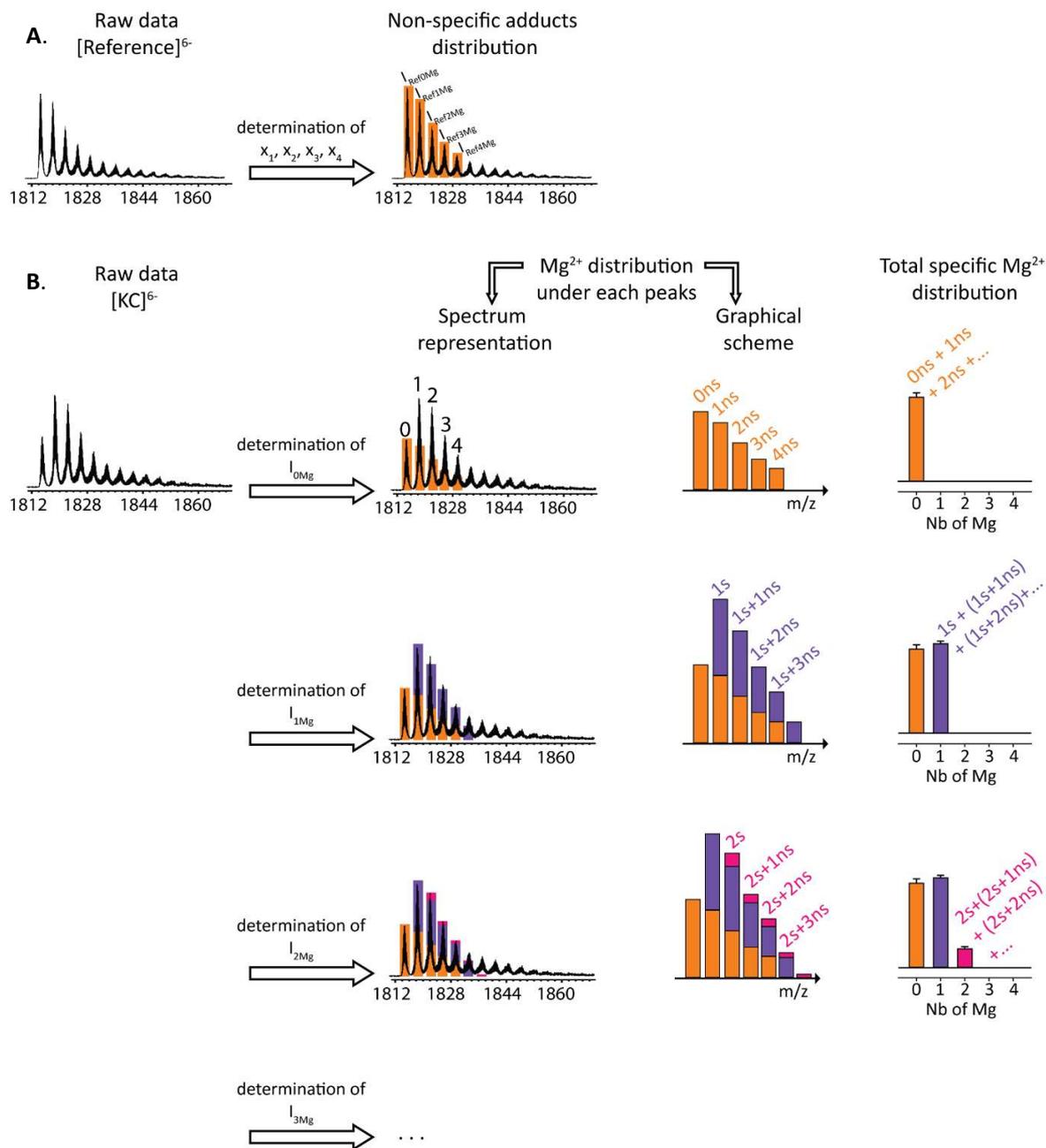


Figure 30: Schematic representation of all the steps involved in the adduct cleaning treatment. “ns” means non-specific and “s” specific.

### VII.3.3. Determination of equilibrium binding constants

Here we used MS to determine binding constant for all studied kissing complexes models.

$K_D$  is the equilibrium dissociation constant related to equilibrium (3) and can be defined as followed:



$$K_D = \frac{[M][L]}{[ML]} \quad (3a)$$

Here we present an example of the equilibrium between a compound M and a ligand L, which form the complex ML to simplify the equations. The following section are describing all the procedure to determine dissociation constant,  $K_D$ .

#### VII.3.3.1. Determination of response factors and concentrations

Peak intensities can be related to the concentrations of the species presents in solution. In MS, concentration ratios are not necessarily equal to the intensity ratios. Indeed, different molecules may not respond in the same way to desolvation and ionization by the ESI source. We need to determine the response factors, which are the proportionality coefficients between the concentrations and the intensities of the different species (4).

$$I_M = R_M \times [M] \text{ and } I_{ML} = R_{ML} \times [ML] \quad (4)$$

If we take the example of equilibrium (3), by combining equations (3a) and (4) we have the relation between concentration ratios, intensity ratios and response factor ratios.

$$\frac{[M]}{[ML]} = \frac{R_{ML}}{R_M} \times \frac{I_M}{I_{ML}} \quad (5)$$

Where  $I_M$  and  $I_{ML}$  are the intensities of M and ML respectively, and  $R_M$  and  $R_{ML}$  their respective response factor.

The complete method to have access to the response factors was described by Gabelica *et al* in 2009.<sup>9</sup> Briefly, by doing titrations of M by L, ratios of intensities between the species M or ML compared to an internal standard are determined. For our study, the internal standard is dT<sub>6</sub>. This standard does not have any interactions with the different partner present in solution, and is thus present at the same free concentration in all solutions. Ratios of response factors will be then accessible and intensity ratio can be related to concentrations ratio at each point of the titration (6).

$$\frac{[M]}{[Std]} = \frac{I_M}{I_{Std}} \times \frac{R_{Std}}{R_M} = \frac{I_M}{I_{Std}} \times R_f^M \quad \frac{[ML]}{[Std]} = \frac{I_{ML}}{I_{Std}} \times \frac{R_{Std}}{R_{ML}} = \frac{I_{ML}}{I_{Std}} \times R_f^{ML} \quad (6)$$

The total concentration of M,  $[M]_{tot}$ , can be expressed as the sum of  $[M]$  and  $[ML]$ .

$$[M]_{tot} = [M] + [ML] \quad (7)$$

By combining equation (6) and (7) we can then have:

$$\frac{[M]_{tot}}{[Std]} = \frac{I_M}{I_{Std}} \times R_f^M + \frac{I_{ML}}{I_{Std}} \times R_f^{ML} \quad (8)$$

As we know  $[M]_{tot}$ ,  $[Std]$ ,  $\frac{I_M}{I_{Std}}$  and  $\frac{I_{ML}}{I_{Std}}$ , we can use matrix calculations to solve the system of linear equations for  $R_f^M$  and  $R_f^{ML}$ . Knowing the response factors, we can then go back to the corrected concentrations of M and ML at each point of the titration.

The application of the method requires to respect some conditions. First it is important that the internal standard concentration does not change along the titration and that it does not interact with the other species in solution. Secondly, the complexes formed should not be dissociated during the gas phase and all the complexes have to be detected by the mass spectrometer.

Experimentally, one of the two hairpins involved in a loop-loop interaction is assimilated as the ligand L and the other one as the compound M. The kissing complex formed is thus the complex ML. The titrations were done at least once each way, meaning hairpin 1 by hairpin 2 and the other way round. The concentration of M is fixed at 10  $\mu\text{M}$ , so  $[M]_{tot} = 10 \mu\text{M}$  and  $[dT_6] = 2 \mu\text{M}$ . The ligand is added from 0 to 2 equivalents. The ionic strength is fixed at 150 mM  $\text{NH}_4\text{OAc}$ . All the titrations M by L are done at several concentrations of  $\text{Mg}(\text{OAc})_2$ : 0, 50, 200 and 600  $\mu\text{M}$ .

#### VII.3.3.2. Determination of $K_D$ values

Now that concentrations are determined for each point of the titration, the set of data are evaluated and fitted using the DynaFit software (4.07.109, BioKin Ltd, Watertown M.A, USA)<sup>10</sup>.

The input is all concentrations of M and ML as a function of the concentration of the ligand L. The software will use iterations to fit the data and gives as output the binding constants and the best fit of the data. The software is based on small scripts where the user gives the following information:

- [task] defines the type of data (equilibrium in this case) and what the user wants the software to do (here, fit).

- [mechanism] defines the equation of the equilibrium which have to be taken into account to calculate the dissociation constant.
- [constants] defines the starting point of the iterations to find the constant. The question mark means that the software has to evaluate this value.
- [concentrations] is the section where the user has to precise the starting concentrations of the compound M.
- The [data] section contains the set of data the software is going to fit. The variable is the concentration of free L. The data given as .txt files are formed by two columns, one with the concentration of total L added and the corresponding  $M_xL_y$  concentration. Also, the user can specify the response of the molecule. Here, the response is equal to 1 as the correction has been already made to obtain the concentrations.

Figure 31 gives an example of a script based on a simple 1:1 equilibrium. DynaFit can be used to fit various equilibrium from a simple model to more complicated once (successive equilibria for example).

```
[task]
data = equilibrium
task = fit
[mechanism]
M + L <==> ML : K1 dissociation
[constants]
K1 = 1 ?
[concentrations]
M = 10
[data]
variable L
file ...\M.txt | response M = 1
file ...\ML.txt | response ML= 1
[output]
```

Figure 31: Example of script for DynaFit Software.

## VIII. Ion Mobility Spectrometry: how to obtain CCS values?

Experimentally, an IM-MS experiment gives the arrival time distribution (ATD) of the ions of interest. As explained in the introduction, this ATD can be related to the Collision Cross Section (CCS) of the molecule. The CCS is related to the orientationally averaged momentum transfer collision integral. From the ATD, the CCS of the center of the distribution can be determined and the ATD can be converted into a CCS distribution.

### VIII.1. Determination of the CCS of the ATD peak center

The ions are accelerated by the electric field until the friction forces counterbalance it. In these conditions, the ions will drift at a constant velocity ( $v_d$ ).

$$v_d = K \times E \quad (9)$$

This velocity is proportional to the electric field ( $E$ ) and to the mobility ( $K$ ).  $E$  is defined by:

$$E = \frac{\Delta V}{L} \quad (10)$$

Where  $\Delta V$  is the voltage applied and  $L$  the length of the drift tube. The mobility can be also connected to the reduced mobility ( $K_0$ ) which is the mobility normalized at  $p_0 = 760$  Torr and  $T_0 = 273.15$  K. We have then:

$$K_0 = K \frac{p T_0}{p_0 T} \quad (11)$$

By combining equation (9), (10) and (11), we can express  $K_0$  as a function of  $t_d$  (drift time of the ion):

$$K_0 = \frac{v_d p T_0}{E p_0 T} = \frac{L p T_0}{t_d E p_0 T} = \frac{1 L^2 p T_0}{t_d \Delta V p_0 T} \quad (12)$$

This reduced mobility can be associated to the collision cross section using the Mason-Schamp equation (kinetic theory of ion transport).<sup>11</sup>

$$CCS = \frac{3}{16} \sqrt{\frac{2\pi}{\mu k_B T}} \times \frac{ze}{N_0} \times \frac{1}{K_0} \quad (13)$$

Where  $k_B$  is the Boltzmann constant,  $T$  the temperature,  $\mu$  is the reduced mass of the ion ( $\mu = \frac{m_{gas} m_{ion}}{m_{gas} + m_{ion}}$ , in our case the gas is Helium so  $m_{gas} = 4.002$ ),  $ze$  the charge in Coulomb,  $N_0$  is the buffer gas number density at  $T_0$  and  $p_0$ . By combining equation (11) and (12) we have the expression of CCS:

$$CCS = \frac{t_d}{L} \frac{3}{16} \sqrt{\frac{2\pi}{\mu k_B T}} \times \frac{zeE}{N_0} \times \frac{p}{p_0} \times \frac{T_0}{T} \quad (14)$$

Experimentally, we do not have directly access to the drift time  $t_d$ . Indeed, the arrival time measured  $t_a$  is a sum between the drift time and the time spent outside the mobility cell ( $t_0$ ).

$$t_a = t_d + t_0 \quad (15)$$

We have then:

$$CCS = \frac{(t_a - t_0)}{L} \frac{3}{16} \sqrt{\frac{2\pi}{\mu k_B T}} \times \frac{zeE}{N_0} \times \frac{p}{p_0} \times \frac{T_0}{T} \quad (16)$$

In practice, IMS experiments are carried out at five different drift tube voltages. Using a Gaussian fit of the arrival time distribution peak, we can determine  $t_a$  of the center of each peak. The Peakfit 4.12 software was used to fit each peak of the ATDs. Then by plotting  $t_a$  as a function of  $\frac{1}{\Delta V}$  (or  $\frac{P}{\Delta V}$ ), we can have access to  $K_0$  (from the slope) and to  $t_0$  (Figure 32). The  $CCS_{\text{peakcenter}}$  can be calculated for each of the 5 voltages. Error on the CCS values can be estimated from the relative error on the slope.

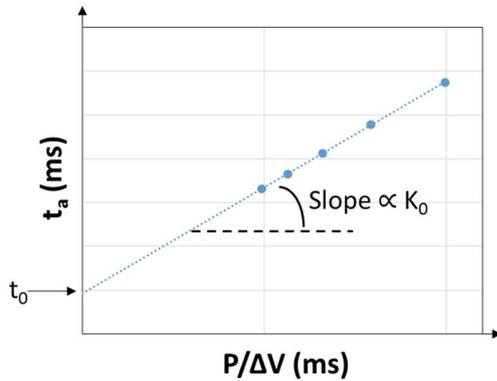


Figure 32: Example of a plot showing  $t_a$  as a function of  $p/\Delta V$ .

## VIII.2. CCS distribution reconstruction

The calculation and reconstruction of collision cross section distributions have been recently published by our team.<sup>12</sup> In the following section, we will explain briefly the two methods used for CCS distribution (CCSD) reconstruction.

### VIII.2.1. Single-field reconstruction method

Marchand *et al*<sup>12</sup> have shown that  $t_A$  is on first approximation linearly proportional to  $CCS \frac{\sqrt{\mu}}{z}$ . Knowing this proportionality factor ( $a$ ) at a given voltage, it is possible to reconstruct the CCS distribution using Equation (17).

$$CCS = a \frac{z}{\sqrt{\mu}} \times t_A \quad (17)$$

The CCS distribution obtained then reflect the good position of each peaks. This method is called the single-field reconstruction method.

### VIII.2.2. $FWHM_{step}$ reconstruction method

The width of the arrival time distribution is related to the conformations of the molecule but also to the diffusion of the molecule inside the drift cell. Because the latter contribution should not be translated into the CCS distribution, it means that the distributions reconstructed with Equation (16) are actually too broad. Marchand *et al*<sup>12</sup> have developed a method to obtain the theoretical Gaussian CCS distribution if no diffusion is observed. The method uses the full width at half maximum (FWHM) of the peaks. Experimentally we have access to  $FWHM_{ATD,exp}$ , which is the experimental FWHM of the Gaussian fit of the arrival time distribution. We can calculate the  $FWHM_{diff}$  due to the diffusion in and outside the drift cell based on equation (18).

$$FWHM_{diff} = 4t_D \sqrt{\ln(2)} \sqrt{\frac{k_B T}{zeV}} \quad (18)$$

Then we will have access to  $FWHM_{step}$  :

$$FWHM_{step} = \sqrt{FWHM_{system}^2 + FWHM_{diff out IMS}^2} \quad (19)$$

It is then possible to plot  $t_A$  as a function of  $\frac{p}{dV}$  for  $t_{Acenter} \pm \frac{1}{2}FWHM_{step}$ . After we will determine the CCS of the point at  $\pm \frac{1}{2}FWHM_{step}$ , and reconstruction of the Gaussian curve related to the CCS distribution only due to the molecule can be done.

## IX. How to analyze MS/MS data

As said previously, when doing MS/MS, the precursor ion is going to be fragmented into various fragments which we need to identify. To do so, we used Mongo Oligo Mass Calculator v2.08 (software hosted by the RNA Institute, University of New York, Albany)<sup>13</sup> and Nuke-Nucleotide calculator v.2.12 (done by Arne Leisner, 1999). These softwares calculate the possible resulting fragments using the nomenclature proposed by McLuckey et al<sup>14</sup> (Figure 33). They will also calculate masses for base losses or internal fragmentation.

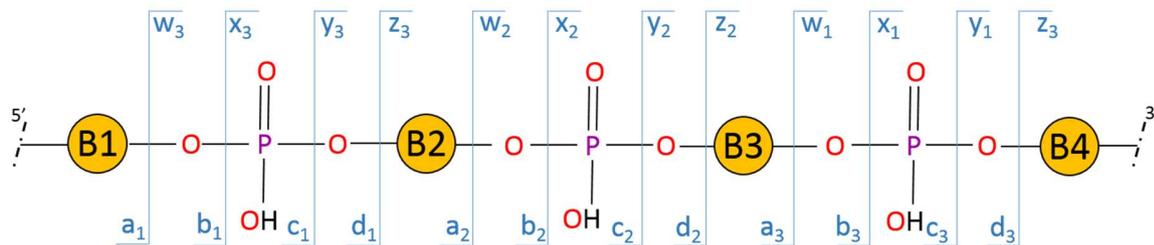


Figure 33: Schematization of the nomenclature of oligonucleotides described by McLuckey et al<sup>14</sup>.

## References

- (1) Cavaluzzi, M. J., and Borer, P. . (2004) Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res.* 32, 13e–13.
- (2) Mergny, J.-L., and Lacroix, L. (2002) Des Tms, encore des Tms, toujours des Tms ! 1–54.
- (3) Wyttenbach, T., and Bowers, M. (2001) Design of a new electrospray ion mobility mass spectrometer. *Int J Mass Spectrom* 212, 13–23.
- (4) Belov, M. ., Ibrahim, Y. ., Clowers, B. ., Prior, D. ., and Smith, R. . (2011) United States Patent: IOn funnel ion trap and process.
- (5) Barinaga, C. J., Koppenaal, D. W., and McLuckey, S. A. (1994) Ion-trap mass spectrometry with an inductively coupled plasma source. *Rapid Commun. Mass Spectrom.* 8, 71–76.
- (6) March, R. E. (1997) An Introduction to Quadrupole Ion Trap Mass Spectrometry. *J. Mass Spectrom.* 32, 351–369.
- (7) Stafford, G. (2002) Ion trap mass spectrometry: A personal perspective. *J. Am. Soc. Mass Spectrom.* 13, 589–596.
- (8) Sun, J., Kitova, E. N., Wang, W., and Klassen, J. S. (2006) Method for distinguishing specific from nonspecific protein-ligand complexes in nanoelectrospray ionization mass spectrometry. *Anal. Chem.* 78, 3010–8.
- (9) Gabelica, V., Rosu, F., and De Pauw, E. (2009) A simple method to determine electrospray response factors of noncovalent complexes. *Anal. Chem.* 81, 6708–15.
- (10) Kuzmic, P. (1996) Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.* 237, 260–273.
- (11) Revercomb, H. E., and Masson, E. A. (1975) Theory of Plasma Chromatography / Gaseous Electrophoresis — A Review. *Anal. Chem.* 47, 970–983.
- (12) Marchand, A., Livet, S., Rosu, F., and Gabelica, V. (2017) Drift Tube Ion Mobility: How to Reconstruct Collision Cross Section Distributions from Arrival Time Distributions? *Anal. Chem.* (13) Mongo Oligo Mass Calculator v2.08 : <http://mods.rna.albany.edu/Masspec-Toolbox>.
- (14) Mcluckey, S. A., Berkel, J. Van, and Glish, L. (1992) Tandem Mass Spectrometry of Small Multiply Charged Oligonucleotides. *J. Am. Soc. Mass Spectrom.* 3, 60–70.

### 3. RESULTS AND DISCUSSION

#### PART I: NEW DEVELOPMENTS IN MASS SPECTROMETRY, NOVEL INSIGHTS INTO KISSING-COMPLEXES STRUCTURE

## Chapter 1. Characterization of the interaction between Mg<sup>2+</sup> and kissing complexes by *Native* mass spectrometry

### X. *Native* MS from Magnesium solutions: optimization of the method

#### X.1. Introduction

Studies of RNA by mass spectrometry were first focused on the analysis of RNA sequence and on the identification of post-transcriptional modifications.<sup>1</sup> Then, MS was used to analyze RNA-protein complexes. In particular, several studies were performed on the Tat-TAR system. In 1997, Sannes-Lowery and co-workers studied the interaction between Tat and TAR sequence using *Native* MS. They obtained different charge states of the complex and they determined binding stoichiometries.<sup>2</sup> In a more recent study, Schneeberger et al analyzed the binding sites of Tat to TAR by *Native* Top-Down MS, and showed that it is a suitable technique for such analysis.<sup>3</sup> The team of Daniele Fabris has studied the interaction of the dimerization initiation site of the HIV-1, which forms an RNA kissing complex, with the nucleocapsid protein.<sup>4-6</sup> All these experiments involving RNA-protein or RNA-RNA-Protein complexes were done in solutions containing ammonium acetate with addition of methanol or isopropanol prior the experiments. Even if Mg<sup>2+</sup> has been shown to be involved in RNA folding and stabilization of complexes<sup>7,8</sup>, none of the previously published MS experiments were done using solutions doped with magnesium.

One of the main challenges when adding salts in MS is the formation of adducts that can degrade the quality of the mass spectra. NH<sub>4</sub>OAc is commonly used in the field of *Native* MS, owing to its ability to transfer protons and be volatile. Indeed, NH<sub>4</sub><sup>+</sup> will transfer a proton the DNA or RNA molecule and will be released as NH<sub>3</sub>. Ammonia and acetate are volatile. Because ammonium competes with sodium around the nucleic acids, the number of non-specific sodium adducts is reduced, making NH<sub>4</sub>OAc a solution of choice for ESI-MS.<sup>9</sup> NH<sub>4</sub><sup>+</sup> non-specific adducts can remain bound if the internal energy is not sufficient for all proton transfer reactions to occur, and this can become predominant for long RNAs.<sup>10</sup> Na<sup>+</sup>, even present as traces in the solution, is not volatile so it cannot be removed and forms non-specific adducts.<sup>11</sup> However, Marchand *et al.* have shown that alkali cations are not necessarily incompatible with MS. Indeed, they were able to spray solutions doped with up to at least 1 mM K<sup>+</sup> cations, to study G-quadruplexes folding.<sup>12,13</sup>

Our goal here is to study the possibility of doing mass spectrometry from solutions doped with magnesium, and find the best method to analyze RNA-RNA complexes. As presented in the introduction,

kissing complexes have been shown to interact with magnesium cations<sup>14-16</sup> and we will use TAR-R06 and RNAIi-RNAIi kissing complex as models.

The first step is to verify that kissing complexes can be formed in MS-compatible solution, and then optimize the experimental setup to prove that Mg<sup>2+</sup>-containing solutions are not incompatible with MS. This section describes the methods used to achieve this goal.

## X.2. Results

### X.2.1. Characterization in solution

It is important to check if kissing complexes can be formed in MS-compatible solutions. To do so, we used thermal denaturation to evaluate first the stability of kissing complexes in ammonium acetate (NH<sub>4</sub>OAc) solutions, without magnesium. For this experiment, a concentration of 150 mM NH<sub>4</sub>OAc is used. The results are presented in Figure 34.

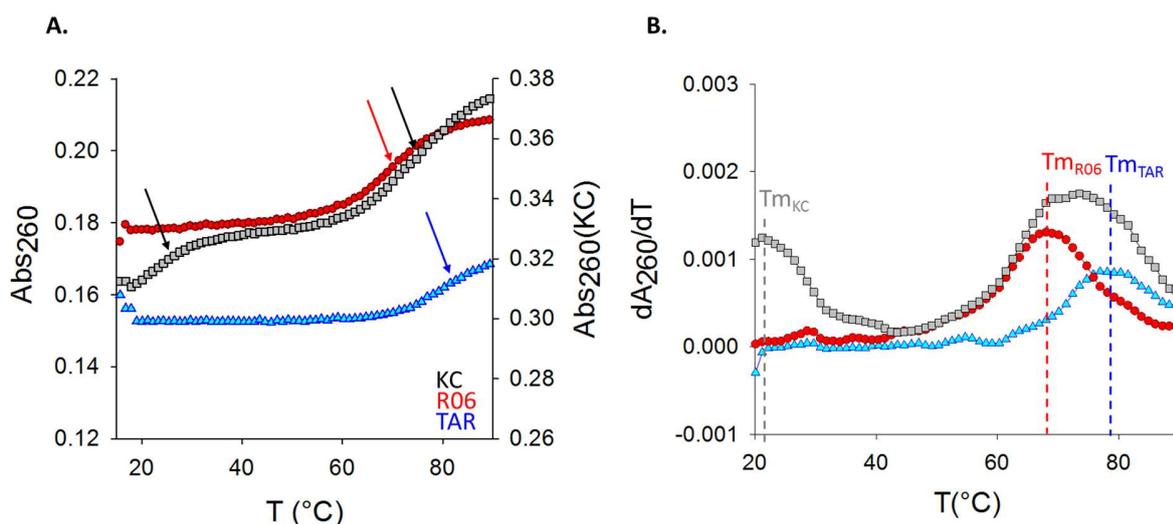


Figure 34: Thermal denaturation curves obtained for TAR alone (blue curve), R06 alone (red curve) and the kissing complex formed from an equimolar mixture of TAR and R06 (KC, in black). Solutions were containing 1  $\mu$ M of hairpins and 150 mM NH<sub>4</sub>OAc. A) Absorbance at 260 nm as a function of temperature. The arrows are showing the position of the transitions. For an easier readability, a second axis for the Abs<sub>260</sub> of the KC is added (right axis). B) First derivative of the curves showing the T<sub>m</sub> value determination.

The two hairpins alone present only one transition each, indicated by an increase of the absorbance. This transition corresponds to the opening of the hairpin by the rupture of the H-bonds between complementary bases. On the contrary, when the solution containing both hairpins is heated, we observe two transitions. The first one corresponds to the dissociation of the kissing complex into its two monomers, and the second one is the dissociation of the hairpins. The melting temperature (T<sub>m</sub>) of a

kissing complex is the temperature at which half of the KC remains formed. As all baselines cannot be determined, we used the first derivatives of the curves to determine the melting temperature of both transitions. The  $T_m$  values are around 80°C and 75°C respectively for TAR and R06, and around 20°C for the kissing complex. These values agree with the literature in which  $T_{m(\text{TAR})} \approx 85.3^\circ\text{C}$ ,  $T_{m(\text{R06})} \approx 71.5^\circ\text{C}$  and  $T_{m(\text{KC})} \approx 31^\circ\text{C}$  (Lebars et al<sup>17</sup>) where the experiments were carried out using solutions containing higher salt concentrations (20 mM NaCl, 140 mM KCl and 300  $\mu\text{M}$   $\text{MgCl}_2$ ). The differences observed between their values and ours can come from the salt concentrations, which influence the KC stabilization.

TAR-R06 kissing complexes can be formed in  $\text{NH}_4\text{OAc}$  and at 22°C (room temperature) TAR-R06 kissing complex should be formed at least at 50%.

We also wanted to check the effect of adding magnesium on the stability of the kissing complexes in 150 mM  $\text{NH}_4\text{OAc}$ . The results are presented on Figure 35 as the first derivative curves of the absorbance by the temperature. Only the region between 15 and 50°C is shown.

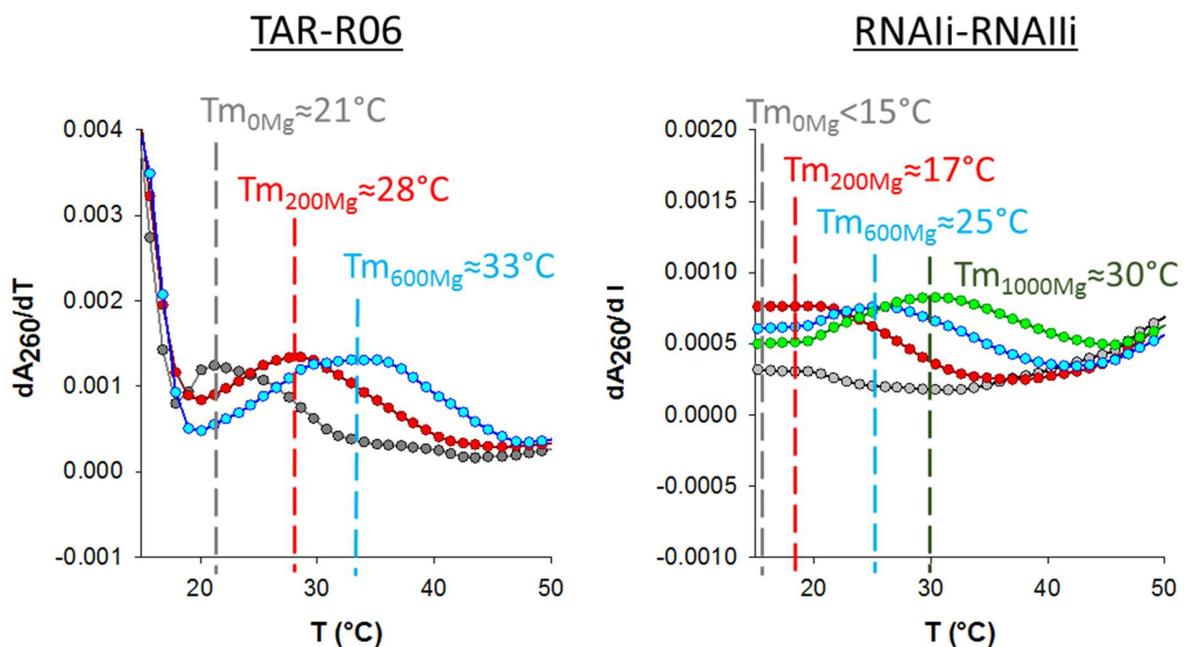


Figure 35: UV melting of TAR-R06 and RNAi-RNAIi at different  $\text{Mg}(\text{OAc})_2$  concentrations. First derivative of the absorbance at 260 nm in function of the temperature. The local maximum on the curves represent the dissociation of the KC into hairpins. At  $T > 50^\circ\text{C}$  (not shown on the graphs), the maximum corresponds to the melting of the individual hairpins. In grey: solution without  $\text{Mg}(\text{OAc})_2$ , red with 200  $\mu\text{M}$ , blue with 600  $\mu\text{M}$  and green with 1 mM.

When the concentration of magnesium increases, the melting temperature of the KC also increases, for both systems. This means that the KC becomes more and more stable with an addition of  $\text{Mg}^{2+}$ . In the

literature, one can find values of  $T_m$  for both systems:  $T_{m(\text{TAR-R06})} \approx 31.3^\circ\text{C}$ <sup>17</sup> at 300  $\mu\text{M}$   $\text{MgCl}_2$  and  $T_{m(\text{RNAIi-RNAIii})} \approx 59^\circ\text{C}$ <sup>18</sup> at 5 mM  $\text{MgCl}_2$ . Our values agree with the literature. These results show clearly the effect of magnesium on the KC stability. Also, RNAIi-RNAIii needs more magnesium than TAR-R06 to have the same  $T_m$  (600  $\mu\text{M}$  vs 1 mM).

### X.2.2. IMS Q-TOF optimization

Different instruments were used over the 3 years. The first step was to test different tuning parameters to have soft enough conditions to keep the kissing complex intact, but also find conditions in which a good signal is obtained with solutions containing magnesium. Our starting point, for each instrument used, was the set of tuning parameters optimized to analyze a specific G-quadruplex called  $G_4T_4G_4$ . This DNA sequence folds into a G-quadruplex having three specific  $\text{NH}_4^+$  between the G quartets.<sup>19</sup> An evaluation of the number of cations detected will be a sign of the softness of our parameters. The tune parameters are described in the supporting information of Marchand *et al.*'s papers.<sup>12,13</sup>

As optimization is needed whatever the instrument used, the following sections are focused on the optimization of TAR-R06 detection on the AGILENT 6560 IMS Q-TOF. Optimization experiments on the LCT premier are described in appendix A1.1.

#### X.2.2.1. What does the spectrum look like?

Our starting point was to use a set of trapping voltages that is soft enough to preserve the kissing complex. The parameters used are presented on the material and method section and can be found in the literature, as they were used on G-quadruplexes and duplexes DNA.<sup>13,20</sup> This set is called the "SOFT tune". The spectrum on Figure 36 is acquired using the SOFT tune.

We can observe both hairpin monomers at different charge states, 4- and 3-, where the 4- is the major charge state. The kissing complex is detected at charge states 6- (major one) and 5-.  $\text{NH}_4^+$  adducts are present on all the different species detected. Also we note a difference in the number of adducts between  $[\text{TAR}]^{4-}$  and  $[\text{R06}]^{4-}$ . As R06 is longer than TAR we can expect to have more  $\text{NH}_4^+$  adducts in identical activation conditions.

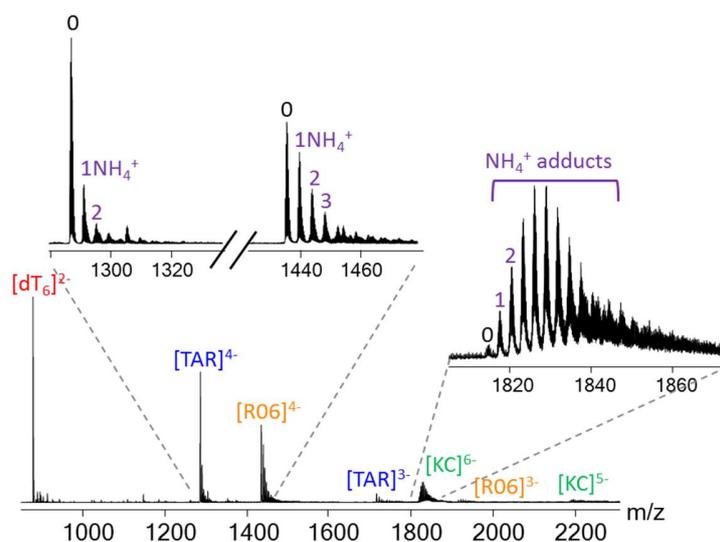


Figure 36: ESI-MS spectra of TAR-R06 obtained with SOFT parameters on the Agilent 6560 IMS-Q-TOF. Zooms on the major charge states for TAR, R06 and the Kissing complex show the ammonium adducts distribution.

#### X.2.2.2. How to influence declustering?

As said in the introduction,  $\text{NH}_4^+$  non-specific adducts and other adducts like  $\text{Na}^+$  can remain on RNA and so influence the final spectra obtained. To avoid non-specific adducts detection, we may have to change different parameters to activate the ions in order to decluster them. In negative ion mode, we cannot lose permanent positive ions such as  $\text{Na}^+$  or  $\text{Mg}^{2+}$ , but we can get rid of  $\text{NH}_4^+$  thanks to proton transfer and  $\text{NH}_3$  loss. Different ways to activate the ions prior to the ion mobility cell are possible when using the IMS Q-TOF: trap entrance grid delta (TEGD), the fragmentor and other trapping voltages. Changing those different parameters will allow us to change the pre-IMS activation conditions. We will first see the influence of the set of trapping voltages on declustering and then emphasize on the influence of the trap entrance grid delta and fragmentor voltage on the ion activation.

- Influence of the other trapping voltages on declustering

In addition to the SOFT set of trapping voltages, we used a second set of parameters corresponding to harsher conditions called HARD. The parameters of this tune is described on the materials and methods. Figure 37 shows the ESI-MS spectra for TAR-R06 at the two different tuning sets.

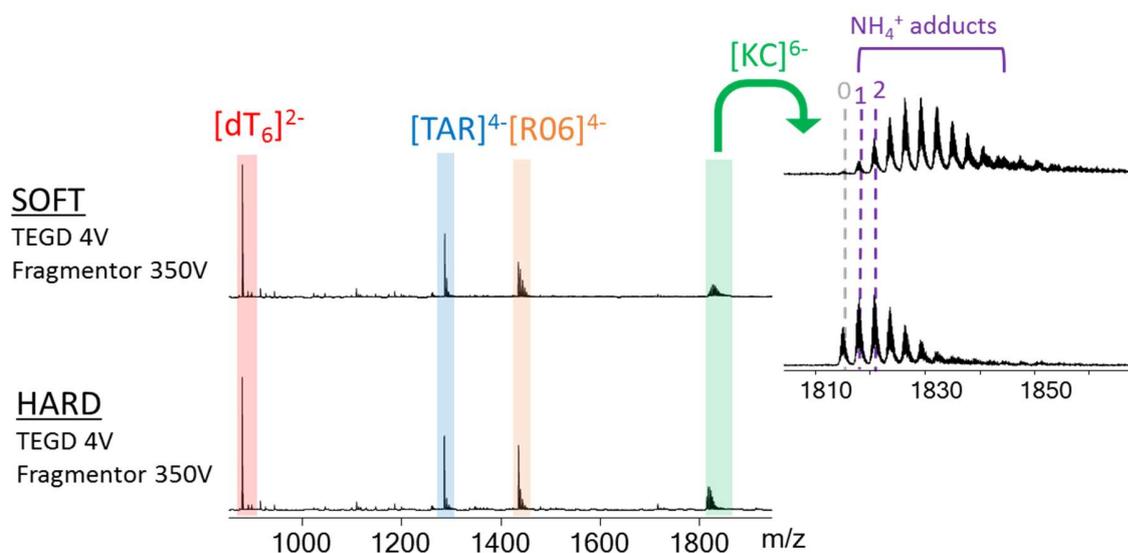


Figure 37: ESI-MS spectra of TAR-R06 obtained with the two different sets of parameters: SOFT (top) and HARD (bottom). Only the major charge states of each ion are identified. On the right, zoom on the  $KC^{6-}$  showing the difference in ammonium adducts between each spectrum.

On all spectra, we observe the same charge states for  $dT_6$  (internal standard), TAR, R06 and the kissing complex. Changes in trapping voltages have no impact on the charge state distribution. The number of  $NH_4^+$  adducts decreases with the use of the HARD voltages. More energy was given to allow a more efficient declustering. With the last tune, the peak of the KC without  $NH_4^+$  is increasing, which means that it is possible to get rid of ammonium ions. Also, the kissing complex does not dissociate when the trapping voltages are increased.

- Influence of TEGD and fragmentor voltages on the ESI-MS spectra

In addition to modification of the trapping voltages, it is also possible to change the Trap Entrance Grid Delta voltage and the fragmentor voltage.

### Influence of the Trap Entrance Grid Delta (TEGD)

To evaluate the effect of TEGD on ammonium adducts distribution, TEGD was varied from 4 V to 20 V. Tune SOFT and HARD trapping voltages were tested in addition to a change in TEGD value. The results are presented as 3D graphs showing the m/z distribution of a specific ion, here  $[KC-6H^+]^{6-}$ , as a function of the TEGD voltages (Figure 38). The third dimension, color-coded, is the number of counts.

Whatever the tune parameters used, the TEGD influences declustering. Indeed,  $NH_4^+$  adducts are progressively decreased, which corresponds to losses of  $NH_3$ . By comparing the two graphs, the HARD

tune is a continuity of the SOFT one in terms of pre-IMS ion activation. The peak of KC without  $\text{NH}_4^+$  is progressively increased until it becomes the major one in the harshest conditions. Also, even if the conditions become harsher, we are still in native mode as far as the kissing complex remains formed.

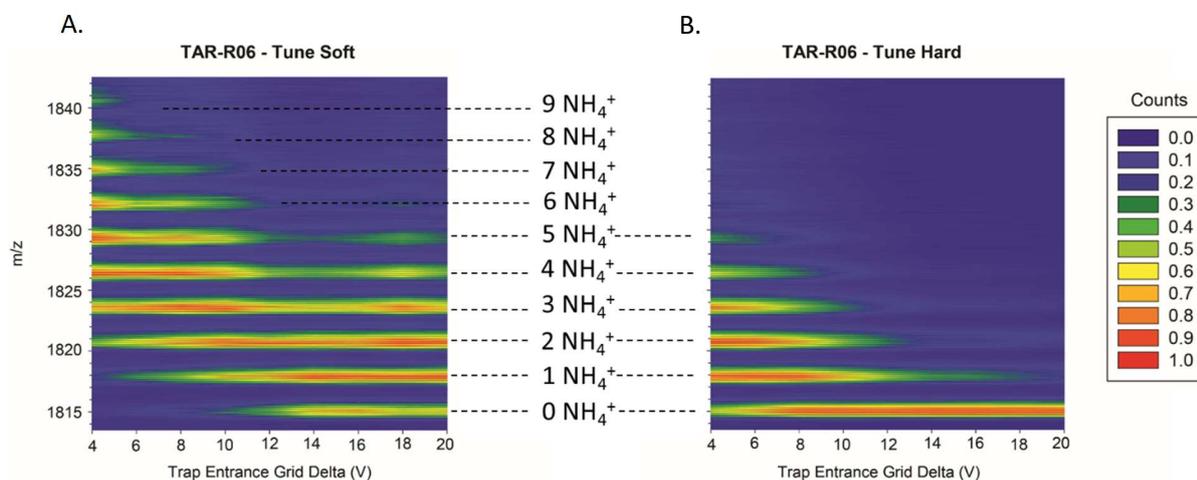


Figure 38: 3D plots showing the influence of the TEGD voltage on the ammonium distribution. A) Graph obtained with the SOFT tune. B) Graph obtained with the HARD tune. Fragmentor is set at 350 V.

### Influence of the Fragmentor

Same type of experiments were performed to check the influence of the fragmentor. For those experiments, the fragmentor was varied from 350 V to 600 V. The maximum fragmentor value imposed by the manufacturer is 600 V. Figure 39 presents the range of  $m/z$  of the ion of interest, here  $[\text{KC-6H}^+]^6$ , as a function of the fragmentor voltage.

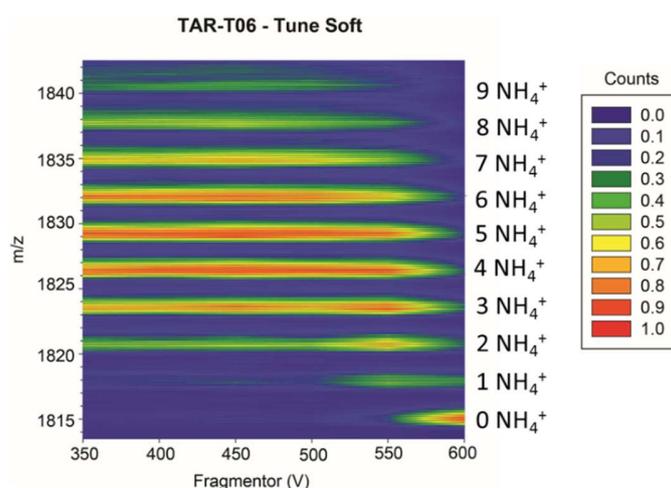


Figure 39: 3D plots showing the influence of the fragmentor voltage on the ammonium distribution using the SOFT tune. TEGD is set at -4 V.

On the contrary to the progressive effect of the TEGD on declustering, the effect of the fragmentor is seen from 500 V. Based on these results, we usually test 350 V and 600 V to test the fragmentor voltage's influence.

- Comparison between spectra with and without Magnesium

We injected solutions doped with magnesium acetate up to 1 mM. The results (Figure 40) show that it is still possible to inject solutions containing  $\text{Mg}(\text{OAc})_2$  without a total loss of signal. Contrary to the LCT results, the data obtained with the AGILENT IMS-Q-TOF show that solutions containing  $\text{MgCl}_2$  at a concentration at least equal to 1 mM are still amenable to MS analysis.

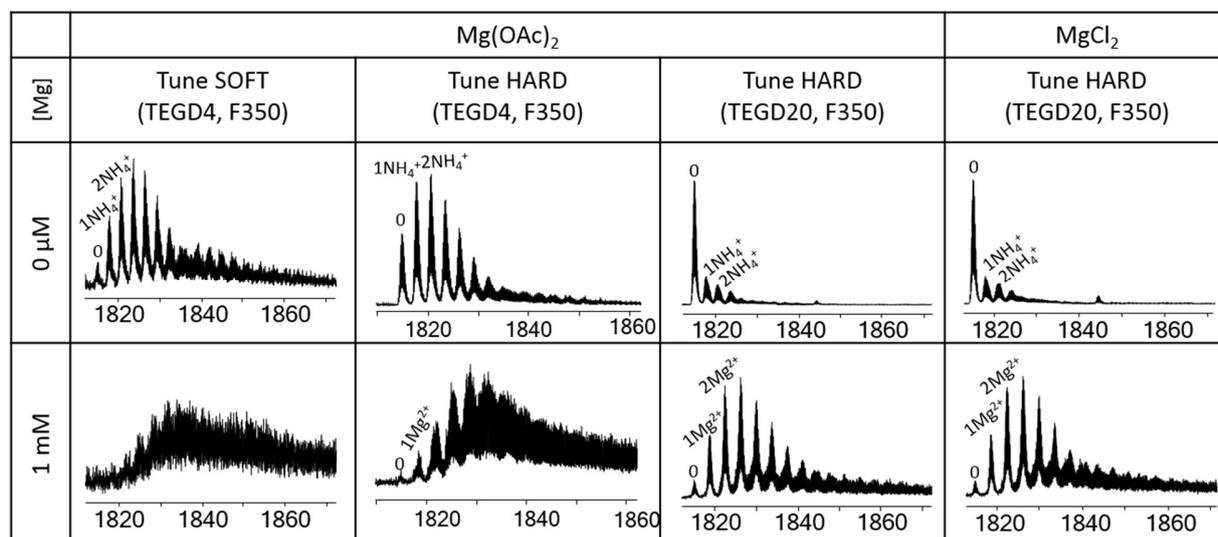


Figure 40: Zoom on TAR-R06 KC charge state 6- at different concentrations of  $\text{Mg}(\text{OAc})_2$  or  $\text{MgCl}_2$ . The tune and parameters used are indicated on the figure.

Nevertheless, tuning parameters still have a great importance. Indeed, depending on the tuning parameters used, the magnesium adducts distribution is not always distinguishable. With soft conditions, only a hump is observed for peaks representing the kissing complex plus magnesium and ammonium ions. On the contrary, with harsher conditions, different peaks corresponding to magnesium adducts are distinct. The ammonium adduct distribution is hidden under the magnesium adduct distribution, because they are overlapping.

### X.2.3 Choice of solution conditions

$\text{NH}_4\text{OAc}$  is the most commonly used electrolyte in *Native MS*,<sup>9</sup> but other electrolytes can be used as well, like triethylammonium acetate (TEAA)<sup>21</sup> or trimethylammonium acetate (TMAA)<sup>12</sup>. For G-quadruplexes,

Marchand *et al.* have shown that the use of TMAA allows the detection of specific  $K^+$  adducts and improve mass spectral quality.<sup>12</sup> We asked ourselves what can be the best solution conditions to study kissing complexes with and without magnesium.

Starting from  $NH_4OAc$ , we chose different ammonium salts to test the effect of the counterion on the mass spectrum:  $NH_4Cl$ ,  $NH_4NO_3$ ,  $NH_4Br$ ,  $(NH_4)_2SO_4$ ,  $NH_4PF_6$ ,  $NH_4CF_3SO_3$  and  $HCOONH_4$ . We also performed the experiment with TMAA. All the experiments were performed first without magnesium and with 10  $\mu M$  of each hairpin, 2  $\mu M$  of  $dT_6$  and 150 mM of the desired ammonium salt. The results are summarized in the appendix A1.2.

The full scan mass spectra obtained for  $(NH_4)_2SO_4$ ,  $NH_4PF_6$ ,  $NH_4Br$  and  $NH_4CF_3SO_3$  do not display the characteristic peaks of TAR-R06. Only peaks corresponding to  $NH_4^+$ - counterion clusters are detected. For  $NH_4NO_3$ , the two hairpins at charge state 3- are detected, but not the kissing complex charged 6-. The range of  $m/z$  used was under that of the 5- charge state. The best signals and spectra are obtained using  $NH_4OAc$ ,  $NH_4Cl$  and  $HCOONH_4$ , even if the intensity of  $KC^{6-}$  was low. Using these three solutions, kissing complexes 6- is detected in addition to the hairpin monomers. Concerning  $HCOONH_4$  the intensity of the monomers 3- is increased compared to  $NH_4OAc$ . When using  $NH_4Cl$ , we can see that the signal is noisier compared to the spectra obtained with  $NH_4OAc$  and  $HCOONH_4$ . The type of ammonium salt used has thus an influence on the charge states detected and on the formation or detection of kissing complex. (1) With  $Cl^-$ ,  $HCOO^-$  and  $CH_3COO^-$ , kissing complex 6- is detected. (2) With  $NO_3^-$ , only the hairpins are detected. (3) With  $SO_4^{2-}$ ,  $PF_6^-$  and  $SO_3CF_3^-$ , salt clusters hamper the observation of RNA signals.

Concerning TMAA, no kissing complex is detected. Only the two hairpins at charge state 4 and 3- are detected.  $TMA^+$  has an ionic radius of 0.322 nm compared to 0.137 nm for  $NH_4^+$ .<sup>22</sup> The hypothesis is that monovalent cations are involved in the loop-loop interaction and stabilize the kissing complex, but as  $TMA^+$  is bulkier than  $NH_4^+$ , the stabilization cannot happen.

Based on these results we decided to continue the experiments using only ammonium acetate as the main supporting electrolyte.

### X.3. Discussion

For all instrument, the SOFT conditions used previously in the group to obtain ~24-base long G-quadruplexes mass spectra were too soft, and thus declustering of the kissing complex was incomplete.  $NH_4^+$  can form hydrogen bonds with the RNA. Due to its anionic nature, RNA tends to retain  $NH_4^+$  in its ion atmosphere<sup>22</sup>, and thus a lot of ammonium adducts can be retained if the ionization is too soft. The

involvement of tightly bound  $\text{NH}_4^+$  ions is in line with the fact that kissing complexes cannot be formed when  $\text{TMA}^+$  ions are used instead. Monovalent salts are not working just to maintain a certain ionic strength, but they seem to play a more direct role in kissing complexes stabilization. To study the kissing complexes with magnesium and distinguish the magnesium adducts, reducing as much as possible the detected adducts of ammonium and sodium ions in the final mass spectra is important. Using harsher conditions will help to get rid of  $\text{NH}_4^+$  adducts, and a careful preparation of the sample to reduce as much as possible the number of  $\text{Na}^+$  adducts. Using this method, distinguishing each  $\text{Mg}^{2+}$  adduct is possible.

Our results showed that solutions doped with  $\text{Mg}(\text{OAc})_2$  up to 1 mM can be injected and analyzed by *Native MS*. The concentrations of magnesium used in our study are below the *total* concentration of magnesium present in the cells, which is around 17-20 mM.<sup>23</sup> However the concentration of *free* magnesium available in the cell has been estimated to be in between 0.25 to 1 mM depending on the cell type, the majority of  $\text{Mg}^{2+}$  being bound to proteins, nucleic acids or other biomolecules.<sup>24</sup> Based on this statement, the use of concentrations of  $\text{Mg}^{2+}$  below 1 mM can fairly mimic *in vitro* the free magnesium available in the cells.

## XI. Determination of specific cations number

### XI.1. Introduction

In the literature (see introduction), two types of  $Mg^{2+}$  binding have been defined: “diffuse ions” and “site-bound” (also called “chelated ions”). The first type of ions can be considered as non-specific as they are surrounding all RNA molecules, whatever their structure. They are still hydrated and are binding through weak interactions. The second type of ions involves stronger interactions at a specific location on the RNA. Both ions contribute to the stability and folding of the RNA.<sup>25,7,26</sup>

Here we propose a direct method, using *native* MS, to count the number of cations which are specific to the loop-loop motif. As seen in the previous chapter, when spraying solutions doped with  $Mg^{2+}$ , many  $Mg^{2+}$  adducts are detected. In this section, we will discuss a technique that allows us to distinguish between specific and non-specific adducts. In addition, this chapter will also focus on the comparison between the two models of kissing complex, TAR/R06 and RNAIi/RNAIli, and a comparison between magnesium and manganese.

### XI.2. Results

#### *XI.2.1. Characteristic spectra of TAR-R06 and RNAIi-RNAIli*

The DTIMS-Q-TOF is used for most of the experiments described in this chapter. In order to have clear spectra with clear distribution of adducts, we used the HARD tune, a TEGD voltage of 20 V and a Fragmentor voltage of 350 V. Using these parameters, the  $Mg^{2+}$  adducts distribution is well- defined, and the extraction of peak integrals is easy. Let's first describe the full spectra of TAR-R06 and RNAIi-RNAIli (Figure 41).

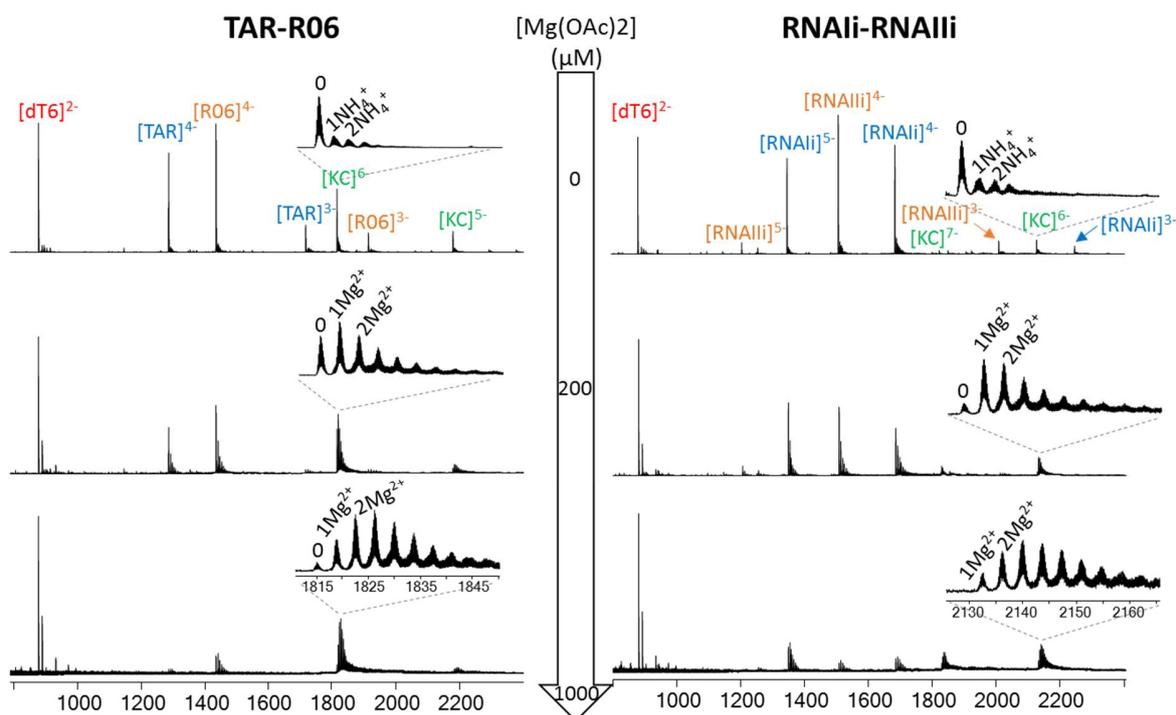


Figure 41: ESI-MS spectra obtained using the DTIMS Q-TOF of A) TAR-R06, B) RNAIi-RNAIii. The trapping voltages are set as HARD, TEGD at 20 V and Fragmentor at 350 V. Samples were containing 10  $\mu\text{M}$  of each hairpin, 150 mM of  $\text{NH}_4\text{OAc}$  and 2  $\mu\text{M}$  of  $\text{dT}_6$ . On the picture, there is a zoom of the  $\text{KC}^{6-}$  with the adducts distribution.

For both kissing complexes, we can observe the same type of ions. Monomers (hairpin alone) at different charge states are observed in both cases (4- and 3-). The charge state 5- is present for RNAIi and RNAIii. We can see on both spectra that the KC is detectable at different charge states, 6- and 5- for TAR-R06, and 7- and 6- for RNAIi-RNAIii. In all cases, the major charge state is 6-. The difference in charge state distribution can come from the number of bases of each system: 36 for TAR-R06 and 40 RNAIi-RNAIii. When the hairpin or the complex is bigger, there are more chances to have a higher charge state. When magnesium is added to the solution, the intensities of the hairpins decrease, for both KC tested. Also, the intensity of the major charge state of the KC, i.e 6-, is not decreasing as much as the monomer signals do when the concentration of  $\text{Mg}(\text{OAc})_2$  is increasing. On the contrary, the total peak integral looks larger, as there are more adducts. Concerning the lower charge state of the KC, for  $[\text{TAR-R06}]^{5-}$  at high concentration of magnesium, this peak seems lost in the background, which increases. The intensity of  $[\text{RNAIi-RNAIii}]^{7-}$  is however clearly higher with magnesium than without.

To quantify our observations, we calculated the integrals of all charge states detected of the KC. We calculated then the ratio of the calculated integrals over the integral of our internal reference which is  $\text{dT}_6$ . The results are shown on Figure 42.

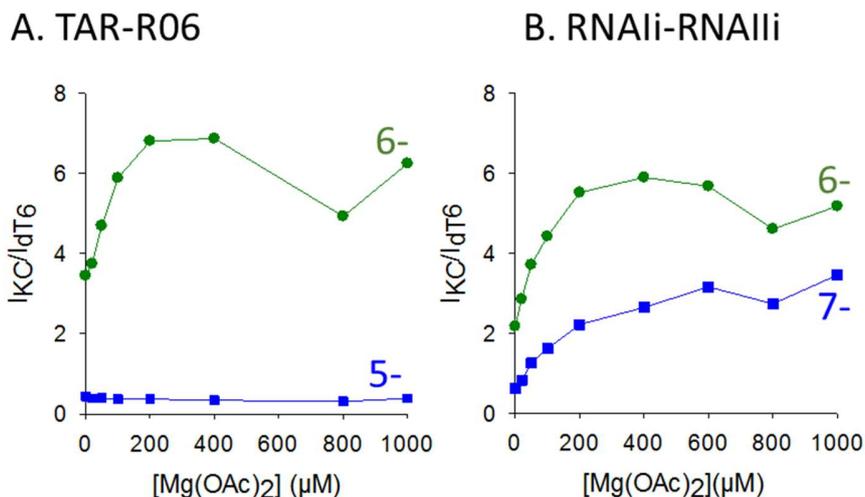


Figure 42: Relative integrals of the different charge state detected for A) TAR-R06 kissing complex and B) RNAIi-RNAIii kissing complex.

Increasing the concentration of magnesium has the same effect on  $[\text{TAR-R06}]^{6-}$ ,  $[\text{RNAIi-RNAIii}]^{6-}$  and  $[\text{RNAIi-RNAIii}]^{7-}$ : the relative intensity of each of these charge states is increasing all along the titration, until it reaches a plateau at 200-400  $\mu\text{M}$  of  $\text{Mg}(\text{OAc})_2$ . The TAR-R06 kissing complex 5- signal is not exploitable quantitatively. For RNAIi-RNAIii, the 6- is about 2-fold higher than the 7-, and both seem exploitable.

Next, we show the relative signal percentage of both hairpin monomers and the KC. To do that, all observed charge states were summed to calculate the peak area of each species. Then, we calculate the percentage of each species with regard to the total ion signal. The results are presented in Figure 43.

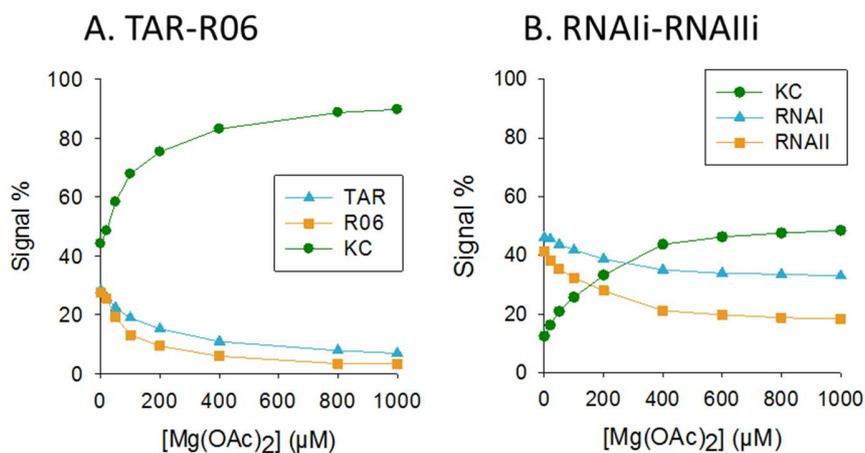


Figure 43: Signal percentage of each species in function of the  $\text{Mg}(\text{OAc})_2$  concentration for A) TAR-R06 and B) RNAIi-RNAIii.

For both models, the KC total signal increases and the hairpin's signal decreases along the titration. In ammonium acetate alone (0  $\mu\text{M}$  of  $\text{Mg}(\text{OAc})_2$ ), TAR-R06 KC already represents 44% of the signal, whereas the RNAIi-RNAIii KC represents 13% of the signal. This result have to be compared with UV meting experiments (chapter 1.X.2.1). The experiments are performed at room temperature which is around 23°C. Based on the results of UV melting,  $T_{\text{mTAR-R06}} \approx 22^\circ\text{C}$  which means that at room temperature most of the complexes are still formed. On the contrary,  $T_{\text{mRNAIi-RNAIii}} \leq 20^\circ$  which means that at 23°C the complexes have begun to unfold. At higher concentration of magnesium, the KC signal increases, up to 80% for TAR-R06 and 50% for RNAIi-RNAIii. This means that magnesium favors the formation of KC. Note however that, to relate the peak areas to concentrations, we will have to determine the response factors (see Section XII.2.1).

The formation of a KC is based on the following equilibrium:



Using mass spectrometry, we can identify all the partners involved in this equilibrium, and determine their respective proportions. Several hypotheses can be studied: firstly, maybe magnesium does not have the same affinity for both kissing complexes, secondly the different species may not respond in the same way to the ionization process and that they have different response factors.

When the concentration of magnesium increases, the number of adducts also increases. The distributions differ for both systems, and differ from charge state to charge state. At 200  $\mu\text{M}$  of  $\text{Mg}(\text{OAc})_2$ , there is no more KC without magnesium for RNAIi-RNAIii, whereas for TAR-R06 we still have this peak at 1 mM. Here several hypotheses can be made. First maybe those two systems have a different number of specific cations and/or these cations have a different affinity for each system. In the following section, we will investigate the number of specific  $\text{Mg}^{2+}$  binding to each KC.

#### *XI.2.2. Adduct cleaning: TAR-R06 vs RNAIi-RNAIii*

In order to understand if the observed  $\text{Mg}^{2+}$  adducts are specific to the kissing complex motif or not (or how many are specific), we used a mathematical treatment which was first described by the Klassen's group for proteins<sup>27</sup> and then adapted to G-quadruplex nucleic acids by our group<sup>12</sup>. Briefly, the non-specific adducts distribution is determined, at each magnesium concentration, using a reference compound to which the cation of interest should not bind specifically. This intensity distribution is then subtracted from the total  $\text{Mg}^{2+}$ -bound distribution of the corresponding charge state of the molecule of interest. More details can be found in the materials and methods section VII.3.2. Thanks to this method

we can deduce the specific distribution of magnesium for each point of the titration of the sequence of interest by magnesium.

### Specific vs non-specific and choice of the reference sequence

An important concept before discussing the choice of the reference sequence is to define what is specific and what is not. Specific  $Mg^{2+}$  ions are magnesium ions which are binding specifically *to the motif of interest*, here the kissing loop motif. Non-specific adducts are binding presumably on the external phosphates on low-affinity binding sites. These non-specific cations can come from the solution, but can also be formed during the ESI process. To subtract these two types of non-specific adducts we need to choose a control sequence. In our case, the control sequence is a canonical duplex RNA which has the same length and composition as our motifs of interest. It forms base pairs, but not via a kissing loop motif.

Two web programs were used to conceive the duplex sequences before ordering them. First, Zbio.net<sup>28</sup> will generate sequences by mixing initial sequences, i.e TAR and R06 and RNAIi and RNAlli. The RNAfold webserver (v.2.4.3), part of the ViennaRNA suite<sup>29</sup> from the Institute for Theoretical Chemistry (University of Vienna), is then used to check if the complex formed is a duplex and if this duplex is stable (calculation of free energy  $\Delta G^\circ$ ). The sequences created are the duplex 17R1-17R2 and 20R1-20R2 respectively for TAR-R06 and RNAIi-RNAlli (Table 3).

RNA reference duplex for TAR-R06	17R1	5'- GGA GCU CCC AGA CGA CC -3'
	17R2	5'- GGU CGU CUG GGA GCU CC -3'
RNA reference duplex for RNAIi-RNAlli	20R1	5'- GUG AGC UCC CAG ACG ACC UG -3'
	20R2	5'- CAG GUC GUC UGG GAG CUC AC - 3'

Table 3: Sequences of the RNA duplexes used as references for the adduct cleaning procedure.

### Comparison of the mass spectra of the kissing complexes and the duplex references

To apply the mathematical treatment, we first do titration of each kissing complexes and duplexes of reference by  $Mg(OAc)_2$ . Raw spectra for all sequences (kissing complexes and reference duplexes) are on Figure 44 .

One can follow visually the decrease of all hairpins intensities. For the KC 6- and Duplexes 6-, we can clearly see the increase of the number of magnesium adducts along the titration. A difference between the distribution of magnesium adducts is nevertheless seen when comparing  $[KC]^{6-}$  with  $[Duplexes]^{6-}$ .

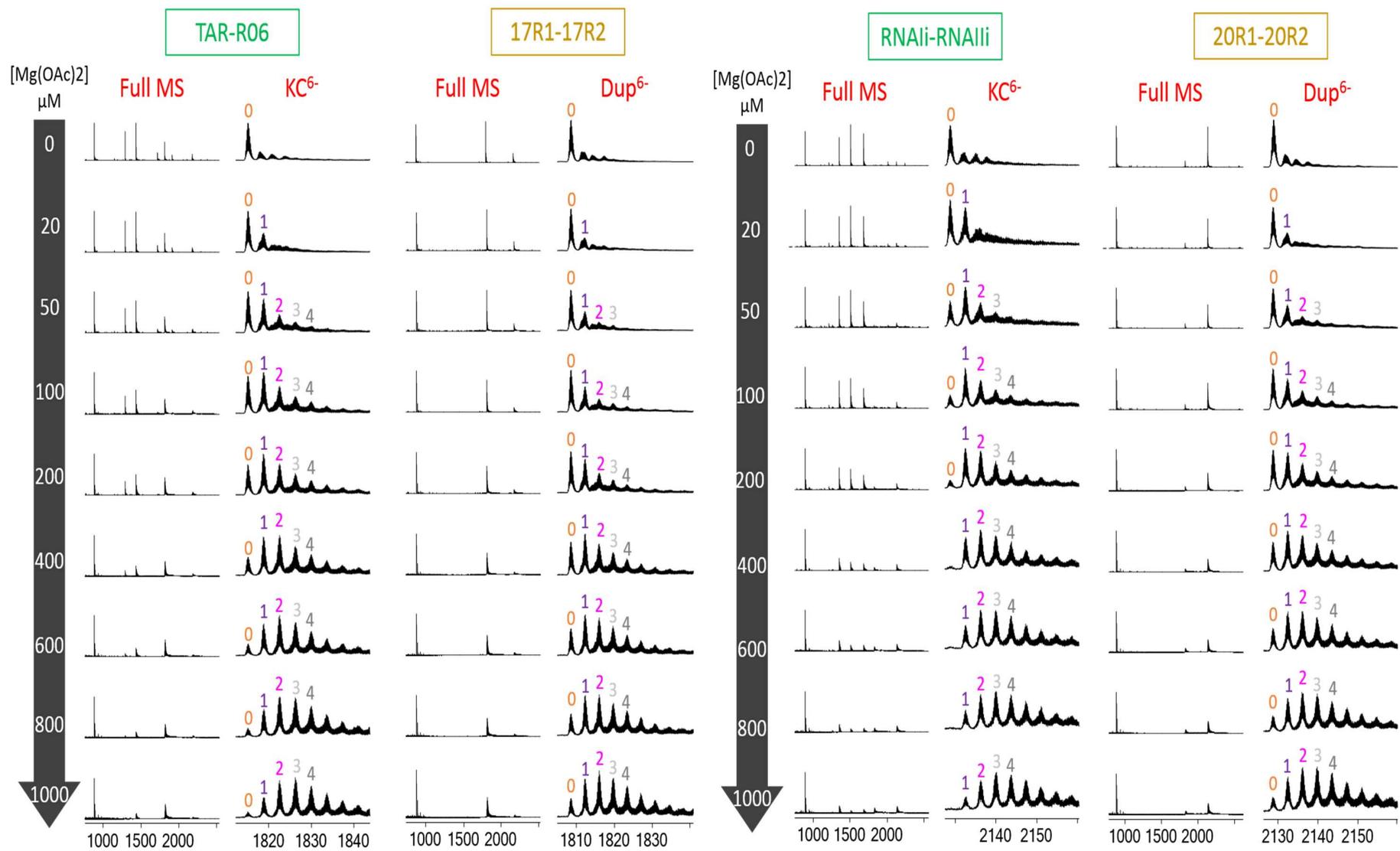


Figure 44 : Mass spectrometry titrations of 2 μM dT<sub>6</sub>, 10 μM of each complementary strands or hairpin, 150 mM NH<sub>4</sub>OAc and an increasing concentration of Mg(OAc)<sub>2</sub>. Full MS spectra and zoom on the KC/duplex major charge state. These data are used for adduct cleaning.

## After data treatment

The analysis is done on all charge states of the KC, i.e. 6- and 5- for TAR-R06 and 6- and 7- for RNAIi-RNAIi. The result of the adduct cleaning should be the same for each charge state of a species. For the mathematical treatment, 4 non-specific adducts are taken into account. The resulting graphs, after data treatment, show the normalized specific intensity of each adducts (here 4) as a function of the concentration of  $\text{Mg}(\text{OAc})_2$  (Figure 45).

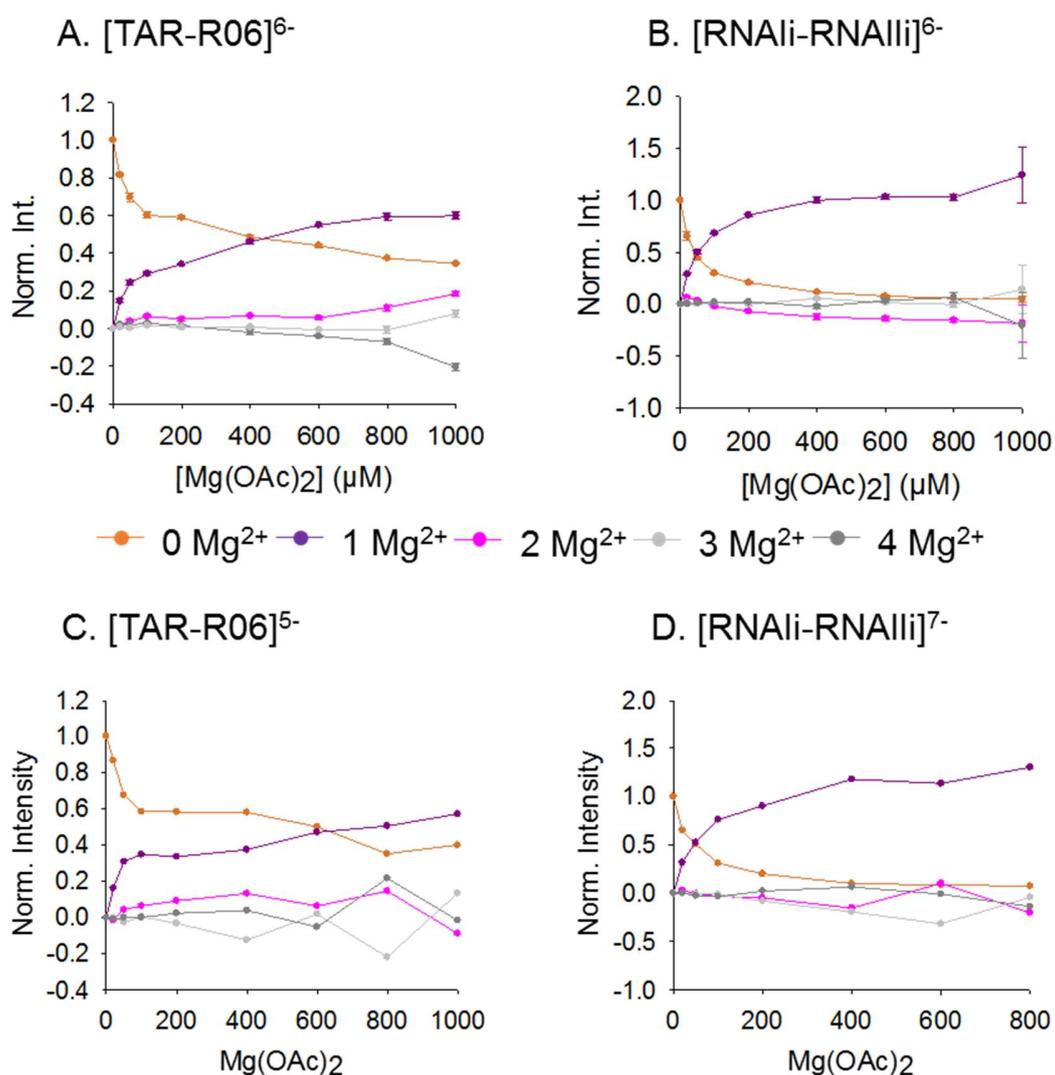


Figure 45: Distribution of specific  $\text{Mg}^{2+}$  distribution after data treatment for: A) [TAR-R06]<sup>6-</sup>, B) [TAR-R06]<sup>5-</sup>, C) [RNAIi-RNAIi]<sup>6-</sup> and D) [RNAIi-RNAIi]<sup>7-</sup>. Errors bars on graph A and B shows the difference between triplicates. No error bars are present on graph C and D. For [TAR-R06]<sup>5-</sup> the intensity was too low. For [RNAIi-RNAIi]<sup>7-</sup> the intensity was for some replicates too low.

On Figure 45A., we have the results for [TAR-R06]<sup>6</sup>. As the concentration of magnesium acetate increases, the proportion of 1 Mg<sup>2+</sup> increases from 0.2 at 20 μM up to 0.6 at 1 mM. This means that at least one magnesium ion is specific to the TAR-R06 loop-loop motif. From 600 μM Mg(OAc)<sub>2</sub>, we see that a second specific Mg<sup>2+</sup> is appearing but in low proportion (0.2 at 1 mM). Looking at the 3- and 4-Mg<sup>2+</sup> adducts, we see that there is no particular increase of these two curves. Nevertheless, it has to be noticed that the proportion of 4 Mg<sup>2+</sup> is going to negative values and seems to be compensated by an increase of the 3 Mg<sup>2+</sup>. Those results could be explained by the fact that more magnesium adducts may be present on the RNA reference duplex in comparison to the kissing complex. It is also possible that the actual error at 1000 μM Mg<sup>2+</sup> is larger than the error bar from the triplicate suggests. The results for [TAR-R06]<sup>5</sup> look similar to those obtained for [TAR-R06]<sup>6</sup>. However, as the intensity of this charge state is quite low it is difficult to extract accurately the area of each peak corresponding to the 5- distribution.

In 2000, Ducongé *et al.*<sup>30</sup> used thermal denaturation and enthalpy-based calculations to determine the number of Mg<sup>2+</sup> cations (called ΔMg<sup>2+</sup>) that bind upon the formation of the kissing complex TAR-R06<sub>24</sub>. The enthalpy change (ΔH) of kissing complex formation was deduced from the dependence of the  $T_m$  from the total RNA concentration, then the number of magnesium binding upon the formation of the kissing complex is determined from the Mg<sup>2+</sup> concentration dependence of the  $T_m$  according respectively to equation (21) and (22).

$$\frac{1}{T_m} = \frac{R}{\Delta H} \ln[RNA]_{tot} + \frac{\Delta S - R \ln 4}{\Delta H} \quad (21)$$

$$\frac{\partial(\frac{1}{T_m})}{\partial \ln[Mg^{2+}]} = \Delta Mg^{2+} \frac{R}{\Delta H} \quad (22)$$

By plotting  $T_m$  as a function of [RNA]<sub>tot</sub> at a certain [Mg<sup>2+</sup>], they have access to ΔH and by plotting  $T_m$  as a function of [Mg<sup>2+</sup>] at a certain [RNA]<sub>tot</sub>, they have access to ΔMg<sup>2+</sup>. R06<sub>24</sub> is the R06 sequence with 3 additional base pairs in the stem. They performed the experiments at 3 mM of MgCl<sub>2</sub> to determine ΔH, and then found ΔMg<sup>2+</sup> = 1.7 ± 0.1. These results agree with ours, even though the concentration used in our study are lower.

Note that on the mass spectrometry side, we do not need to reach binding site saturation in order to reveal the stoichiometries, as opposed to many other biophysical techniques. On the melting analysis side, the determination is rather indirect. We should also note that the extracted numbers of magnesium ions

bound is not defined in the same way. In the solution melting analysis,  $\Delta Mg^{2+}$  is defined between the kissing complex and the separated hairpins, whereas in the mass spectrometry analysis the specific magnesium uptake is defined between the kissing complex and a reference duplex.

The results obtained for  $[RNAIi-RNAIi]^{6-}$  (Figure 45, B) show that when the concentration of  $Mg(OAc)_2$  increases, the proportion of 1 specific  $Mg^{2+}$  ion increases too, until it reaches a plateau with a proportion equal to 1. From these results, it seems that at high concentration of  $Mg^{2+}$  almost 100% of the kissing complex detected binds to a single magnesium ion. At 1 mM, the error bars are larger; the noise is largest at high concentrations of magnesium. The results obtained for  $[RNAIi-RNAIi]^{6-}$  and  $[RNAIi-RNAIi]^{7-}$  are in good agreement: only one magnesium ion seems specific to the RNAIi-RNAIi kissing loop motif.

In 1995, Gregorian *et al*<sup>18</sup> used the same method as described for TAR-R06<sub>24</sub> by Ducongé *et al*<sup>30</sup>. They used  $\Delta H$  obtained at  $[Mg^{2+}] = 5$  mM. They found that the formation of the kissing complex is coupled to the uptake of two magnesium ions. Our results show that one magnesium is specific to RNAIi-RNAIi, not two. We discussed above the difference between the two approaches. This difference in our results can also be due to the sample buffer. Indeed, in MS experiment the concentration of magnesium could not be increased to 5 mM and this may play a role in the number of magnesium detected. Maybe several binding sites are co-existing but cannot be populated at the same time that is why by MS we only “see” a maximum of one magnesium ion bound at a time.

### *XI.2.3. Influence of the choice of the reference duplex*

With this method the number of specific magnesium is compared to an RNA duplex of reference, which is made by shuffling the sequence of interest. Here we wanted to check the quantitative robustness of the method using another reference. This reference is still an RNA duplex of the same length and composition than the sequence of interest. For this test, we used 17R3-17R4 and 20R3-20R4 RNA duplexes as references for respectively TAR-R06 and RNAIi-RNAIi (Table 4).

New RNA reference duplex for TAR-R06	17R3	5'- GAG CCG CAA AUG CCC CG -3'
	17R4	5'- CGG GGC AUU UGC GGC UC -3'
New RNA reference duplex for RNAIi-RNAIi	20R3	5'- GGA GCC GCA AAU GCC CCG UG -3'
	20R4	5'- CAC GGG GCA UUU GCG GCU CC -3'

*Table 4: Sequences of the new RNA duplexes used as references for the adduct cleaning procedure.*

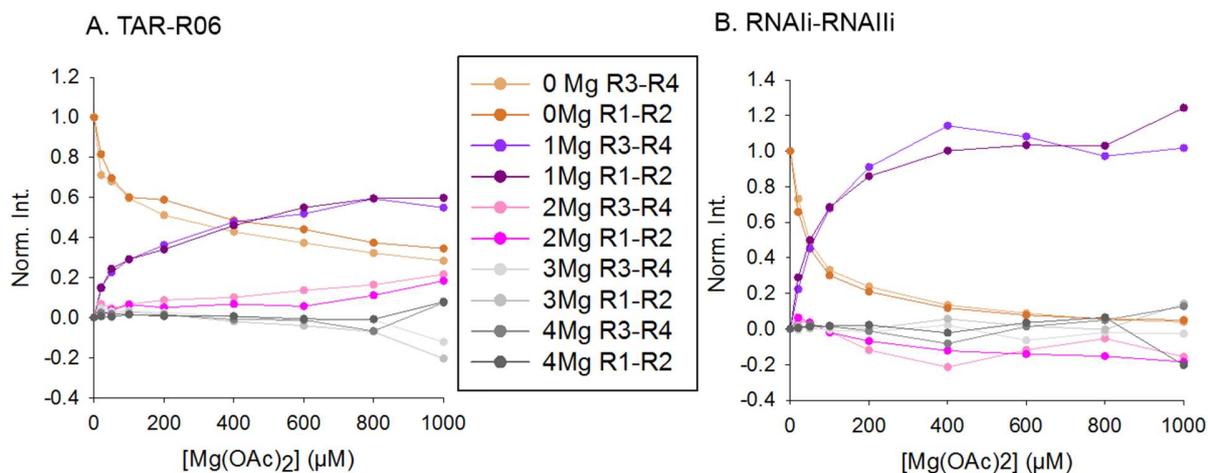


Figure 46: Comparison of the specific Mg<sup>2+</sup> distribution determined using two different references in function of Mg(OAc)<sub>2</sub> concentrations for A) [TAR-R06]<sup>6-</sup> with 17R1-17R2 or 17R3-17R4 as reference and B) [RNAIi-RNAIi]<sup>6-</sup> with 20R1-20R2 or 20R3-20R4 as reference.

Figure 46 summarizes the results obtained when comparing the different reference sequences. Whether for [TAR-R06]<sup>6-</sup> or [RNAIi-RNAIi]<sup>6-</sup>, whatever the reference sequence used, the same results are obtained, both qualitatively and quantitatively. The curves follow the same trend for each magnesium, from 0 to 4. These results confirm that an RNA duplex which have the same length and composition than our sequence studied is a good reference for the adduct cleaning methodology applied to kissing complexes.

#### Xi.2.4. Comparison between cations: Mg<sup>2+</sup> vs Mn<sup>2+</sup>

Another interesting point is to compare divalent cations and see if the same number of cations bind. For this analysis, we decided to compare magnesium to manganese for several reasons. First, manganese is found as traces in bacteria and mammalian cells. It has been shown since the 1970's that some enzymes can bind specifically to manganese<sup>31</sup>. Also, in several biochemical and biophysical studies magnesium is replaced by manganese because of their close ionic radius (0.72 Å for Mg<sup>2+</sup> and 0.83 Å for Mn<sup>2+</sup> for a coordination number of 6)<sup>32</sup> and it seems that they have similar coordination preferences to Oxygen and Nitrogen<sup>33</sup>, even if Mg<sup>2+</sup> preferentially binds to O. Secondly, during sample preparation for MS experiment we do desalting to remove as much as possible sodium which is present. Nevertheless, we cannot completely get rid of Na<sup>+</sup> adducts. When the resolution and sensitivity of the instrument available in the lab, peaks of Na<sup>+</sup> and Mg<sup>2+</sup> appear unfortunately at the same position. Using manganese, which have an atomic weight of 55,<sup>34</sup> it is easier to distinguish the distribution caused by manganese from sodium one.

To compare both divalent cations, we performed the adduct cleaning experiment on titrations of TAR-R06 and RNAIi-RNAIi by Mn(OAc)<sub>2</sub>. We used the same concentrations as with Mg(OAc)<sub>2</sub>, i.e from 0 to 1 mM.

The references used were 17R1-17R2 and 20R1-20R2 for TAR-R06 and RNAIi-RNAIii respectively. The analysis is done on the 6- charge state for both models. The results are shown as bar graphs on Figure 47. Results as line and scatter plot can be found on appendix A1.3.

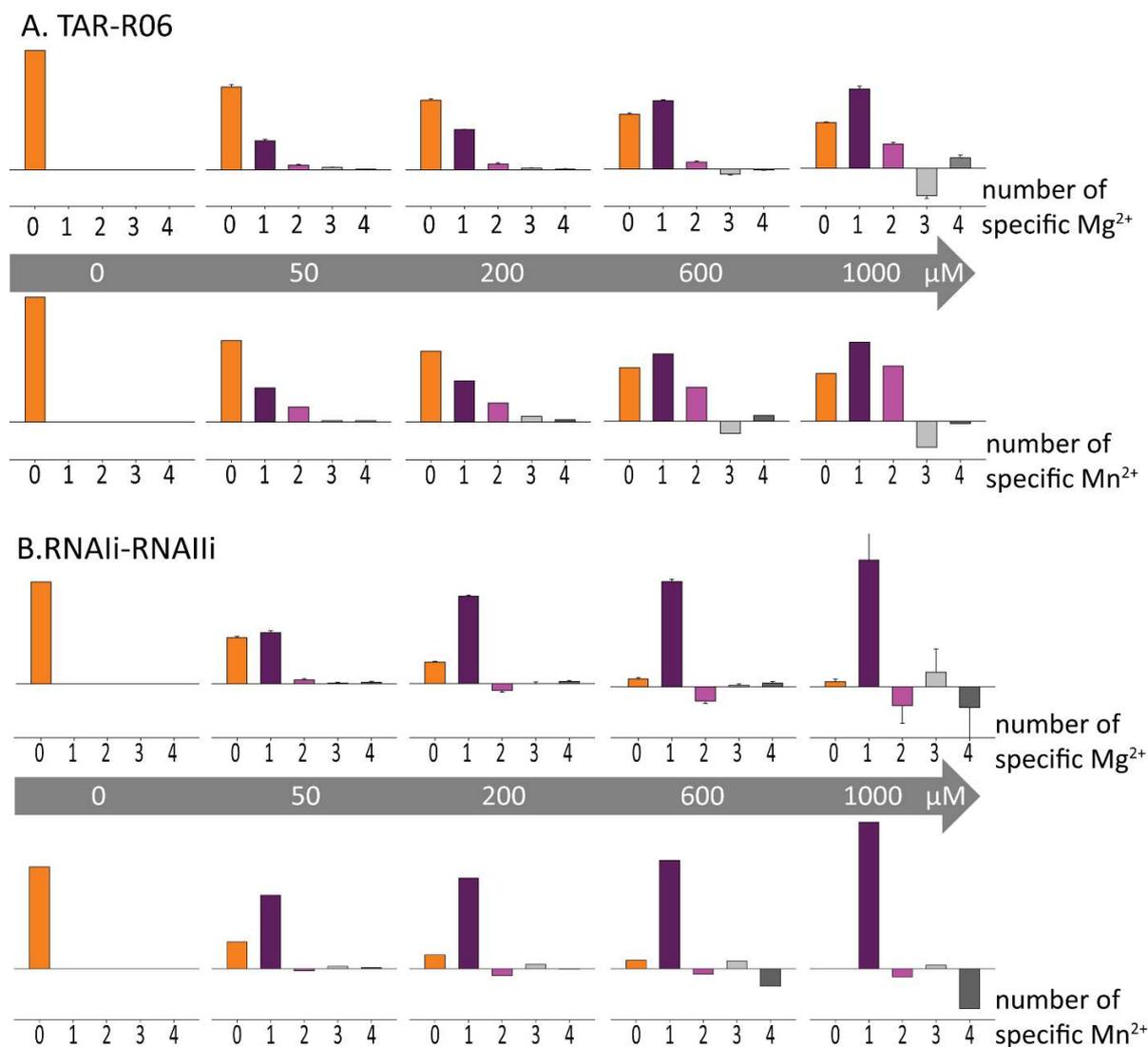


Figure 47: Specific distribution of  $Mn^{2+}$  compared to the one of  $Mg^{2+}$ . A) Results obtained for  $[TAR-R06]^{6-}$  with on the top the  $Mg^{2+}$  specific distribution and at the bottom the  $Mn^{2+}$  specific distribution. B) Results obtained for  $[RNAIi-RNAIii]^{6-}$  with on the top the  $Mg^{2+}$  specific distribution and at the bottom the  $Mn^{2+}$  specific distribution.

The bar charts show the evolution of the proportions of the specific cation distribution. Figure 47A presents the results for  $[TAR-R06]^{6-}$  with the  $Mg^{2+}$  distribution (top) and  $Mn^{2+}$  distribution (bottom). The numbers of specific adducts are the same for both cations (2 for TAR-R06 and 1 for RNAIi-RNAIii). Also, the proportions of one specific cation for  $Mg^{2+}$  and  $Mn^{2+}$  are similar. Concerning the second specific cation

of TAR-R06, there is a clear difference in proportions. For  $Mn^{2+}$  the second specific cation proportion increases more significantly as the concentration of  $Mn(OAc)_2$  increases. So, either the two cations have different affinities for the kissing complex, or the binding sites for  $Mg^{2+}$  and  $Mn^{2+}$  are different.

The results for  $[RNAIi-RNAIi]^{6-}$  are shown on Figure 47B. Only one specific cation is bound, and the proportions increase faster for  $Mn^{2+}$  than for  $Mg^{2+}$ . Here also, the hypothesis that the two cations have different affinities for the KC, or different binding sites have to be considered.

### XI.3. Discussion

First, we have seen that in typical native MS buffer conditions (only ammonium acetate), TAR-R06 kissing complex as well as RNAIi-RNAIi kissing complex are formed. Magnesium is thus not absolutely necessary for kissing complex formation.

In this chapter, we have shown that mass spectrometry is a powerful tool to determine directly the stoichiometries of divalent ion binding to RNA. However, due to the detection of nonspecific adducts superimposed to the specific ones, a control sequence and a mathematical subtraction is necessary to have information on the number of specifically bound cations onto kissing complexes. It is then possible to compare sequences or cations. The two kissing complexes used as models seem to behave differently with respect to magnesium.

- 1) First, the amount of kissing complexes detected is influenced by the concentration of magnesium added. Although magnesium is not necessary for kissing complex formation, it helps displacing the equilibrium (1) towards the formation of KC. In that sense, magnesium is stabilizing kissing complexes. In chapter 1.XII, we will explain the influence of magnesium on kissing complexes dissociation constant by dissecting the different contributions to the global equilibrium.
- 2) The two kissing complexes do not have the same number of specifically bound magnesium ions. TAR-R06 kissing complex is binding to two specific magnesium, but these two  $Mg^{2+}$  do not have the same affinity. These results agree with the literature. For RNAIi-RNAIi, this kissing complex binds to one specific magnesium. Unfortunately, just based on these two kissing complexes, no rules can be drawn. The differences in the number of cations specifically bound could be due to the number of bases involved in the loop-loop interaction, due to the length of the sequences and/or the composition of the hairpins and the type of bases presents in the loops. It is also important to remember here that the specificity of the magnesium binding is defined by using RNA duplexes as references. Also, it does not mean that additional magnesium ions are not bound

in solution. It just means that their abundance is not significantly higher in the kissing complex compared to a duplex containing the same bases. The results have shown that when the RNA duplex has the same length and the same composition, the exact sequence choice does not have an influence on the results. In chapter 1.XIII. we will use tandem mass spectrometry as a way to obtain details on  $Mg^{2+}$  location.

- 3) Using mass spectrometry, we compared  $Mn^{2+}$  and  $Mg^{2+}$ . The results show that the same stoichiometries of specific cation form for TAR-R06 and RNAIi-RNAIii. However, it seems that  $Mg^{2+}$  and  $Mn^{2+}$  do not have the same affinity. We cannot rule out the possibility that binding sites differ. Indeed, it has been shown that  $Mn^{2+}$  affinity for N7 is higher than  $Mg^{2+}$  <sup>(33,35)</sup> and that N7 atoms is not always the preferred binding site for magnesium to RNA or DNA structures<sup>36,35</sup>. Nevertheless, based on the stoichiometries, manganese can be considered as a good substitute for magnesium.

## XII. The extent of kissing complex stabilization by Mg<sup>2+</sup> is correlated to the cation affinity, due to a displacement of equilibria

### XII.1. Introduction

In this section we will investigate the effect of magnesium on kissing complex stabilization. Kissing complex formation in the absence of magnesium is described by the equilibrium below:



With the equilibrium constant of formation ( $K_{fKC}$ ) and of dissociation ( $K_{dKC}$ ) of KC (for the next equilibria we will just show the equation related to the dissociation):

$$K_{fKC} = \frac{1}{K_{dKC}} = \frac{[\text{KC}]}{[\text{HP1}][\text{HP2}]} \quad (23b)$$

The presence of magnesium displaces this equilibrium towards the formation of higher total amounts of kissing complex (comprising magnesium-free and magnesium-bound forms). In presence of an amount  $x$  of Mg, the apparent kissing complex formation equilibrium is:



With the apparent equilibrium constant of formation of KC:

$$K_{fKC,app} = \frac{1}{K_{dKC,app}} = \frac{[\text{KC}]_{total}}{[\text{HP1}]_{total}[\text{HP2}]_{total}} \quad (24b)$$

We call it “apparent” because a true chemical equilibrium does not change with the environment. Some textbooks call these constant the “conditional” equilibrium constants, because they depend on the experimental *conditions*.

We saw in the previous section that magnesium binds to the kissing complex. It thus consumes the product of reaction (23).

We can define new binding equilibria:



With consecutive association constants of Mg<sup>2+</sup> to the KC:

$$K_{a,Mg1} = \frac{1}{K_{d,Mg1}} = \frac{[KC \bullet Mg]}{[KC][Mg^{2+}]} \quad (25a)$$

$$K_{a,Mg2} = \frac{1}{K_{d,Mg2}} = \frac{[KC \bullet 2Mg]}{[KC \bullet Mg][Mg^{2+}]} \quad (26a)$$

Note that the affinity that we define here is the affinity by stoichiometry, not by binding site. We can also define:

$$[KC]_{total} = [KC] + [KC \bullet Mg] + [KC \bullet 2Mg] \quad (27)$$

Hereafter, we will discuss the dissociation constants, which are more commonly used in biochemistry because they directly indicate the concentrations at which the complexes are formed. Our objective is to dissect the apparent  $K_{dKC,app}$  in terms of  $K_{dKC}$ ,  $K_{dMg1}$  and  $K_{dMg2}$ . For that, we have to determine the absolute quantities of each species (hairpins, KC, KC•Mg, KC•2Mg and free  $Mg^{2+}$ ) from the mass spectra.

## XII.2. Results and discussion

### XII.2.1. Determination of $K_{dKC,app}$

Mass spectrometry has already been used to determine binding constants and stoichiometries between RNA sequences and ligands.<sup>37,38</sup> Most past studies assumed that the response factors of the different species present in solution, i.e. RNA and RNA+L, are equal. The response factor is the proportionality factor between the concentration of the species in solution and the intensity of the respective ions in the mass spectra. To determine the concentrations and then have access to binding constant values, we need to know if we can assume that the response factors are equal. A way to check this is to plot the ratio of the total RNA intensities over the intensity of  $dT_6$ , as a function of the concentration of  $Mg(OAc)_2$ . Hexathymine serves as an internal intensity standard. It does not interact and interfere with the other molecules present in solution. The relative proportions of hairpins and KC vary along the titration, but if their response factors were equal, the total signal relative to  $dT_6$  would be constant. Results are presented on Figure 48.

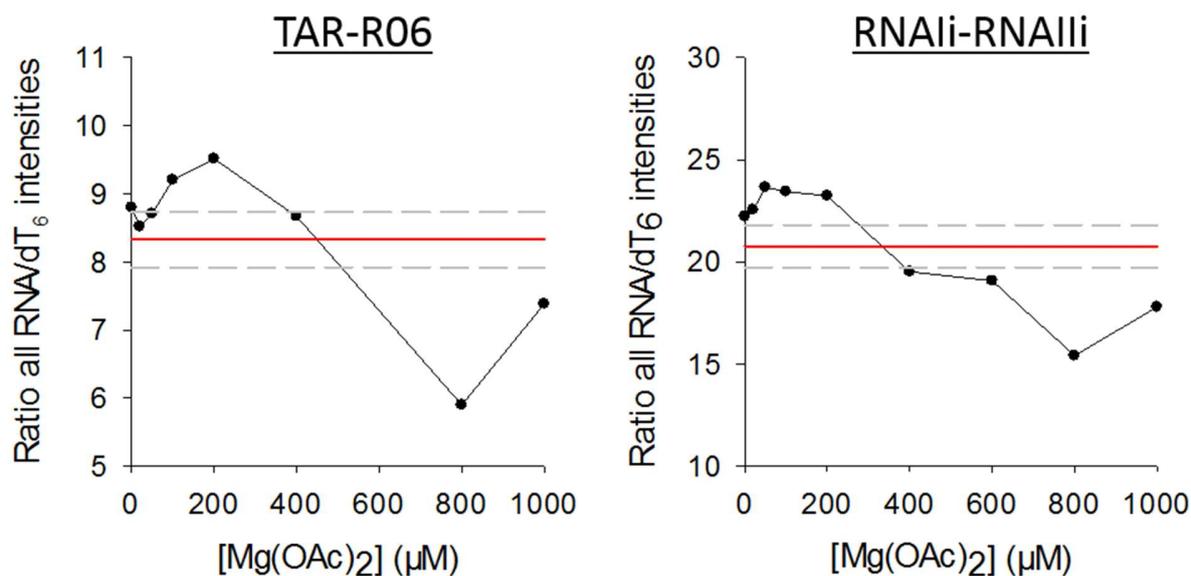


Figure 48: Ratio between the total RNA signal over dT<sub>6</sub> in function of Mg(OAc)<sub>2</sub> concentration for TAR-R06 and RNAIi-RNAIi. In black the experimental points, in red the average and in grey  $\pm 5\%$ .

For both TAR-R06 and RNAIi-RNAIi titrations by Mg(OAc)<sub>2</sub>, the signal is decreasing above 400 μM Mg(OAc)<sub>2</sub>. So, the KC and hairpins likely have different response factors, and we need to determine them in order to determine the concentrations from the mass spectral intensities.

The method used to determine response factors has been described by Gabelica et al<sup>39</sup> and is explained on the material and method section VII.3.3.1. Briefly, by doing titrations of the molecule of interest by its ligand and using a reference that does not interact with either partner, ratios between intensities will allow the calculations of response factors. Then concentrations can be determined.

We first tried to titrate the 1:1 mixture of both hairpins by Mg(OAc)<sub>2</sub>, to determine simultaneously the response factors of each of the RNA forms. However, it was not possible to obtain values. Indeed, too many species had to be considered: both hairpins and the KC but also each KC complex with magnesium. We decided then to do titrations by fixing the concentration of one of the two hairpin and increasing progressively the concentration of the other one. Here, the total signal of KC is integrated. We will thus have the response of KC<sub>total</sub>, and access to the apparent KC dissociation constant  $K_{dKC,app}$ . The titrations were carried out at different concentrations of magnesium. This methodology allows us to determine the response factors of one hairpin per titration experiment, and of the KC.

## TAR-R06

Titration curves are performed by fixing either the concentration of one hairpin (acting as receptor) at 10  $\mu\text{M}$ , and increasing the concentration of the complementary hairpin (acting as ligand) from 0 to 2 equivalents. Distinct titrations were carried out with TAR as receptor and R06 as ligand (to obtain the response factor of TAR and  $\text{KC}_{\text{total}}$ ), then with R06 as receptor and TAR as ligand (to obtain the response factor of R06 and  $\text{KC}_{\text{total}}$ ). The titrations are performed in 150 mM  $\text{NH}_4\text{OAc}$ , 2  $\mu\text{M}$  of  $\text{dT}_6$  and 0, 50, 200 or 600  $\mu\text{M}$   $\text{Mg}(\text{OAc})_2$ . Here we can reasonably assume that the response factors will not change when the complementary hairpin is added to the solution. Response factors are calculated and then the correction is applied to obtain the concentrations. Finally, the concentrations of total KC and receptor are fitted as a function of the ligand concentration using the DynaFit software (4.05.103 BioKin Ltd Watertown, MA, USA).<sup>40</sup> The equilibrium used for the fitting is:



This equilibrium is used whatever the concentration of magnesium in the solution. This method allows us to have information on the total quantity of KC.

Response factors and  $K_{\text{dTAR-R06,app}}$  are presented in

[Mg(OAc) <sub>2</sub> ] ( $\mu\text{M}$ )	RfTAR/RfT6	RfR06/RfT6	RfKCtotal/RfT6	RfHP/RfKC (HP = TAR or R06)	KdTAR- R06. <sub>app</sub> (*10-6 M)	Average KdTAR- R06. <sub>app</sub> (*10-6 M)	Std error
0	1.85		1.58	1.17	1.65	<b>2.42</b>	0.25
0	2.16		1.73	1.25	2.38		
0		1.89	1.39	1.37	2.62		
0		2.07	1.63	1.27	3.01		
50	2.51		1.53	1.64	0.79	<b>1</b>	0.08
50	2.48		1.80	1.38	1.08		
50		2.57	1.79	1.44	1.12		
200	2.42		1.53	1.57	0.36	<b>0.46</b>	0.07
200		2.47	1.77	1.39	0.56		
600	2.13		1.37	1.55	0.07	<b>0.17</b>	0.07
600		2.07	1.54	1.34	0.27		

Table 5. Concentrations were calculating to plot titration curves (Appendix A1.4) and fit them. The concentration calculations are done in two steps:

1) Calculation of  $[HP]_{corr}$  and  $[KC]_{corr}$  by applying the response factor correction.

$$[HP]_{corr} = \frac{Rf_{HP}}{Rf_{T6}} \times \frac{A_{T6}}{A_{HP}} \quad (29)$$

$$[KC]_{corr} = \frac{Rf_{KC}}{Rf_{T6}} \times \frac{A_{T6}}{A_{KC}} \quad (30)$$

Where  $A_{HP}$ ,  $A_{KC}$  and  $A_{T6}$  are respectively the area of the peaks corresponding to the HP, KC and  $dT_6$ .

2) Normalization by  $[HP]_{tot}$  which is the initial fixed concentration of one of the hairpin (10  $\mu\text{M}$  in our experiments).

$$[HP] = \frac{[HP]_{corr}}{[HP]_{corr} + [KC]_{corr}} \cdot [HP]_{tot} \quad (31)$$

$$[KC] = \frac{[KC]_{corr}}{[HP]_{corr} + [KC]_{corr}} \cdot [HP]_{tot} \quad (32)$$

$[\text{Mg}(\text{OAc})_2]$ ( $\mu\text{M}$ )	$Rf_{TAR}/Rf_{T6}$	$Rf_{R06}/Rf_{T6}$	$Rf_{KCtotal}/Rf_{T6}$	$Rf_{HP}/Rf_{KC}$ (HP = TAR or R06)	$K_{dTAR-R06.app}$ (* $10^{-6}$ M)	Average $K_{dTAR-R06.app}$ (* $10^{-6}$ M)	Std error
0	1.85		1.58	1.17	1.65	<b>2.42</b>	0.25
0	2.16		1.73	1.25	2.38		
0		1.89	1.39	1.37	2.62		
0		2.07	1.63	1.27	3.01		
50	2.51		1.53	1.64	0.79	<b>1</b>	0.08
50	2.48		1.80	1.38	1.08		
50		2.57	1.79	1.44	1.12		
200	2.42		1.53	1.57	0.36	<b>0.46</b>	0.07
200		2.47	1.77	1.39	0.56		
600	2.13		1.37	1.55	0.07	<b>0.17</b>	0.07
600		2.07	1.54	1.34	0.27		

Table 5 : Summary of determined response factors and  $K_{dTAR-R06.app}$  obtained after fitting for TAR-R06. Several replicates are done. The standard error is the standard error of the mean.

The ratio between the response factors obtained ( $\frac{Rf_{HP}}{Rf_{KC}}$ ) presents a small variation of 3% between all the experiments. The ways ions are produced and transmitted are similar from one experiment to the other and the results are reproducible. If we compare the response factors obtained for TAR and R06, they are

similar that means we can assume that they are equal. Also, the response factors determined at each concentration of magnesium are varying of about 5% for the two monomers and 2% for the KC. Given that the abundance ratio of KC, KC•Mg and KC•2Mg are varying, it also suggests that the response factors of KC, KC•Mg and KC•2Mg are almost equal. We will thus use this assumption to calculate the concentrations of each, based on the relative proportions determined in the previous section.

The results show a decrease of  $K_{dTAR-R06,app}$  when magnesium is added to the solution. A decrease of 14-fold is observed between the  $K_{dTAR-R06,app}$  at 0  $\mu\text{M}$   $\text{Mg}(\text{OAc})_2$  and the one at 600  $\mu\text{M}$   $\text{Mg}(\text{OAc})_2$ . The affinity between the two hairpins is thus increased in presence of magnesium.

### RNAIi-RNAIIi

The same procedure is applied to RNAIi-RNAIIi. Titrations are performed by fixing the concentration of either RNAIi or RNAIIi, and titrating by the other hairpin in solutions containing 0, 50, 200 or 600  $\mu\text{M}$   $\text{Mg}(\text{OAc})_2$ . After correction, the concentrations of KC and receptor are fitted as a function of the concentration of ligand added. The equilibrium used to fit the data using the DynaFit software is defined as follows, the peak areas of all adducts on KC being summed:



Response factors and  $K_{d\text{RNAIi-RNAIIi},app}$  are presented on

[Mg(OAc) <sub>2</sub> ] ( $\mu\text{M}$ )	RfRNAI/RfT6	RfRNAII/RfT6	RfKC/RfT6	RfHP/RfKC	$K_{d\text{RNAIi-RNAIIi},app}$ (*10 <sup>-6</sup> M)	Average $K_{d\text{RNAIi-RNAIIi},app}$ (*10 <sup>-6</sup> M)	Std error
0	2.61		1.66	1.58	26.8	<b>21.3</b>	1.4
0	1.90		0.76	2.49	19.3		
0	1.89		0.86	2.21	21.2		
0		2.56	1.18	2.17	19.8		
0		1.73	0.76	2.29	19.2		
50	4.81		1.20	4.00	3.66	<b>8.6</b>	2
50	1.83		0.75	2.46	13.5		
50	2.48		1.00	2.48	9.55		
50		1.89	0.59	3.18	7.79		
200	3.57		1.42	2.51	2.23	<b>2.9</b>	0.5
200		1.68	0.54	3.14	3.58		
600	2.92		1.53	1.90	0.84	<b>0.96</b>	0.08

600		1.27	0.43	2.97	1.08		
-----	--	------	------	------	------	--	--

Table 6. Fitting results are presented on appendix A1.5.

[Mg(OAc) <sub>2</sub> ] ( $\mu$ M)	Rf <sub>RNAi</sub> /Rf <sub>T6</sub>	Rf <sub>RNAII</sub> /Rf <sub>T6</sub>	Rf <sub>KC</sub> /Rf <sub>T6</sub>	Rf <sub>HP</sub> /Rf <sub>KC</sub>	$K_{dRNAi-RNAII,app}$ (*10 <sup>-6</sup> M)	Average $K_{dRNAi-RNAII,app}$ (*10 <sup>-6</sup> M)	Std error
0	2.61		1.66	1.58	26.8	<b>21.3</b>	1.4
0	1.90		0.76	2.49	19.3		
0	1.89		0.86	2.21	21.2		
0		2.56	1.18	2.17	19.8		
0		1.73	0.76	2.29	19.2		
50	4.81		1.20	4.00	3.66	<b>8.6</b>	2
50	1.83		0.75	2.46	13.5		
50	2.48		1.00	2.48	9.55		
50		1.89	0.59	3.18	7.79		
200	3.57		1.42	2.51	2.23	<b>2.9</b>	0.5
200		1.68	0.54	3.14	3.58		
600	2.92		1.53	1.90	0.84	<b>0.96</b>	0.08
600		1.27	0.43	2.97	1.08		

Table 6: Summary of determined response factors and  $K_{dRNAi-RNAII,app}$  obtained after fitting for RNAi-RNAII. Several replicates are done.

Although the ratio between response factors of RNAs and dT<sub>6</sub> differed in the replica, the ratio between the response factors of monomers, i.e. RNAi or RNAII, and the one of the KC presents a standard deviation of  $\pm 0.096$  between the different experiments performed. Still, this tells us that the electrospray ionization conditions may have changed between the different experiments. Also, for several experiments made at the same concentration of magnesium, a standard deviation of  $\pm 0.081$  is observed, except at 600  $\mu$ M which is at 0.094. Magnesium may have also an impact on the ionization of the species present in solution. In addition, the response factors determined for RNAi and RNAII presents variations of around 10% between the different experiments done at different magnesium concentrations. This confirms that magnesium changes the response factors of the different species. These conclusions are also valuable for the KC.

The determination of  $K_{dRNAi-RNAII,app}$  shows that adding magnesium in the solution will decrease the equilibrium binding constant. As a result, the affinity between RNAi and RNAII is increasing due to the presence of magnesium. A decrease of 22-folds is observed from  $K_{dRNAi-RNAII,app}$  at 0  $\mu$ M of magnesium and the one at 600  $\mu$ M.

## Comparison TAR-R06 and RNAi-RNAIi

The comparison between the determined  $K_{Dglob}$  for TAR-R06 and RNAi-RNAIi is presented on Figure 49.

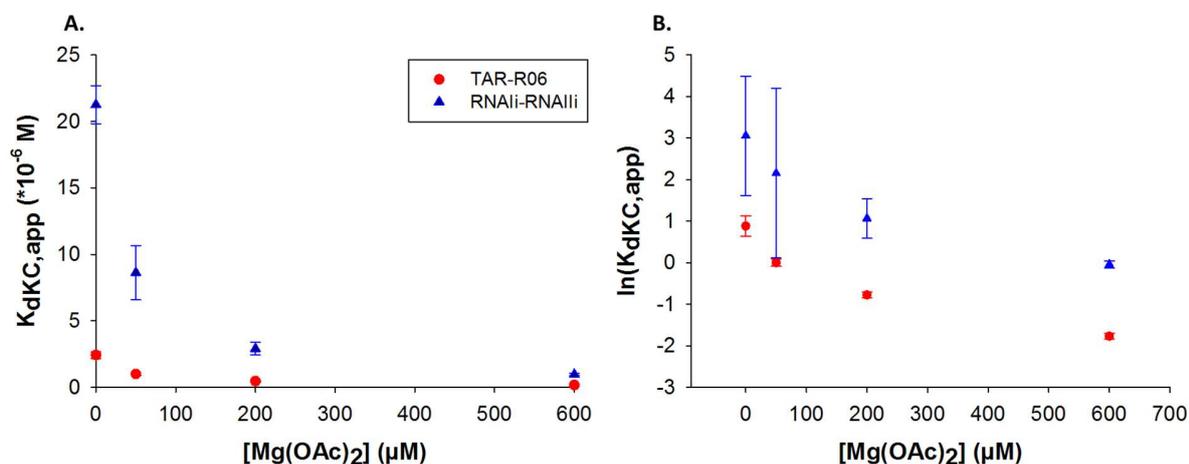


Figure 49: Comparison between the  $K_{dKC,app}$  determined for TAR-R06 (red filled circle) and RNAi-RNAIi (blue filled triangle) as a function of the magnesium acetate concentration.

For both systems when magnesium is added to the solution, a decrease of  $K_{dKC,app}$  is observed. The affinity for the two hairpins is increased. Without magnesium, the affinity for TAR and R06 is 10-folds higher than the one between RNAi-RNAIi.  $K_{dKC,app}$  for RNAi-RNAIi is decreased of about 22 folds compared to 14 folds for TAR-R06. All these results show clearly that RNAi-RNAIi is more magnesium dependent than TAR-R06. By calculating  $\ln(K_{dKC,app})$ , one can have an idea of the evolution of  $\Delta G^\circ$ . As  $\ln(K_{dKC,app})$  decreases,  $\Delta G^\circ$  is decreasing too meaning that the complexes are more and more stable.

### XII.2.2. Specific $Mg^{2+}$ binding constant

Now we want to determine the equilibrium binding constant of each specific magnesium on these kissing complexes. By combining the data of adduct cleaning (proportion of each specific adduct in  $KC_{total}$ ) and of the titration experiments (amount of  $KC_{total}$ ), it is possible to reconstruct the evolution of the concentrations of the different species with specific magnesium as a function of the concentration of magnesium. With the adduct cleaning data we have access to the proportion of specific magnesium at different concentrations of magnesium. The titrations give access to the total concentration of kissing complex at the 1:1 point. We also saw that we can reasonably assume that the response factors of the different magnesium adducts are equal.

From the adduct cleaning experiments we know that there are potentially two specific magnesium ions binding to the kissing complex. We extracted the concentration of all HP and KC at the equivalent point,

i.e. TAR and R06 (or RNAIi, RNAIIi) at the same concentration, at the 4 concentrations of  $Mg(OAc)_2$  used for  $K_{dTAR-R06,app}$  determination: 0, 50, 200 and 600  $\mu M$ . The concentrations are calculated for each different experiments taking into account each different response factor found (the response factors were not averaged before calculating  $K_D$  values). As a remark, here the concentration could have been calculated by using the average of the response factors found for each species. If averages of  $R_f$  are calculated it could be possible to extrapolate the concentrations of HP and KC at all the concentrations of Magnesium. But to do so it is important that the standard error between all the  $R_f$  is not exceeding 5%. It is the case for TAR-R06 system where the standard error does not exceed 5%, but this is not true for RNAIi-RNAIIi system ( $\approx 10\%$  standard error).

Using the results of adduct cleaning we have access to the relative proportion of  $0Mg$ ,  $1Mg_{spe}$  and  $2Mg_{spe}$  (for TAR-R06) at the four  $Mg(OAc)_2$  concentrations of interest. Assuming that the response factors of all magnesium adducts are the same, we can find the concentration of  $[KC]$ ,  $[KC\bullet Mg]$  and  $[KC\bullet 2Mg]$  respectively the KC without magnesium, with one and then two specific magnesium. Then the graph showing the evolution of the concentrations of  $[HP1]$ ,  $[HP2]$ ,  $[KC]$ ,  $[KC\bullet Mg]$  and  $[KC\bullet 2Mg]$  can be reconstructed.

Using the DynaFit software it is possible to fit all these data to find the equilibrium dissociation constants of each specific magnesium. The equilibria taken into account to fit the data are:

1) For TAR-R06:



The model will be annotated as the **0-1-2** model where the binding of the two specific magnesium is successive. The results are presented on Figure 50A and Table 7.

2) For RNAIi-RNAIIi:



The model will be annotated as the **0-1** model. The results are presented on Figure 50B and Table 7.

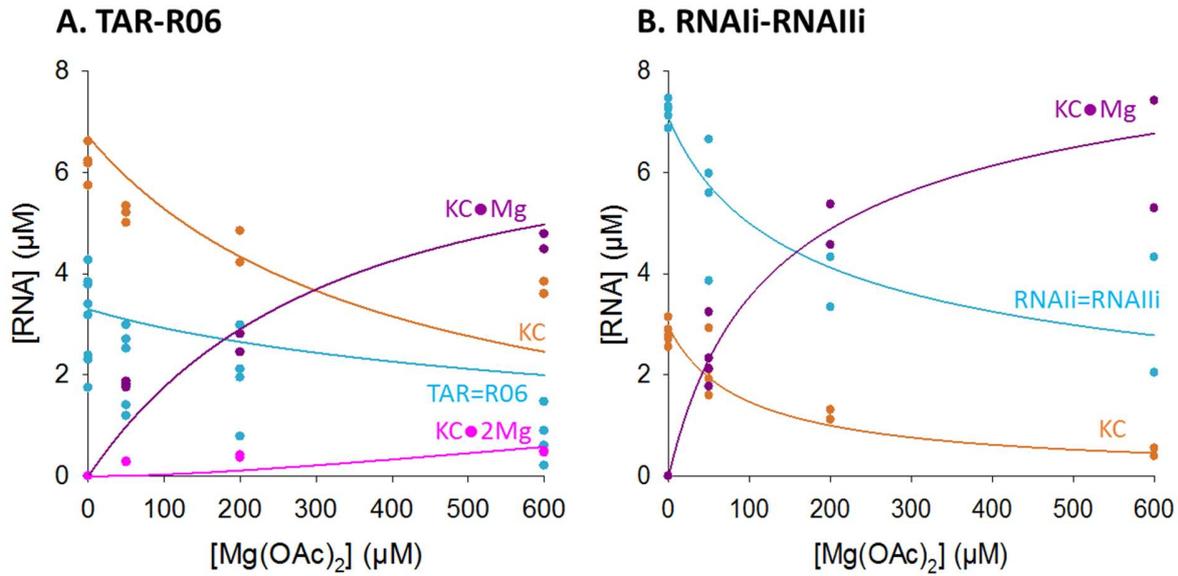


Figure 50: Determination of the  $K_{Mg}$  of each specific magnesium binding to A) TAR-R06 B) RNAIi-RNAIii. The graph shows the concentrations of each species in function of  $Mg(OAc)_2$  concentration. Fits are presented as lines. All the points represent a value found for a replicate of the experiment.

	TAR-R06	RNAIi-RNAIii
$K_{dKC}$	$1.62 \pm 0.16 \mu M$	$17.0 \pm 1.5 \mu M$
$K_{dMg1}$	$290 \pm 40 \mu M$	$40 \pm 5 \mu M$
$K_{dMg2}$	$5100 \pm 3500 \mu M$	-

Table 7: Summary of the values found for  $K_{dKC}$ ,  $K_{Mg1}$  and  $K_{Mg2}$  for both Kissing complexes used.

$K_{dTAR-R06}$  is equal to the value found for  $K_{dTAR-R06,app}$  at 0  $\mu M$  of  $Mg(OAc)_2$  which make the results consistent. From  $K_{dMg1}$  and  $K_{dMg2}$ , we can say that the affinity of the first magnesium is much higher than the one of the second magnesium. This can explain the results found for the adduct cleaning where the second magnesium appears at high concentrations of magnesium only. However due to the huge standard error found on  $K_{Mg2}$  we cannot say that this calculation is accurate. We can just discuss trends concerning the affinity of the 2 magnesium. We can see that the fits of the data do not pass by all the data points, which

result in high standard error and percentage of deviation for the calculated  $K_{dMg}$ .  $K_{dRNAIi-RNAIi}$  found after fitting is consistent with  $K_{dRNAIi-RNAIi,app}$  found without magnesium and the fits are better than for TAR-R06.

### **Comparison between TAR-R06 and RNAIi-RNAIi**

If we compare the specific  $K_{dMg}$  obtained for the first magnesium for both model systems, we note that the affinity of the specific magnesium for KC is much higher for RNAIi-RNAIi than TAR-R06. This explains why RNAIi-RNAIi is more magnesium-dependent than TAR-R06. The affinity of  $Mg^{2+}$  for kissing complexes accounts for the variation of the apparent binding constant, and hence the “stabilization” of kissing complexes in the presence of magnesium.

Several critical points can be addressed. First, the number of concentrations of magnesium tested is limited. Due to time constraints, the data were fitted based only on 4 set of points at 0, 50, 200 and 600  $\mu M$  of  $Mg(OAc)_2$ . We instead preferred to replicate the experiments at these four concentrations. It could be interesting to do this analysis on other magnesium concentrations to increase the number of points and have better fits, or use the average of response factors to extrapolate the concentrations of HP and KC at any concentration. If we compare the quality of the fits obtained for both systems, the fits for RNAIi-RNAIi are passing through the data points for all concentrations of magnesium whereas for TAR-R06 it is not the case. This may be due to larger error propagation when extracting the contributions of specific adducts, or from variations in response factors that we could not take into account. Yet, even if our quantification is not perfect, it is the first time such  $K_D$  values are determined for magnesium binding to kissing complexes.

Also, we have seen that  $K_{dKC,app}$  is influenced by the concentration of  $Mg^{2+}$  ions, that an addition of  $Mg^{2+}$  ions increases of the proportion of KC detected and not HP (Figure 43). In the calculation of  $K_{dKC}$  we do not take into account the non-specific  $Mg^{2+}$  that may stabilize also the kissing loop motif. Indeed, it has been shown in the literature that diffuse monovalent and divalent cations, i.e. non-specific ions, are also helping to stabilize RNA structure in addition to specific ions.<sup>41</sup> Moreover, divalent cations were shown to have a stronger competitiveness than monovalent cations. For example, in a solution with 0.48 mM of  $Mg^{2+}$  and 20 mM of  $Na^+$ , “ $Mg^{2+}$  ions associate 40-fold more strongly than  $Na^+$  ions”<sup>22</sup>, which means that in our solution it is inevitable to have non-specific  $Mg^{2+}$  that are replacing  $NH_4^+$  and will stabilize even more the structure. The non-specific  $Mg^{2+}$  influence the concentration of KC at high concentration of  $Mg^{2+}$ . The new equilibrium (33), which consider all type of  $Mg^{2+}$  ions, should be taken into account.



### XII.2.3. Comparison between $Mg^{2+}$ and $Mn^{2+}$

In Section XI.1.2.4., we have seen that  $Mg^{2+}$  or  $Mn^{2+}$  adopt the same stoichiometries for the respective KCs. We also saw from the relative intensities that the affinity of manganese is different than the one of magnesium. In the following part, we quantified the relative affinities.

We have first investigated the effect of manganese on the affinity of TAR for R06 and RNAIi for RNAIi. We used the same method as described previously, i.e. titrations of one hairpin by the other to determine response factors, concentrations and so  $K_{dKC,app}$ . Response factors and  $K_{dKC,app}$  values are summarized on Table 8. As the experiments were done only by titrating TAR by R06 and RNAIi by RNAIi, values for  $\frac{Rf_{R06}}{Rf_{T6}}$  and  $\frac{Rf_{RNAIi}}{Rf_{T6}}$  cannot be obtained.

		TAR-R06				RNAIi-RNAIi			
		$\frac{Rf_{TAR}}{Rf_{T6}}$	$\frac{Rf_{KCtot}}{Rf_{T6}}$	$\frac{Rf_{HP}}{Rf_{KCtot}}$	$K_{dTAR-R06,app}$ (* $10^{-6}$ M)	$\frac{Rf_{RNAIi}}{Rf_{T6}}$	$\frac{Rf_{KCtot}}{Rf_{T6}}$	$\frac{Rf_{HP}}{Rf_{KCtot}}$	$K_{dRNAIi-RNAIi,app}$ (* $10^{-6}$ M)
[Mg] <sup>2+</sup> ( $\mu$ M)	50	2.49 $\pm 1.4E-3$	1.65 $\pm 0.034$	1.49 $\pm 0.04$	1.00 $\pm 0.08$	2.60 $\pm 0.039$	0.94 $\pm 0.12$	2.83 $\pm 0.042$	8.6 $\pm$ 2.0
	200	2.42	1.53	1.57	0.46 $\pm 0.07$	3.57	1.42	2.51	2.9 $\pm$ 0.5
	600	2.13	1.37	1.55	0.17 $\pm 0.07$	2.92	1.53	1.90	0.96 $\pm$ 0.08
[Mn] <sup>2+</sup> ( $\mu$ M)	50	2.58	1.91	1.35	1.02	2.87	1.86	1.54	4.05
	200	2.24	2.33	0.96	0.084	2.80	2.00	1.40	0.76
	600	1.80	1.81	0.99	3.87E-03	2.73	1.36	2.00	0.064

Table 8: Summary of the response factors and  $K_{dKC,app}$  values obtained for TAR-R06 and RNAIi-RNAIi with  $Mn^{2+}$  ions. For comparison averages are done on the response factors found with  $Mg^{2+}$  ions. Titrations R06 by TAR and RNAIi by RNAIi are not considered in this table for calculating the response factors.

Response factors for the hairpins are similar between  $Mn^{2+}$  and  $Mg^{2+}$ , on the contrary KC response factors are higher with  $Mg^{2+}$  than  $Mn^{2+}$ . Figure 51 shows the evolution of  $\ln(K_{dKC,app})$  as a function of the divalent cation concentration.

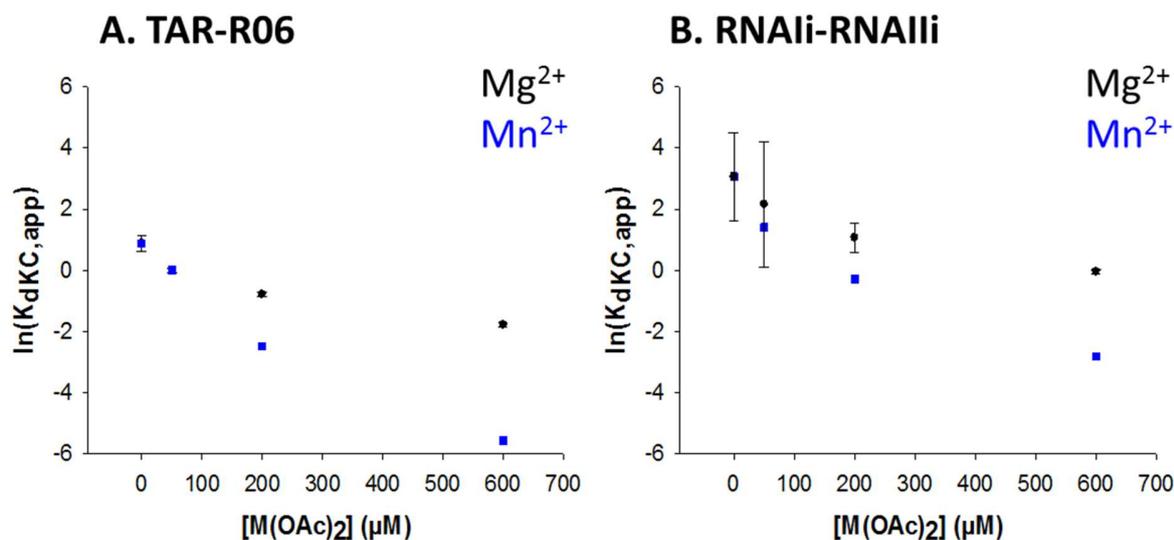


Figure 51: Comparison of the  $K_{dKC,app}$  obtained in presence of  $Mn(OAc)_2$  and  $Mg(OAc)_2$  for A) TAR-R06 and B) RNAIi-RNAIii. Errors bars on magnesium experiments comes from the standard error calculated on at least 3 experiments. Experiments with Manganese were done just once.

For TAR-R06, the evolution of  $\ln(K_{dKC,app})$  as a function of the concentration of manganese acetate follows the same trend as with magnesium acetate at low concentration of cations ( $< 200 \mu M$ ). However, at concentrations equal or higher than  $200 \mu M$ , the affinity between TAR and R06 is higher in  $Mn^{2+}$  than in  $Mg^{2+}$ . For RNAIi-RNAIii, the difference between  $\ln(K_{dRNAIi-RNAIii,app})$  values in  $Mg^{2+}$  or  $Mn^{2+}$  appears from  $50 \mu M$ . For both systems, at high concentrations of divalent cations ( $\geq 600 \mu M$ ), we have  $K_{dKC,app}^{Mn} > K_{dKC,app}^{Mg}$ . These results suggest that  $Mn^{2+}$  reinforces more the interaction between the two hairpins, and even more at high concentration, than  $Mg^{2+}$ . In 1995, Crothers and co-workers have determined the thermal stability of RNAIi-RNAIii in different solutions containing different divalent cations. They showed that the stability of the KC was higher with  $Mn^{2+}$  ions than  $Mg^{2+}$  ions ( $\Delta T_m \approx 5^\circ C$ ).<sup>18</sup> Our results suggest the same conclusion. The experiments with manganese should be replicated.

Next, we wanted to have information on the specific equilibrium binding constant of each specific cation. As previously, we combined data extracted from the adduct cleaning experiments and  $K_{dKC,app}$  experiments to extract  $K_{Mn}$ . Concentrations are calculating using Rf values presented on Table 8 for  $Mn^{2+}$  ions. For the point at  $0 \mu M$ , we took the values found for the point at  $0 \mu M$  of  $Mg^{2+}$  ions (

[Mg(OAc) <sub>2</sub> ] ( $\mu M$ )	RfTAR/RfT6	RfR06/RfT6	RfKCtotal/RfT6	RfHP/RfKC (HP = TAR or R06)	KdTAR- R06. <sub>app</sub>	Average KdTAR- R06. <sub>app</sub>	Std error
--	------------	------------	----------------	-----------------------------------	-------------------------------	--	--------------

					(*10-6 M)	(*10-6 M)	
0	1.85		1.58	1.17	1.65		
0	2.16		1.73	1.25	2.38	<b>2.42</b>	0.25
0		1.89	1.39	1.37	2.62		
0		2.07	1.63	1.27	3.01		
50	2.51		1.53	1.64	0.79		
50	2.48		1.80	1.38	1.08	<b>1</b>	0.08
50		2.57	1.79	1.44	1.12		
200	2.42		1.53	1.57	0.36		
200		2.47	1.77	1.39	0.56	<b>0.46</b>	0.07
600	2.13		1.37	1.55	0.07		
600		2.07	1.54	1.34	0.27	<b>0.17</b>	0.07

Table 5 and 6). Here also, we make the assumption that the response factors of all manganese adducts are equal. We used the 0-1-2 model for TAR-R06 and the 0-1 model for RNAIi-RNAIii to fit the data.  $K_{Mx}$  values obtained are summarized on Table 9.

	<b>Mg<sup>2+</sup></b>		<b>Mn<sup>2+</sup></b>	
	<b>TAR-R06</b>	<b>RNAIi-RNAIii</b>	<b>TAR-R06</b>	<b>RNAIi-RNAIii</b>
$K_{dKC}$	1.62 ± 0.16 μM	17.0 ± 1.5 μM	1.62 ± 0.16 μM	17.9 ± 1.0 μM
$K_{dM1}$	290 ± 40 μM	40 ± 5 μM	210 ± 50 μM	10.4 ± 1.1 μM
$K_{dM2}$	5100 ± 3500 μM	-	1000 ± 400 μM	-

Table 9: Summary of  $K_{dKC}$  and  $K_{Mg}$  values obtained after fitting by DynaFit for TAR-R06 and RNAIi-RNAIii with  $Mg^{2+}$  or  $Mn^{2+}$ . Errors come from the fit.

Concerning TAR-R06, the specific  $K_{dM}$  obtained for each manganese follow the same trend as the ones of the specific  $Mg^{2+}$ . The affinity of the second cation is lower than the one for the first cation. Considering the error found on  $K_{dM1}$ , we can estimate that the affinity of the first  $Mn^{2+}$  or  $Mg^{2+}$  are similar. The most important difference comes from  $K_{M2}$ . Indeed, the affinity of the KC for the second  $Mn^{2+}$  is 5 times higher than for the second  $Mg^{2+}$ . But we should be careful because the standard errors are high.

For RNAIi-RNAIii,  $K_{dKC}$  are equal in both cases and naturally similar. However,  $K_{dM1}$  with  $Mn^{2+}$  is almost 4 times lower than with  $Mg^{2+}$ . The affinity of the specific manganese for RNAIi-RNAIii KC is thus higher than the affinity of the specific magnesium.

### XII.3. Conclusion and perspectives

*Native* mass spectrometry allows us to determine apparent equilibrium dissociation constant of kissing complexes. Calculations of response factors were needed to obtain the concentrations from the peak areas. Our method then combines the determination of specific stoichiometries and of  $K_{dKC,app}$  to have access to specific cations dissociation constant,  $K_{dMg}$ .

It is the first time such level of detail in RNA kissing complex equilibria is reached, taking the difference in response factors of hairpins and kissing complexes into account. Still, the quantification method is not perfect and could be further optimized. For example, concentrations were determined replicate by replicate, and the average was done on the  $K_{dKC,app}$  obtained. An alternative could be to do the average of the response factors determined and calculate the concentrations from these new Rf. However, we saw fluctuations on the Rf values along the different experiments and we do not know yet all the parameters that are influencing these Rf. Currently, as reproducibility is not seen, this method cannot be used. If we assume that the response factors of the hairpins and KC are not changing with  $Mg^{2+}$  addition, which seems to be the case for TAR-R06 but not RNAIi-RNAIi, it could be possible to extrapolate the concentrations of HP and KC whatever the concentration of  $Mg^{2+}$  ions used.

Another important point is the way specific and non-specific adducts are distinguished. We saw that, for TAR-R06, maybe a fraction of the adducts that is not specific compared to the duplex reference may still contribute to the stabilization of the kissing loop motif compared to the separate hairpins. Equilibrium (33) should be considered when fitting the data for the determination of  $K_{dKC}$ ,  $K_{Mg1}$  and  $K_{Mg2}$ .



Yet, even if our quantification is not perfect, it is the first time such  $K_D$  values are determined for magnesium binding to kissing complexes. We also show that the method is adapted to the study of several cations.

## XIII. Mg<sup>2+</sup> localization

### XIII.1. Introduction

In the previous chapter, we have seen that it is possible to deduce, from MS titration and comparison with duplex references, the number of specific magnesium ions bound to kissing loop motifs such as TAR-R06 and RNAIi-RNAIi. As discussed in the literature, misattribution of Mg<sup>2+</sup> binding localization onto RNA structures can happen.<sup>36</sup> Here, we explore the use of tandem mass spectrometry to localize specifically bound magnesium onto kissing complexes structures. The goal is to give enough energy to fragment the kissing complex in its two hairpins, to see on which hairpin the magnesium cations are bound. Next, we want to fragment again the hairpins where Mg<sup>2+</sup> is present, until we can have access to the precise localization.

One challenge was to perform MS/MS on kissing complexes with the instruments available in the laboratory. Also, depending on the instrument, the fragmentation channels could be different. The dissociation of kissing complexes was not reported a lot in the literature. In 2007, the group of Daniele Fabris, has performed sustained off-resonance irradiation collision induced dissociation (SORI-CID) using a Fourier Transform Ion Cyclotron Resonance (FTICR) mass spectrometer to study the dissociation of the dimerization initiation site (DIS) of HIV-1 mRNA sequence.<sup>6</sup> The DIS sequence has the ability to dimerize forming a kissing complex which then can isomerize into a complementary RNA duplex. The authors compared the dissociation pathway of the KC and the duplex using different CID activation and compared their relative stability. They found that, using their instrumentation, the kissing complex dissociates into its two hairpin monomers, while the duplex needs more activation to be fragmented.

Firstly, in this chapter, we will see the optimization of the magnesium salt and buffer in order to maximize the fraction of specific adducts. Then, we will study the dissociation of TAR-R06 and RNAIi-RNAIi kissing complexes using collision-induced dissociation (CID) fragmentation in different instruments. We will also compare their fragmentation pathway to one of the RNA reference duplexes used for the adduct cleaning experiments. The results are discussed in light of molecular dynamics simulations, which were performed to obtain a theoretical positioning of specific magnesium ions.

### XIII.2. Results

#### *XIII.2.1. Influence of the magnesium salt*

In order to perform MS/MS on kissing complexes with magnesium, we need to take into account the fraction of specific and non-specific adducts under a peak of given  $m/z$ . Based on the results found with

the adduct cleaning, we need to maximize the proportion of specific magnesium and so reduce the extent of non-specific adduction. Initially we performed the experiment using magnesium acetate ( $\text{Mg}(\text{OAc})_2$ ) where the counter-ion is the same as the electrolyte ( $\text{NH}_4\text{OAc}$ ). Here, we first explored if the use of other salts could influence the number of non-specific adducts.

We tried magnesium chloride ( $\text{MgCl}_2$ ) and magnesium bromide ( $\text{MgBr}_2$ ). These experiments are done on the Agilent 6560 DTIMS-Q-TOF. The concentrations of magnesium salt used for the titration and the method are the same as in chapter XI.2.2. Adduct cleaning: TAR-R06 vs RNAIi-RNAIli. Global spectra of TAR-R06 and zoom on the kissing complex charged 6- are shown on Figure 52.

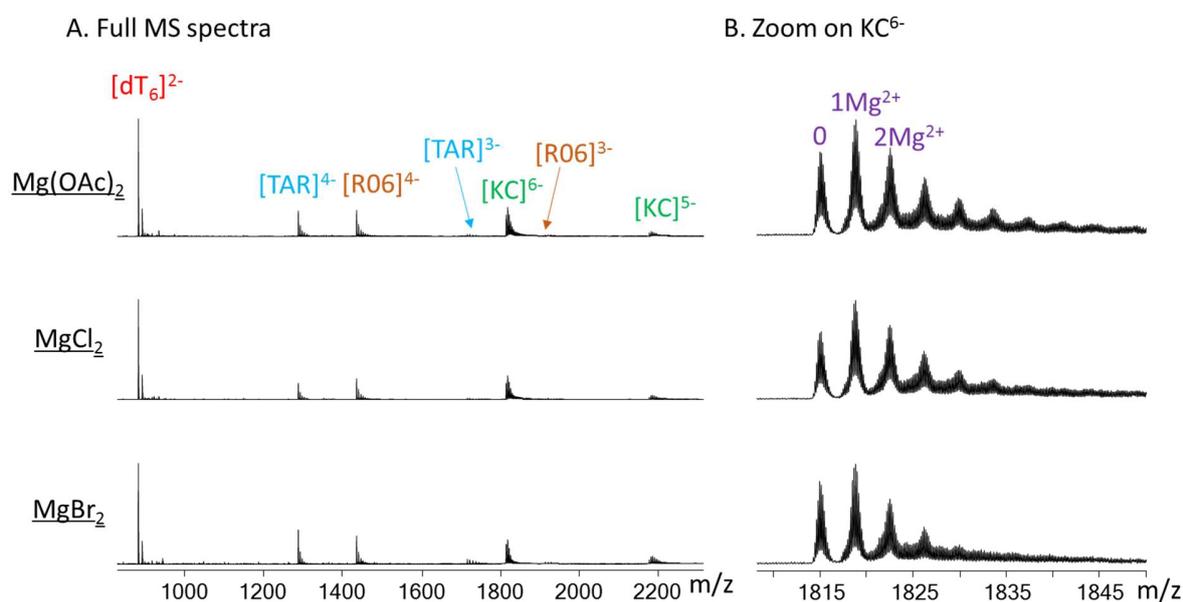


Figure 52: ESI-MS spectra of TAR-R06 kissing complex in 150 mM  $\text{NH}_4\text{OAc}$  and 200  $\mu\text{M}$  of  $\text{Mg}(\text{OAc})_2$ ,  $\text{MgCl}_2$  or  $\text{MgBr}_2$ . A) Full MS spectra. B) Zoom on the kissing complex 6-. Data were acquired on the Agilent 6560 DTIMS-Q-TOF.

When zooming on the kissing complex's major charge state for TAR-R06, it seems that fewer adducts are detected with  $\text{MgBr}_2$  than with  $\text{MgCl}_2$  or  $\text{Mg}(\text{OAc})_2$ . With  $[\text{RNAIi-RNAIli}]^{6-}$ , such difference is observed as well.

Data treatment focuses on the major charge state of each KC. Results after data treatment are presented on Appendix A1.6. The curves corresponding to the three different salts are mostly superimposed. Whatever the kissing complex studied, i.e. TAR-R06 or RNAIi-RNAIli, the type of counterion used does not seem to modify the total proportion of specific adducts.

For MS/MS we need to select a precursor ion. As the one-magnesium is present even at low concentrations of magnesium, we will focus on the first  $Mg^{2+}$  adduct. Before doing MS/MS, we need to have access to the proportion of specific and non-specific magnesium under the peak of interest, i.e. the peak of  $KC+1Mg^{2+}$ . Using duplex reference experiments, we know the fraction of  $[KC+1Mg^{2+}]^{6-}$  intensity which is specifically bound to  $Mg^{2+}$  (equation 34 below).

$$I_{1Mg} = i1Mg - i0Mg * x_1 \quad (3)$$

Where  $I_{1Mg}$  is the intensity of specifically bound  $Mg^{2+}$ ,  $i1Mg$  is the relative intensity of the peak and  $i0Mg * x_1$  is the part of non-specific  $Mg^{2+}$  under this peak (measured on the reference duplex). The fraction of  $1Mg^{2+}$  is thus:

$$P(I_{1Mg}) = \frac{I_{1Mg}}{i1Mg} \quad (354)$$

This method is applied to  $[TAR-R06+1Mg^{2+}]^{6-}$  and  $[RNAIi-RNAIi+1Mg^{2+}]^{6-}$  for all concentrations of magnesium used and for each magnesium salt. The goal here is to see which conditions maximize both the total signal-to-noise ratio and the fraction of specific cation under the peak. The results are presented as bar charts on Figure 53.

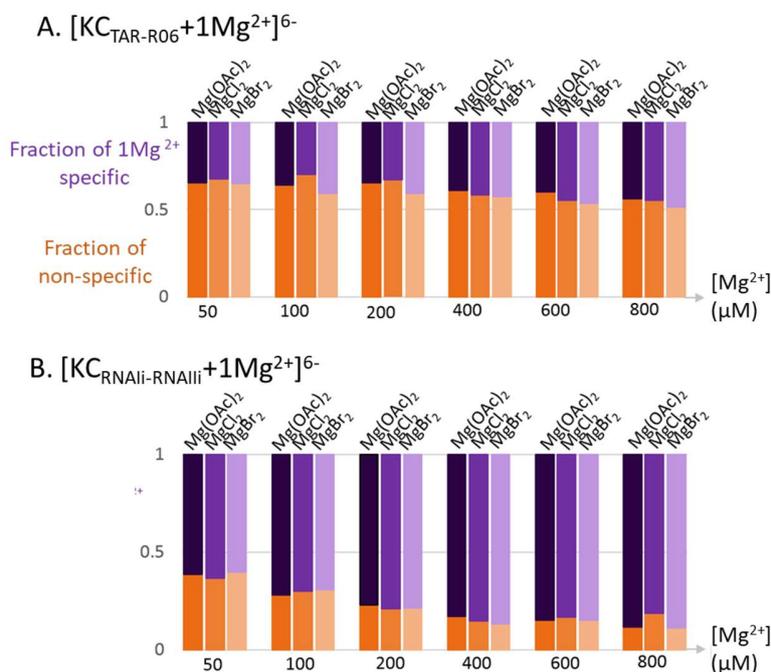


Figure 53: Proportions of specific and non-specific  $Mg^{2+}$  under the peak of A)  $[TAR-R06+1Mg^{2+}]^{6-}$  and B)  $[RNAIi-RNAIi+1Mg^{2+}]^{6-}$  at different magnesium salts and different concentration.

The proportion of  $1\text{Mg}^{2+}$  under the peak  $[\text{KC}_{\text{TAR-R06}}+1\text{Mg}^{2+}]^{6-}$  (Figure 53,A) does not reach 50% even at 800  $\mu\text{M}$  of  $\text{Mg}^{2+}$  ions. This proportion does not increase a lot when magnesium is added to the solution. The results observed are the same whatever the magnesium salt used. The determined fraction means that, at 100  $\mu\text{M}$  of magnesium for example, less than 50% of the total signal of this peak is due to specific  $\text{Mg}^{2+}$  binding. Consequently, when using MS/MS, < 50% of the signal will reflect a specific  $\text{Mg}^{2+}$  and > 50% will reflect a non-specific magnesium.

In the case of  $[\text{KC}_{\text{RNAIi-RNAIii}}+1\text{Mg}^{2+}]^{6-}$  (Figure 53,B), more than 50% of the signal of the peak is due to a specific magnesium. This is true for all three magnesium salts used. Also, this proportion increases from 50% at 50  $\mu\text{M}$  of magnesium, until  $\approx 80\%$  at 800  $\mu\text{M}$  of magnesium. When doing MS/MS, the probability to have a specific magnesium is higher for this kissing complex than for TAR-R06.

In summary, whatever the magnesium salt used, the fraction of specific magnesium bound is similar. As this salt is in negligible concentration compared to electrolyte, less than 1 mM against 150 mM of  $\text{NH}_4\text{OAc}$ , it could be of interest to vary the type of buffer used and see if it can have a greater influence on the adduct distribution.

### *XIII.2.2. MS/MS*

We chose to perform the experiments using 150 mM  $\text{NH}_4\text{OAc}$  and a concentration of 100  $\mu\text{M}$  of  $\text{Mg}(\text{OAc})_2$ . MS/MS is performed on TAR-R06 and RNAIi-RNAIii kissing complexes. In this section, we will try to answer to the following questions: what are the dissociation channels of kissing complexes? Do we fragment it into its two forming hairpins? What kind of fragments do we obtain after collision activation? Can we have information on the possible binding sites of magnesium? What type of information can we learn from MS/MS experiments?

#### *XIII.2.2.1. MS/MS using the AGILENT 6560 IMS-Q-TOF*

- **TAR-R06**

First, experiments without magnesium are done in order to understand the behavior of TAR-R06 kissing complex under fragmentation conditions. The concentration of the two hairpins is increased to 20  $\mu\text{M}$  to increase the signal for MS/MS.  $[\text{TAR-R06}]^{6-}$  is selected ( $m/z$  of 1815) as precursor ion by the quadrupole and then several collision voltages are applied into the collision cell: 0, 20, 30, 40 and 50 V. Instrument trapping parameters are set as SOFT in order to avoid pre-collision cell activation. The isolation window is set at 4  $m/z$ . We remind here that the collision time is less than 1 ms and cannot be changed by the user. The results are shown on Figure 54.

By increasing the activation voltage, the kinetic energy of the ions is increased, which leads to more energetic ion-gas collisions. This will increase the fragmentation efficiency. At 20 V, we can observe that  $[KC]^{6-}$  has begun to fragment by losing a guanine, and has begun to dissociate into its two monomers as well. At 50 V, the ion corresponding to the  $[KC]^{6-}$  has been completely depleted. These results confirm that  $[TAR-R06]^{6-}$  can fragment into its two monomers. When the voltage increases further, base losses from the two monomers are also observed. Predominantly guanine (at 30 V) and cytosine (at 40 V) are lost, and we presume these could come from the stems of TAR and R06.

Solution containing 100  $\mu\text{M}$   $\text{Mg}(\text{OAc})_2$  are then injected and the ion  $[TAR-R06+1\text{Mg}^{2+}]^{6-}$  ( $m/z$  of 1818.5) is isolated and fragmented using the same collision voltages as previously. Results are shown on Figure 55.

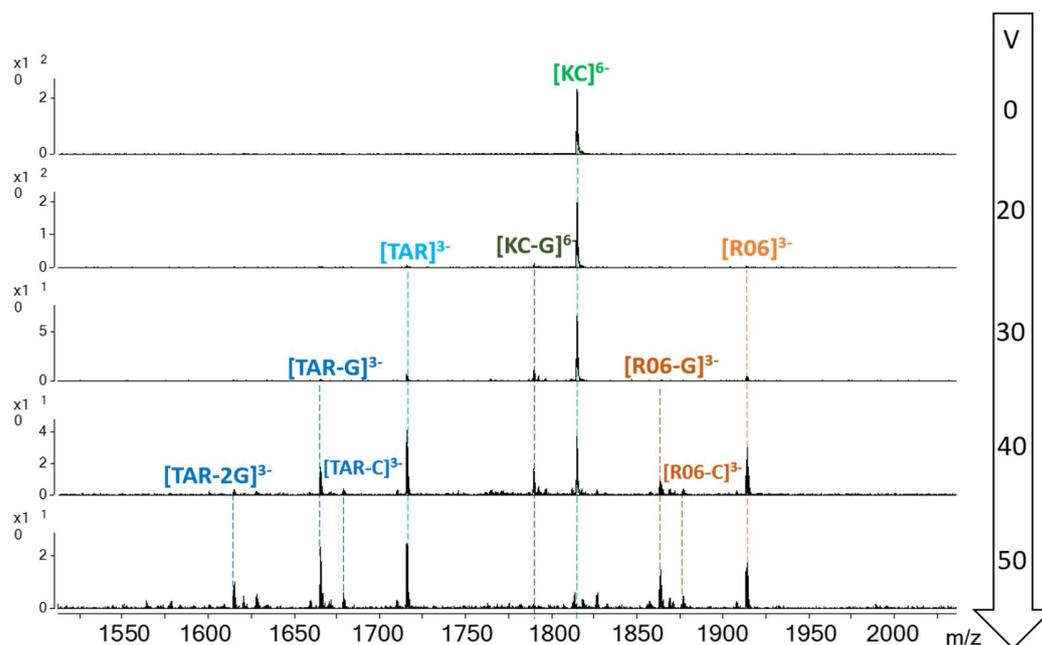


Figure 54: MS/MS spectra obtained after selecting  $[TAR-R06]^{6-}$  and application of different activation voltages. Data are obtained using the AGILENT 6560 DTIMS Q TOF instrument. Activation time < 1ms. Sample: 20  $\mu\text{M}$  of each hairpins and 150 mM  $\text{NH}_4\text{OAc}$ .

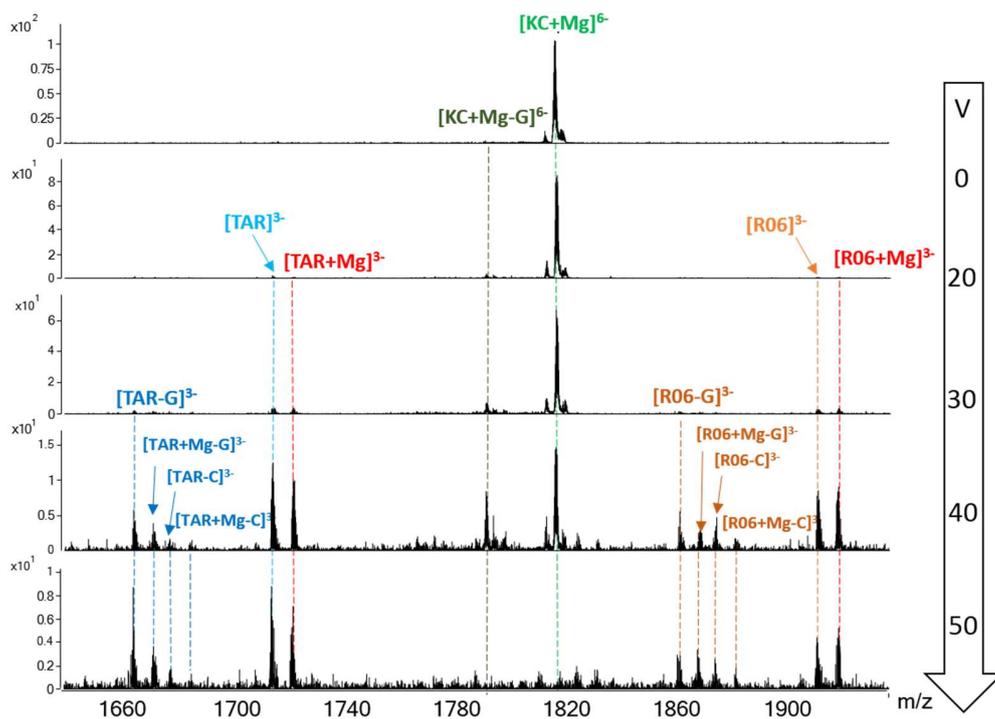


Figure 55: MS/MS spectra obtained after selecting  $[TAR-R06]^{6-}$  and application of different activation voltages. Data are obtained using the AGILENT 6560 DTIMS Q TOF instrument. Activation time < 1ms. Sample: 20  $\mu$ M of each hairpins, 150 mM  $NH_4OAc$  and 100  $\mu$ M  $Mg(OAc)_2$ .

As previously, at 0 V we observe the precursor ion.  $[KC_{TAR-R06}+1Mg^{2+}]^{6-}$  is dissociated into  $[TAR]^{3-}$  and  $[R06]^{3-}$  but also into  $[TAR+1Mg^{2+}]^{3-}$  and  $[R06+1Mg^{2+}]^{3-}$  from 20 V. At 40 V, as the number of fragments increased, we have identified peaks corresponding to the monomers plus a loss of a G but with 1  $Mg^{2+}$  and then the same but with a loss of a C. At 50 V the peak of the precursor ion has been completely fragmented.

What is even more interesting here is the presence of both monomers charge 3- plus one magnesium. The relative intensities of  $[R06+1Mg^{2+}]^{3-}$  look higher. However, we need to remember that R06 is longer than TAR so non-specific  $Mg^{2+}$  could be more abundant. We saw before that only < 50% of the signal is due to KC-specific magnesium. If this magnesium is binding to a single site and if this site is located only on one hairpin, with the MS/MS spectra we should show a bias in magnesium retention on one of the hairpins. These results lead to several hypotheses: 1) what was considered as specific is actually not so the mathematical treatment is mistaken, 2) the magnesium is binding to a single binding site which is exactly in-between the two hairpins, meaning that  $Mg^{2+}$  after fragmentation has equal chances to go on one hairpin or the other, or 3) there are two specific binding sites, one on each hairpins, which have equivalent affinities. Each hypothesis will be reviewed.

- 1) Concerning the first hypothesis, we have demonstrated that whatever the magnesium salt used and whatever the RNA reference duplex used, we still find the same number of specific adducts. So we can have some confidence into these data. We recall that “specific” means “specific to the KL motif”.
- 2) If the magnesium is coordinated right in-between the two hairpins, we anticipate that more energy is necessary to disrupt the kissing complex, because the  $Mg^{2+}$  ion would then act as a bridge between the two negatively charged strands. To check this, we can calculate the relative intensity of the precursor ion in comparison of the total intensity of all ions (fragments + precursor). To do so we have extracted the integrals of each peak and then do the following calculation:

$$I_{rel(\text{parent ion})} = \frac{I_{\text{parent ion}}}{\sum(I_{\text{fragments}} + I_{\text{parent ion}})} \quad (5)$$

Then we can draw the graph representing  $I_{rel(\text{parent ion})}$  in function of the collision voltage. The results are shown on Figure 56.

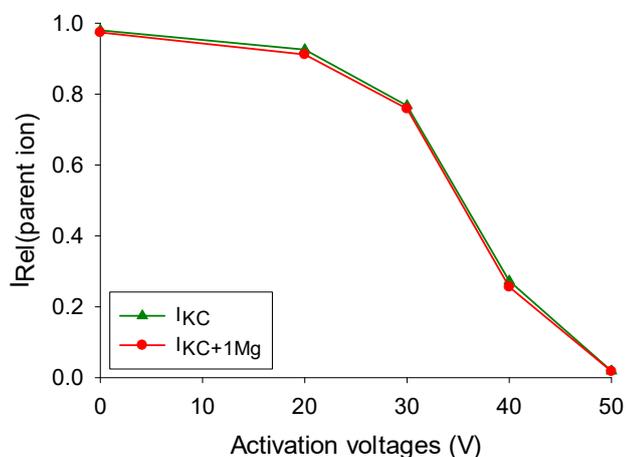


Figure 56: Evolution of the intensity of the precursor ion in function of the activation voltage. In green, the intensity of  $[KC]^{6-}$  and in red the intensity of  $[KC+1Mg^{2+}]^{6-}$ .

The evolution of the relative intensity of  $[KC_{TAR-R06}]^{6-}$  and  $[KC_{TAR-R06}+1Mg^{2+}]^{6-}$  is the same which means that the magnesium does not have an impact on the collision energy needed to fragment the kissing complex. This is strongly in disfavor of hypothesis number 2.

- 3) The third hypothesis seems more plausible. If there are two binding sites with equivalent affinity, this can explain why there is as much  $[\text{TAR}+1\text{Mg}^{2+}]^{3-}$  as  $[\text{R06}+1\text{Mg}^{2+}]^{3-}$ . Also, equivalent binding sites are not necessarily incompatible with the fact that the  $K_D$  of the second magnesium binding is lower than for the first magnesium binding. However, it means that the sites are equivalent, but not independent, and negative cooperativity is at play.

Based on these promising results, it could be of great interest to re-fragment the fragment ions  $[\text{TAR}+1\text{Mg}^{2+}]^{3-}$  and  $[\text{R06}+1\text{Mg}^{2+}]^{3-}$  to potentially localized the magnesium observed. Unfortunately it is not possible to do  $\text{MS}^n$  with the AGILENT instrument. In order to possibly do  $\text{MS}^n$  we performed the experiments on the Thermo LCQ Fleet.

- **RNAIi-RNAIii**

The same types of experiments were performed with the kissing complex RNAIi-RNAIii. First, experiments without magnesium are carried out, isolating  $[\text{RNAIi-RNAIii}]^{6-}$  as precursor ion. We used voltages from 0 to 50 V to fragment this ion. RNAIi-RNAIii  $[\text{KC}]^{6-}$  does not undergo fragmentation, except the loss of a G at 30 V, contrary to  $[\text{TAR-R06}]^{6-}$ . The dissociation of the KC into its two monomers is not observed. RNAIi-RNAIii is bigger than TAR-R06 in terms of number of bases (40 vs 34 respectively). This means that the charge density and the Coulomb repulsion is lower for  $[\text{RNAIi-RNAIii}]^{6-}$  than for  $[\text{TAR-R06}]^{6-}$ . To overcome this problem, we used  $[\text{RNAIi-RNAIii}]^{7-}$  as precursor ion, and because its signal was not sufficient we used a supercharging agent, sulfolane. This additive increases the charge states of the different species detected.<sup>42</sup> The full scan MS spectrum using 1% of sulfolane is presented on top of Figure 57A. This spectrum shows perfectly the effect of sulfolane on species charge states: charge state up to 7- are detected for RNAIi and RNAIii. Also,  $[\text{KC}]^{9-}$ ,  $[\text{KC}]^{8-}$  and  $[\text{KC}]^{7-}$  are observed.

We performed MS/MS on  $[\text{KC}_{\text{RNAIi-RNAIii}}]^{7-}$  using the same activation voltages as previously (from 0 to 50 V). MS/MS spectra are presented on Figure 57B. At 20 V, fragments corresponding to  $[\text{RNAIi}]^{4-}$ ,  $[\text{RNAIii}]^{3-}$  and  $[\text{RNAIi}]^{3-}$ ,  $[\text{RNAIii}]^{4-}$  are detected. When the activation voltage increases, the intensities of those peaks are increasing too. At 30 V, base losses from  $[\text{RNAIi}]^{4-}$  and  $[\text{RNAIii}]^{3-}$  begin to appear. The lost base is guanine. We can thus clearly see the dissociation of the kissing complex into its two monomers.

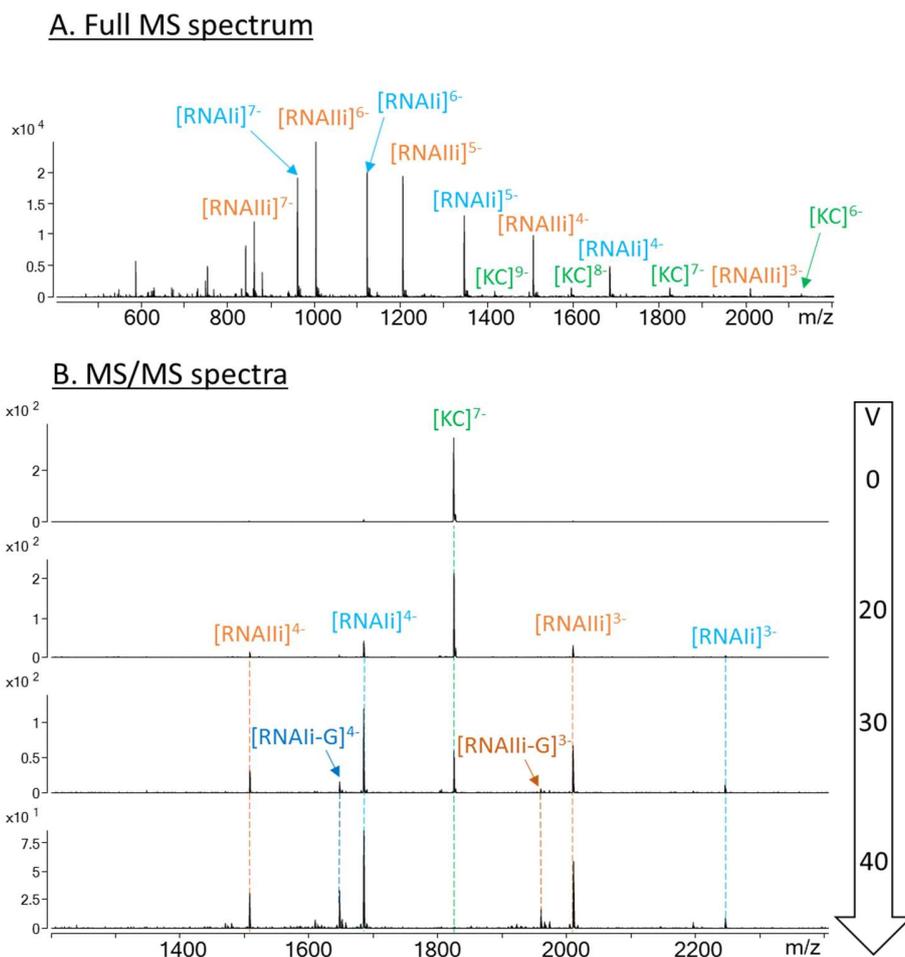
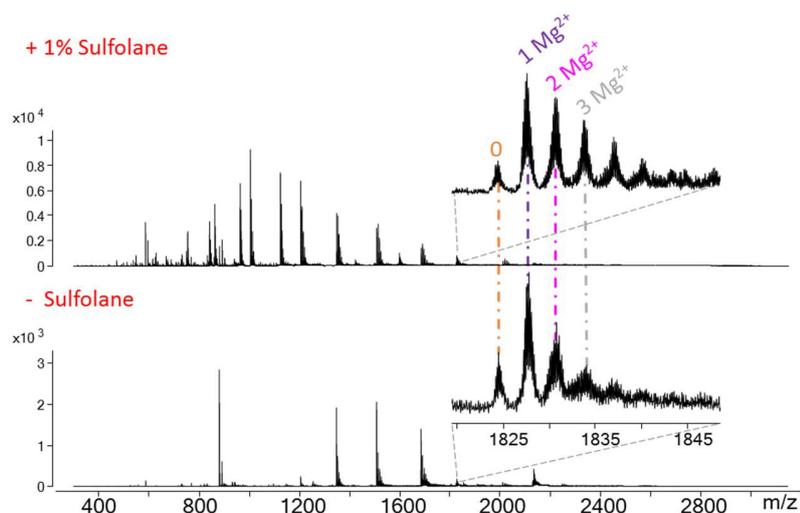


Figure 57: A) Full MS spectrum of 20  $\mu\text{M}$  RNAlI, 20  $\mu\text{M}$  RNAlII, 150 mM  $\text{NH}_4\text{OAc}$  and 1% Sulfolane. B) MS/MS spectra of the same solution and after selecting  $[\text{RNAlI-RNAlII}]^{7-}$  as precursor ion, at different activation voltage. Activation time < 1ms.

MS/MS experiments with 100  $\mu\text{M}$   $\text{Mg}(\text{OAc})_2$  are then performed to potentially localize the magnesium position. The solution is still prepared with an addition of 1% sulfolane to increase the charge states of the kissing complex. The first step is to check the spectra obtained with both magnesium and sulfolane. Full MS spectra of a solution containing 100  $\mu\text{M}$   $\text{Mg}(\text{OAc})_2$  with or without 1% sulfolane are shown on Figure 58, A. The spectrum shows an increase of the charge states of the different species present in solution. The presence of magnesium does not prevent sulfolane from playing its supercharging role. However, if we have a look at the zoom of  $[\text{KC}]^{7-}$  on both spectra, we can see that the distribution of magnesium adducts is slightly changed in presence of sulfolane. Adduct cleaning have not been performed on solution where 1% of sulfolane and the fraction of specific  $\text{Mg}^{2+}$  under the peak  $[\text{KC}_{\text{RNAlI-RNAlII}+1\text{Mg}^{2+}}]^{7-}$

was not determined. We will make the assumption that sulfolane would not affect too much the proportion of 1-specific  $\text{Mg}^{2+}$ , and that he may just have an impact on the number on non-specific adducts. MS/MS experiments on  $[\text{KC}+1\text{Mg}^{2+}]^{7-}$  are then performed (Figure 58B). As for  $[\text{TAR-R06}]^{6-}$ , the kissing complex dissociates in its two monomers charged 4 and 3-, with or without 1  $\text{Mg}^{2+}$ . The intensity of  $[\text{RNAIi}+1\text{Mg}^{2+}]^{4-}$  seems slightly higher than  $[\text{RNAIi}+1\text{Mg}^{2+}]^{3-}$  but as RNAIi is longer than RNAIii, more non-specific adducts can be retained.

### A. Full MS spectra



### B. MS/MS spectra

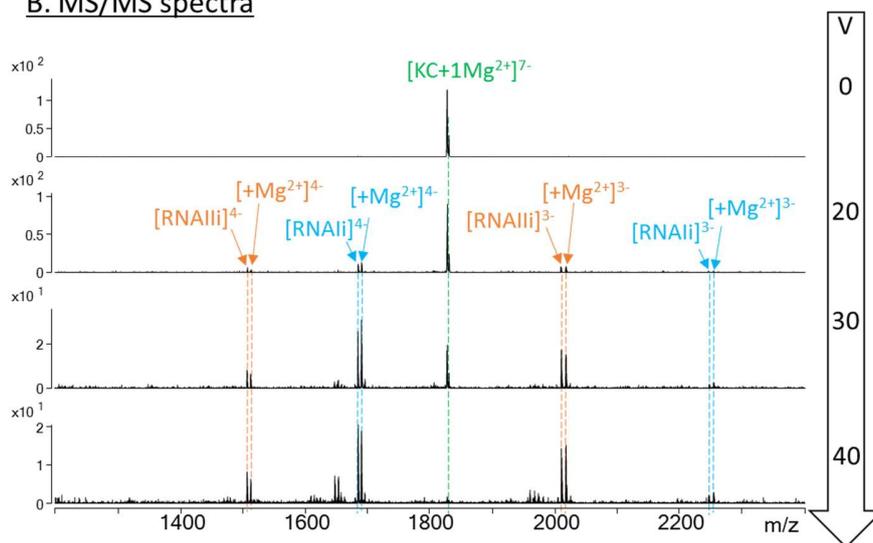


Figure 58: A) Full MS spectra of solutions containing RNAIi and RNAIii (10 or 20  $\mu\text{M}$ ), 150 mM  $\text{NH}_4\text{OAc}$ , 100  $\mu\text{M}$   $\text{Mg}(\text{OAc})_2$ , with or without 1% sulfolane. Spectra are acquired using the HARD tune, TEGD 20, F350. A zoom of  $[\text{KC}]^{7-}$  is also shown. B) MS/MS spectra of the same solution and after selecting  $[\text{RNAIi-RNAIii}]^{7-}$  as precursor ion, at different activation voltages. Activation time < 1ms.

As for TAR-R06, the same hypotheses can be evoked. We previously found that the same specific  $Mg^{2+}$  number was found on  $KC_{RNAIi-RNAIIi}$  charge 6<sup>-</sup> and 7<sup>-</sup>. Also we have checked that the magnesium salt do not have an influence on the number of specific  $Mg^{2+}$  on  $[RNAIi-RNAIIi]^{6-}$ . We assume it is the same for the 7-charge state and that Sulfolane will not change the number of specific ions on  $RNAIi-RNAIIi$ . We can then refute the first hypothesis saying that the mathematical treatment is wrong. The second hypothesis is that the specific magnesium ion is as much coordinated on  $RNAIi$  as  $RNAIIi$ . It could be considered as a salt bridge. To check this hypothesis, the evolution of the relative intensity of  $[KC]^{7-}$  and  $[KC+1Mg^{2+}]^{7-}$  in function of the activation voltage is determined. Results are shown on Figure 59.

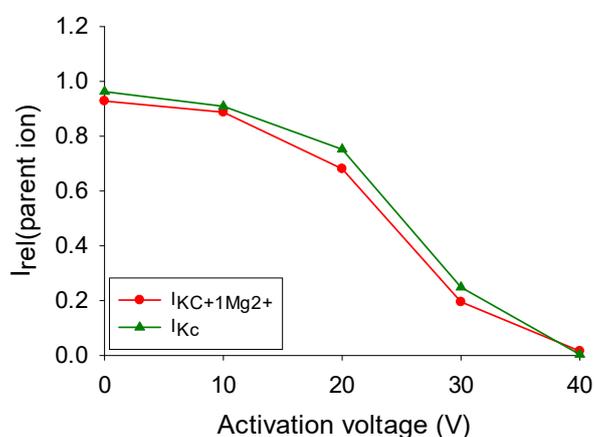


Figure 59: Evolution of the intensity of the precursor ion in function of the activation voltage. In green, the intensity of  $[RNAIi-RNAIIi]^{7-}$  and in red the intensity of  $[RNAIi-RNAIIi+1Mg^{2+}]^{7-}$ .

$[KC+1Mg^{2+}]^{7-}$  does not need more energy to dissociate than  $[KC]^{7-}$ . The second hypothesis can be refuted. Indeed, if magnesium was a bridging salt between the two hairpins, we should give more energy to the kissing complex to dissociate.

Our last hypothesis is that there are actually two preferential binding sites, one on  $RNAIi$  and the other on  $RNAIIi$ , but those two sites cannot be populated at the same time. Crothers *et al*<sup>18</sup> have determined the uptake of two  $Mg^{2+}$  ions for  $RNAIi-RNAIIi$  formation, at 5 mM of  $Mg^{2+}$  ions. Maybe at high concentration of magnesium, the two sites can be populated at the same time but we cannot see them in our solution conditions (100  $\mu$ M  $Mg(OAc)_2$ ). An extremely negative cooperativity could also explain why only one magnesium ion at a time is detected to bind specifically to  $RNAIi-RNAIIi$ .

When using the AGILENT 6560, it has been possible to fragment both kissing complexes into their respective monomers. Nevertheless, it is complementary and interesting to use the LCQ Fleet to obtain an in-depth research of magnesium localization.

#### XIII.2.2.2. MS/MS using LCQ Fleet

The LCQ Fleet is the only instrument available to do MS<sup>n</sup> in the lab. With the LCQ, the activation time can be chosen. We decided to perform the experiment at 30 ms (long time) and 3 ms (short time). The short time is selected in order to have a conditions closer to the ones of the AGILENT 6560. The results obtained for TAR-R06 allowed us to understand the fragmentation pathway of kissing complexes at two different activation times. At 30 ms, base losses were predominant compared to 3 ms where both hairpin monomers are seen after activation. Figure 60 is showing full MS spectra and MS/MS spectra obtained at 3 ms of activation time and with 0 and 45 V applied. From the full MS spectra, we can see that the intensity of the detected  $[KC_{\text{TAR-R06}}]^{6-}$  is quite low, so when doing MS/MS the intensities are even lower. Same results are seen on the spectrum when 100  $\mu\text{M}$  of  $\text{Mg}(\text{OAc})_2$  is added. As the relative intensities of  $[\text{TAR}+1\text{Mg}^{2+}]^{3-}$  and  $[\text{R06}+1\text{Mg}^{2+}]^{3-}$  are too low, MS<sup>n</sup> cannot be performed.

The MS/MS study using the LCQ Fleet will not be described in more details in the main text of this manuscript. However, the results can be found on appendix A1.7 as supporting information.

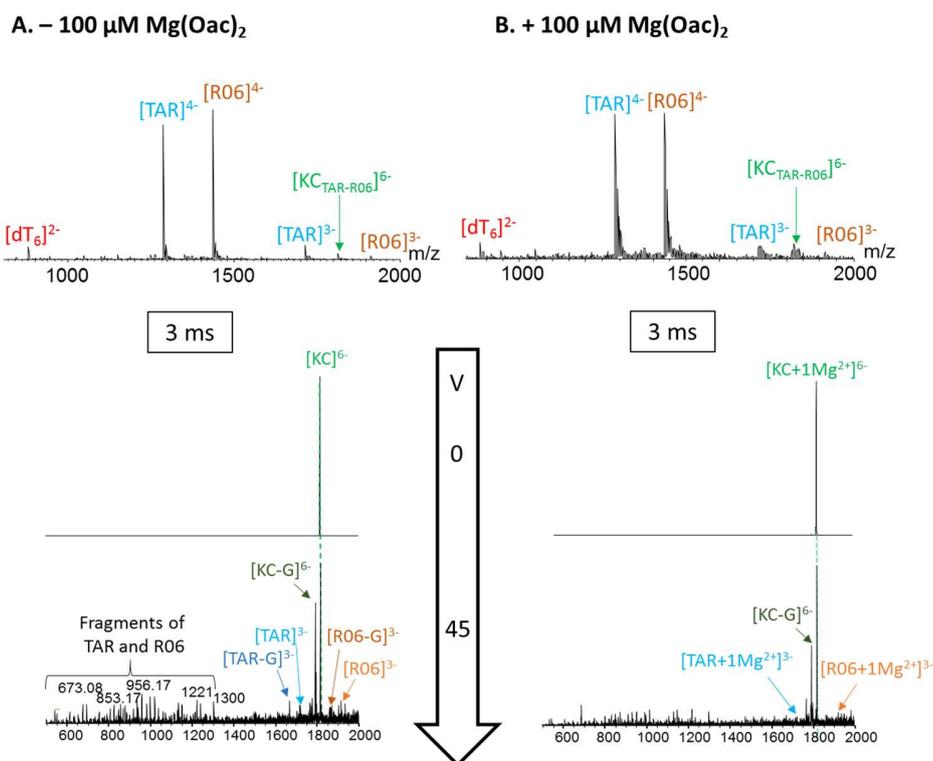


Figure 60: Full ESI-MS and MS/MS spectra of TAR-R06 using the LCQ Fleet. A. Solution was containing only 150 mM NH<sub>4</sub>OAc. B. 100 μM of Mg(OAc)<sub>2</sub> added to the solution. Acquisition of MS/MS spectra are done at 3 ms activation time. Hairpins were diluted at 20 μM each.

### XIII.3. Discussion

First, we have seen that whatever the magnesium salt used, it does not have an impact on the final results. Indeed, there is still the same number of specific cation determined, i.e. two for TAR-R06 KC and one for RNAIi-RNAIi KC. Also, we have demonstrated that the type of magnesium salt used does not have an influence on the proportion of magnesium specific detected under the peak of interest, which is [KC+1Mg<sup>2+</sup>]<sup>6-</sup>. We showed that the type of salts would not affect MS/MS results and that there are several critical points to take into account for kissing complexes CID fragmentation.

#### Critical points in MS/MS of kissing complexes

The experiments have shown that the way kissing complexes fragment depends on different parameters: the activation time linked to the type of instruments, the amount of energy given to the ions and the charge density.

First, the use of two different instruments have shown that the results depends a lot of the type of instrument. Depending on the activation time of the ions, the fragmentation channels will differ. With the

AGILENT 6560, which has a lower activation and fragmentation time, the dissociation of the kissing complexes into their monomers is favored. In comparison, at longer activation time ( $\geq 3$  ms), the base losses predominate. We propose a graphical representation to summarize these results (Figure 61) based on Gabelica *et al.*'s<sup>43</sup> representation.

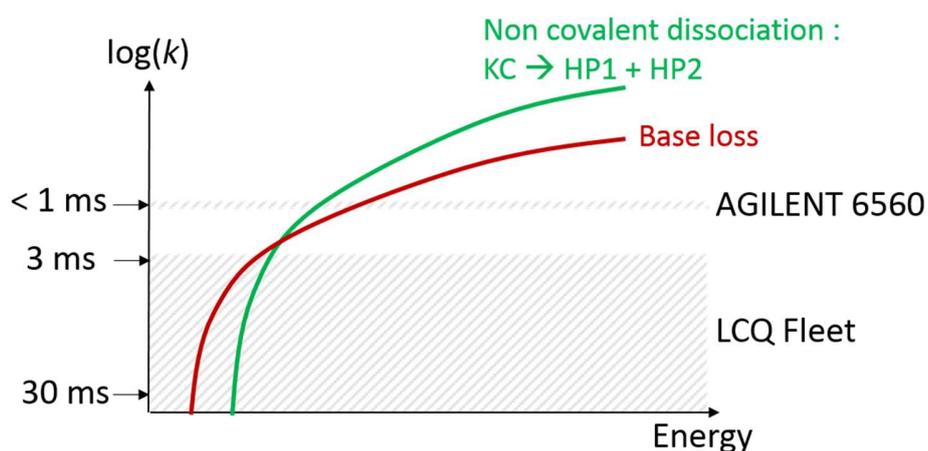


Figure 61: Schematic representation of the internal energy necessary to fragment a kissing complex depending on the activation time and the instrument. This scheme is not at scale.

More energy is needed to fragment the kissing complex at short activation time, because to detect fragments the rate of fragmentation ( $\log(k)$ ) must be higher. In all the experiments, fragments coming from non-covalent fragmentation and base losses are observed. In all cases, the bases which are lost first are guanines. This base loss can come from the 5' end of both TAR and R06.

We have learned also that charge density is very important for kissing complex fragmentation. Isolation of higher charge state for long kissing complexes will be necessary to observe fragmentation. The use of supercharging agent, like sulfolane, could be considered. However, the way sulfolane affects species charge states is still unclear and several studies on its link with ESI process are ongoing.<sup>42</sup> What could be of interest in our study, is to use the adduct cleaning methodology with solutions containing 1% sulfolane, to prove if there is a change or not in magnesium specific number and deduce then the fraction of specific and non-specific under the  $KC+1Mg^{2+}$  peak.

#### Localization of $Mg^{2+}$ on KC by MS/MS, and discussion in light of molecular modeling

Using MS/MS helped us to bring new insights in the localization of magnesium cations onto kissing complexes. The dissociation of the two kissing complexes in their respective hairpins plus  $Mg^{2+}$  let us draw several hypotheses. By showing that no differences in the amount of energy needed to fragment the

kissing complex with or without  $Mg^{2+}$  are seen, we can consider that two binding sites are present on both KC studied.

To obtain more information on the localization of the magnesium, molecular dynamics simulations were performed by Dr. Joséphine Abi-Ghanem, a former member of the team. Molecular dynamics (MD) in solution with  $Mg^{2+}$  cations were performed on TAR-R06 and RNAIi-RNAIi kissing complexes, and also on each monomer. Starting structures come from crystallographic structures (PDB: 2JLT<sup>17</sup> for TAR-R06 and 2BJ2<sup>15</sup> for RNAIi-RNAIi). Final structure were obtained after 300 ns of MD. The final structure for each hairpin is an average of 25 MDs of 10 ns. MDs were performed on the 6- charge state for both kissing complexes and on the 4- charge state for each monomer. The modellings were done with 2  $Mg^{2+}$  ions in each trajectory.  $Na^+$  ions were added to obtain the good charge state. Two criteria were used to define a binding site for  $Mg^{2+}$ : the ion density should be the highest and the contact between the base and  $Mg^{2+}$  has to be lower than 5 Å. The results are gathered in Figure 62.

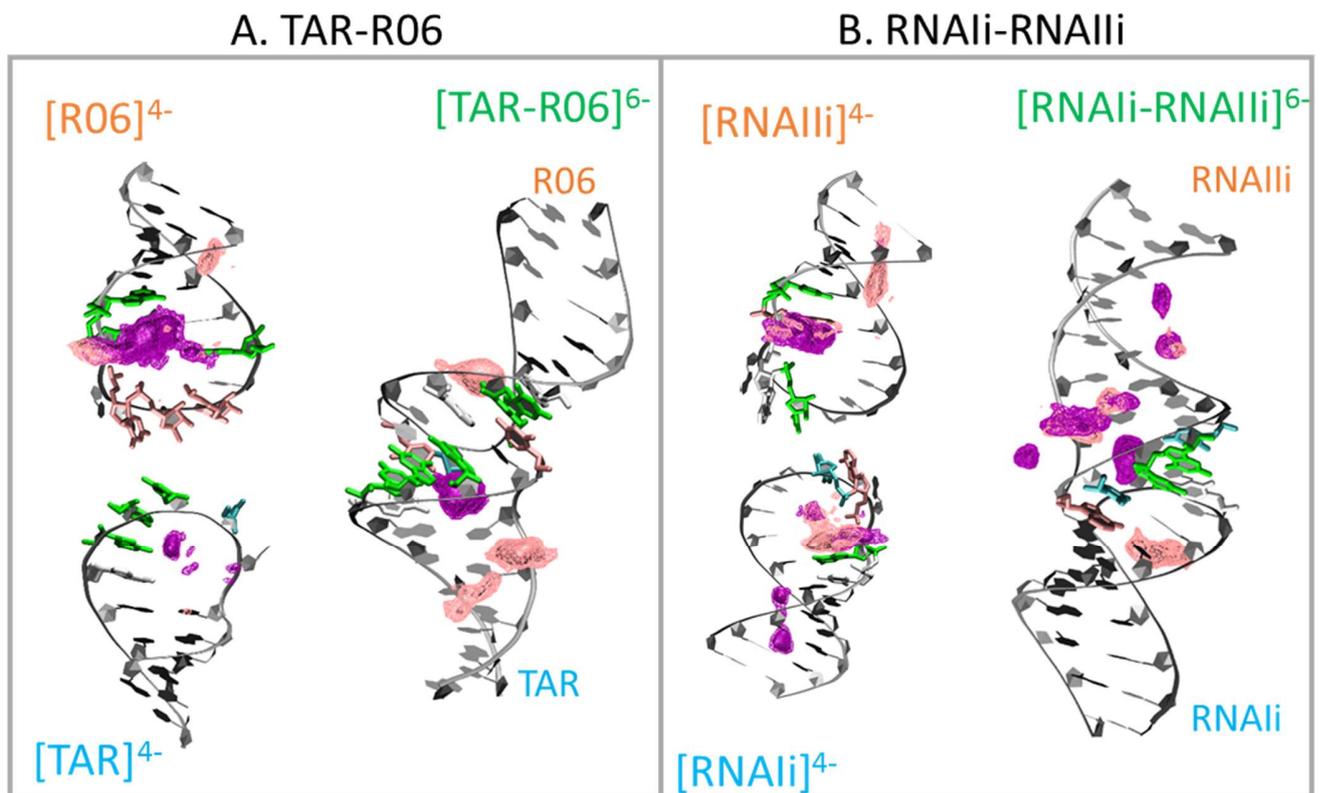


Figure 62: Resulting MD structures for A. TAR-R06 and B. RNAIi-RNAIi. MDs are performed during 300 ns. Force field and parameters used F99 + parmbsc0 +  $\chi OL3$ . In violet, density of the first  $Mg^{2+}$ . In pink, density of the second  $Mg^{2+}$ .

Figure 62A shows the results obtained for TAR, R06 and the KC. For the TAR hairpin, one magnesium is positioned in the loop, close to the three G's. For R06, the position of the magnesium is fluctuating more, but is mainly between the phosphates of the cytosines presents in the loop, plus in contact with the G's in the stem. In the TAR-R06 kissing complex, one of the two magnesium ions is not moving (violet one) and this one corresponds to the one on TAR. The other  $Mg^{2+}$  (in pink) is moving a lot between the two hairpins; one of the main sites corresponds to the one in R06, but it also is occasionally located on TAR. These results confirm that two magnesium are specifically pertaining to TAR-R06 motif. Also, one site seems to be specific and the other can be qualified as "preferential". As the binding sites on the kissing complex are the same as on the hairpins alone, it let us think that magnesium is binding first to the two hairpins and then there is formation of the kissing complex. This could explain why the KC can form without  $Mg^{2+}$ .

Figure 62B presents the results on RNAIi, RNAIIi, and the corresponding KC. The MDs on the individual hairpin indicate that  $Mg^{2+}$  is binding on each hairpin.  $Mg^{2+}$  is contacting the guanines in RNAIIi loop and cytosines present in RNAIi loop. Now, for  $KC_{RNAIi-RNAIIi}$ , we can see that one  $Mg^{2+}$  ion is moving only in RNAIIi (in violet) and the other one (in pink) is present on both hairpins. On RNAIi, one can describe the presence of one potential binding site as it is the main site of one of  $Mg^{2+}$  (in pink). As one of the  $Mg^{2+}$  is moving only on RNAIIi loop, it can constitute a preferential binding pocket formed between the stem and the loop. As this same  $Mg^{2+}$  is present only on RNAIIi, it agrees with the hypothesis that the two binding sites cannot be populated at the same time.



## Chapter 2. Mass spectrometry to screen kissing complex affinities

### XIV. Introduction

In the previous chapter, we showed that mass spectrometry is a powerful tool to study the interaction between kissing complexes and magnesium. Here, we use mass spectrometry as a screening tool to rank kissing complexes differing by one base pair. We want to quantify the impact of a modification of a non-canonical base pair, present in the kissing loop, on the kissing complex stability. The G●U wobble base pair can be more stabilizing than a canonical base pair.<sup>44</sup> Here too magnesium can influence the KC stability, and therefore some experiments were carried out with magnesium to check for its influence. This work contributed to the general understanding of why a non-canonical base pair has an impact on KC stability. Mass spectrometry and other biophysics results presented hereafter were recently published in a *ChemPhysChem*<sup>45</sup> paper, in combination with molecular dynamics and thermodynamics calculations performed by Josephine Abi-Ghanem.

In the following section, we will detail the experimental part of this work, focusing on the mass spectrometry and biophysical study. The *ChemPhysChem*<sup>45</sup> accepted manuscript is shown in Appendix A2.

### XV. Kissing complexes used

In some kissing complexes, the presence of a non-canonical base pair in the kissing loop motif makes it more stable than with a canonical base pair. This is the case for example of R06 where the closing base pair G●A makes the kissing complex formed with TAR more stable than with a G-C base pair.<sup>30</sup> In this study we used a system called K1-K1' previously described by Durand et al.<sup>46</sup> This system has been selected by SELEX (Systematic Evolution of Ligands by Exponential Enrichment), which generates a library of oligonucleotides and selects an aptamer (molecule with a high affinity and specificity for its target) by several rounds *in vitro*. The SELEX buffer contains 20 mM HEPES and several salt concentrations (20 mM NaAc, 3 mM Mg(OAc)<sub>2</sub> and 140 mM KAc). The aptamer found, K1', has a Uracil on position 7 which form a G●U base pair with G<sub>12</sub> of its target K1, and a Cytosine on position 12 (Figure 63A). To assess the importance of this base pair for the kissing complex stability, mutations are done on bases 7 and 12 (last bases of the loop) of K1'. The names of the mutated hairpins are given by the mutation on those two bases N<sub>7</sub>N<sub>12</sub>, so we have: UC (original one), CC, CU and UU (Figure 63). It will be then possible to see the impact of the localization of the G●U base pair (5'end for UC vs 3'end CU), of a double G●U base pair at each end (UU) and make comparison with a Watson-Crick base pair (CC). After the SELEX selection, the stability of the different kissing complexes formed is ranked as follow: UC > CC > CU > UU.

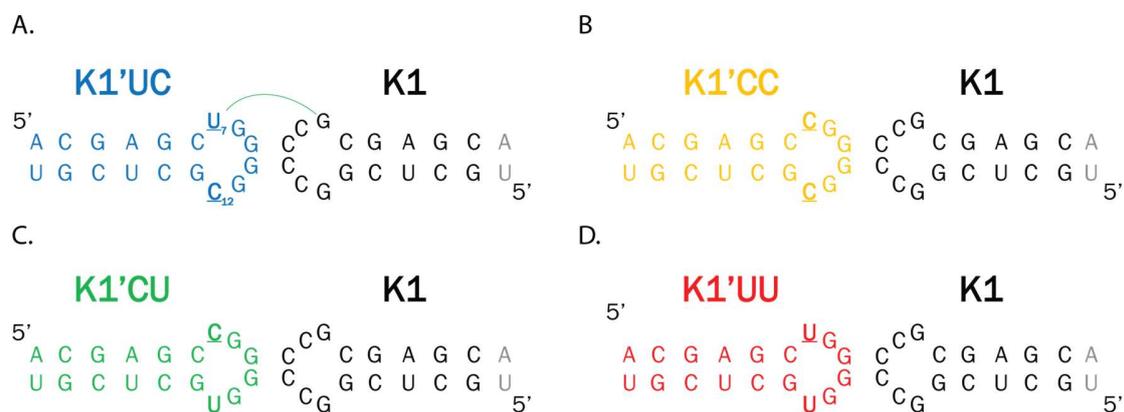


Figure 63: Sequences of the different variant of K1-K1': A) UC kissing complex, B) CC kissing complex, C) CU kissing complex and D) UU kissing complex.

## XVI. Results

### XVI.1. Mass spectrometry experiments

We used native mass spectrometry to have information on the relative stability of the four kissing complexes. As a consequence and for simplification, we assume that the response at the ESI process of all four KCs is the same, this will allow us to have access to the experimental free energy  $\Delta G^{\circ}_{assoc}$ . Based on the equilibrium (37), it is possible to express  $\Delta G^{\circ}_{assoc}$  (38).



$$\Delta G^{\circ}_{assoc} = -RT \ln \left( \frac{[KC]}{[K1][Ligand]} \right) \quad (38)$$

Where the Ligand are UC, CC, CU or UU. If the response factors of each KCs are equal, the peak area will be equal to the concentration of the species. We will then a proportional link between  $\Delta G^{\circ}_{assoc}$  and the ratio of areas, as follows:

$$\Delta G^{\circ}_{assoc} \propto -\ln \left( \frac{A(KC)}{A(K1)} \right) \quad (39)$$

Where  $A(KC)$  and  $A(K1)$  are the peak area of the KC and K1 respectively. To calculate  $A$ , all observed charge states are summed.

We used the AGILENT 6560 to perform the experiments, using the HARD tune (see material and method), a TEGD of 12 or 20 V and a fragmentor of 350 V. The experiments are carried out using solutions of 10  $\mu$ M of each hairpin, 150 mM  $\text{NH}_4\text{OAc}$  and 2  $\mu$ M  $\text{dT}_6$ .

The first experiments are done using K1 and the four different complementary hairpins (UC, CC, CU and UU) with various concentrations of  $\text{Mg}(\text{OAc})_2$  i.e. 0, 200 and 800  $\mu\text{M}$ . The results are shown on Figure 64.

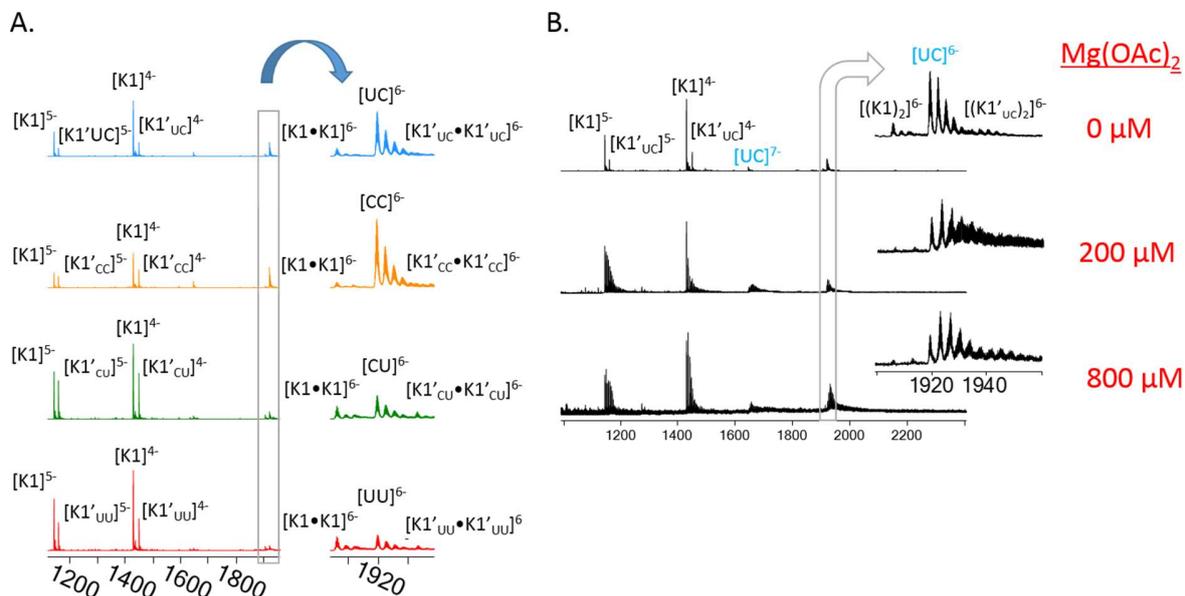


Figure 64: ESI-MS spectra of the original sequences. A) MS spectra of UC (blue), CC (yellow), CU (green) and UU (red) without  $\text{Mg}(\text{OAc})_2$ . In addition, a zoom on the 1880-1960  $m/z$  area. B) MS spectra of  $\text{K1-K1}'_{\text{uc}}$  at different  $\text{Mg}(\text{OAc})_2$  concentrations.

For all four systems, heterodimers (kissing complexes) are detected at two different charge states: 7- and 6- (which is the major one). In addition, homodimers of both K1 and the K1' are detected at charge state 6- (Figure 64, A). These results revealed the formation of another equilibrium that we have to take into account: Hairpin + Hairpin  $\rightleftharpoons$  HD (homodimer). The spectra of each individual hairpin always show the presence of a certain amount of homodimers (data not shown). By comparing the global intensity of homodimers when the hairpins are alone or mixed, we have seen that the intensity of homodimers decreases when the complementary hairpin is added. From this, we can hypothesize that the equilibrium  $\text{HP} + \text{HP} \rightleftharpoons \text{HD}$  is displaced towards the heterodimer formation.

K1 and K1' are close on the  $m/z$  scale, and some adducts are superimposed. This is even more problematic when magnesium is added to the solution (Figure 64B), and complicates the integration of the peaks. To overcome the problem, we truncated K1 (called  $\text{K1}_{\text{short}}$ ) by deleting the terminal base pair A-U, to allow a better mass separation. The experiments were repeated with  $\text{K1}_{\text{short}}$  and the extraction of the peak areas could be performed. Then, we calculated  $-\ln\left(\frac{A_{\text{KC}}}{A_{\text{K1sho}}}\right)$ . The calculations were done at least once on the original sequence for comparison. The results are presented on Figure 65.

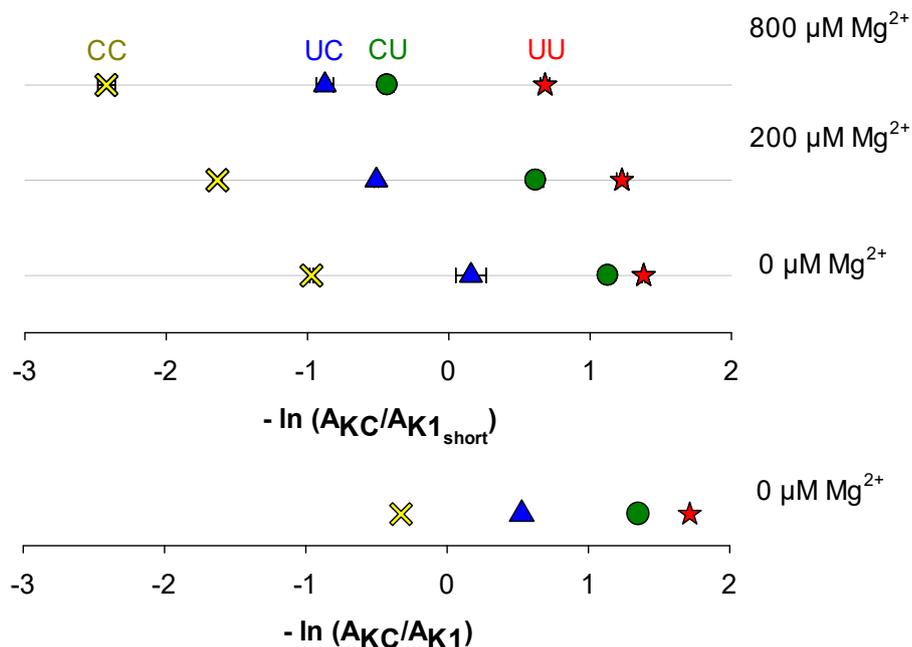


Figure 65: Relative quantification of the peak areas of the KC over the free  $K1_{short}$  or  $K1$ . Most stable complexes are on the left, the less stable are on the right. CC in yellow, UC in blue, CU in green and UU in red.

Similarly to what we found for TAR-R06 and RNAIi-RNAIli kissing complexes, heterodimers can form without magnesium. Also, the stability order found with  $K1$  or  $K1_{short}$  is the same, which shows that deleting the last A-U base pair of  $K1$  does not have a major impact on the kissing complex stability. In all the conditions tested, the most stable kissing complex is CC, followed by UC, CU and UU. Magnesium increases the stability of all KC but we still observe the same stability order. We observe that Watson-Crick base pairs are more stable than a G•U base pair at the 5' end. A kissing complex with a wobble base pair at the 5' end seems also more stable than when this base pair is at the 3' end (UC vs CU).

To validate the MS results, we used other biophysical techniques to characterize the four kissing complexes: UV melting experiments and Surface Plasmon Resonance (SPR).

## XVI.2. UV melting experiments

Thermal denaturation experiments are performed to check if thermal stability in solution of the kissing complexes reflects the MS results. UV melting experiments are done in the MS buffer, i.e. 150 mM  $\text{NH}_4\text{OAc}$  and two concentrations of  $\text{Mg}(\text{OAc})_2$  are tested: 0 and 800  $\mu\text{M}$ . The experiments are done using  $K_{short}$  and all four complementary hairpins. Results are presented on Figure 66.

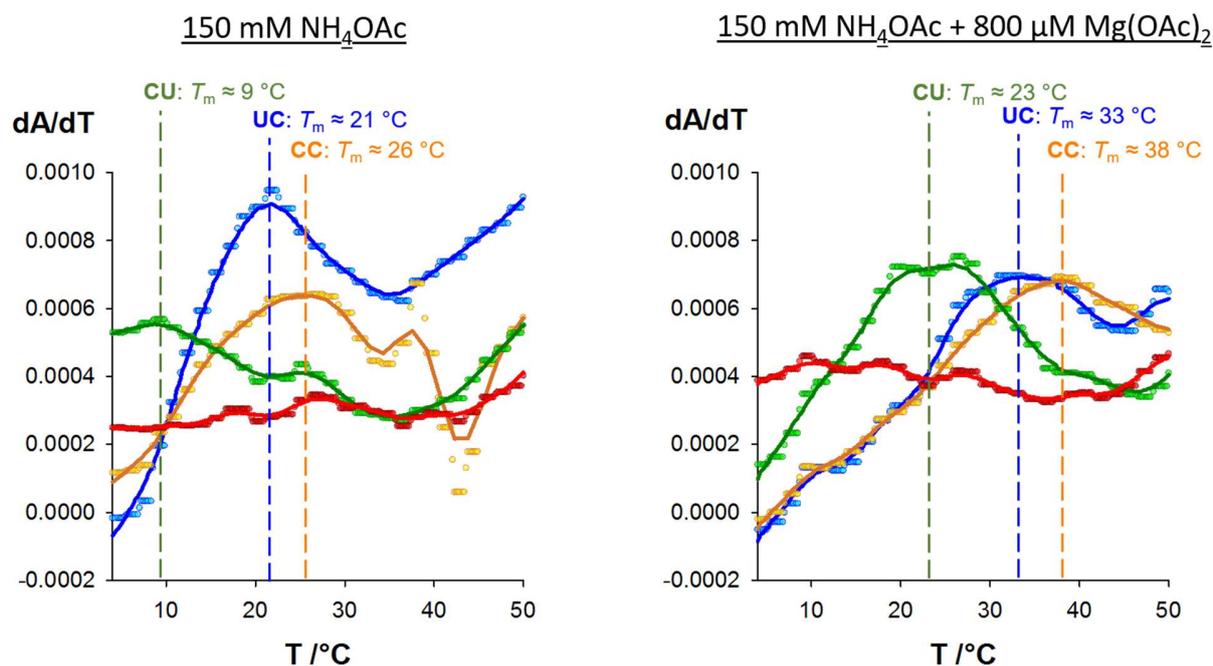


Figure 66: UV melting of UC (blue), CC (yellow), CU (green) and UU (red): first derivative of the absorbance at 260 nm as a function of the temperature. The local maximum on the curves represent the dissociation of the KC into hairpins. At  $T > 50^\circ\text{C}$  (not shown on the graphs), the maximum corresponds to the melting of the individual hairpins.

At 0 and 800 μM, the stability ranking is  $\text{CC} > \text{UC} > \text{CU} > \text{UU}$ . For UU, accurate value of  $T_m$  cannot be determined as melting is not observed. When magnesium is added, the stability of each hairpin increases, keeping the same stability ranking. These results perfectly agree with MS results.

### XVI.3. Surface Plasmon Resonance Experiments

Surface Plasmon Resonance (SPR) is another biophysical technique allowing the determination of binding constants between two partners. SPR is based on the measurement of refractive index changes. Briefly, one partner (called the ligand) is immobilized on a sensorchip surface and the other one (called the analyte) is injected in the flow cell. When there is binding between the analyte and the ligand, the change in mass will be detected using the total internal reflection of a laser light on the surface. The angle at which there is plasmon resonance will change due to the change in refractive index caused by the binding of the two partners.

By using SPR, we would like to obtain binding constants between the different partners and then rank the kissing complexes in function of their  $K_D$ s.

## Experimental setup

Biotinylated K1 is immobilized on a streptavidin sensorchip. A BIAcore T200 (GE Healthcare, Uppsala, Sweden) is used. The MS buffer (150 mM  $\text{NH}_4\text{OAc}$ ) was used for all experiments containing 0, 1 or 3 mM  $\text{Mg}(\text{OAc})_2$ . All complementary hairpins, i.e. UC, CC, CU and UU are diluted and injected from 4  $\mu\text{M}$  to 31.25 nM with a flow rate of 20  $\mu\text{L}\cdot\text{min}^{-1}$ . Multi-cycle kinetics experiments are performed, meaning that each concentration of the analyte is injected separately as a single run. Regeneration is necessary between each cycle to remove all still bound analytes. Here, we used EDTA at 3 mM.

The BIAeval 3.1 software, Table Curve 2D 5.01 (Systat Software, Inc.) and Sigma Plot 12.5 are used to fit and analyze the data.

## Results

First the resulting sensorgrams are fitted using a 1:1 model. Another variable parameter, called bulk signal (RI), is considered. This RI can correct the signal when there is a drop due to a difference in the refractive index between the flow buffer and the analyte solution. The results are presenting on Figure 67.

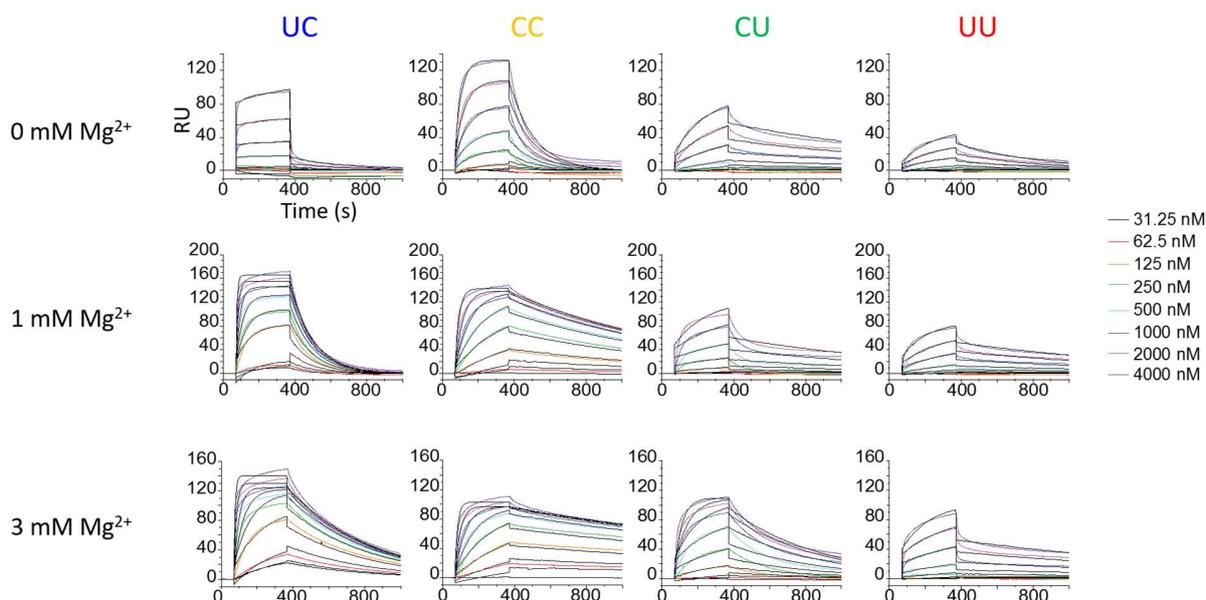


Figure 67: SPR results. Sensorgrams obtained for the titration of hairpin K1 attached to the sensorship, and each of the four variants of the hairpin K1' injected in the microfluidics system. The fits with the 1:1 model are in black.

The BIAeval fits corresponding to a 1:1 model (in black) are not fitting the data well, which means that the binding is more complicated. Also, by looking at the association phase of each curve, it seems that they can be characterized by two components: a fast association and a slower one. If there are several rate constants, it means that several equilibria are detected. The curves are changing in presence of magnesium, meaning that magnesium affects the binding constants. One explanation of the double association rate constants can be due to the presence of homodimers detected by MS. It is possible that there are competitions between the formation of heterodimers and the dissociation of homodimers. Maybe the curves are representatives of a heterogeneity of both ligand and analyte.

Even if these data could not be fitted by a 1:1 model, it is possible to extract the  $R_{eq}$  values at which the steady state is reached. To obtain the  $R_{eq}$ , we had to extrapolate the association curves to an infinite time. The association parts of the sensorgrams were analyzed with Table curve 2D v5.01 (Systat Software Inc.), and fitted with two first-order formation kinetics (equation:  $y = A + B(1 - e^{-Cx}) + D(1 - e^{-Ex})$ ). The two first order model was chosen because of the two phases observed during the association step. The total extrapolated  $R_{eq}$  was obtained by summing the fitting parameters  $A+B+D$ . Then  $R_{eq}$  is plotted as a function of the analyte concentrations, and fitted with a 1:1 Langmuir binding model (Equation:  $R_{eq} = \frac{R_{max}x}{K_D+x}$ ). The results are shown on Figure 68.

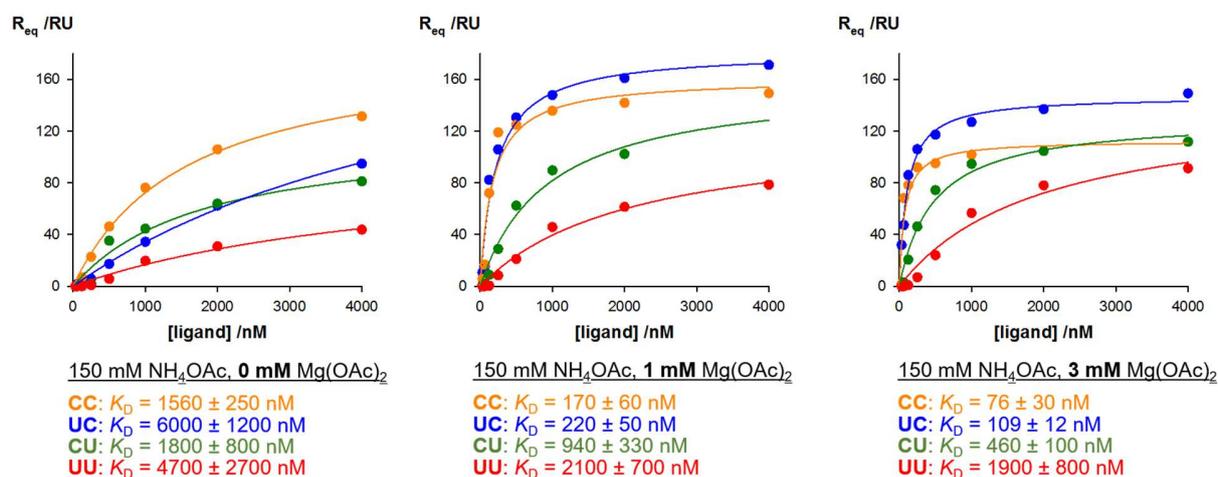


Figure 68:  $K_D$  determination using  $R_{eq}$  extrapolation. Equilibrium dissociation constants obtained by analyzing the steady-state of SPR experiments. The steady-state RU of each sensorgram was obtained by fitting the association part with two first-order association rates, and  $R_{eq}$  is calculated as the sum of the RUs extrapolated at infinite time.  $R_{eq}$  is plotted as a function of the ligand concentration, and fitted by a 1:1 binding model to obtain the  $K_D$  values indicated on the figure.

Based on the  $K_D$ s values, we have obtained a stabilization order: CC > UC > CU > UU, i.e. at 1 and 3 mM of Mg(OAc)<sub>2</sub>, but not at 0 μM. Without magnesium, the SPR results extrapolated to steady state mean that UC would be the least stable kissing complex, which is not in agreement with the MS and UV melting experiments. These results are obtained using an extrapolation at the infinite time to have  $R_{eq}$  which include an error that we should consider. Also, the standard error found on each value of  $K_D$ s are > 20%, meaning that the values obtained are giving an order of magnitude of the real  $K_D$ .

$K_D$  values are extracted by fitting thanks to the curvature of the curves. However, we also have an information about the amount of ligands bounds given by  $R_{eq,max}$ .  $R_{eq,max}$  is different for the four different systems, meaning that different amount of ligands is binding, although the same amount of analyte is immobilized. At 3 mM of magnesium, based on  $R_{eq,max}$  the ranking found is similar to the one found after SELEX, i.e. UC > CC ≥ CU > UU. The amount bound is not directly proportional to the  $K_D$  values obtained. The hypothesis here is again that the 1:1 model is too simple for the systems studied. The involvement of homodimers revealed by mass spectrometry may be part of the complexity.

## XVII. Discussion and conclusion

In this chapter, we have used mass spectrometry to determine the stabilization order of several kissing complexes differing from only one or two bases. We have shown that MS is complementary to other biophysical techniques such UV melting and SPR. Also, although the MS readout is done in the gas phase, the MS relative stabilities are comparable to the ones obtained in solution.

The results have shown that the Watson-Crick base pair G-C makes the kissing complex more stable than with another Watson-Crick base pair. It has been shown in the case of these sequences, that the G•U base pair in the 5'-end of the kissing loop is also leading to more stable complexes (almost as stable as the GC base pair) than at the 3'-end.

MS can thus be used as a screening tool to check for stability (like we did here). However, in biology what is important also is the binding kinetics, how fast a molecule can bind to another and be dissociated. For drug design these data is of great importance as the drug as to live long enough to assess its function but it also has to be removed.

## References

- (1) Thomas, B., and Akoulitchev, A. V. (2006) Mass spectrometry of RNA. *Trends Biochem. Sci.* **31**, 173–181.
- (2) Sannes-Lowery, K. a, Hu, P., Mack, D. P., Mei, H. Y., and Loo, J. a. (1997) HIV-1 Tat peptide binding to TAR RNA by electrospray ionization mass spectrometry. *Anal. Chem.* **69**, 5130–5135.
- (3) Schneeberger, E. M., and Breuker, K. (2017) Native Top-Down Mass Spectrometry of TAR RNA in Complexes with a Wild-Type tat Peptide for Binding Site Mapping. *Angew. Chemie - Int. Ed.* **56**, 1254–1258.
- (4) Hagan, N., and Fabris, D. (2003) Direct mass spectrometric determination of the stoichiometry and binding affinity of the complexes between nucleocapsid protein and RNA stem-loop hairpins of the HIV-1 TAR-recognition element. *Biochemistry* **42**, 10736–10745.
- (5) Hagan, N. A., and Fabris, D. (2007) Dissecting the protein-RNA and RNA-RNA interactions in the nucleocapsid-mediated dimerization and isomerization of HIV-1 stemloop 1. *J. Mol. Biol.* **365**, 396–410.
- (6) Turner, K. B., Hagan, N. A., and Fabris, D. (2007) Understanding the isomerization of the HIV-1 dimerization initiation domain by the nucleocapsid protein. *J. Mol. Biol.* **369**, 812–28.
- (7) Draper, D. E. (2004) A guide to ions and RNA structure. *RNA* **10**, 335–343.
- (8) Woodson, S. A. (2005) Metal ions and RNA folding: A highly charged topic with a dynamic future. *Curr. Opin. Chem. Biol.* **9**, 104–109.
- (9) Konermann, L. (2017) Addressing a Common Misconception: Ammonium Acetate as Neutral pH “Buffer” for Native Electrospray Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **28**, 1827–1835.
- (10) Giessing, A. M. B., and Kirpekar, F. (2012) Mass spectrometry in the biology of RNA and its modifications. *J. Proteomics* **75**, 3434–3449.
- (11) Kebarle, P., and Verkerk, U. H. (2009) Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass Spectrom. Rev.* **28**, 898–917.
- (12) Marchand, A., and Gabelica, V. (2014) Native electrospray mass spectrometry of DNA G-quadruplexes in potassium solution. *J. Am. Soc. Mass Spectrom.* **25**, 1146–1154.
- (13) Marchand, A., and Gabelica, V. (2016) Folding and misfolding pathways of G-quadruplex DNA. *Nucleic*

*Acids Res.* 44, 10999–11012.

(14) Chang, K. Y., and Tinoco, I. (1994) Characterization of a “kissing” hairpin complex derived from the human immunodeficiency virus genome. *Proc. Natl. Acad. Sci. U. S. A.* 91, 8705–9.

(15) Lee, A. J., and Crothers, D. M. (1998) The solution structure of an RNA loop–loop complex: the ColE1 inverted loop sequence. *Structure* 6, 993–1007.

(16) Jossinet, F., Paillart, J. C., Westhof, E., Hermann, T., Skripkin, E., Lodmell, J. S., Ehresmann, C., Ehresmann, B., and Marquet, R. (1999) Dimerization of HIV-1 genomic RNA of subtypes A and B: RNA loop structure and magnesium binding. *RNA* 5, 1222–34.

(17) Lebars, I., Legrand, P., Aimé, A., Pinaud, N., Fribourg, S., and Di Primo, C. (2008) Exploring TAR-RNA aptamer loop-loop interaction by X-ray crystallography, UV spectroscopy and surface plasmon resonance. *Nucleic Acids Res.* 36, 7146–7156.

(18) Gregorian, R. S., and Crothers, D. M. (1995) Determinants of RNA hairpin loop-loop complex stability. *J. Mol. Biol.* 248, 968–984.

(19) Balthasart, F., Plavec, J., and Gabelica, V. (2013) Ammonium ion binding to DNA G-quadruplexes: Do electrospray mass spectra faithfully reflect the solution-phase species? *J. Am. Soc. Mass Spectrom.* 24, 1–8.

(20) Porrini, M., Rosu, F., Rabin, C., Darré, L., Gómez, H., Orozco, M., and Gabelica, V. (2017) Compaction of Duplex Nucleic Acids upon Native Electrospray Mass Spectrometry. *ACS Cent. Sci.* 3, 454–461.

(21) Potier, N., van Dorselaer, A., Cordier, Y., Roch, O., and Bischoff, R. (1994) Negative electrospray ionization mass spectrometry of synthetic and chemically modified oligonucleotides. *Nucleic Acids Res.* 22, 3895–3903.

(22) Bai, Y., Greenfeld, M., Travers, K. J., Chu, V. B., Lipfert, J., Doniach, S., and Herschlag, D. (2007) Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *J. Am. Chem. Soc.* 129, 14981–8.

(23) Romani, A. M. P. (2011) Cellular magnesium homeostasis. *Arch. Biochem. Biophys.* 512, 1–23.

(24) Grubbs, R. D. (2002) Intracellular magnesium and magnesium buffering. *BioMetals* 15, 251–259.

(25) Pyle, A. M. (2002) Metal ions in the structure and function of RNA. *J. Biol. Inorg. Chem.* 7, 679–690.

- (26) Bowman, J. C., Lenz, T. K., Hud, N. V., and Williams, L. D. (2012) Cations in charge: Magnesium ions in RNA folding and catalysis. *Curr. Opin. Struct. Biol.* 22, 262–272.
- (27) Sun, J., Kitova, E. N., Wang, W., and Klassen, J. S. (2006) Method for distinguishing specific from nonspecific protein-ligand complexes in nanoelectrospray ionization mass spectrometry. *Anal. Chem.* 78, 3010–8.
- (28) Zbio.net [http://www.zbio.net/eng/scripts/01\\_16.html](http://www.zbio.net/eng/scripts/01_16.html).
- (29) Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., and Hofacker, I. L. (2008) The Vienna RNA websuite. *Nucleic Acids Res.* 36, 70–74.
- (30) Ducongé, F., Di Primo, C., and Toulmé, J. J. (2000) Is a closing “GA pair” a rule for stable loop-loop RNA complexes? *J. Biol. Chem.* 275, 21287–21294.
- (31) Keele, B. B., McCord, J. M., and Fridovich, I. (1970) Superoxide Dismutase from *Escherichia coli* B: a new manganese-containing enzyme. *J. Biol. Chem.* 245, 6175–6181.
- (32) (1991) Ionic radii in crystals 14–16.
- (33) Bock, C. W., Katz, A. K., Markham, G. D., and Glusker, J. P. (1999) Manganese as a replacement for magnesium and zinc: Functional comparison of the divalent ions. *J. Am. Chem. Soc.* 121, 7360–7372.
- (34) Sansonetti, J. E., and W.C. Martin, W. C. Basic Atomic Spectroscopic Data at <https://www.nist.gov/pml/handbook-basic-atomic-spectroscopic-data>.
- (35) Leonarski, F., D’Ascenzo, L., and Auffinger, P. (2016) Binding of metals to purine N7 nitrogen atoms and implications for nucleic acids: A CSD survey. *Inorganica Chim. Acta* 452, 82–89.
- (36) Leonarski, F., D’Ascenzo, L., and Auffinger, P. (2016) Mg<sup>2+</sup> ions: do they bind to nucleobase nitrogens? *Nucleic Acids Res.* 45, 987–1004.
- (37) Sannes-Lowery, K. A., Griffey, R. H., and Hofstadler, S. A. (2000) Measuring dissociation constants of RNA and aminoglycoside antibiotics by electrospray ionization mass spectrometry. *Anal. Biochem.* 280, 264–271.
- (38) Griffey, R. H., Sannes-Lowery, K. A., Drader, J. J., Mohan, V., Swayze, E. E., and Hofstadler, S. A. (2000) Characterization of low-affinity complexes between RNA and small molecules using electrospray ionization mass spectrometry. *J. Am. Chem. Soc.* 122, 9933–9938.

- (39) Gabelica, V., Rosu, F., and De Pauw, E. (2009) A simple method to determine electrospray response factors of noncovalent complexes. *Anal. Chem.* *81*, 6708–15.
- (40) Kuzmic, P. (1996) Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.* *237*, 260–273.
- (41) Misra, V. K., and Draper, D. E. (1998) On the Role of Magnesium Ions in RNA Stability. *Biopolymers* *48*, 113–135.
- (42) Metwally, H., McAllister, R. G., Popa, V., and Konermann, L. (2016) Mechanism of Protein Supercharging by Sulfolane and m -Nitrobenzyl Alcohol: Molecular Dynamics Simulations of the Electrospray Process. *Anal. Chem.* *88*, 5345–5354.
- (43) Gabelica, V., and De Pauw, E. (2002) Comparison of the collision-induced dissociation of duplex DNA at different collision regimes: Evidence for a multistep dissociation mechanism. *J. Am. Soc. Mass Spectrom.* *13*, 91–98.
- (44) Varani, G., and McClain, W. H. (2000) The G·U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep.* *1*, 18–23.
- (45) Abi-Ghanem, J., Rabin, C., Porrini, M., Dausse, E., Toulmé, J. J., and Gabelica, V. (2017) Electrostatics Explains the Position-Dependent Effect of G·U Wobble Base Pairs on the Affinity of RNA Kissing Complexes. *ChemPhysChem* *18*, 2782–2790.
- (46) Durand, G., Dausse, E., Goux, E., Fiore, E., Peyrin, E., Ravelet, C., and Toulmé, J. J. (2016) A combinatorial approach to the repertoire of RNA kissing motifs; Towards multiplex detection by switching hairpin aptamers. *Nucleic Acids Res.* *44*, 4450–4459.
- (47) Mcluckey, S. A., Berkel, J. Van, and Glish, L. (1992) Tandem Mass Spectrometry of Small Multiply Charged Oligonucleotides. *J. Am. Soc. Mass Spectrom.* *3*, 60–70.

## 4. RESULTS AND DISCUSSION

### PART II: STUDY OF RNA AND DNA COMPLEXES BY ION MOBILITY SPECTROMETRY-MASS SPECTROMETRY REVEALS COMPACTION IN THE GAS PHASE

## Chapter 3. DNA and RNA duplexes structures in the gas phase

### XVIII. Introduction

We have seen in the Part I of the results that the ions surrounding nucleic acids and forming the so-called “ion atmosphere” are important for nucleic acids structures. Indeed, they reduce the repulsion between the phosphates which allow DNA or RNA to adopt a specific structure.<sup>1-4</sup> If the ionic environment is important for nucleic acids structure, the binding of specific cations and ligands is also important to assure their biological role. In some cases, a change in conformation happens upon ligand binding and is responsible of the biological answer, like in RNA riboswitches.<sup>5-7</sup>

In mass spectrometry, molecules are desolvated, ionized and analyzed in the gas phase. Neither the water nor the ion atmosphere is present anymore. So, the questions asked here are “what is the structure of biomolecules in the gas phase? To what extent do they keep a memory of their solution structure?”. *Native* mass spectrometry alone cannot bring detailed knowledge about the structures in the gas phase. Ion mobility coupled to mass spectrometry (IM-MS) adds a new dimension which is the separation of the molecules according to their shape, based on their electrophoretic mobility in a buffer gas.<sup>8</sup> A lot of studies were focused on proteins and how they behave in the gas phase. It has been shown that proteins keep a long-time memory of their *native* state in the gas phase.<sup>9</sup> However, proteins can undergo rearrangement, collapsing and sometimes unfolding in the gas-phase.<sup>9,10</sup> Sharon and co-workers have shown that in Waters Synapt™ instruments, the T-wave height and the velocity have an impact on the folding of proteins, like ovalbumin or Gβγ.<sup>11</sup> When the T-wave height is increased it will induce a compaction of the overall structure. On the contrary, the inverse phenomenon is observed when increasing the T-wave velocity. By this work, they show the importance of the different instrument parameters onto protein structures.

Alternatively, by doing “collision induced unfolding” experiments, one can follow the evolution of the CCSs of a specific structure as a function of pre-IMS activation and deduce its gas-phase unfolding patterns. Such experiments were done on the BSA protein<sup>12</sup>, avidin and aldolase<sup>13</sup>, or ubiquitin<sup>14</sup>. Structure of proteins will be different from solution structure and will depend mostly on the energy given to the ion. Our question for the following work is “Do nucleic acids structures behave like proteins when spraying in the gas phase?”. Our starting point will be the study of DNA and RNA duplexes to understand the behavior of nucleic acids in the gas phase. Investigations on DNA duplexes in the gas phase using mass spectrometry have been performed since the early 2000's. These studies suggest that Watson-Crick base pairing and stacking are partially preserved from solution to the gas phase.<sup>15-17</sup> Study of (GC)<sub>n</sub> duplexes by IM-MS has

been performed by the Bowers lab. They showed, using short scale molecular dynamics, that short duplexes (<16-mer) structures look like the A-helix and that longer duplexes (>18-mer) adopt the B-helix structure in the gas phase. They reveal also that the longer duplexes better preserve their structure.<sup>18,19</sup> However, all these studies have been performed on solutions containing MeOH, and for the IM-MS study NH<sub>4</sub>OH was added to the solutions, which can be considered as non-*native* conditions. The resulting charge states were high (1 negative charge per couple of base pairs). Another study performed by Burmistrova and co-workers have shown that duplexes with a high percentage of GC base pair are more stable and that GC tracts are more compact than AT-tracts.<sup>20</sup>

In this chapter, we will describe the analysis of DNA and RNA duplexes by *Native* IM-MS to probe conformational details. Studying duplexes and comparing DNA to RNA will be a starting point, before analyzing RNA-RNA complexes. This work served as basis for molecular dynamics simulations and is the object of a recently published paper by Porrini M, Rosu F, Rabin C, et al in *ACS Central Science* (appendix A3) on which we showed that duplex nucleic acids undergo compaction from the solution to the gas phase when analyzed by *Native* IM-MS.

## XIX. Results and discussion

### XIX.1. ESI-MS spectra

#### XIX.1.1. Experimental aspects

We performed all the experiments on the AGILENT 6560 IMS-Q-TOF. The mobility cell is a drift tube filled with helium. The instrument was tuned to have the softest conditions where the peak of the duplex without NH<sub>4</sub><sup>+</sup> adducts is observed.

For this study, several DNA and the corresponding RNA are used. The detailed results of the sequences presented in Table 10 will be shown in this chapter. The first sequences are based on the Dickerson-Drew dodecamer<sup>21</sup> (here noted 12-d<sub>66</sub> and 12-d<sub>100</sub>). Derived sequences made by mixing 12-d<sub>66</sub> and 12-d<sub>100</sub> are also used in order to have a range of duplexes from 12-base pair (bp) to 36-bp duplexes. The corresponding RNA duplexes are also studied. We used the duplexes from Lippens et al<sup>22</sup>. Only 14ab is used as DNA and RNA, noted d<sub>14ab</sub> and r<sub>14ab</sub> respectively. We also analyzed by IM-MS the sequences r<sub>17R1-17R2</sub>, d<sub>17R1-17R2</sub> and r<sub>20R1-20R2</sub> and d<sub>20R1-20R2</sub>. Sequences are summarized in Table 10.

	Name + Sequences	%GC
12 bp	12-d <sub>66</sub> : (dCGCGAATTCGCG) <sub>2</sub>	66 %
	12-d <sub>100</sub> : (dCGCGGGCCCGCG) <sub>2</sub>	100 %
	12-r <sub>66</sub> : (rCGCGAAUUCGCG) <sub>2</sub>	66 %
	12-r <sub>100</sub> : (rCGCGGGCCCGCG) <sub>2</sub>	100 %
14 bp	d <sub>14ab</sub> : (dTAATACGACTTAAC)•(dGTTAAGTCGTATTA)	29 %
	r <sub>14ab</sub> : (rUAAUACGACUUAAC)•(rGUUAAGUCGUUUA)	29 %
17 bp	d <sub>17R1-17R2</sub> : (dGGAGCTCCCAGACGACC)•(dGGTCGTCTGGGAGCTCC)	71 %
	r <sub>17R1-17R2</sub> : (rGGAGCUCCCAGACGACC)•(rGGUCGUCUGGGAGCUCC)	71 %
18 bp	d <sub>18a-18b</sub> : (dCTACTCGTTACCTTCTT)•(dAAGAAGGTAACGAGTAGG)	39 %
20 bp	d <sub>20R1-20R2</sub> : (dGTGAGCTCCCAGACGACCTG)•(dCAGGTCGTCTGGGAGCTCAC)	65 %
	r <sub>20R1-20R2</sub> : (rGUGAGCUCCCAGACGACCUG)•(rCAGGUCGUCUGGGAGCUCAC)	65 %
22 bp	d <sub>22a-22b</sub> : (dCCTACTCGTTACCTTCTTCTGA)•(dTCAGAAGAAGGTAACGAGTAGG)	45 %
24 bp	d <sub>24a-24b</sub> : (dCCTACTCGTTACCTTCTTCTGATA)•(dTATCAGAAGAAGGTAACGAGTAGG)	42 %
	d <sub>66+66</sub> : (dCGCGAATTCGCGCGCGAATTCGCG) <sub>2</sub>	66 %
	d <sub>100+100</sub> : (dCGCGGGCCCGCGCGGGCCCGCG) <sub>2</sub>	100 %
	r <sub>66+66</sub> : (rCGCGAAUUCGCGCGCGAAUUCGCG) <sub>2</sub>	66 %
	r <sub>100+100</sub> : (rCGCGGGCCCGCGCGGGCCCGCG) <sub>2</sub>	100 %
36 bp	d <sub>100+100+100</sub> : (dCGCGGGCCCGCGCGGGCCCGCGCGGGCCCGCG) <sub>2</sub>	100 %
	r <sub>100+100+100</sub> : (rCGCGGGCCCGCGCGGGCCCGCGCGGGCCCGCG) <sub>2</sub>	100 %
	d <sub>36a-36b</sub> : (dCCTACTCGTTACCTTCTTCTGACTTCCCTTTTCTT)• (dAAGAAAGAGGGAAGTCAGAAGAAGGTAACGAGTAGG)	44 %

Table 10: Name and sequences of the different DNA and RNA duplexes extensively used in this study. The percentage of GC base pair is also indicated.

#### XIX.1.2. ESI-MS spectra: DNA vs. RNA

The spectra on Figure 69 are obtained using soft conditions to see the 0-adduct peak of each duplex. DNA and RNA spectra are obtained using the same tuning parameters for comparison. The figure shows various spectra from 12-bp duplexes to 36-bp duplexes.

The major charge states detected are generally the same for DNA and RNA: 5- for 12-bp, 6- for 17-bp and 20bp and 7- for 24-bp. For duplexes from 12-bp to 20-bp, at least three charge states are observed for DNA, whereas for RNA just two charge states are observed. Also, the intensities of each charge state changes from DNA to RNA. The charge states detected for DNA or RNA duplexes, are always low compared to the number of phosphate groups present. For example, with a 12-bp duplex, we should have 22 negative charges. In solution, nucleic acids are surrounded by counterions that counter-balance the high repulsion between phosphates and thus reduce the net charge<sup>4</sup>. Determining the charge of duplexes in solution is not so intuitive because the counterions have to be considered. It has been proposed that 76% neutralization of phosphate is effective when the distance between ions and the nucleic acid is at 18 Å.<sup>23</sup> As the ESI droplets are charged, the final charge state may reflect the thickness of the layer of counterions.

If we make a zoom on the different charge state for DNA and RNA, one can see that more  $\text{NH}_4^+$  adducts are present on RNA than DNA for the same instrumental parameters. One hypothesis can be that RNA and DNA do not receive the same internal energy even if the activation parameters are equal. As a consequence desolvation and declustering will be different.

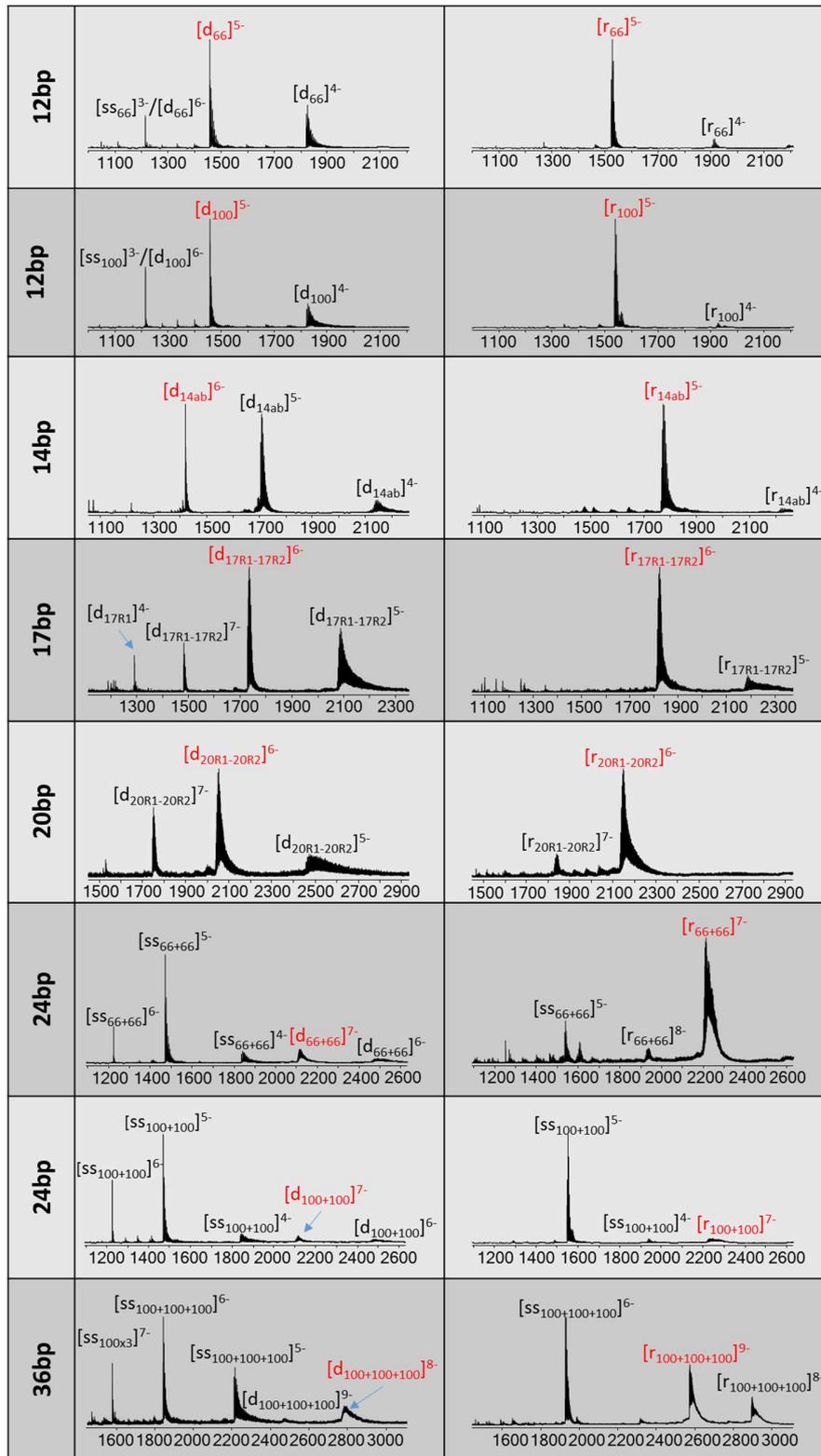


Figure 69: ESI-MS of DNA and RNA duplexes recorded in 150 mM  $\text{NH}_4\text{OAc}$  from soft source conditions. Duplexes from 12-bp to 36-bp are represented on the figure. Major charge state for each duplex analyzed is in red.

## XIX.2. IM-MS results

### XIX.2.1. Reconstruction of CCS distributions

The main value obtained when doing IM-MS is the arrival time distribution. From this arrival time distribution, we have seen in the material and method section that we can determine the CCS of the center of the peak, which will be the  $CCS_{exp}$  value. But an important point is the reconstruction of the CCS distribution that will give us information about the broadness of the peak. The single-field reconstruction and Step field width analysis ( $FWHM_{step}$ ) are described in details in the material and methods section.

To reconstruct the CCS distributions of the duplexes, we chose to use the single-field reconstruction which reflects the correct position of the different peaks. It also makes possible to reconstruct the CCS distribution for non-gaussian arrival time distributions or for noisy data. Here, we tested the  $FWHM_{step}$  method on different duplexes to see if the broadness observed for the different peaks is due to diffusion or if the width of the distributions are reflecting only the flexibility of the molecules. The results for  $d_{100}/r_{100}$ ,  $d_{17R1-17R2}/r_{17R1-17R2}$  and  $d_{66+66}/r_{66+66}$  are presented on Figure 70. We compare the results to the CCS distribution obtained for  $d(TG_4T)_4$ , described on figure 3 of the paper from Marchand et al.<sup>24</sup>  $d(TG_4T)_4$  is a G-quadruplex DNA that has a narrow CCS distribution due to the rigidity of its structure.<sup>25</sup>

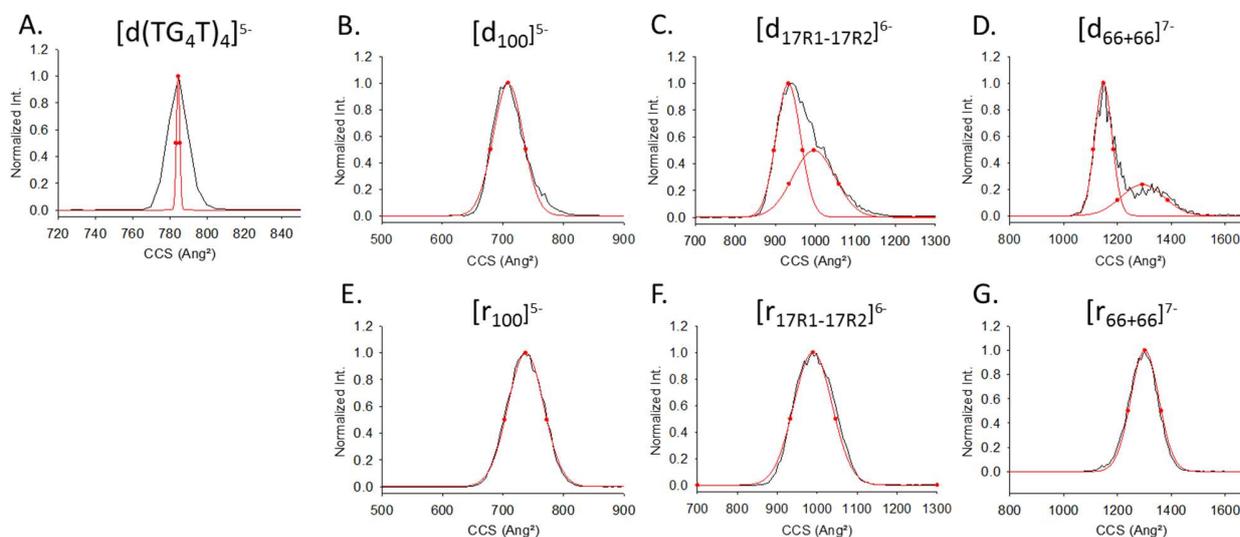


Figure 70: Example of reconstructed CCS distributions of  $d(TG_4T)_4$  several DNA and RNA duplexes (from B to G). Black line: reconstruction obtained using the method 2. Red curve: Gaussian curves reconstructed using the method 3.

The results indicate that for a thin peak, like for  $[d(TG_4T)_4]^{5-}$ , the contribution of the diffusion is large compared to the contribution of the conformational diversity (comparison between black and red lines), meaning that the two ways of reconstructed the CCS distribution are not the same. In this case the

FWHM<sub>step</sub> method is correct compared to the single field reconstruction method. For DNA and RNA duplexes, we can see, based on the examples of the Figure 70, that the two methods used to reconstruct the CCS distribution give the same results. Most of the broadness of the peaks is due to the variety of structures that are coexisting. The single field reconstruction is then an excellent approximation for the CCS distribution reconstruction.

We will assume that these results can be applied to all the different duplexes, DNA and RNA, studied in this manuscript. We will then use only the single field reconstruction method to reconstruct the CCS distributions.

### *XIX.2.2. Comparison between <sup>DT</sup>CCS<sub>He</sub> of DNA and RNA*

#### *XIX.2.2.1. Raw arrival time distributions*

All the experiments presented below were carried out using soft enough conditions to preserve the duplexes, and avoid as much as possible modifications of their structures. To do so, we used a low pre-IMS activation using the SOFT trapping parameters, a trap entrance grid delta voltage of 4V and a fragmentor voltage of 350V. The raw heat maps showing the arrival time distribution as a function of the *m/z* are presented on Figure 71. These heat maps show both the mass spectrum, and information on the arrival time of the peaks with different numbers of adducts.

As said in Section II.1.2 ESI-MS spectra: DNA vs. RNA, and confirmed by the heat maps presented on Figure 71, ammonium ions adducts are retained on both DNA and RNA duplexes, but more are seen on the RNAs. For long duplexes (>20 bp), the arrival time distributions are complex to describe because a different one is seen for each number of ammonium ion adducts. The two-dimensional heat maps show more clearly that the arrival time distribution with ammonium adducts is shifted to higher average arrival time compared to the fully declustered duplexes, for several RNA duplexes, for example  $[r_{14ab}]^{5-}$ ,  $[r_{17R1-17R2}]^{6-}$  or  $[r_{100+100+100}]^{9-}$ . As cations take space, they can increase the determined CCS (and so the arrival time), but we will see below that the progressions are not regular according to the number of adducts. Duplexes without adducts are more compact so several conformations can be detected: one with cations and another without. Because nucleic acids are surrounded by ions in solution and the final declustering occurs in the gas phase, the structure without cations may be the furthest away from the structure in solution. As a working hypothesis, it could be of interest to keep counter-ions to reflect the structure in solution, even though it will be detrimental to the quality of the spectrum and the MS sensitivity.

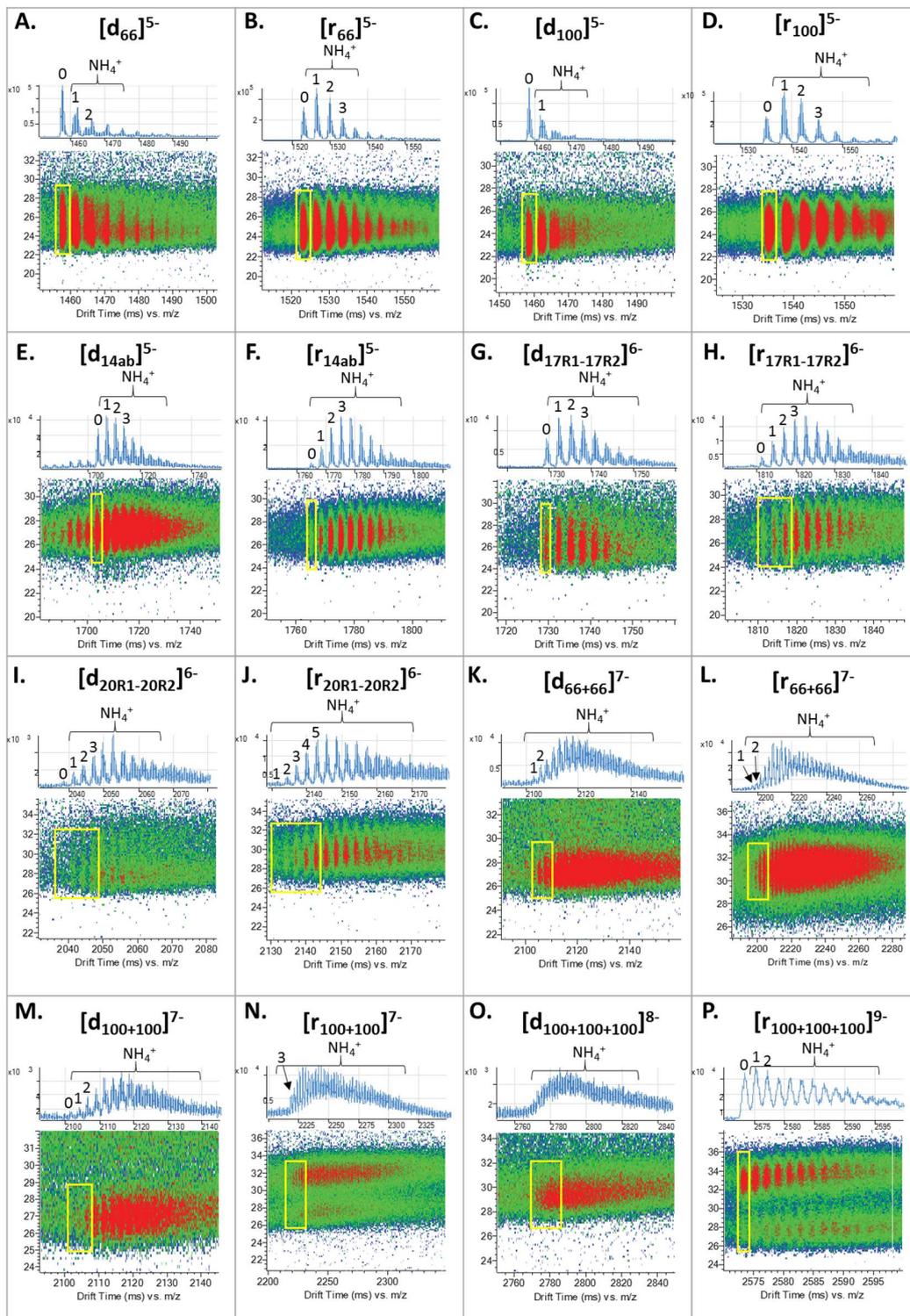


Figure 71: Total arrival time distribution of the major charge state of  $d_{66}$ ,  $r_{66}$ ,  $d_{100}$ ,  $r_{100}$ ,  $d_{14ab}$ ,  $r_{14ab}$ ,  $d_{17R1-17R2}$ ,  $r_{17R1-17R2}$ ,  $d_{20R1-20R2}$ ,  $r_{20R1-20R2}$ ,  $d_{66+66}$ ,  $r_{66+66}$ ,  $d_{100+100}$ ,  $r_{100+100}$ ,  $d_{100+100+100}$  and  $r_{100+100+100}$  (from A to P respectively). We can see the effect of the ammonium adducts distribution in soft conditions on the arrival time. Ammonium adducts are identified on the spectra. The yellow rectangle correspond the zone used to calculate the  $DTCCS_{He}$ .

Here, to calculate  $^{DT}CCS_{He}$ , we used the arrival time distribution of the peak of interest without ammonium adducts, whenever possible. When the duplex without ammonium ions was not observed at low pre-IMS activation, we checked if the arrival time distribution with ammonium was changing dramatically. As the arrival time was usually not changing, we used the first few detectable adducts to calculate the  $^{DT}CCS_{He}$ . The arrival time distributions taken into account are shown on Figure 71 as yellow squares.

#### XIX.2.2.2. $^{DT}CCS_{He}$ distributions: DNA vs RNA

- Comparison of the CCS values corresponding to the center of the ATDs

By determining the  $CCS_{exp}$  of the center of the arrival time distributions, one can first compare the value for DNA and RNA duplexes. A summary of the results is presented on Table 11.

		NH <sub>4</sub> OAc					NH <sub>4</sub> OAc		
	Sequences	Charge State	$^{DT}CCS_{He}$ Peak 1	$^{DT}CCS_{He}$ Peak 2		Sequences	Charge State	$^{DT}CCS_{He}$ Peak 1	$^{DT}CCS_{He}$ Peak 2
12-bp	d <sub>66</sub>	4	697		22-pb	d <sub>22ab</sub>	6	1078	
		5	735	793			7	1112	1202
		6	788	876			8	1338	
	r <sub>66</sub>	4	695		24-bp	d <sub>66+66</sub>	6	1119	
		5	733				7	1148	1293
	d <sub>100</sub>	4	706			r <sub>66+66</sub>	7	1300	
		5	730				8	1386	
		6	784				d <sub>100+100</sub>	6	1106
	r <sub>100</sub>	4	696			7	1142	1336	
		5	737			r <sub>100+100</sub>	7	1162	1331
14-bp	d <sub>14ab</sub>	5	822			d <sub>24ab</sub>	6	1126	
		6	873				7	1160	
	r <sub>14ab</sub>	5	818			7	1367		
17-bp	d <sub>171-17R2</sub>	6	921		32-bp	d <sub>32ab</sub>	8	1410	
		7	957				8	1449	
	r <sub>17R1-17R2</sub>	6	937			36-bp	d <sub>100+100+100</sub>	9	1526
7		992		8	1488				
18-bp	d <sub>18ab</sub>	5	935		r <sub>100+100+100</sub>		9	1523	1856
		6	983				8	1495	
20-bp	d <sub>20R1-20R2</sub>	6	1056		d <sub>36ab</sub>		9	1556	
		7	1160						
	r <sub>20R1-20R2</sub>	6	1070						
		7	1168						

Table 11: Summary of the experimental  $^{DT}CCS_{He}$  determined for DNA and RNA duplexes of different sizes. Soft conditions were used to acquire the data. Samples were prepared in 150 mM NH<sub>4</sub>OAc.

For a same duplex, when several charge states are detected, the  $^{DT}CCS_{He}$  increases when the charge state increases too. This is true for all duplexes, whatever the size, and the type of nucleotide. Between two consecutive charge states, a compaction of  $\approx 5\%$  is most of the time observed for charge states naturally present in *native* conditions.

$^{DT}CCS_{He}$  values obtained for DNA and RNA are very similar, despite the fact that in solution, DNA duplexes adopt a B-Helix structure whereas RNA duplexes adopt an A-Helix. Also, the results show that for a same duplex, higher charge states have a  $^{DT}CCS_{He}$  higher than the lower charge states. For higher charge states, these results can be explained by the fact that if more charges are present, the electrostatic repulsion will be higher and so the structure should be more extended.

In a study from Burmistrova et al<sup>20</sup>, the percentage of GC content was linked to the stability and compaction of duplexes using MS/MS and IM-MS experiments. Their sequences were duplexes containing GC tracts in-between AT tracts (or the contrary), and the percentage of GC was varied. The authors showed that the GC tracts undergo more compaction than AT tracts. For a given length of duplexes, i.e. 12-mer or 24-mer for example, and for a given charge state, our observation that  $CCS_{d_{100}} < CCS_{d_{66}}$ ,  $CCS_{d_{100+100}} < CCS_{d_{66+66}} < CCS_{d_{24ab}}$  are in line with the results found by Burmistrova. The percentage of compaction seems to depend on the percentage of GC content for DNA. However, for RNA no rule can be dictated as for the 12-mer, the CCS obtained are the same and  $r_{100+100}$  is less compact than  $r_{66+66}$ . Studying more sequences could bring some insight into this question.

- **Comparison of CCS distributions**

The  $^{DT}CCS_{He}$  distributions obtained from arrival time conversion are represented as violin plots on Figure 72. Only the charge states in common between DNA and RNA are represented for comparison. DNA and RNA duplexes are covering the same range of CCSs and in all cases, the peaks are broad. For the 12-bp duplexes until the 20-bp duplex, mainly one broad peak is observed for all charge states.  $[d_{66}]^{5-}$  is not following this rule because we can see that two CCS distributions can be distinguished (Figure 72, A). The more compact distribution is similar to the one of the RNA. For  $d_{20R1-20R2}$  (Figure 72, E), the intensity of the peaks corresponding to the DNA duplex charge states 7- and 6- was very low, reason why the distribution is noisy.

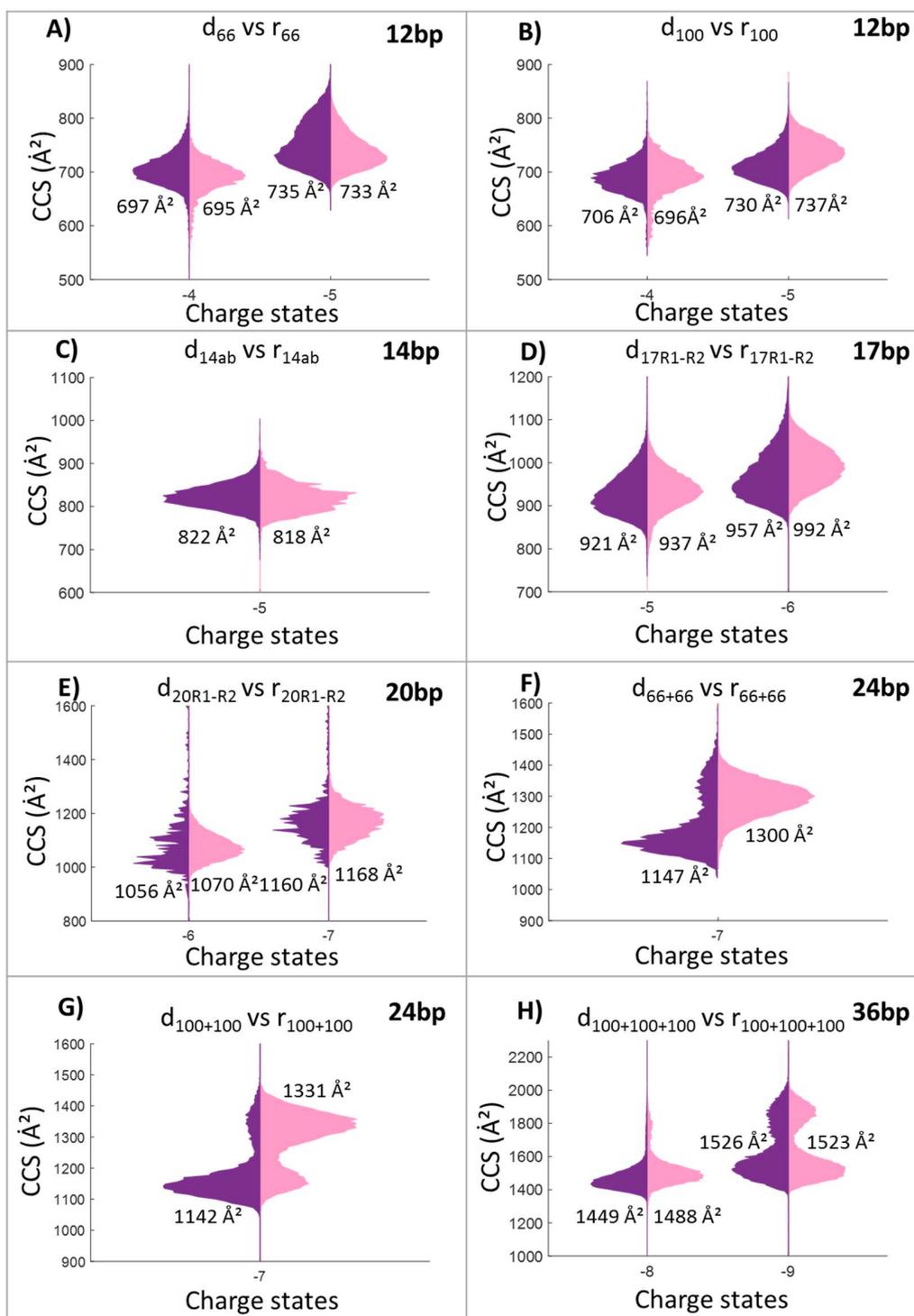


Figure 72: Violin-plots showing the difference in  $^{DT}CCS_{He}$  between DNA (in violet, on the left) and RNA (in pink, on the right). The charge states showed are in common between DNA and RNA duplexes. Only the  $^{DT}CCS_{He}$  value of the major peak (when two are present) is written on the figure. A)  $d_{66}$  vs  $r_{66}$ , B)  $d_{100}$  vs  $r_{100}$ , C)  $d_{14ab}$  vs  $r_{14ab}$ , D)  $d_{17R1-17R2}$  vs  $r_{17R1-17R2}$ , E)  $d_{20R1-20R2}$  vs  $r_{20R1-20R2}$ , F)  $d_{66+66}$  vs  $r_{66+66}$ , G)  $d_{100+100}$  vs  $r_{100+100}$  and H)  $d_{100+100+100}$  vs  $r_{100+100+100}$ .

The CCS distributions obtained for the duplexes studied are quite broad, meaning that several conformations coexist. Less diverse conformations are present when the peak is thinner. For some longer sequences, such as the 24-bp  $d_{100+100}$  and  $r_{100+100}$  (Figure 72, G) and the charge state 9- for the 36-bp (Figure 72, H), two conformational ensembles are observed for both DNA and RNA. This means that at least two families of conformations exist. For the 36-bp, the conformational ensembles observed for DNA and RNA have similar CCS distributions. Concerning the 24-bp (Figure 72, F and G), we can see that the same two populations are present for both DNA and RNA, but the proportions are inverted. Maybe, for all duplexes studied we have formation of several conformations, but for small duplexes we cannot separate them due to the broadness of the peak (on the contrary to long duplexes).

For the longer duplexes, two hypotheses could explain the presence of two peaks: 1) the two peaks correspond to B-Helix vs A-Helix which are well defined in this case, or 2) the two peaks correspond to different structures in solution, for example a kissing complex and a perfect duplex. Because the two strands are self-complementary, the formation of a kissing complex is possible. Figure 73 presents an example with the 24-mer  $d_{100+100}$ .



Figure 73: Example of the 24-mer  $d_{100+100}$ , showing that the formation of a kissing complex is possible. Size of the loops are random. Hydrogen bonds are not represented.

We have seen in Chapter I that kissing complexes are not formed in presence of TMAA. So to check if one of the possible structure is a kissing complex, we performed the experiment using 150 mM of TMAA, instead of  $NH_4OAc$ , on  $r_{100+100}$ ,  $d_{100+100}$ , and  $d_{66+66}$ .

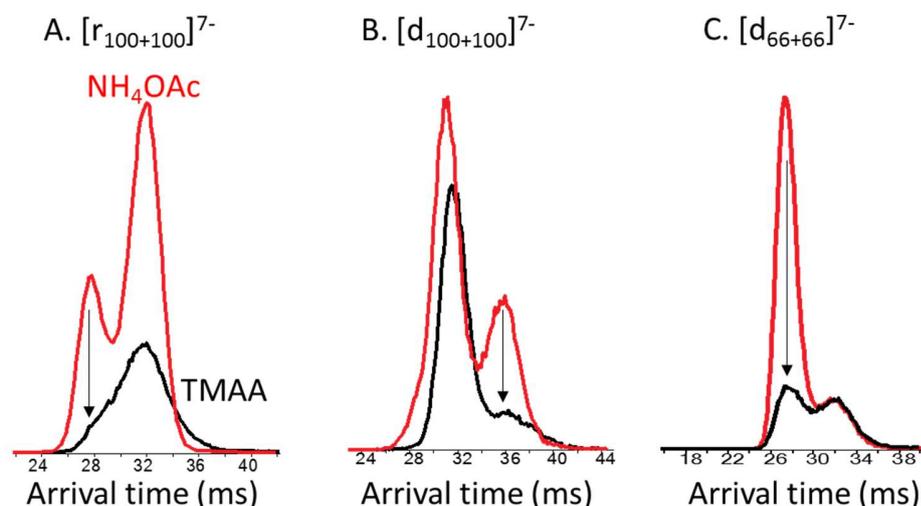


Figure 74: Raw arrival time distribution obtained in 150 mM of TMAA for A)  $[r_{100+100}]^{7-}$ , B)  $[d_{100+100}]^{7-}$  and C)  $[d_{66+66}]^{7-}$

One of the peaks for each 24-mer is decreasing in presence of TMAA, meaning that this conformation could correspond to a kissing complex. Yet the peaks do not completely disappear. For  $r_{100+100}$  and  $d_{100+100}$ , the disappearing conformation corresponds to the minor one. On the contrary, for  $d_{66+66}$ , it is the major peak so the major conformation that is affected. With TAR-R06, we saw that TMAA was completely preventing the kissing complex to form. The results here are inconclusive. Using TMAA instead of NH<sub>4</sub>OAc changes the proportions of the two conformations, but we can still not assign them. Molecular modeling may help us to decipher which peak corresponds to which structure. We'll see in the next chapter that kissing complexes cannot be differentiated from regular duplexes based on the CCS values.

- Contribution of molecular dynamics and  $CCS_{theo}$  calculations

The experimental data alone cannot give much more information about the structure in the gas phase. To investigate more deeply the structure of the detected ions, molecular dynamics (MD) and theoretical calculation of CCSs were performed, in collaboration with Dr. M. Porrini. All the  $CCS_{theo}$  were calculated using the Exact Hard Sphere Scattering (EHSS) model mentioned in the introduction. The development of the method is based on the 12-mers duplexes.

All the results are summarized on Figure 75. We will explain the reasoning of the method by briefly going step by step. All the steps are described in more details in the paper and supporting information.<sup>26</sup>

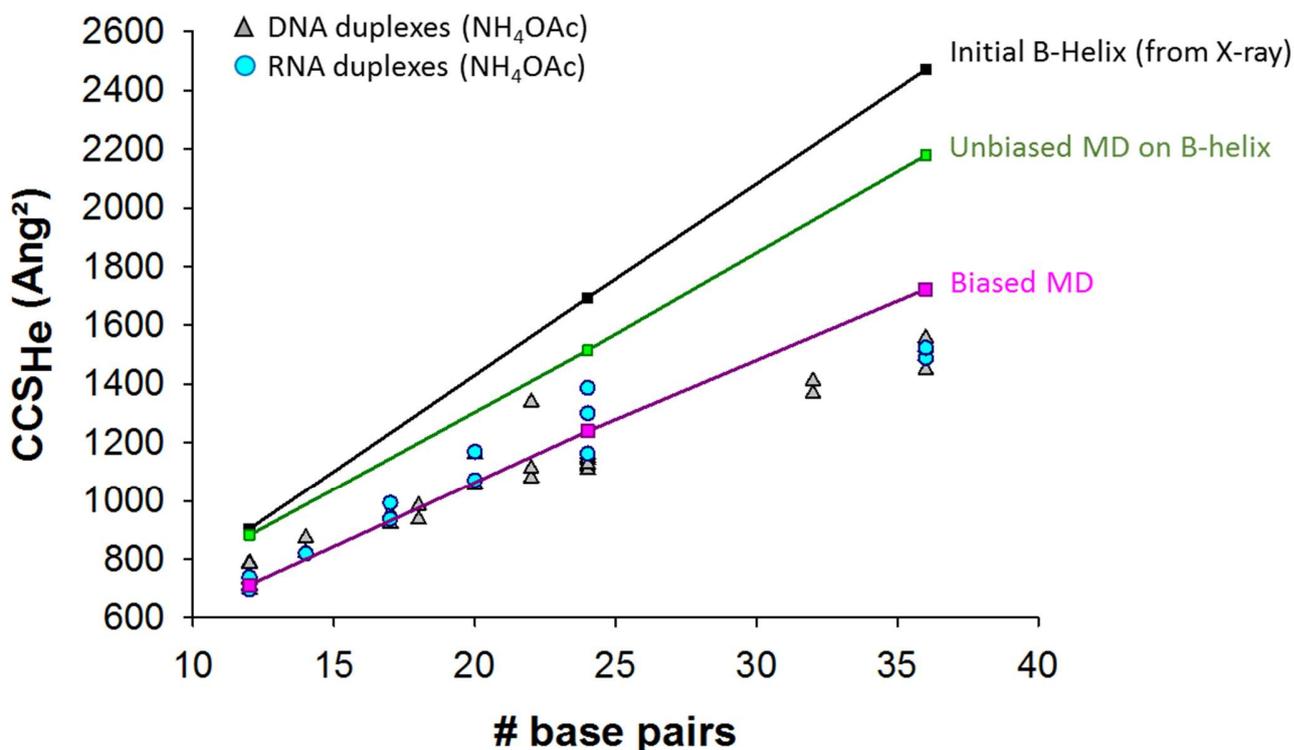


Figure 75: Comparison between  $^{DT}CCS_{He}$  and  $CCS_{calc}$  for all the different DNA and RNA duplexes studied. The black curve refers to the calculated CCS in solution, the blue curve to the  $CCS_{calc}$  using MD on B-Helix and the purple curve to the zipped helix MD. For clarity, only the major peak is represented here, even if several are present.

The first step was to calculate the theoretical CCS value of the X-ray structure of the 12-bp B- or A-helix (Figure 75, black line).  $CCS_{theo}$  of 24- and 36-mers were also calculated, based on Hyperchem models with canonical bond lengths and angles known from X-ray crystallography. Calculations revealed that the  $^{DT}CCS_{He}$  determined experimentally for the charge states naturally produced by *Native MS*, are at least 20% more compact. As example,  $CCS_{calc}$  for  $d_{66}$  is at  $903 \text{ \AA}^2$  compared to  $735 \text{ \AA}^2$  experimentally and  $d_{100}$ - $CCS_{calc}=908 \text{ \AA}^2$  compared to  $730 \text{ \AA}^2$  experimentally. These first results show that compaction is happening when going from the solution to the gas phase, at charge states naturally occurring in *native MS*.

As these calculated values did not match with the experimental ones, several molecular dynamics (MD) simulations were performed to find the kind of structures that could match with the experimental CCS values. First, unbiased MD simulations were performed directly on “naked” B-helix meaning that the solvent was removed. The results still did not match with the experiment values either (Figure 75, green line). With this model, the structures were forming spontaneously H-bonds between the phosphates of each side of the minor groove, making a “zipping” of the structure. Then, Temperature Replica Exchange

MD (T-REMD) were performed but here also the data did not match the experimental values. However, these results gave information about the starting structure which seems to be more globular than an helix. The last step was to take into account the desolvation process that occur during the ESI process. DNA are desolvated through the Charge Residue Model and so presence of  $\text{NH}_4^+$  in both grooves have to be considered. As explained in the paper, during solvent evaporation, phosphate-ammonium-phosphate bonds are formed. During declustering, H-bond between phosphates will remain and then a zipping of both grooves is happening. The final biased MD are taking into account the formation of these H-bonds. Then after calculation of the  $\text{CCS}_{\text{theo}}$  of the structures obtained by biased MD, the CCSs were matching (Figure 75, pink curve).

Comparison between all the experimental  $\text{DTCCS}_{\text{He}}$  values and the theoretical values using the different simulation allows us to check if all the charge states and all duplexes are compacting following the same trend. The results indicate that the low charge states, i.e. charge states obtained during *native* ESI-MS, follow the trend of  $\text{CCS}_{\text{calc}}$  using the biased MD. However, we can see that when two peaks are obtained in the CCS distribution it is more difficult to assign a structure. Figure 76 is presenting the results only for the four duplexes where two peaks were present.

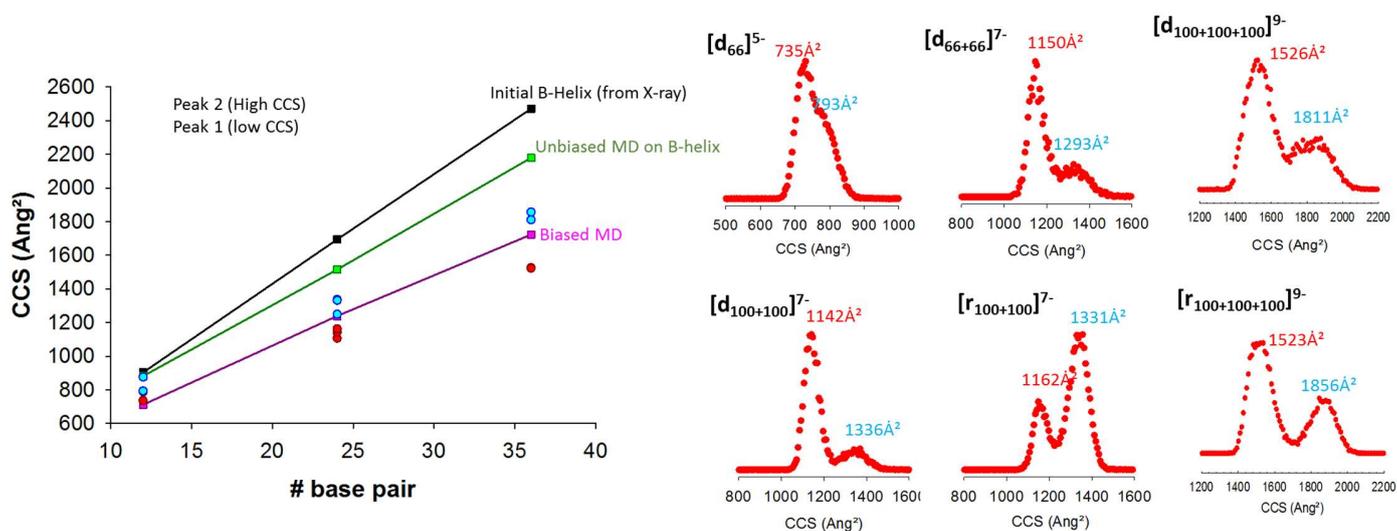


Figure 76: Comparison of the  $\text{CCS}_{\text{He}}$  obtained for the four duplexes having two peaks:  $d_{66}$ ,  $d_{66+66}$ ,  $d_{100+100}$  and  $r_{100+100}$ . CCS distribution were shown on the right, with the CCS of the peak centers written.

### *XIX.2.3. Collision-induced compaction of DNA and RNA duplexes*

#### *XIX.2.3.1. Arrival time distributions at low and high pre-IMS activation*

We have seen in the Chapter 1 that one can change several pre-IMS parameters to change ion activation: the Trap Entrance Grid Delta (TEGD), the Fragmentor and trap voltages. Changing these parameters will act on the declustering step and the ions will be more or less activated. The calculated  $^{DT}CCS_{He}$  in the previous section are obtained using a fragmentor voltage at 350V, and a TEGD adjusted depending on the size of the sequence. We have seen in the introduction that proteins can undergo conformational changes in the gas phase due to collision induced by ion activation. In the following results, we asked ourselves what will happen to duplexes structure depending on pre-IMS activation.

We performed the experiments at two different fragmentor voltages: 350V and 600V. These two values are based on the results obtained in chapter 1.X.2.2.2 “How to influence declustering”. The raw heat maps of DNA and RNA duplexes tested at 350V and 600V are presented on Figure 77.

The results show clearly that the fragmentor voltage has an effect not only on the mass spectrum but also on the arrival time distribution. Setting the fragmentor voltage at 600V reduces the number of ammonium adducts present. Also, for short duplexes from 12 to 17-bp, a small distribution of  $Na^+$  ions adducts becomes visible. When  $NH_4^+$  adducts are numerous, the distribution of  $NH_4^+$  and  $Na^+$  adducts are overlapping at 350V and it is difficult to distinguish clearly between the two. Using harsher conditions modifies the arrival time distribution. A shift to shorter arrival time is observed, and this effect is larger on RNA than DNA duplexes. If the ions arrive earlier, this implies that the conformation of the duplex is more compact at 600V than at 350V. The structures are collapsing when the duplexes had received more internal energy in the source region.

At a fragmentor voltage of 600 V, the number of ammonium adducts is reduced (as expected from our previous results on kissing complexes (chapter 1.I.)). It increases also the global arrival time shift observed when considering all  $NH_4^+$  adducts for all RNAs, and also for  $[d_{100+100}]^7$ . It thus seems that presence of ions helps to maintain a more elongated structure than without. Ions may thus prevent a high collapsing of the structure in the gas phase.

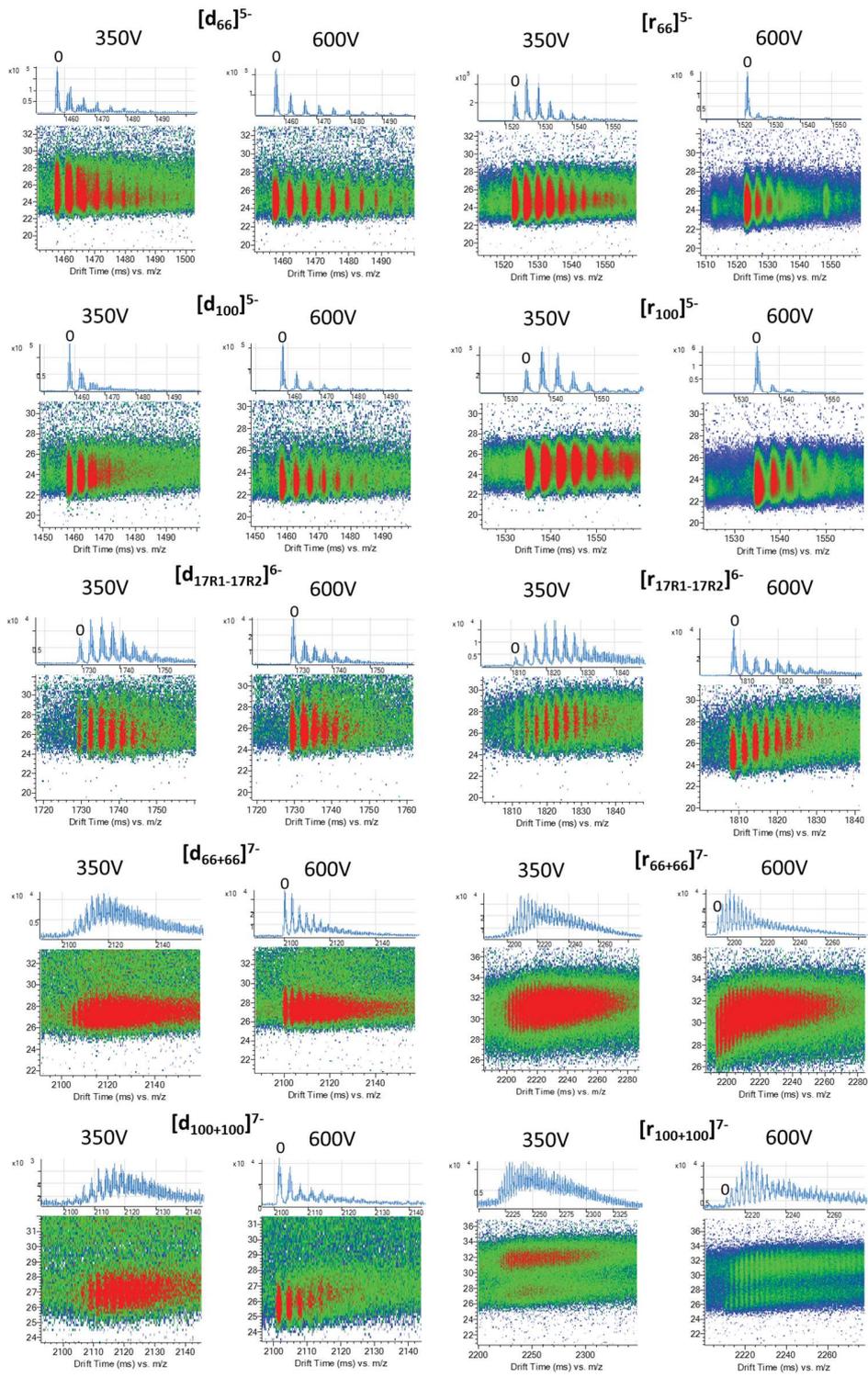


Figure 77: Total arrival time distribution of the major charge state of  $d_{66}$ ,  $r_{66}$ ,  $d_{100}$ ,  $r_{100}$ ,  $d_{17R1-17R2}$ ,  $r_{17R1-17R2}$ ,  $d_{66+66}$ ,  $r_{66+66}$ ,  $d_{100+100}$  and  $r_{100+100}$ . Fragmentor is set at 350 or 600V. We can see the effect of the fragmentor on the arrival time distribution.

### XIX.2.3.2. Comparison between $^{DT}CCS_{He}$ using pre-IMS activation

To confirm that DNA and RNA duplexes are collapsing when using pre-IMS activation,  $^{DT}CCS_{He}$  are determined at fragmentor voltages of 350V and 600V. The determination of the  $^{DT}CCS_{He}$  at 600V is made on the peak corresponding to the duplex without any ammonium or sodium adduct. The results at 350V are the ones described in section I.2.2. Final results are presented on Figure 78.

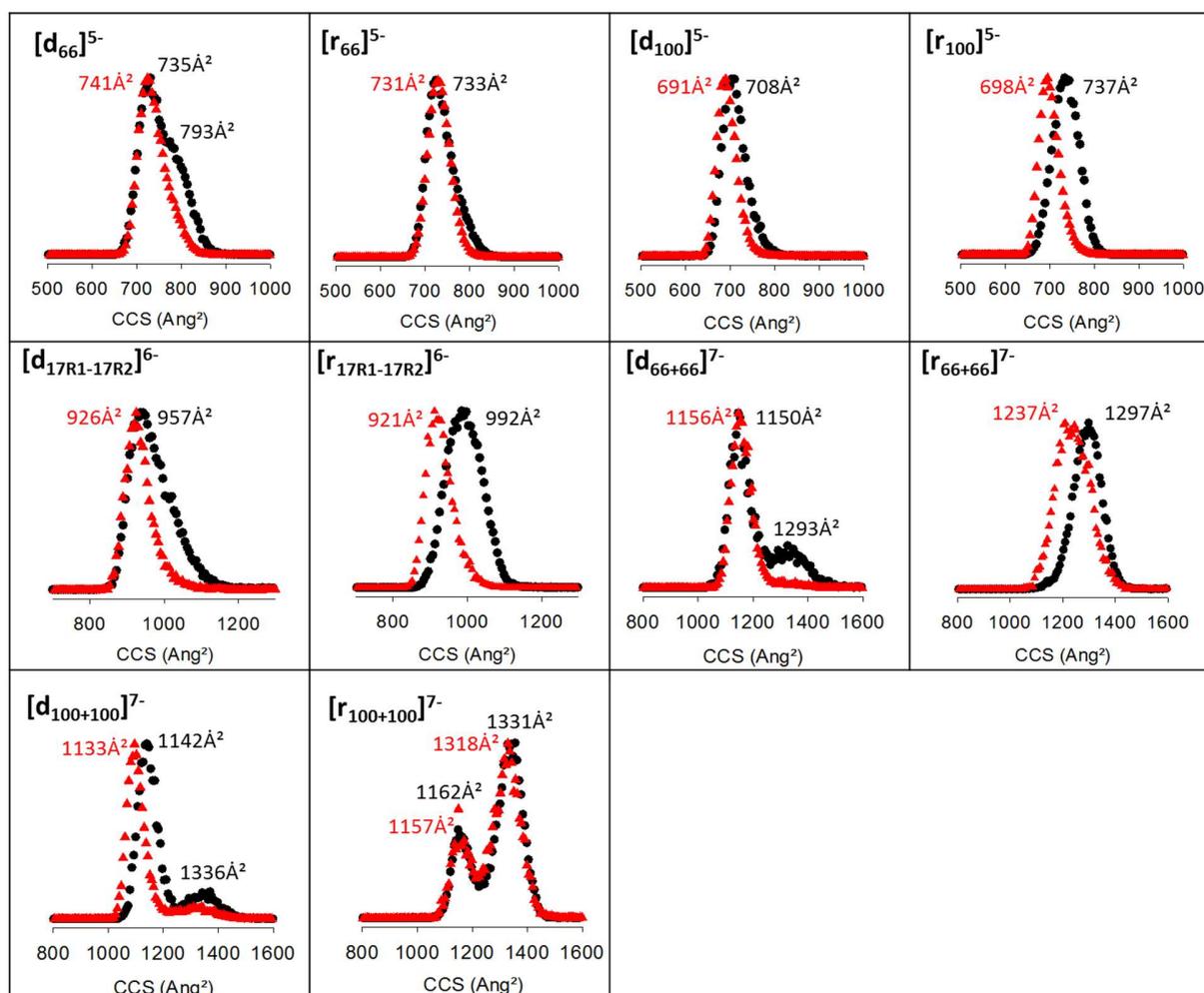


Figure 78: Effect of the source fragmentor on the  $^{DT}CCS_{He}$  distribution on the major charge state of  $d_{66}$ ,  $r_{66}$ ,  $d_{100}$ ,  $r_{100}$ ,  $d_{17R1-17R2}$ ,  $r_{17R1-17R2}$ ,  $d_{66+66}$ ,  $r_{66+66}$ ,  $d_{100+100}$  and  $r_{100+100}$ . In black circle, fragmentor at 350V and in red triangle, fragmentor at 600V. At 600V no  $NH_4^+$  adducts were taken into account for CCS reconstruction. The CCS values indicated on the graph are the one of the peak center. Reconstruction are obtained from step field reconstruction from  $\Delta V=390V$ . The table summarize the percentage of compaction and the GC content.

The results indicate that with pre-IMS activation (fragmentor at 600 V),  $^{DT}CCS_{He}$  distributions are changing, which is in correlation with the first observations done on arrival time distribution. For  $d_{66}$  and  $d_{66+66}$  the high CCS peak is lost in favor of the low CCS distribution. The CCS distribution of  $r_{66}$  is losing its tail at 600

V, and the same happens for  $d_{17R1-17R2}$ . For the 24-mer  $d_{100+100}$ , the first peak is compacted and enhanced compared to the high-CCS peak. In some cases like for  $d_{66}$ ,  $r_{66}$  or  $d_{66+66}$ , activation seems to not have an impact on the CCS distribution of the major peak. One hypothesis could be that the highest compaction of that ensemble was already reached even in the softest conditions used. In all other cases, the CCS distribution is shifted towards lower CCS values.

When compaction is occurring for both DNA and RNA for a same sequence, the compaction is more important for RNA. For example, a compaction of 2% is observed for  $d_{100}$  compared to 5% for  $r_{100}$ . In the same way,  $d_{17R1-17R2}$  undergoes a compaction of 3% and 7% for  $r_{17R1-17R2}$ . These results show that activation induces compaction of the structures of the different duplexes at the charge state obtained in *Native* conditions.

Pre-IMS activation induces a compaction of the structures. It also contribute to the loss of high-CCS ensembles of structures when several populations are present.

## XX. Conclusion and perspectives

To summarize our results, by using *native* IM-MS, we obtained information about the conformations of duplexes present in the gas phase which are revealed by the  $^{DT}CCS_{He}$  value and the broadness of the peaks. DNA and RNA duplexes have similar CCS distribution even if they do not have the same structure in solution (B-Helix vs A-helix respectively). We also show that duplexes DNA and RNA undergo compaction from the solution to the gas phase, at charge states naturally produced by *native* ESI-MS.

In terms of structural study on RNA or DNA, one thus needs to be extremely careful when drawing conclusions from CCS determinations. The compaction effect due to the transition from the solution to the gas phase has to be taken into account.

### XX.1. Perspectives: the effect of the charge state on $^{DT}CCS_{He}$

To increase the charge state of some duplexes, we performed the experiment using 0.5% sulfolane which is a supercharging agent.<sup>27,28</sup> Results are summarized in Table 12.

As explained in Section II.1. , the charged states detected are low compared to the size, and so the number of phosphates per duplexes. When duplexes are long, the maximum charge state detected is higher, which is logical as the number of bases is increased. However, one can see that only a difference of 3 charges is detected between 12-bp and 36-bp duplexes. The addition of sulfolane is increasing the charge states detected but few supplementary charge states are detected. Here we noticed that the addition of

sulfolane may change the CCS of a specific charge state (example with  $[d_{100}]^{6-}$ ) or different populations of conformations are appearing (example with  $d_{14ab}$ ,  $d_{18ab}$ , or  $d_{24ab}$ ). The effects are different depending on the sequence and its size. Several studies are done to understand the effect of sulfolane as supercharging agent as it is still not well known.<sup>27,28</sup> For a same molecule, lower charge states have lower CCS values than higher charge states.

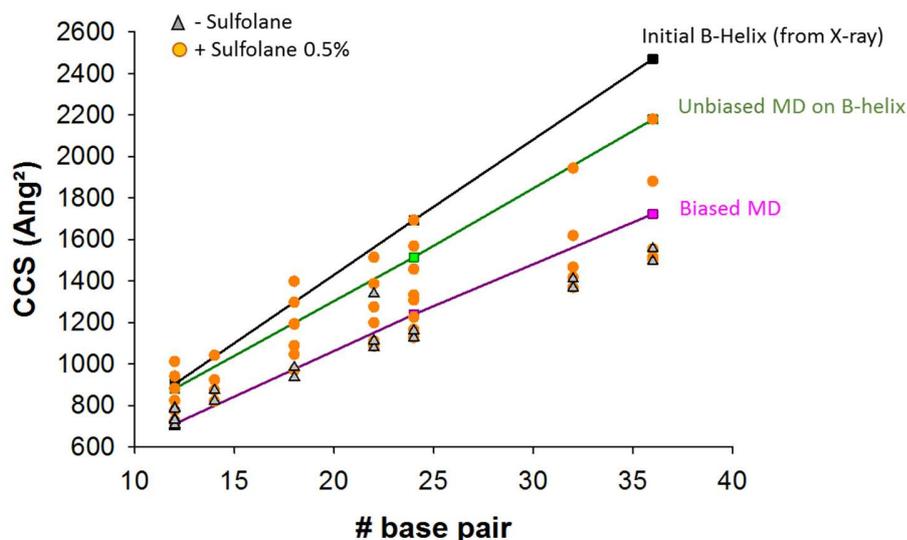


Figure 79: Comparison between the theoretical CCS obtained using different models and the experimental ones obtained for duplexes with or without sulfolane. Grey triangle are the CCS of duplexes without sulfolane, and orange circles with sulfolane added.

Interestingly, when the charge state is increased for the largest duplexes, the  $^{DT}CCS_{He}$  of new peaks follow preferentially the trend of the  $CCS_{theo}$  obtained from unbiased MD of B-Helix (Figure 79). As the charge density increases, the repulsion between phosphates may be higher too and so the structure will tend to be more elongated than at charge states naturally present in *native* MS. If collapse is prevented over the entire electrospray process, we may have reached our goal of minimally perturbing the solution structures during their transition from solution to gas phase, by actually increasing the charge states. However, if compaction occurs first and extension occurs next due to Coulomb repulsion, the collision cross sections may be similar to unbiased B-helix simulations just by coincidence. Further work on a larger panel of long duplexes would help test these hypotheses.

Also we do not have much information on the effect of sulfolane on the structure in solution of duplexes. In particular, although the volume percentage of sulfolane in the injected solution is only 1%, the concentration may be much higher in the final droplets from which the high charge state duplexes are

produced. Again, a more extensive study should be performed on the impact of sulfolane on duplexes structures both in solution and in the gas phase.

		NH <sub>4</sub> OAc		NH <sub>4</sub> OAc + 0,5% Sulfolane				NH <sub>4</sub> OAc		NH <sub>4</sub> OAc + 0,5% Sulfolane				
	Sequences	Charge State	<sup>DT</sup> CCS <sub>He</sub> Peak 1	<sup>DT</sup> CCS <sub>He</sub> Peak 2	<sup>DT</sup> CCS <sub>He</sub> Peak 1	<sup>DT</sup> CCS <sub>He</sub> Peak 2		Sequences	Charge State	<sup>DT</sup> CCS <sub>He</sub> Peak 1	<sup>DT</sup> CCS <sub>He</sub> Peak 2	<sup>DT</sup> CCS <sub>He</sub> Peak 1	<sup>DT</sup> CCS <sub>He</sub> Peak 2	
12-bp	d <sub>66</sub>	4	697				24-bp	d <sub>66+66</sub>	6	1119				
		5	735	793	777				7	1148	1293			
		6	788	876	881			r <sub>66+66</sub>	7	1300				
		7			1012				8	1386				
	r <sub>66</sub>	4	695					d <sub>100+100</sub>	6	1106	1249			
		5	733						7	1142	1336			
	d <sub>100</sub>	4	706					r <sub>100+100</sub>	7	1162	1331			
		5	730		726				d <sub>24ab</sub>	6	1126		1127	
		6	784		824					7	1160		1166	1308
	7			941		8					1225	1456		
	r <sub>100</sub>	4	696					9				1332	1568	
		5	737					10			1694			
14-bp	d <sub>14ab</sub>	5	822		820		32-bp	d <sub>32ab</sub>	7	1367		1368		
		6	873		873	923			8	1410		1416		
		7			1041				9			1467		
	r <sub>14ab</sub>	5	818						10			1619	1944	
18-bp	d <sub>18ab</sub>	5	935				36-bp	d <sub>100+100+100</sub>	8	1449				
		6	983		970	1088			9	1526	1811			
		7			1046	1192		r <sub>100+100+100</sub>	8	1488				
		8			1193	1297			9	1523	1856			
		9			1398				d <sub>36ab</sub>	8	1495		1503	
20-bp	d <sub>20R1-20R2</sub>	6	1056				9	1556			1554			
		7	1160				10				1516	1881		
20-bp	r <sub>20R1-20R2</sub>	6	1070				11			2180				
		7	1168											
22-pb	d <sub>22ab</sub>	6	1078		1085									
		7	1112	1202	1110	1275								
		8	1338		1198	1387								
		9			1514									

Table 12: Summary of the experimental <sup>DT</sup>CCS<sub>He</sub> determined for DNA and RNA duplexes of different sizes. Soft conditions were used to acquire the data. Samples were prepared in 150 mM NH<sub>4</sub>OAc with, when indicated, addition of 0.5% of Sulfolane.

## Chapter 4. Kissing complexes structure in the gas phase

### XXI. Introduction

As said in the introduction of Chapter 3, studies have been performed on proteins and it was shown that they undergo rearrangement, collapsing or unfolding in the gas phase depending on the activation parameters.<sup>9,10,11</sup> Concerning nucleic acids, we have shown in the previous chapter, that DNA and RNA duplexes undergo compaction at charged states naturally obtained when analyzed by *native* IM-MS.<sup>26</sup> In 2008, Bowers team has performed IM-MS on 16-mer DNA hairpin with 4 bases in the loop, 25-mer pseudoknots and 19 base pair cruciform DNA.<sup>29</sup> Using molecular dynamics of 2ns, they show that the DNA hairpin sequence 4- and pseudoknot DNA 6- are keeping their native structure in the gas phase, and that the cruciform DNA studied is undergoing a conversion of the B-Helix into an A-helix.

Based on the different studies on nucleic acids and proteins, we asked ourselves if kissing complexes will keep their native structure in the gas phase or if they will undergo compaction, as seen with duplexes. In the following chapter, We will discuss the use of IM-MS to probe the structure of kissing complexes in the gas phase, and compare our results to the ones obtained for duplexes.

### XXII. Results and discussion

#### XXII.1. Kissing complexes in the gas phase

##### XXII.1.1. Raw arrival time distributions

Data are first acquired using the softest conditions to preserve kissing complexes from dissociation and to avoid possible rearrangement due to pre-IMS activation. To do so, we used the SOFT tune, a TEGD of 4V and a fragmentor voltage at 350V (as described in chapter 1). Heat maps of the different peaks of interest are presented on Figure 80. These raw data show the arrival time distribution corresponding to each ammonium adduct.

The two hairpins forming both TAR-R06 and RNAlI-RNAlII kissing complexes do not have the same length. TAR is a 16-mer, R06 a 18-mer, RNAlI a 21-mer and RNAlII a 19-mer. More ammonium adducts are present on longer hairpins for a same charge state. For example [TAR]<sup>4+</sup> has fewer adducts than [R06]<sup>4+</sup>. The same holds for [RNAlI]<sup>5-</sup> and [RNAlII]<sup>4-</sup>. For all charge states detected for each hairpin and kissing complexes, only one arrival time distribution is seen, which means that only one conformation ensemble is detected. When looking at the average arrival time between the peak without ammonium and the ones with, for all species, no shift is seen. The hypothesis is a few ammonium

adducts do not change the final structure of hairpins in the gas phase. Also, for a same species, the lower the charge state will be, the more ammonium adducts will be present (see RNAIi and RNAIli charge state 4 and 5-).

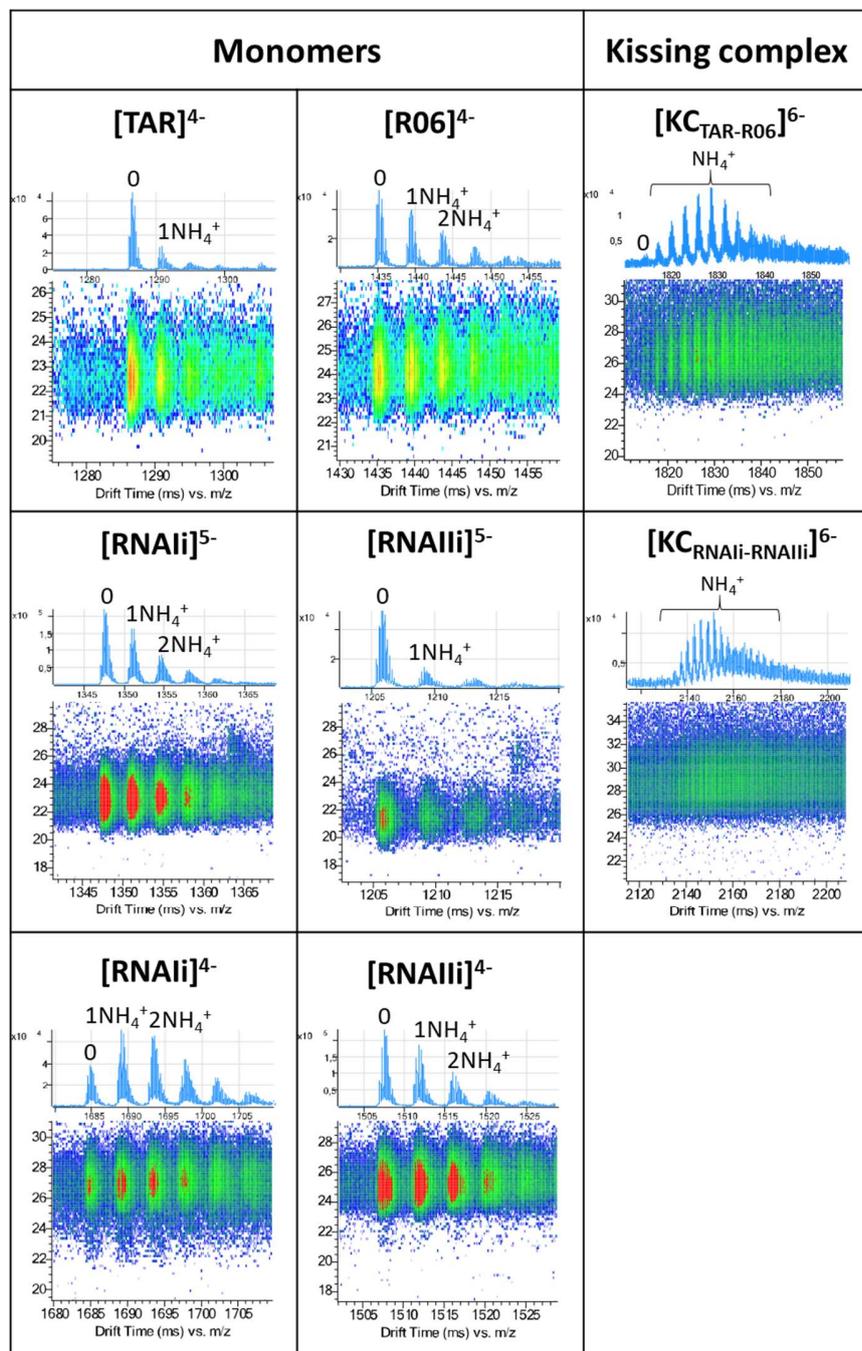


Figure 80: Total arrival time distribution of hairpins and kissing complexes using the SOFT parameters, TEGD of 4V and fragmentor set at 350V. For TAR, R06 and both KC only the major charge state is represented. As the major charge state for RNAIi and RNAIli are not the same, both charge states are shown for comparison.

### XXII.1.1.2. Analysis of CCS distributions

The  $^{DT}CCS_{He}$  distribution obtained for monomers (hairpins) and kissing complex are presented on Figure 81. The reconstruction of the distribution is done using the single field reconstruction method described in the material and method section.

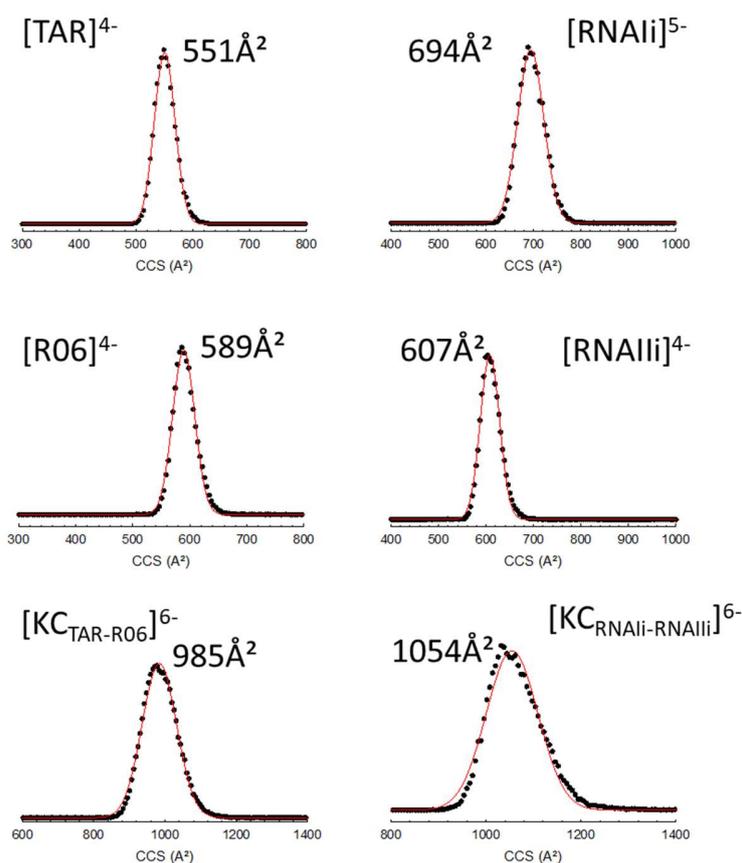


Figure 81:  $^{DT}CCS_{He}$  distributions of hairpins and KC. Only major charge states are represented. Red curve represent the Gaussian fit of the CCS distribution. For the determination of hairpin CCSD, only the peak without  $NH_4^+$  was considered. For the KC, 5 to 10 adducts were taken into account to obtain the CCSD presented here.

Figure 81 shows the  $^{DT}CCS_{He}$  distribution of the major charge states of all hairpins and kissing complexes. The results show clearly that only one broad distribution is observed for the hairpins alone, i.e. TAR, R06, RNAlI and RNAlIi, and for kissing complexes. As the peak of the distribution is broad, several related conformations coexist in the gas phase. These results suggest that the structure of hairpins and kissing complexes are flexible enough to access a large range of conformations upon desolvation and declustering.

### XXII.1.1.3. Contribution of molecular dynamics

As for the study of DNA and RNA duplexes, molecular dynamics and theoretical calculations of CCS were done to obtain information on the most probable structure of the hairpins and kissing complexes. A specific methodology to model the changes in RNA structures in the gas phase was developed by J. Abi-Ghanem. The EHSS method was used to calculate the theoretical CCS values. The method is applied to the TAR-R06 system, and the charge states considered are the ones obtained naturally by *native* MS.

As for duplexes, calculations of the theoretical CCS from the X-ray structure with solvent is done. The  $CCS_{exp}$  are found to be 25 to 30% more compact than the theoretical  $CCS_{theo}$  obtained. Here also the different structures are undergoing compaction from solution to gas phase.

	$CCS_{theo}$ in solution	$CCS_{exp}$
$[TAR]^{4-}$	$695A^2$	$551A^2$
$[R06]^{4-}$	$772A^2$	$589A^2$
$[KC_{TAR-R06}]^{6-}$	$1251A^2$	$985A^2$

Tableau 1: Theoretical values of CCS obtained for the structure of TAR, R06 and the kissing complex from the X-ray structure.

Then the new methodology developed by J-Abi-Ghanem is employed. The starting structure for the MD simulations in the gas phase is submitted to heating in 50 distinct trajectories, and the most probable protonated structures are obtained after random protonation and energy calculation. By applying this methodology to TAR and R06, both charge 4-, the final  $CCS_{theo}$  are matching with the  $CCS_{exp}$ . The results (Figure 82) show that the structure of the hairpin loop is collapsing on the 5' stem of the hairpin.

Concerning the  $CCS_{theo}$  for the  $KC_{TAR-R06}$ , they are matching with the  $CCS_{exp}$  found when SOFT parameters are applied (SOFT trapping voltage, TEGD of 4V and fragmentor at 350V). Figure 83 shows the comparison between  $CCS_{theo}$  and  $CCS_{exp}$ . The  $CCS_{theo}$  are matching the  $CCS_{exp}$  found for SOFT trapping conditions. However, the compaction of the structure is even more pronounced after ion activation by changing the trapping voltage conditions to hard ones.

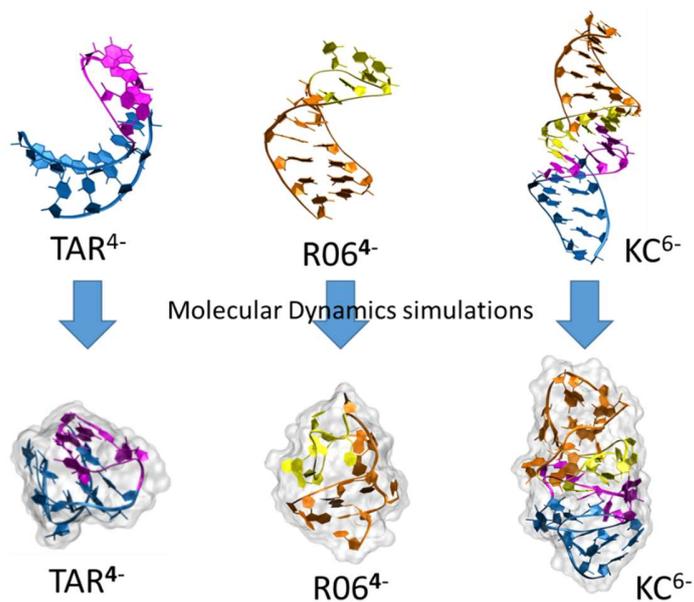


Figure 82: Structures of  $TAR^{4-}$ ,  $R06^{4-}$  and  $KC_{TAR-R06}^{6-}$  before and after molecular dynamics methodology application.

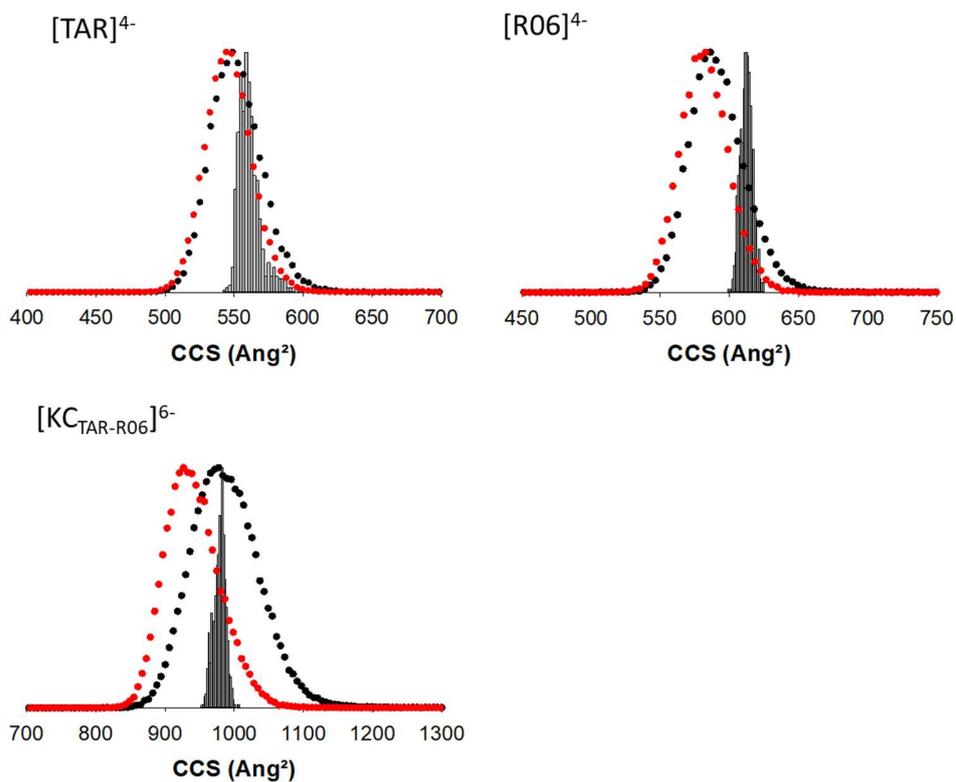


Figure 83: Comparison between  $CCS_{theo}$  obtained for the most probable structures obtained by MD and the  $DTCCS_{He}$  obtained experimentally. CCS distribution are reconstructed from the single field reconstruction method at  $\Delta V=390V$ . Black dots: SOFT trapping parameters, TEGD 4V, Fragmentor set at 350V, and red dots: HARD trapping parameters, TEGD 12 and Fragmentor at 350V.

### XXII.1.2. Collision-induced compaction of RNA kissing complexes

As we did for DNA and RNA duplexes, we wanted to have information on the conformations detected by the IM-MS when pre-IMS activation is performed. One can change three main parameters to activate the ions prior to the IMS drift tube: the Trap Entrance Grid Delta, the fragmentor and the trapping voltages. In this section, we check the effect of all these parameters on the CCS of each hairpin and kissing complex.

First, we will evaluate the effect of the fragmentor voltage on CCS distributions.  $^{DT}CCS_{He}$  are determined using soft conditions, a TEGD of -4V and the fragmentor was set at 350V or 600 V. Reconstruction of CCS distribution for the major charge states of the hairpins and kissing complexes are presenting on Figure 84.

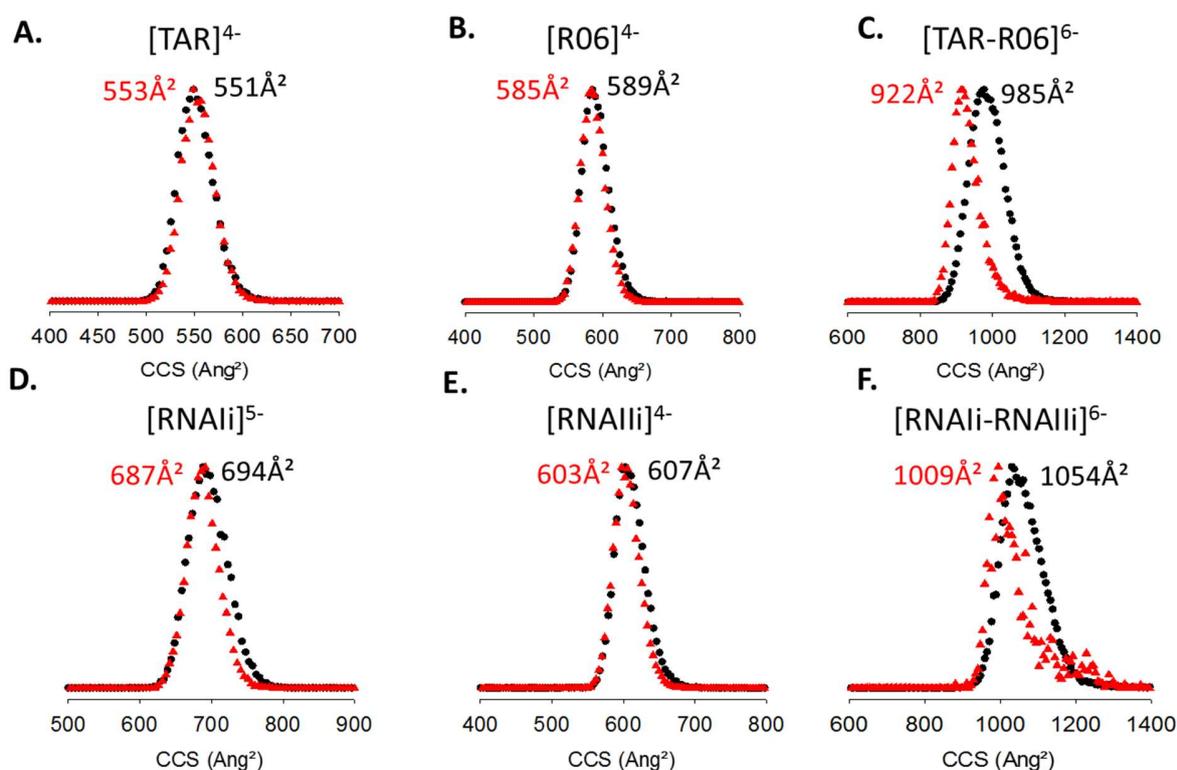


Figure 84: Effect of the source fragmentor on the  $^{DT}CCS_{He}$  distribution on the major charge state of TAR, R06,  $KC_{TAR-R06}$ , RNAIi, RNAIli and  $KC_{RNAIi-RNAIli}$  (from A to F respectively). In black dots, fragmentor at 350V and in red triangles, fragmentor at 600V. SOFT trapping parameters and TEGD at 4V were used.

Pre-IMS activation using the fragmentor voltage does not have an impact on TAR, R06 and RNAIli. No compaction or elongation is seen. A small compaction (1%) is observed for RNAIi. On the contrary, biggest changes are observed between the CCS distribution of kissing complexes at 350V and 600V. For both studied systems, i.e TAR-R06 and RNAIi-RNAIli, a compaction of the structure is detected. A decrease of about 5% is observed between the CCS distributions at 350V and 600V. In chapter 3, we saw that the same

percentage of compaction was observed for duplexes, like  $r_{100}$  or  $r_{17R1-17R2}$ . Large structures such as kissing complexes or duplexes undergo changes when activated in the gas phase.

To evaluate the effect of the Trap Entrance Grid Delta on the CCS distribution, we performed CIU experiments by increasing progressively the value of the TEGD from 4 to 20V. This method allows us to follow the eventual changes in CCS during pre-IMS activation. CIU plots are then done for each hairpin and each kissing complex. We also tested the effect of the set of trapping voltage (SOFT vs HARD) in addition to an increase of the TEGD. CIU plots are gathered on Figure 85. The  $^{DT}CCS_{He}$  values determined for a TEGD at 4V are indicated on each CIU plot.

The  $^{DT}CCS_{He}$  distribution for  $[TAR]^{4-}$ ,  $[R06]^{4-}$ ,  $[RNAIi]^{5-}$  and  $[RNAIi]^{4-}$  is not changed when pre-IMS activation is performed by increasing the TEGD. Indeed, only one conformation is detected and remains all along the experiment. However even if no change is occurring during activation by the TEGD, a compaction of less than 5% is observed between the SOFT conditions at TEGD -4V and HARD conditions at TEGD -4V.

For TAR-R06 and RNAIi-RNAIi kissing complexes, the  $^{DT}CCS_{He}$  distribution of the 6- charge state is broad at low pre-IMS activation and SOFT conditions. CIU plots reveal that several conformations may coexist and the activation is shifting the population of conformations to compact ones. At the end, only one broad distribution is observed. A first compaction is observed between SOFT and HARD conditions. A compaction of 4% is observed for  $[KC_{TAR-R06}]^{6-}$  and 3% for  $[KC_{RNAIi-RNAIi}]^{6-}$ . Also, compaction is occurring when the TEGD is increased, both with SOFT and HARD conditions. The CIU plots in SOFT and HARD conditions show a continuity in the compaction of the structure. As compaction is occurring at high pre-IMS activation, rearrangement may occur in the gas phase, as for duplexes.

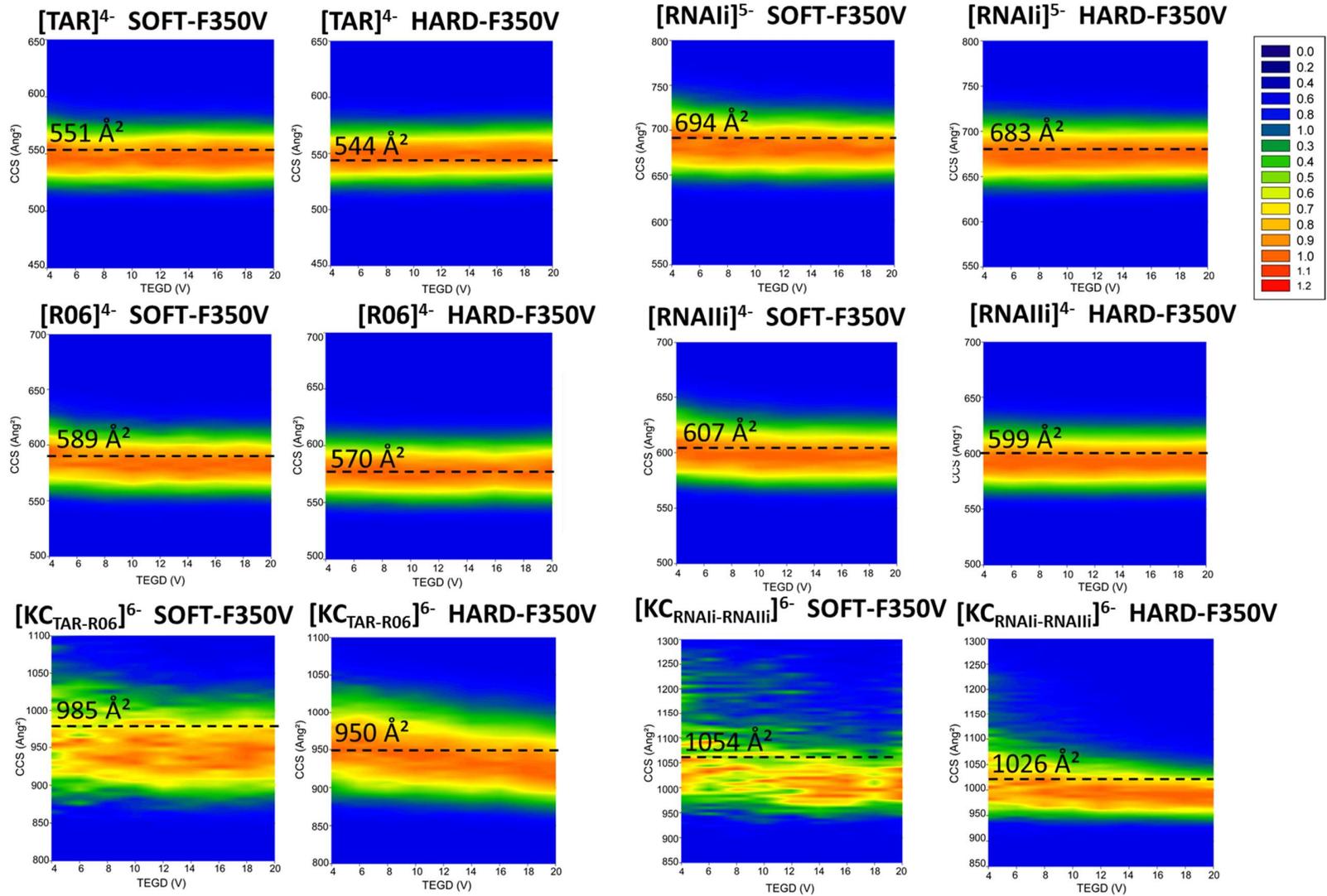


Figure 85: CIU plots showing the  $^{DT}CCS_{He}$  distribution evolution of the major charge state as a function of the TEGD voltage for TAR, R06,  $KC_{TAR-R06}$ , RNAIi, RNAIi and  $KC_{RNAIi-RNAIi}$ . The two sets of trapping voltages are tested: SOFT and HARD conditions on the right. Fragmentor was set at 350V.  $^{DT}CCS_{He}$  value at TEGD 4V are noted on the figure.

#### XXII.1.4. Effect of magnesium ions on the structure

IM-MS is used to depict conformational changes upon activation, but it can also bring knowledge on changes upon cations or ligand binding<sup>36,37</sup>. In the first part of this manuscript, we have studied the impact of  $Mg^{2+}$  ions on loop-loop interactions and begun to study  $Mg^{2+}$  localization. Using IM-MS we can potentially bring new data on the potential effect of magnesium on the structure of kissing complexes. Does magnesium extend the structure of the KC or does it compact it? To answer this question, we performed IM-MS using SOFT conditions to reduce as much as possible the compaction due to pre-IMS activation. However, we have also seen when using too soft conditions the  $Mg^{2+}$  distribution is not well defined and distribution with  $NH_4^+$  can overlap. We decided then to perform the experiments using the SOFT trapping voltages, a TEGD of 4V and the two specific voltages of fragmentor, i.e 350V and 600V. Using these conditions, we can compare the CCS distributions obtained and have information about the effect of magnesium on kissing complexes conformations. The experiments were done on TAR-R06 kissing complex. The results are shown on Figure 86.

Raw heat maps of  $[KC_{TAR-R06}]^{6-}$  at 350V and 600V are represented on Figure 86, A and B respectively. At 350V, the arrival time distribution of the different  $Mg^{2+}$  adducts is unchanged. At 600V, the arrival time distribution looks broader for the peaks with magnesium than the one without. To better visualize this, we reconstructed the  $^{DT}CCS_{He}$  distribution of  $[KC_{TAR-R06}]^{6-}$ ,  $[KC_{TAR-R06}+1Mg^{2+}]^{6-}$  and  $[KC_{TAR-R06}+2Mg^{2+}]^{6-}$  at 350V and 600V (Figure 86C and 18D, respectively). Indeed, the CCS distribution is broader with  $Mg^{2+}$ , with a contribution at higher CCS. It means that  $Mg^{2+}$  increases the number of conformation coexisting, and that the additional ones in presence of  $Mg^{2+}$  are less compact. It could be that the collapse is incomplete when magnesium is retained.

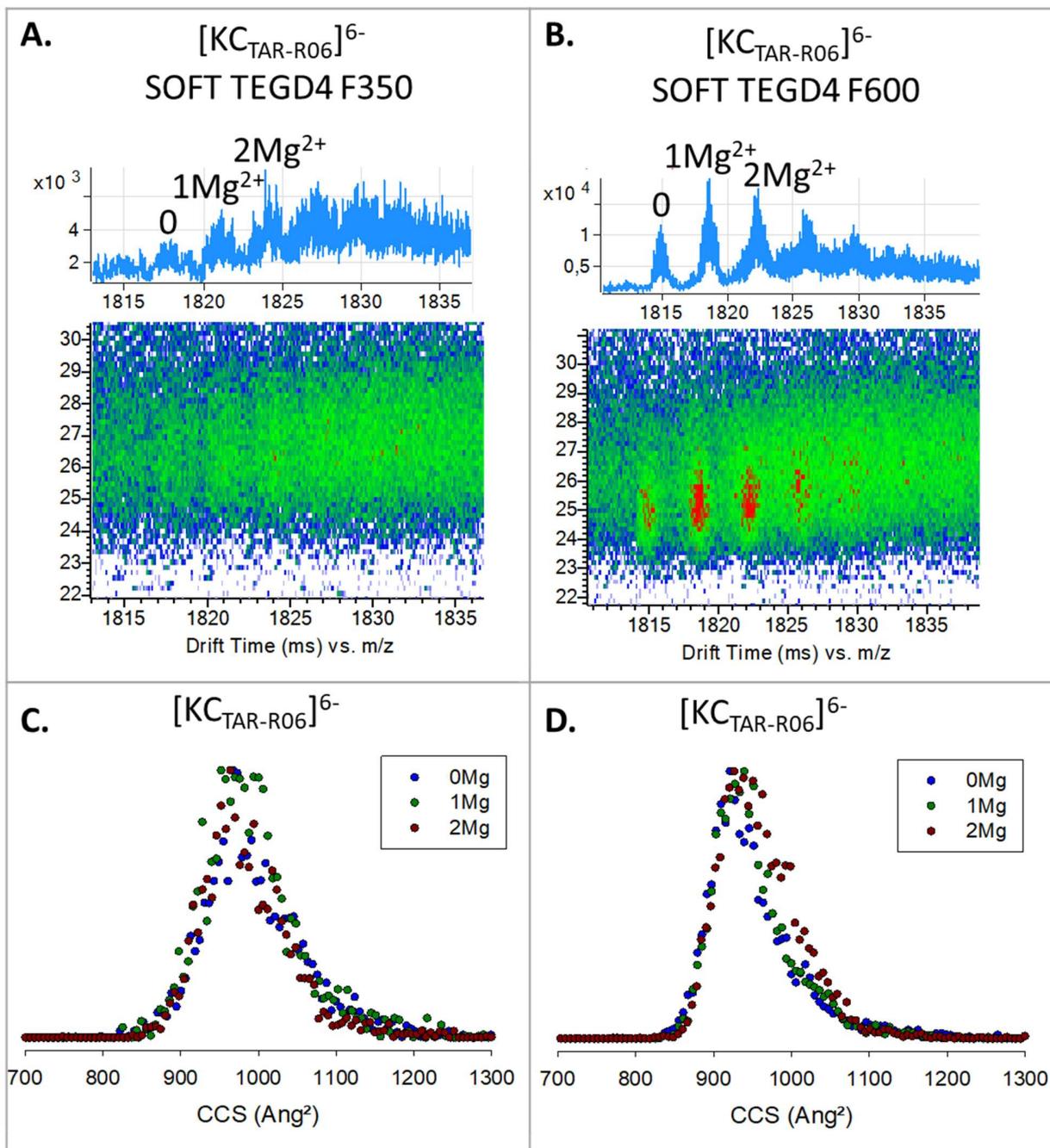


Figure 86: Determination of the effect of  $Mg^{2+}$  on the arrival time and CCS of TAR-R06 kissing complex. A) Raw heat map of  $[KC_{TAR-R06}]^{6-}$  using the SOFT tune, TEGD 4 and F350. B) Raw heat map of  $[KC_{TAR-R06}]^{6-}$  using the SOFT tune, TEGD 4 and F600. C and D) Reconstruction of the  $^{DT}CCS_{He}$  distribution of the peak corresponding to  $[KC_{TAR-R06}]^{6-}$ ,  $[KC_{TAR-R06}+1Mg^{2+}]^{6-}$  and  $[KC_{TAR-R06}+2Mg^{2+}]^{6-}$ .

## XXII.2. Comparison RNA kissing complexes with RNA duplexes

### XXII.2.1. ESI-MS analysis of RNA duplexes and RNA kissing complexes

#### XXII.2.1.1. Experimental aspects

In this chapter we will compare RNA duplexes to RNA kissing complexes having the same number of bases, and the same composition to unbiased the analysis. We chose to compare the TAR-R06 kissing complex with the 17R1-17R2 duplex used for the adduct cleaning in part I. We used both the DNA and RNA sequence of 17R1-17R2 for complete comparison. We also studied RNAIi-RNAIii kissing complex and 20R1-20R2 duplex.

As in the previous study of duplexes, the experiments are performed using the AGILENT DTIMS-Q-TOF where the mobility cell is filled with Helium.

#### XXII.2.1.2. ESI-MS spectra : RNA duplexes vs kissing complexes

The spectra on Figure 87 are obtained using the SOFT instrumental parameters, only to maintain the duplexes and kissing complexes formed. The figure shows the spectra of TAR-R06 compared to 17R1-17R2, and RNAIi-RNAIii compared to 20R1-20R2.

The 6- charge state is the major charge state detected for both couple duplex/kissing complex. Duplexes and kissing complexes are detected at the same charge states. The two hairpins forming each kissing complex are detected at different charge states, whereas no single strand is detected for the RNA duplexes. As the same charge states are detected for all couples duplex/kissing complex, comparison of their respective  $^{DT}CCS_{He}$  can be done.

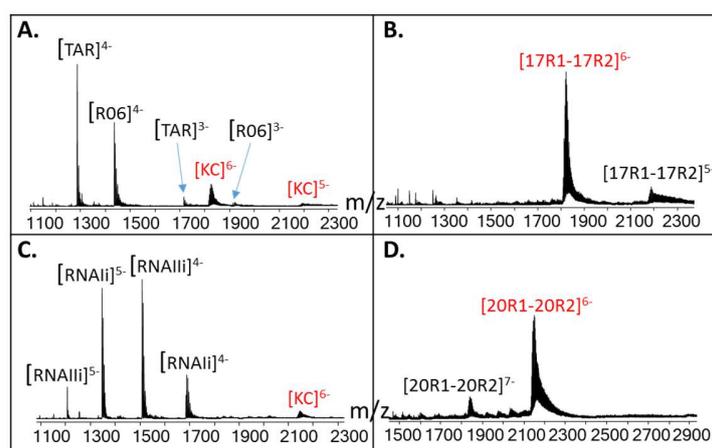


Figure 87: ESI-MS of A) TAR-R06 kissing complex, B) the RNA duplex 17R1-17R2, C) RNAIi-RNAIii kissing complex and D) RNA duplex 20R1-20R2. Spectra are recorded in 150 mM  $NH_4OAc$  from soft source conditions. Major charge state for each complex analyzed is in red.

### XXII.2.1.3. CCS distribution: comparison between RNA duplexes and RNA kissing complexes

CCS distribution reconstruction are obtained by using the single-field reconstruction method. Each kissing complex and its relative duplex possess the same percentage of GC base pair.

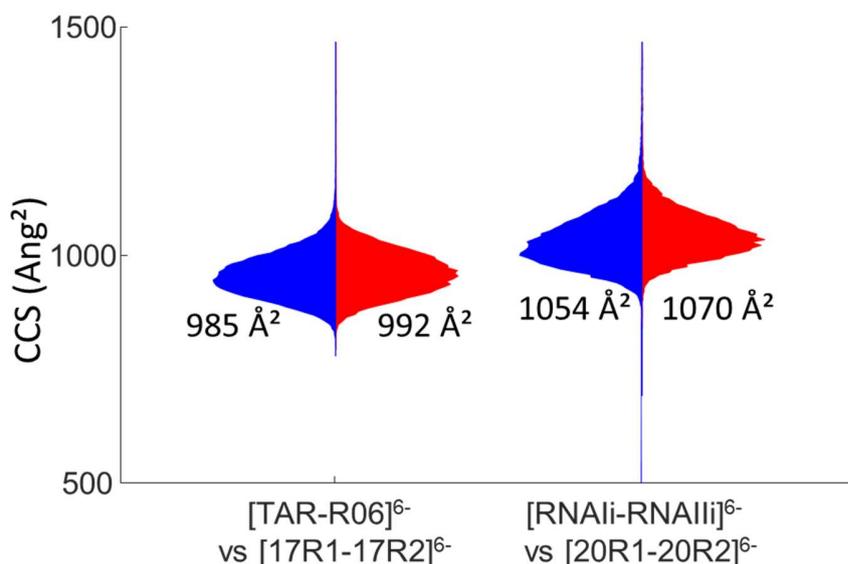


Figure 88: Violin plots comparing the CCS distribution obtained for each kissing complexes and their corresponding duplexes.

On Figure 88, we can visualize the CCS distribution of the kissing complexes and their corresponding duplexes. The distribution of  $[KC_{\text{TAR-R06}}]^{6-}$  is very similar to the one of  $[17R1-17R2]^{6-}$ . Both distributions are equally broad. Similar results are observed for  $[KC_{\text{RNAIi-RNAIii}}]^{6-}$  and  $[20R1-20R2]^{6-}$ . These broad distributions reflect the flexibility of both duplexes and kissing complexes. Also, the  $^{DT}CCS_{\text{He}}$  for the peak center of the distribution is lower for KC than for duplexes. A small difference (less than 2%) remains between the CCS for KC and duplexes. However, this difference may not be significant enough.

### XXIII. Conclusion and perspectives

Like DNA and RNA duplexes, RNA kissing complexes are undergoing compaction at charge states naturally obtained by *native* MS. It could be of interest to study higher charge states of kissing complexes to know if at higher charge states the structure is going to be more elongated, as for duplexes, or if compaction will remain.

## References

- (1) Bai, Y., Das, R., Millett, I. S., Herschlag, D., and Doniach, S. (2005) Probing counterion modulated repulsion and attraction between nucleic acid duplexes in solution. *Proc. Natl. Acad. Sci.* *102*, 1035–1040.
- (2) Bai, Y., Greenfeld, M., Travers, K. J., Chu, V. B., Lipfert, J., Doniach, S., and Herschlag, D. (2007) Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *J. Am. Chem. Soc.* *129*, 14981–8.
- (3) Arcella, A., Portella, G., and Orozco, M. (2014) Structure of Nucleic Acids in the Gas Phase, in *Nucleic Acids in the gas phase*, p 297.
- (4) Lipfert, J., Doniach, S., Das, R., and Herschlag, D. (2014) Understanding nucleic acid-ion interactions. *Annu. Rev. Biochem.* *83*, 813–41
- (5) Montange, R. K., and Batey, R. T. (2008) Riboswitches: emerging themes in RNA structure and function. *Annu. Rev. Biophys.* *37*, 117–33.
- (6) Roth, A., and Breaker, R. R. (2009) The Structural and Functional Diversity of Metabolite-Binding Riboswitches. *Annu. Rev. Biochem.* *78*, 305–334.
- (7) Serganov, A., and Nudler, E. (2013) A decade of riboswitches. *Cell* *152*, 17–24.
- (8) Bohrer, B. C., Merenbloom, S. I., Koeniger, S. L., Hilderbrand, A. E., and Clemmer, D. E. (2008) Biomolecule Analysis by Ion Mobility Spectrometry. *Annu. Rev. Anal. Chem.* *1*, 293–327.
- (9) Meyer, T., Gabelica, V., Grubmüller, H., and Orozco, M. (2013) Proteins in the gas phase. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* *3*, 408–425.
- (10) Breuker, K., and McLafferty, F. W. (2008) Stepwise evolution of protein native structure with electrospray into the gas phase, 10<sup>-12</sup> to 10<sup>-2</sup> s. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 18145–18152.
- (11) Michaelievski, I., Eisenstein, M., and Sharon, M. (2010) Gas-phase compaction and unfolding of protein structures. *Anal. Chem.* *82*, 9484–9491.
- (12) Eschweiler, J. D., Rabuck-Gibbons, J. N., Tian, Y., and Ruotolo, B. T. (2015) CIUSuite: A Quantitative Analysis Package for Collision Induced Unfolding Measurements of Gas-Phase Protein Ions. *Anal. Chem.* *87*, 11516–11522.
- (13) Bornschein, R. E., Niu, S., Eschweiler, J., and Ruotolo, B. T. (2016) Ion Mobility-Mass Spectrometry Reveals Highly-Compact Intermediates in the Collision Induced Dissociation of Charge-Reduced Protein Complexes. *J. Am. Soc. Mass Spectrom.* *27*, 41–49.
- (14) Zhong, Y., Han, L., and Ruotolo, B. T. (2014) Collisional and Coulombic Unfolding of Gas-Phase Proteins: High Correlation to Their Domain Structures in Solution. *Angew. Chemie Int. Ed.* *53*, 9209–9212.
- (15) Schnier, P. D., Klassen, J. S., Strittmatter, E. F., and Williams, E. R. (1998) Activation Energies for Dissociation of Double Strand Oligonucleotide Anions: Evidence for Watson–Crick Base Pairing in Vacuo. *J. Am. Chem. Soc.* *120*, 9605–9613.

- (16) Gabelica, V., and De Pauw, E. (2002) Collision-induced dissociation of 16-mer DNA duplexes with various sequences: Evidence for conservation of the double helix conformation in the gas phase. *Int. J. Mass Spectrom.* 219, 151–159.
- (17) Pan, S., Sun, X., and Lee, J. K. (2006) Stability of complementary and mismatched DNA duplexes: Comparison and contrast in gas versus solution phases. *Int. J. Mass Spectrom.* 253, 238–248.
- (18) Gidden, J., Ferzoco, A., Baker, E. S., and Bowers, M. T. (2004) Duplex formation and the onset of helicity in poly d(CG)<sub>n</sub> oligonucleotides in a solvent-free environment. *J. Am. Chem. Soc.* 126, 15132–15140.
- (19) Baker, E. S., and Bowers, M. T. (2007) B-DNA Helix Stability in a Solvent-Free Environment. *J. Am. Soc. Mass Spectrom.* 18, 1188–1195.
- (20) Burmistrova, A., Gabelica, V., Duwez, A. S., and De Pauw, E. (2013) Ion mobility spectrometry reveals duplex DNA dissociation intermediates. *J. Am. Soc. Mass Spectrom.* 24, 1777–1786.
- (21) Drew, H. R., Wingt, R. M., Takano, T., Broka, C., Tanakt, S., Itakura, K., and Dickerson, R. E. (1981) Structure of a B-DNA dodecamer : Conformation and dynamics. *Biochemistry* 78, 2179–2183.
- (22) Lippens, J. L., Ranganathan, S. V., D’Esposito, R. J., and Fabris, D. (2016) Modular calibrant sets for the structural analysis of nucleic acids by ion mobility spectrometry mass spectrometry. *Analyst* 141, 4084–4099.
- (23) Pasi, M., Maddocks, J. H., and Lavery, R. (2015) Analyzing ion distributions around DNA: Sequence-dependence of potassium ion distributions from microsecond molecular dynamics. *Nucleic Acids Res.* 43, 2412–2423.
- (24) Marchand, A., Livet, S., Rosu, F., and Gabelica, V. Ion Mobility : How to Reconstruct Collision Cross Section Distributions from Arrival Time Distributions ? *Anal Chem.*
- (25) D’Atri, V., Porrini, M., Rosu, F., and Gabelica, V. (2015) Linking molecular models with ion mobility experiments. Illustration with a rigid nucleic acid structure. *J. Mass Spectrom.* 50, 711–726.
- (26) Porrini, M., Rosu, F., Rabin, C., Darré, L., Gómez, H., Orozco, M., and Gabelica, V. (2017) Compaction of Duplex Nucleic Acids upon Native Electrospray Mass Spectrometry. *ACS Cent. Sci.* 3, 454–461.
- (27) Ogorzalek Loo, R. R., Lakshmanan, R., and Loo, J. A. (2014) What protein charging (and Supercharging) reveal about the mechanism of electrospray ionization. *J. Am. Soc. Mass Spectrom.* 25, 1675–1693.
- (28) Metwally, H., McAllister, R. G., Popa, V., and Konermann, L. (2016) Mechanism of Protein Supercharging by Sulfolane and m -Nitrobenzyl Alcohol: Molecular Dynamics Simulations of the Electrospray Process. *Anal. Chem.* 88, 5345–5354.
- (29) Baker, E. S., Dupuis, N. F., and Bowers, M. T. (2009) DNA hairpin, pseudoknot, and cruciform stability in a solvent-free environment. *J. Phys. Chem. B* 113, 1722–7.
- (30) Brunel, C., Marquet, R., Romby, P., and Ehresmann, C. (2002) RNA loop-loop interactions as dynamic

functional motifs. *Biochimie* 84, 925–944.

(31) Barth, A., Kobbe, D., and Focke, M. (2016) DNA-DNA kissing complexes as a new tool for the assembly of DNA nanostructures. *Nucleic Acids Res.* 44, 1502–1513.

(32) Chang, K. Y., and Tinoco, I. (1997) The structure of an RNA “kissing” hairpin complex of the HIV TAR hairpin loop and its complement. *J. Mol. Biol.* 269, 52–66.

(33) Lebars, I., Richard, T., Di primo, C., and Toulmé, J. J. (2007) NMR structure of a kissing complex formed between the TAR RNA element of HIV-1 and a LNA-modified aptamer. *Nucleic Acids Res.* 35, 6103–6114.

(34) Lebars, I., Legrand, P., Aimé, A., Pinaud, N., Fribourg, S., and Di Primo, C. (2008) Exploring TAR-RNA aptamer loop-loop interaction by X-ray crystallography, UV spectroscopy and surface plasmon resonance. *Nucleic Acids Res.* 36, 7146–7156.

(35) Bernacchi, S., Ennifar, E., Tóth, K., Walter, P., Langowski, J., and Dumas, P. (2005) Mechanism of hairpin-duplex conversion for the HIV-1 dimerization initiation site. *J. Biol. Chem.* 280, 40112–21.

(36) Marchand, A., and Gabelica, V. (2014) Native electrospray mass spectrometry of DNA G-quadruplexes in potassium solution. *J. Am. Soc. Mass Spectrom.* 25, 1146–1154.

(37) Göth, M., and Pagel, K. (2017) Ion mobility–mass spectrometry as a tool to investigate protein–ligand interactions. *Anal. Bioanal. Chem.* 409, 4305–4310.



## 4. GENERAL DISCUSSION AND CONCLUSIONS

The goal of this work was to bring new knowledge in the fields of RNA structure and interaction with magnesium cations using *native* mass spectrometry and ion mobility spectrometry. This work aims to contribute to the development of the use of mass spectrometry in RNA research. This work was divided in two parts answering different problematics:

- 1) The first part consisted in the characterization of binding equilibria between magnesium and RNA kissing complexes, used as models. Our goal was to develop a method to quantify the different species by taking into account the non-specific adducts and determine equilibrium binding constants.
- 2) The second part was focused on the characterization of structures of RNA duplexes and kissing complexes by ion mobility-mass spectrometry. Our goal was to investigate the behavior of such complexes in the gas phase, with the aim to detect shape changes upon cation binding.

We based our study on RNA models which were kissing complexes formed between TAR and its aptamer R06, and RNAIi and RNAIli. These two kissing complexes were well characterized by other biophysical techniques which allow us to compare the results obtained by mass spectrometry. For almost all the experiments presented in this manuscript we used the AGILENT DTIMS Q-TOF instrument. Other instruments were tested as comparison like the LCQ Fleet for MS/MS data or the LCT premier.

## XXIV. Discussion about mass spectrometry results (PART I)

### XXIV.1. Mass spectrometry compatible with Magnesium salts

Our results showed that solutions doped with  $\text{Mg}(\text{OAc})_2$ ,  $\text{MgCl}_2$  or  $\text{MgBr}_2$  up to 1 mM can be analyzed by *native* MS. Our results also suggested that optimization of the method was needed to distinguish and quantify all the different adducts of  $\text{Mg}^{2+}$ , whatever the instrument used.

Our starting point for the instrument parameters were the ones usually used in the team to study G-quadruplexes or i-motifs in conditions soft enough to maintain non-covalent interactions and ensure declustering. However, these conditions were too soft for the study of kissing complexes with  $\text{Mg}^{2+}$ . Indeed, a large number of ammonium adducts were retained on hairpin monomers and kissing complexes. Therefore the adduct distributions of the different cations (i.e.  $\text{NH}_4^+$ ,  $\text{Mg}^{2+}$  and  $\text{Na}^+$ ) were overlapping, making quantification difficult. To reduce the sodium adducts, the samples were carefully prepared and desalted prior to the experiments. The use of harsher conditions by changing parameters involved in ion activation allowed us to get rid of  $\text{NH}_4^+$  adducts, and declustering was better. In the DTIMS Q-TOF instrument, trapping voltages are important for ion activation. One can increase the trapping voltages, and/or trap entrance grid delta (TEGD) voltage or the fragmentor voltage to activate more ions. These parameters have to be adjusted depending on the molecule analyzed.

In these conditions, we could analyze kissing complexes in solutions containing up to 1 mM  $\text{Mg}^{2+}$ . Some would argue that concentrations of  $\text{Mg}^{2+}$  below 1 mM does not reflect physiological conditions. Most biophysical studies in the literature are carried out in 3 or 5 mM  $\text{Mg}^{2+}$ . However, while it is true that the concentrations used are below the *total* concentration of  $\text{Mg}^{2+}$  in the cells, they are in the same range than the *free*  $\text{Mg}^{2+}$  in the cells, which means that our experiments can fairly mimic cells environment.

### XXIV.2. Analysis of $\text{Mg}^{2+}$ binding to RNA by *native* mass spectrometry

We have developed a method to quantify each different species, by subtracting the contribution of the magnesium non-specific adducts. It was then possible to determine individual binding constants, and tandem mass spectrometry helped to localize magnesium cations binding to kissing complexes. This treatment can be adapted to different cations, like here with  $\text{Mg}^{2+}$  or  $\text{Mn}^{2+}$ , and also to different sequences and motifs.

### *XXIV.2.1. Method to determine non-specific vs specific adducts distribution: critical points and limitations*

In Chapter 1.XI. we showed how to mathematically extract the specific magnesium distribution from the total magnesium adducts distribution.

The first step is to define what is specific and what is not. Specific  $Mg^{2+}$  ions are magnesium ions which are binding specifically to the motif of interest, here kissing complexes. These  $Mg^{2+}$  ions are pertaining to this motif only. Non-specific adducts are binding presumably on the external phosphates on non-specific locations (i.e. there are numerous low-affinity binding sites). These non-specific cations come from the solution, but can also be formed during the ESI process. We need to find a control sequence to subtract these two types of non-specific adducts. The reference sequence must be a molecule of the same type, i.e. RNA for RNA, with the same length and same composition than the sequence studied, but not containing the specific motif. For the RNA kissing complexes, we chose RNA duplexes as references. As the amount of non-specific adducts can vary from one reference to another, it can be a source of errors that can be controlled by comparing different references.

The mathematical treatment by itself can be a source of error. Indeed, all the equations depend on the proportions of non-specific adducts determined and depend on the extracted integrals of each peaks. Integration of the peaks have to be done carefully and background has also to be taken into account (and assumed to reflect the background of the overall spectrum). The equations depend also on the number of non-specific adducts considered. The error is increasing with the number of adducts considered.

### *XXIV.2.2. Quantitative aspects*

In chapter I.XII, apparent equilibrium binding constants between two hairpins as a function of  $Mg^{2+}$  concentration were determined. We showed also that by combining stoichiometry and global binding constants, it is possible to have access to individual magnesium ion binding constants. We then have shown that if several magnesium ions are specifically binding to a kissing complex, which is the case for TAR-R06, these two magnesium do not necessarily have the same affinities. The individual  $K_D$ s were determined as per individual stoichiometry, and not per binding site.

Equilibrium binding constants are determined using data from a gas phase technique but they reflect solution equilibria. One important step here is to correlate the relative intensities of the different species to their abundance in solution. Different molecules do not respond in the same way to the electrospray process. A correction is sometimes needed based on the determination of ratios of response factors that

are correlated to ratios between intensities. This correction is a source of error for the calculated concentrations as it needs calculations. If response factors cannot be determined, like it was the case for the different  $Mg^{2+}$  adducts, one can make the approximation that the response factors are equal and the relative intensity will be directly proportional to the concentration in solution.

Another important point is the way specific and non-specific adducts are distinguished. We saw that, for TAR-R06, maybe a fraction of the adducts that is not specific compared to the duplex reference may still contribute to the stabilization of the kissing loop motif compared to the separate hairpins. This is something that should be also consider: non-specific adducts are stabilizing the RNA structure in addition to the specific ones.

#### *XXIV.2.3. MS/MS as a tool for $Mg^{2+}$ localization ?*

In Chapter I.XIII, we used tandem mass spectrometry to investigate on the position of the most affine  $Mg^{2+}$  ion. Our results led us to propose a model of equilibrium for each kissing complexes studied. Thanks to the results obtained by MS/MS, we hypothesize a negative cooperativity between two  $Mg^{2+}$  binding sites. For RNAIi-RNAIi, the cooperativity is so negative that the two potential binding sites cannot be populated at the same time. For TAR-R06, the binding of the first  $Mg^{2+}$  ion on one of the two sites, reduces the affinity of the second  $Mg^{2+}$  ion for the second binding site.

We note that MS/MS interpretation has a caveat.  $Mg^{2+}$  is forming an ion pair with the RNA that cannot be removed during the collision activation. However, it is known that the affinity of  $H_2O$  for  $Mg^{2+}$  is high, so we suppose that during desolvation the last water molecules to go are the ones close to magnesium ions. If water molecules exist between  $Mg^{2+}$  and RNA, like it is the case for “outer sphere” binding, one could still argue that if water is removed,  $Mg^{2+}$  could either stay bound to RNA or move to another place. During the desolvation, relocalization of magnesium cations could happen.

However the results obtained are very promising. First, we have brought new knowledge on fragmentation pathways of kissing complexes depending on the instrument used and collision time. Indeed, at long collision time, base loss is predominant. On the contrary, low collision time influences the dissociation of the kissing complex towards the fragmentation in its two monomers. Another important point is the charge density. We showed that depending on the length of the kissing complex, it can be necessary to increase the charge state by addition of a supercharging agent. As  $Mg^{2+}$  cannot be removed, MS/MS is a good technique to probe the potential location of this cation. We have thus opened doors for future investigations of the magnesium binding sites by MS/MS. Higher-resolution mass spectrometers

with more comprehensive backbone fragmentation (for example, with fragmentation of radicals) would be helpful.

### XXIV.3. Mass spectrometry as a screening tool

In Chapter 2, we showed that *native* MS can also be used as a screening tool to quickly rank kissing complexes differing by one base pair depending on their relative stability. We also demonstrate that MS is a complementary method to other biophysical techniques (the MS results were consistent with SPR and UV-melting results), and MS could be used for the characterization of new aptamers or the screening of designed variants.

The advantage of mass spectrometry for screening is that the identification of the different species is quick, it can even allow to detect “unknown” complexes present in solution, like homodimers. Stoichiometries are easily identified in an assumption-free manner, and the relative intensities can be directly correlated to the abundances in solution thanks to a correction with an internal standard. Indeed, as a simplification and to do a quicker analysis, the response factors of all the different kissing complexes studied were assumed to be equal, and the internal standard was the common strand. Making this simple assumption here, helped us to quickly confirm results obtained from other techniques if no absolute determination of the  $K_{DS}$  is needed.

## XXV. Discussion about the ion mobility-mass spectrometry results (PART II)

### XXV.1. Initial compaction in the gas phase

Chapter 3 and 4 were focused on the analysis of the structure of duplexes (RNA and DNA) and RNA kissing complexes in the gas phase. One of our main results is that for the charge states naturally observed in *native* mass spectrometry, a compaction of the structure is observed even when using soft pre-IMS conditions. This result is true for both DNA and RNA duplexes and also for hairpins and kissing complexes. Our study also shows the complementarity between experimental data and theoretical ones obtained by molecular dynamics and theoretical collision cross section calculations.

The charge states naturally observed in *native* MS are low compared to the number of phosphates. In solution, DNA and RNA are surrounded by the “ion atmosphere”, which reduces the net charge of nucleic acids. It is possible that the final charge state reflects the thickness of the counterion layer around the nucleic acids. The comparison between DNA and RNA duplexes also indicates that RNA is more prone to retain ammonium cations.

IM-MS reveals that ensembles of conformations are adopted. Our results showed that the arrival time distribution of both RNA and DNA complexes but also kissing complexes were broad compared for example to a rigid G-quadruplex. To check if this broadness was due to diffusion and/or to coexistence of a population of structures, we reconstructed the CCS peak without the contribution of diffusion. The data showed that the broadness is due to a population of structures coexisting in the gas phase. The use of IM-MS showed us that duplexes and kissing complexes do not adopt just one conformation but an ensemble of related conformations. We also show that DNA and RNA duplexes have similar  $CCS_{exp}$ , so no difference between A- and B-helix is seen by IM-MS.

As native proteins keep their globular shape from the solution to the gas phase, it was thought that nucleic acids could similarly keep their structure intact. But contrary to this anticipation, with the help of molecular dynamics and  $CCS_{theo}$  calculation using the EHSS method, it was possible to show that the structures of DNA and RNA are more compact in the gas phase than their initial structure in solution. Our hypothesis is that, at low charge states produced from native solution conditions (aqueous ammonium acetate at physiological ionic strength), the trend of all biomolecules is to become globular in the gas phase, because the formation of extra intramolecular hydrogen bonds is not counteracted by sufficient Coulomb repulsion. For nucleic acids, the compaction results in a huge change of shape and CCS.

Initially we wanted to use ion mobility spectrometry to study the effect of cations or ligands on the structure of more complex RNAs. But our research revealed a limitation of the use of IM-MS to probe solution structure. One need to be aware that compaction is happening, and that IM-MS may not be suitable for the study of conformational changes. This has to be kept in mind when studying RNA complexes in the gas phase. A direction worth exploring in the future is whether increasing the charge states produced by electrospray could prevent compaction from occurring.

## XXV.2. Collision-induced compaction

IM-MS was also used to monitor the effect of ion activation on RNA duplexes and kissing complexes gas-phase conformations. Our results showed that by activating the ions, a compaction can be observed at all “native” charge states. From these results, one can say that a “double compaction” is happening. The first one due to the transition between solution and gas phase, and the other one due to activation.

For RNA and DNA duplexes this statement is true for all sequences tested but the phenomenon was more pronounced on RNA than DNA. The presence of more  $\text{NH}_4^+$  adducts on RNA could explain this hypothesis. Indeed,  $\text{NH}_4^+$  help to maintain a more elongated structure and when declustering is more pronounced, the structure is further compacted. Activation does not have an impact on the hairpin arrival time distribution, probably because the compaction was already complete even in our softest ion transfer conditions. Based on our results, the compaction happens until the structure cannot be compacted anymore. The size of the RNA, its total charge state, and its microscopic distribution of charges (influenced by cations remaining bound) will influence this step.



## 5. APPENDICES

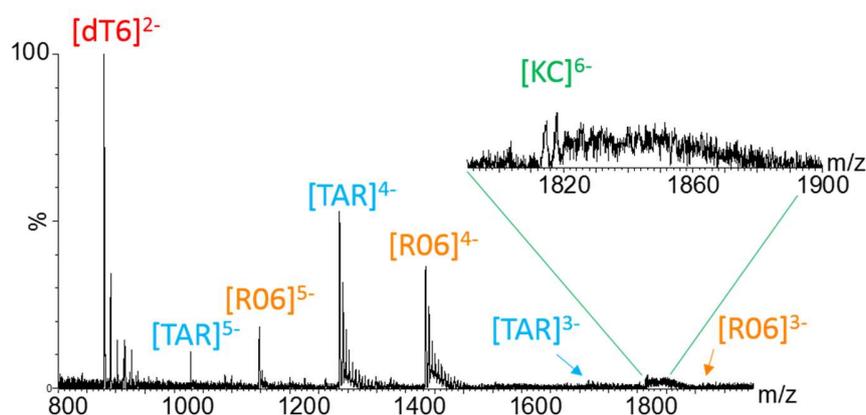
### I. A1. Appendices Chapter 1

#### A1.1. Optimization of the method using the LCT premier

##### **Solutions without magnesium**

The first optimizations were done on the LCT premier (described in the material and methods chapter).

The spectrum on SI 1 was obtained using the soft conditions described in Marchand et al.<sup>12</sup>



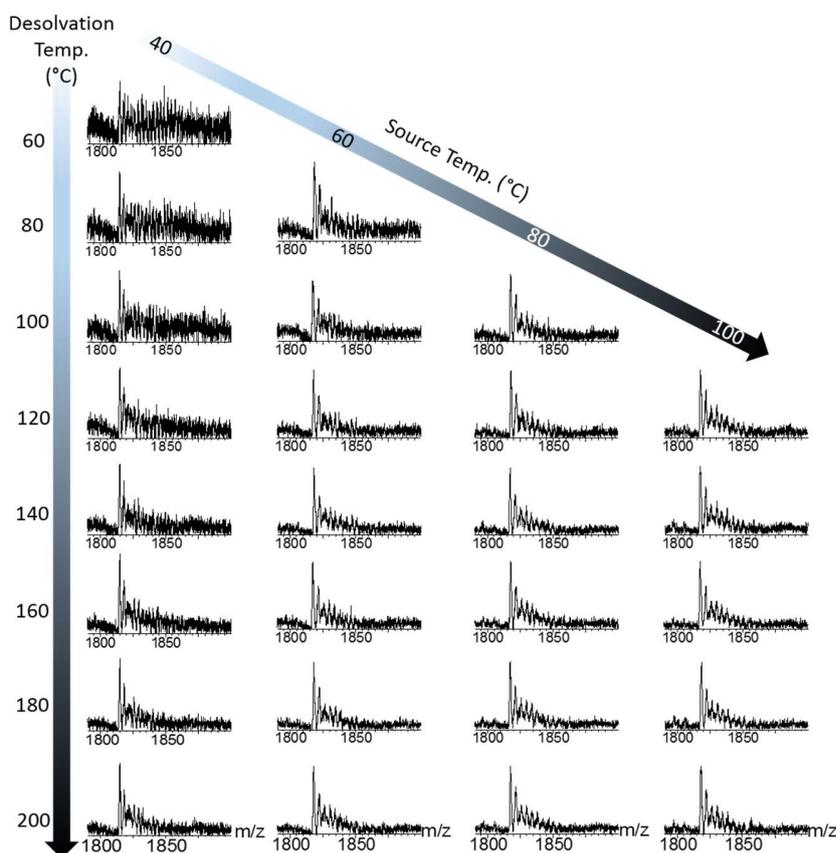
SI 1 : First spectrum obtained with the LCT premier with 5  $\mu\text{M}$  of each hairpin, 100 mM  $\text{NH}_4\text{OAc}$  and 2  $\mu\text{M}$   $\text{dT}_6$ . No desalting performed. The source temperature was set at 40°C and the desolvation temperature at 60°C.

We can observe both hairpin monomers at different charge states, from 5- to 3-, where the 4- is the major charge state. The kissing complex is detected only at charge 6-. As no desalting was performed,  $\text{Na}^+$  adducts are present in addition to  $\text{NH}_4^+$  adducts.

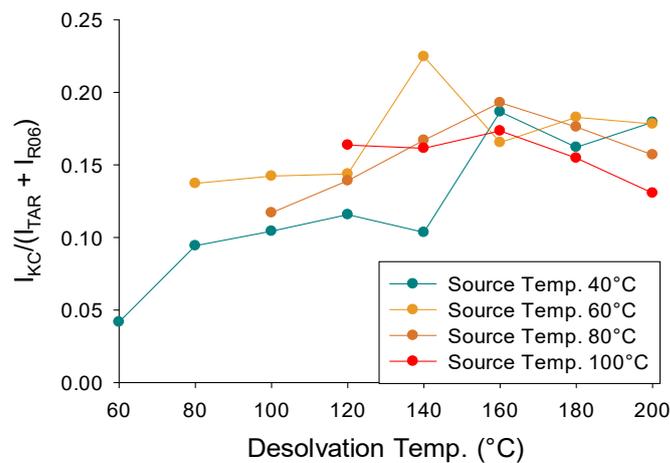
To decrease  $\text{Na}^+$  adducts, the samples are desalted (see material and methods). Also, to decrease the number of  $\text{NH}_4^+$  adducts, it is necessary to activate more the ions to help desolvation and declustering. To do so, we choose to play on the temperature. With the LCT instrument, two different temperatures can be modified: the source temperature (set initially at 40°C) and the desolvation temperature, which is the temperature of the desolvation gas (set initially at 60°C).

Several sets of source/desolvation temperatures are tested to check their effect on declustering without disrupting the non-covalent complexes present in solution. The source temperature is set at 40, 60, 80 and 100°C and the desolvation temperature at 60, 80, 100, 120, 140, 160, 180 and 200°C

We can see that by increasing both temperatures, fewer ammonium adducts are present. Also, the signal to noise ratio increases. It is important to verify the evolution of the fraction of kissing complex. Indeed, if dissociation of the complex is happening, we are not in soft enough conditions anymore. SI 2 presents the evolution of the ratio of the kissing complex area and the total area of both hairpin monomers as a function of the source and desolvation gas temperatures. All charge states are considered. The area of the peak is determined by calculating the integral of the peak taking into account all the adducts. At high desolvation temperature (>160 °C), the ratio  $\frac{A_{KC}}{A_{TAR} + A_{R06}}$  is decreasing meaning that the total intensity detected for TAR and R06 is increasing. Dissociation of the kissing complex into its monomers happens when a too high in source activation is applied. From these results, setting the source temperature at 60°C and the desolvation temperature at 140°C is a good compromise to obtain good enough spectra.



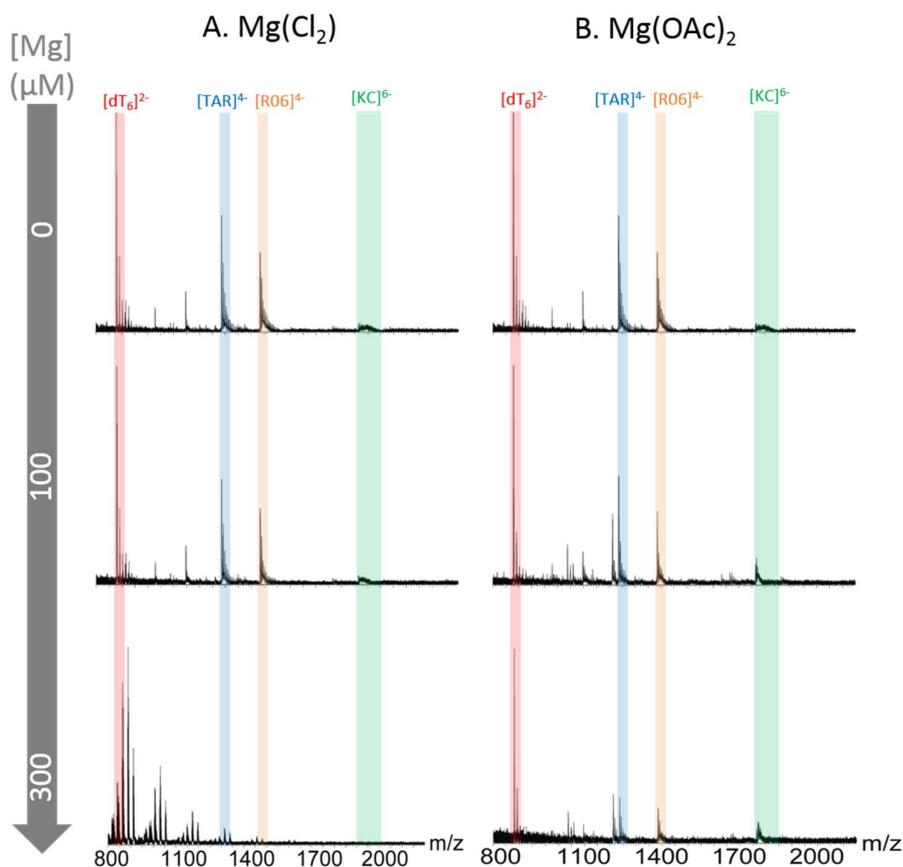
SI 2: Zoom on TAR-R06 KC 6- charge state recorded with the LCT Premier at different source and desolvation temperatures. Sample preparation: 5  $\mu$ M TAR, 5  $\mu$ M R06, 150 mM NH<sub>4</sub>OAc. Desalting was not performed on these samples.



SI 3: Evolution of the intensity of kissing complex detected over hairpin monomers intensities in function of the source and desolvation temperatures.  $I_{KC}$  is the total intensity of detected kissing complex,  $I_{TAR}$  is the relative intensity of TAR,  $I_{R06}$  is the relative intensity of R06

### Solutions doped with magnesium

In biochemistry,  $MgCl_2$  is most commonly used when magnesium is needed in the buffer solutions. A first screening, using increasing concentrations of  $MgCl_2$ , is performed to know if this salt can be used in *Native* MS experiments. SI 4 presents ESI-MS spectra obtained when  $MgCl_2$  salt is added at different concentrations.

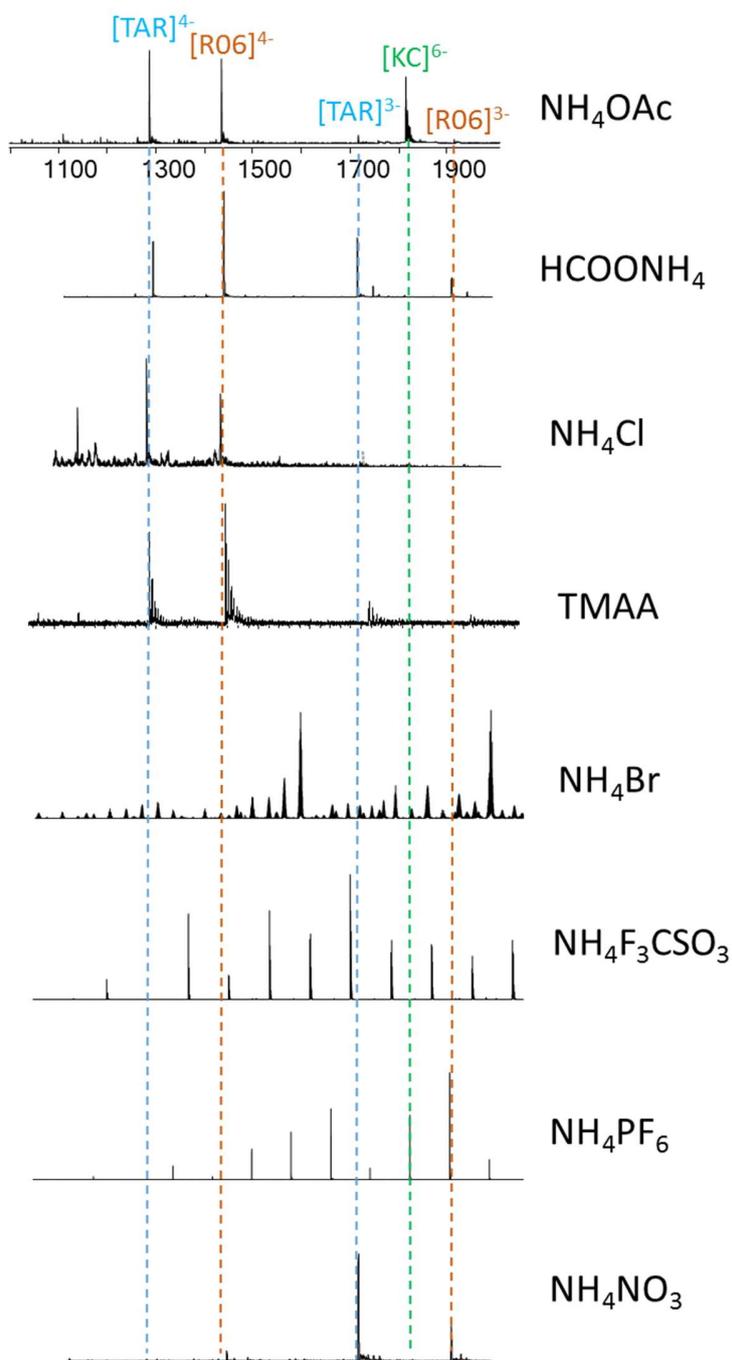


SI 4: ESI-MS spectra obtained with the LCT premier at different concentrations 0, 100 and 300 μM of A) MgCl<sub>2</sub> and B) Mg(OAc)<sub>2</sub>. Solutions were containing: 5 μM of TAR and R06, 150 mM NH<sub>4</sub>OAc.

The kissing complex and its monomers are detected at concentrations below 300 μM of MgCl<sub>2</sub>. Unfortunately, at higher concentration ( $\geq 300 \mu\text{M}$ ), no RNA ions are detected. In the low mass range, the spectrum is showing only high counts of clusters of Mg<sup>2+</sup>/Cl<sup>2-</sup>. To bypass this problem, we used another salt: magnesium acetate (Mg(OAc)<sub>2</sub>).

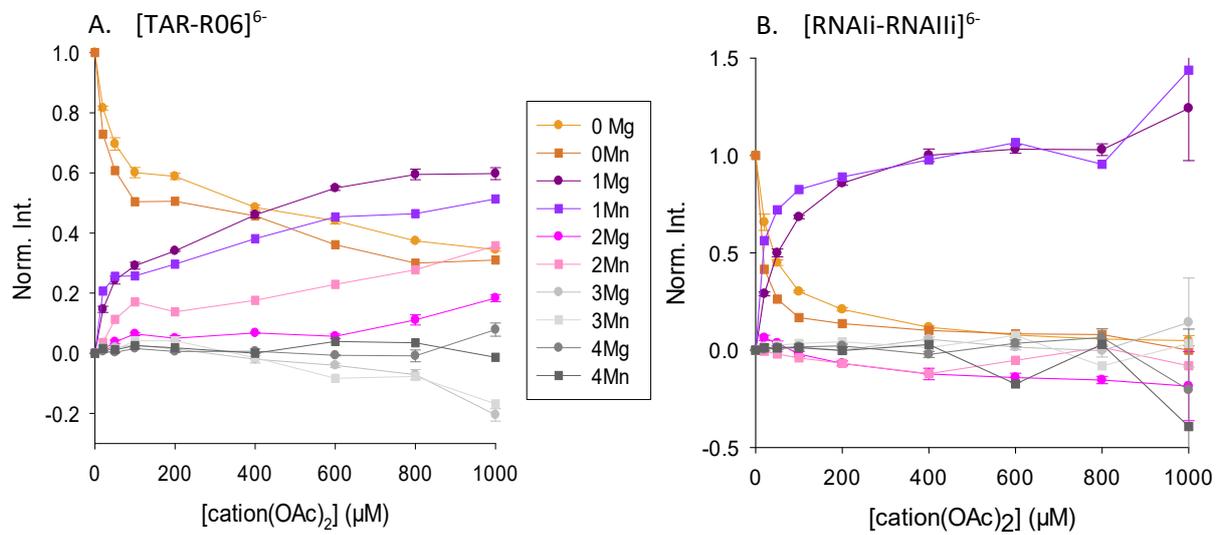
On the contrary to MgCl<sub>2</sub>, solutions can be doped with Mg(OAc)<sub>2</sub> at least until 300 μM (SI 4, B). At this concentration, clusters are appearing, but we can still see the signal of the two hairpins charge states 4- and the kissing complex 6-. Despite the signal to noise ratio decreases, solutions doped with Mg(OAc)<sub>2</sub> until 1 mM have been injected and the data were still analyzable.

## A1.2. Choice of solution conditions: full MS spectra



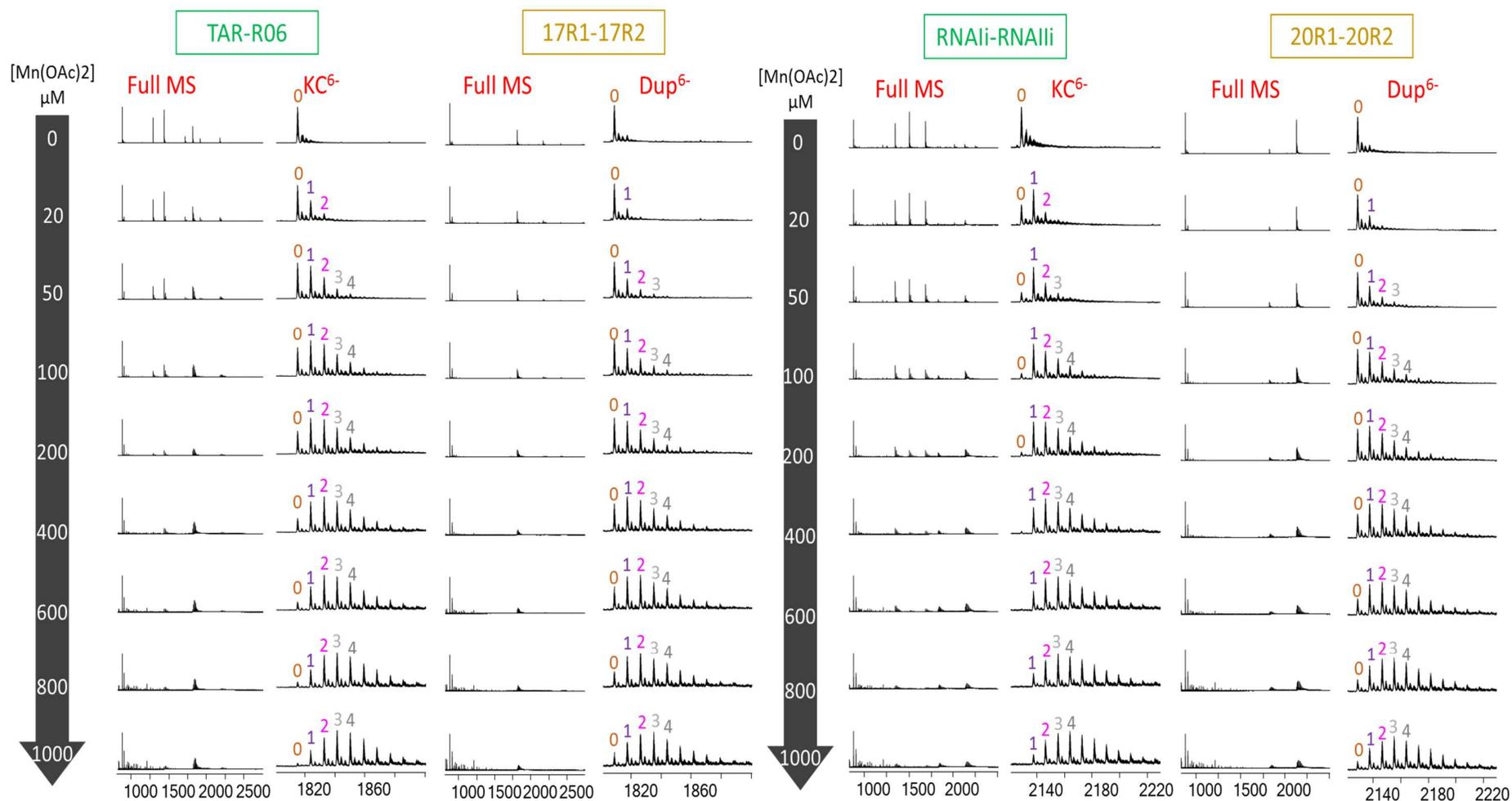
SI 5: Full MS spectra obtained using different electrolyte: NH<sub>4</sub>OAc, HCOONH<sub>4</sub>, NH<sub>4</sub>Cl, TMAA, NH<sub>4</sub>Br, NH<sub>4</sub>F<sub>3</sub>CSO<sub>3</sub>, NH<sub>4</sub>PF<sub>6</sub> and NH<sub>4</sub>NO<sub>3</sub>. MS spectra were acquired on the Agilent IMS Q-TOF, LCQ Fleet and LCT premier depending on their availability.

### A1.3a. Comparison of the number of specific $Mg^{2+}$ and $Mn^{2+}$



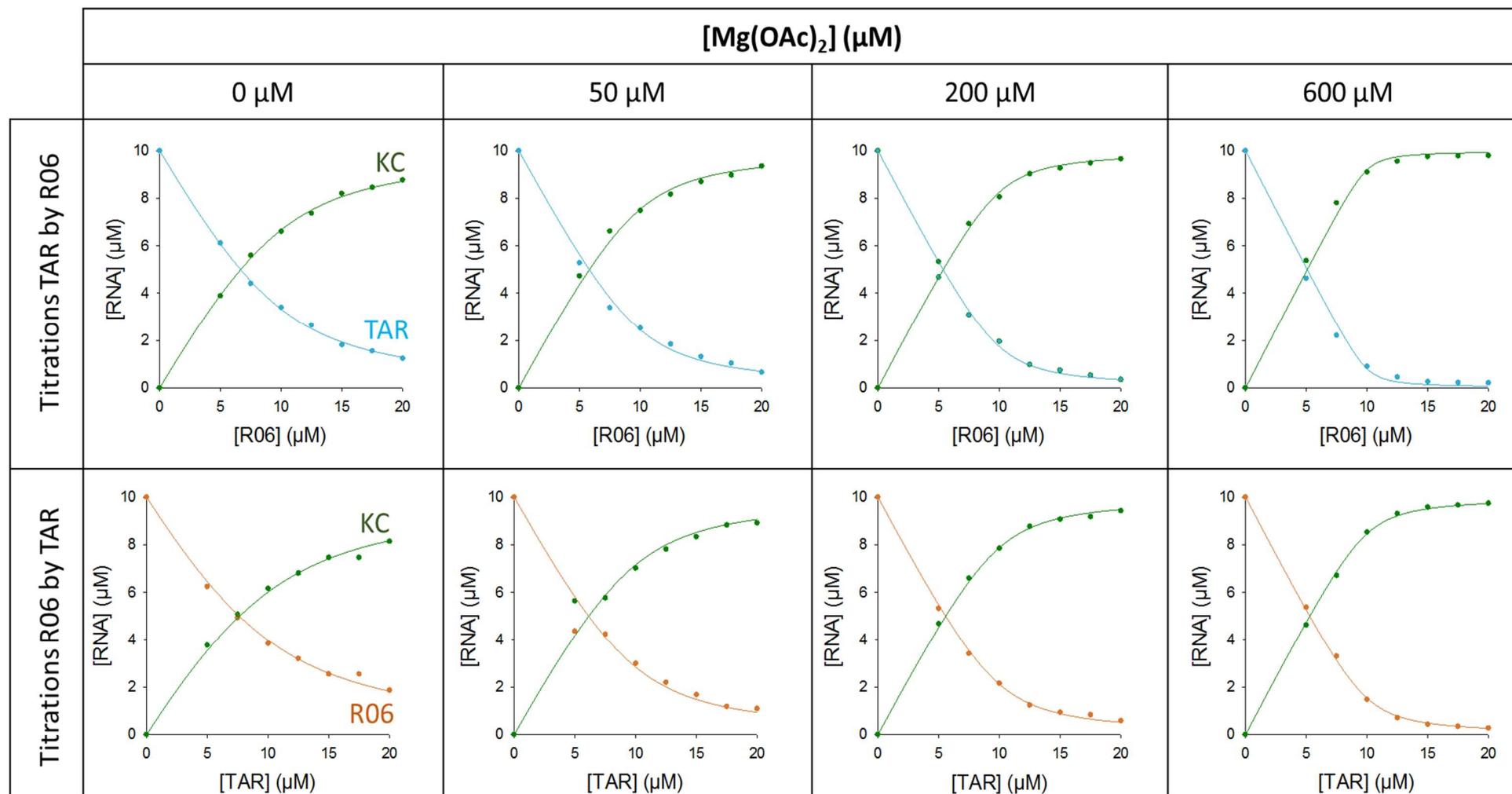
SI 6: Comparison of the specific cation distribution between  $Mg^{2+}$  ( ) and  $Mn^{2+}$  ( ) for A) [TAR-R06]<sup>6-</sup> and B) [RNAIi-RNAIii]<sup>6-</sup>.

A1.3b. Mass spectrometry titration of TAR-R06, 17R1-17R2 and RNAi-RNAii, 20R1-20R2 by  $\text{Mn}(\text{OAc})_2$ : full mass spectra and zoom on the charge state 6-



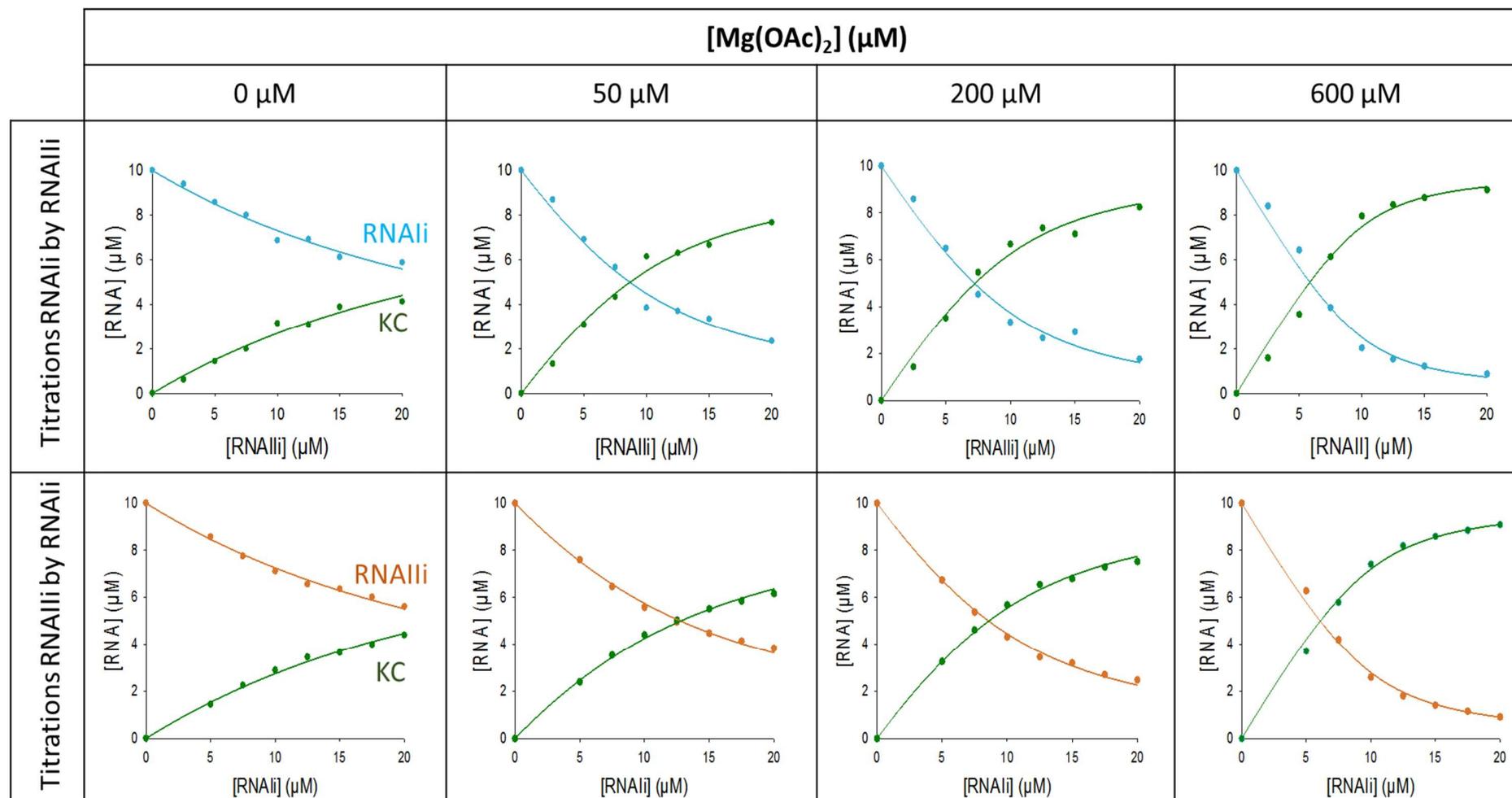
SI 7: Mass spectrometry titrations of 2  $\mu\text{M}$   $dT_6$ , 10  $\mu\text{M}$  of each complementary strands or hairpin, 150 mM  $\text{NH}_4\text{OAc}$  and an increasing concentration of  $\text{Mn}(\text{OAc})_2$ . Full MS spectra and zoom on the KC/duplex major charge state. These data are used for adduct cleaning.

A1.4. Titration curves for the determination of  $K_{d\text{TAR-R06,app}}$



SI 8: Titration curves obtained for (top) TAR by R06 and (bottom) R06 by TAR at 4 different concentrations of  $\text{Mg}(\text{OAc})_2$ . Only one replicate of each experiment is presented.

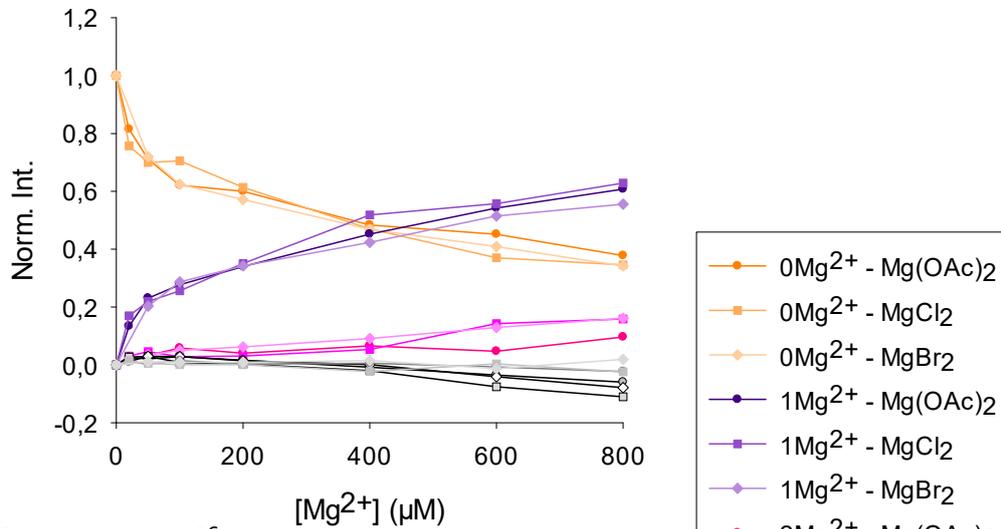
A1.5. Titration curves for the determination of  $K_{d\text{RNAIi-RNAIi,app}}$



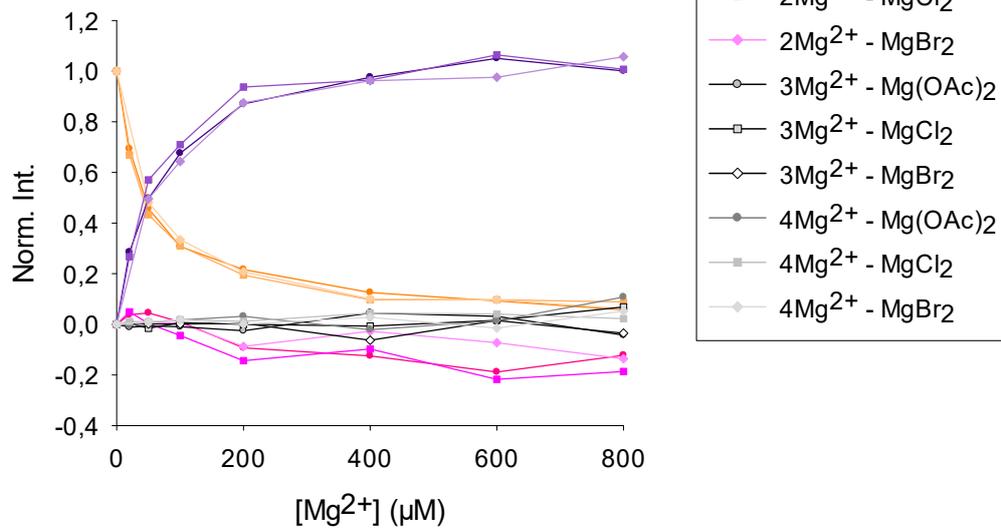
SI 9: Titration curves obtained for (top) RNAIi by RNAIi and (bottom) RNAIi by RNAIi at 4 different concentrations of  $\text{Mg}(\text{OAc})_2$ . Only one replicate of each experiment is presented.

A1.6. Comparison between  $\text{Mg}(\text{OAc})_2$ ,  $\text{MgCl}_2$  and  $\text{MgBr}_2$

A.  $[\text{TAR-R06}]^{6-}$



B.  $[\text{RNAIi-RNAIi}]^{6-}$



SI 10: Comparison of the specific  $\text{Mg}^{2+}$  distribution using  $\text{Mg}(\text{OAc})_2$  (circles),  $\text{MgCl}_2$  (squares) or  $\text{MgBr}_2$  (triangles) for A)  $[\text{TAR-R06}]^{6-}$  and B)  $[\text{RNAIi-RNAIi}]^{6-}$ . The top graph presents the results obtained for  $[\text{TAR-R06}]^{6-}$ , the bottom graph presents  $[\text{RNAIi-RNAIi}]^{6-}$  results.

## A1.7. MS/MS study of TAR-R06 using the LCT Fleet instrument

### **TAR-R06**

The LCQ Fleet mass analyzer is an ion trap where MS/MS can be performed directly (the trap can be considered as similar as the collision cell of the AGILENT instrument). The instrument is tuned in a way to assure good desolvation and a good signal. The capillary temperature is set at 290°C, the capillary voltage at 90 V and the tube lens at 12 V. Here, the activation time can be chosen that is why we decided to perform the experiment at 30 ms (long time) and 3 ms (short time). The short time is selected in order to have a value closer to the one of the AGILENT 6560. The activation voltages are varying depending on the collision time. The samples are prepared with 20  $\mu$ M of each hairpin and 10% MeOH is added to the solution in order to increase and to stabilize the signal.

First the experiments are done without magnesium in order to compare the type of fragmentation obtained between 30 and 3 ms and then compare the results with the AGILENT ones. The results are shown in figure SI 11.

The tests of different activation voltages have shown that, for 30 ms, from 0 to 24 V no fragmentation is occurring. The kissing complex begins to fragment at 24 V and at 28 V has been completely fragmented. Concerning the resulting fragments observed, at 24 V, the KC is losing a Guanine but there is also small peaks corresponding to the two monomers charge 3-. However, we can see that at this voltage, [TAR]<sup>3-</sup> and [R06]<sup>3-</sup> are losing a Guanine too. When the activation voltage is increasing (26-28V) there are more and more fragments of TAR and R06. These fragments are due to the fragmentation of the backbone of the RNA. In these conditions, covalent bounds are breaking. From the identification of some fragments based on McLuckey nomenclature<sup>47</sup> it seems that most of the fragmentation happens next to Guanine and Cytosine.

If we compare the results obtained at 3 ms with the ones at 30 ms, as expected, more collisional energy is needed to fragment the precursor ion. Here, the voltage has to be set at least at 43 V to observe the first fragments. As for the experiment at 30 ms, the [KC]<sup>6-</sup> is fragmented into its two monomers and is losing a G. When the energy is increasing (45V), the fragments corresponding to [TAR-G]<sup>3-</sup> and [R06-G]<sup>3-</sup> are increasing. Also, at the voltage, several other fragments of TAR and R06 are appearing, like at 30 ms. Some fragments are in common between 30 ms and 3 ms. However, as it is difficult to identify all of them precisely, we can just say that the intensities of those fragments are lower at 3 ms than at 30 ms. In addition, the major fragment ions are different between the two activation times. For example, at the

higher voltage, 28 vs 50V, c is the major ion at 30ms whereas at 3 ms R06:  $\gamma_4^{1-}$  is higher. At 50 V, the  $[KC]^6$  has been almost completely fragmented. Based on these results, one can say that the fragmentation profiles obtained at 30 ms and 3 ms are close. With these results we cannot say if the dissociation of the KC into its two monomers is predominant at 3 ms than at 30 ms. If we compare these results with the one obtained with the AGILENT, it seems that at low collision time ( $< 1$ ms) the first channel of dissociation is the dissociation of the complex into its two monomers accompanied by the loss of one G. Then there is base losses and fragmentation of covalent bounds. At high collision time (30 ms) is the LCQ its seems that base losses and covalent bound breaks are predominant. However, we have to be careful when comparing the two instruments as they are not composed of the same mass analyzer and one has a higher resolution and sensitivity (AGILENT 6560) than then other (LCQ Fleet).





used. What we can see is base losses, mostly guanines, and fragmentation of TAR and R06. G base loss at 24 V is observed but  $Mg^{2+}$  remains on the KC.

Concerning the results at 3 ms, there are not really conclusive as well. In this case, before 47 V, no fragmentation is happening. At 47 V, there is the occurrence of the first fragments which correspond to base losses: 1793.92 m/z is a G loss and 1768.67 m/z is a two G loss. At 48 V, small peaks corresponding to the dissociation of  $[KC+1Mg^{2+}]^{6-}$  into  $[TAR+1Mg^{2+}]^{3-}$  and  $[R06+1Mg^{2+}]^{3-}$  are appearing. However, their intensities are quiet low and they are accompanied by fragmentation of TAR and R06 and base losses. As without magnesium in the solution, the fragments observed are similar between 30 and 3 ms, with different intensities.

From these results, it is impossible to do  $MS^n$ . Indeed, the intensities of the monomers plus one magnesium are too low to do at least  $MS^3$ .

### **RNAIi-RNAIii**

Concerning RNAIi-RNAIii, unfortunately the resolution and sensitivity of the instrument does not allow us to observe the fragmentation of  $[KC]^{7-}$  without sulfolane. Also, because of a lack of time experiments with sulfolane were not done. More investigations are needed if we want to perform  $MS/MS$  on RNAIi-RNAIii kissing complex using the LCQ Fleet.



II. A2. Article 1: [Electrostatics Explains the Position-Dependent Effect of G•U Wobble Base Pairs on the Affinity of RNA Kissing Complexes](#), Abi-Ghanem J. et al, *ChemPhysChem*, 2017

This is the author pre-print of a manuscript published in final form in: Abi-Ghanem et al., *ChemPhysChem* (2017), 18(19), 2782-2790. <http://dx.doi.org/10.1002/cphc.201700337>

---

**Electrostatics explains the position-dependent effect of G•U wobble base pairs on the affinity of RNA kissing complexes**

Josephine Abi-Ghanem,<sup>[a]</sup> Clémence Rabin,<sup>[a]</sup> Massimiliano Porrini,<sup>[a]</sup> Eric Dausse,<sup>[b]</sup> Jean-Jacques Toulmé,<sup>[b]</sup> Valérie Gabelica\*<sup>[a]</sup>

[a] Dr. Josephine Abi-Ghanem, Mrs. Clémence Rabin, Dr. Massimiliano Porrini, Dr. Valérie Gabelica  
Laboratoire Acides Nucléiques : Régulations Naturelle et Artificielle (ARNA, U1212, UMR5320)  
Univ. Bordeaux, INSERM, CNRS  
IECB, 2 rue Robert Escarpit, 33607 Pessac, France  
E-mail: [v.gabelica@iecb.u-bordeaux.fr](mailto:v.gabelica@iecb.u-bordeaux.fr)

[b] Eric Dausse, Dr. Jean-Jacques Toulmé  
Laboratoire Acides Nucléiques : Régulations Naturelle et Artificielle (ARNA, U1212, UMR5320)  
Univ. Bordeaux, INSERM, CNRS  
146 rue Léo Saignat, 33076 Bordeaux

**Abstract**

In the RNA realm, non-Watson Crick base pairs are abundant and affect either the RNA 3D structure and/or its function. Here we investigated the formation of RNA kissing complexes where the loop-loop interaction is modulated by non-Watson-Crick pairs. Mass spectrometry, surface plasmon resonance and UV-melting experiments show that the G•U wobble base pair favors the kissing complex formation only when placed at specific positions. We tried to rationalize this effect by molecular modelling, including molecular mechanics Poisson–Boltzmann surface area (MMPBSA) thermodynamics calculation and PBSA

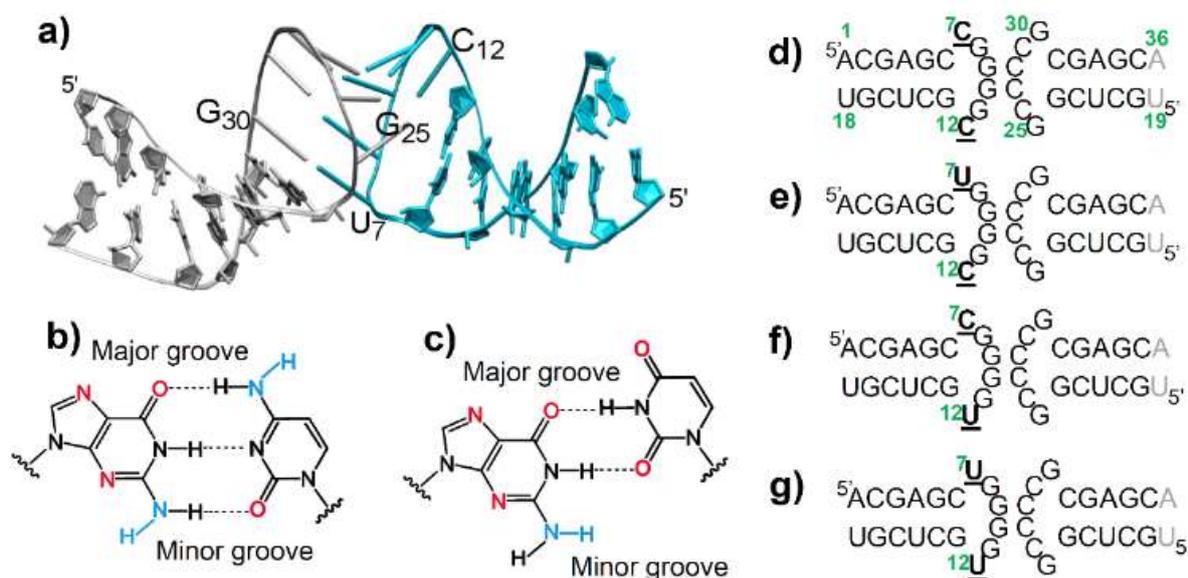
calculation of the electrostatic potential surfaces. Modeling reveals that the G•U stabilization is due to a specific electrostatic environment defined by the base pairs of the entire loop-loop region. The loop is not symmetric, and therefore the identity and position of each base pair matters. Predicting and visualizing the electrostatic environment created by a given sequence can help designing specific kissing complexes with high affinity, for potential therapeutic, nanotechnology or analytical applications.

## Introduction

Aptamers are often referred to as “chemical antibodies”:<sup>[1]</sup> aptamer oligonucleotides can bind to specific targets (such as small molecules, proteins, or nucleic acids) via an adapted tridimensional structure. They have several advantages compared to traditional antibodies. They can be automatically synthesized in large quantities, and they have low or no immunogenicity. The discovery of riboswitches,<sup>[2]</sup> which are mRNA molecules with aptamer components that regulate gene expression in the cell, has revived the interest in RNA-based sensors. Besides, FDA approval of the first aptamer as drug has boosted research in RNA-based therapeutics.<sup>[3-5]</sup>

New RNA aptamers are traditionally found by the SELEX method (Systematic Evolution of Ligands by Exponential enrichment).<sup>[3]</sup> When SELEX is applied to a nucleic acid target, often the resulting aptamer is found to bind its target via a kissing loop (KL) motif: the two nucleic acids (the aptamer and the target) form the so-called “kissing complex” (KC).<sup>[4]</sup> In the kissing complex, the loops of two hairpins interact via base pairing (Figure 1a). Besides aptamers, KL motifs also form in natural sequences, for example in some riboswitches<sup>[5]</sup> or in the dimerization domain of retroviruses.<sup>[6]</sup> Understanding the forces driving the formation of KL motifs can therefore help understanding the principles of natural and artificial selection of RNA-RNA recognition motifs.

Non-canonical base pairs are frequent, both in artificially and naturally selected KL motifs. The selection process can therefore favor non-canonical base pairs over their Watson-Crick counterpart. This can be either for stability, kinetics, or flexibility reasons. For example, in the kissing complex of R06 aptamer with the HIV1 TAR RNA element, a non-canonical G•A base pairs leads to a more stable structure than the fully Watson-Crick analog.<sup>[4b]</sup> In ribozymes, binding and catalysis efficiency decreases when the naturally occurring G•U base pair is mutated to a canonic G•C base pair.<sup>[7]</sup> Finally, in tRNA, mismatches boost aminoacylation and translation, primarily by affecting the conformational flexibility.<sup>[8]</sup>



**Figure 1. Kissing complexes with G•C and G•U base pair variants.** (a) 3D structure of the kissing complex UC: K1 in gray and K1' in cyan. The nucleobases in the kissing loop are shown as sticks. The closing base pairs are U<sub>7</sub>•G<sub>30</sub> and C<sub>12</sub>•G<sub>25</sub> (for base numbering, see panel d). (b) Watson-Crick G•C base pair (3 hydrogen bonds) and (c) non-canonical G•U wobble base pair (2 hydrogen bonds). Electronegatively and electropositively charged atoms are shown in red and blue, respectively. (d–f) Sequences of the four variants. K1 is on the right; the terminal bases in grey have been deleted for some mass spectrometry experiments. K1' is on the left. The kissing complexes are named according to the identity of their bases in positions 7 and 12. Four mutants with C or U at positions 7 or 12 have been studied: (d) CC, (e) UC, (f) CU and (g) UU.

The G•U wobble base pair is found in many biological processes such as recognition (by proteins, ligands or ions),<sup>[9]</sup> or catalytic activities.<sup>[7a]</sup> *In vitro* selection of aptamers against stem-loop (i.e., hairpin) RNA structures also sometimes return hits containing one mismatched G•U base pair in the kissing loop motif.<sup>[10]</sup> The G•U base pair displays distinctive chemical groups in the major and minor groove (Figure 1b, c): the O4 of the uracil protrudes in the major groove and gives an electronegative surface, whereas the NH<sub>2</sub> of the guanine protrudes in the minor groove and gives the base pair a more electropositive surface.<sup>[11]</sup> Replacing G•C by G•U also induces structural changes, because the whole U base needs to protrude towards the major groove in order to pair with G. We wanted to understand how and why natural or artificial selection processes could sometimes favor the G•U base pair over the canonical GC base pair.

Our objective is to understand the interplay between structure and energetics for kissing complexes containing a single G•U wobble base pair, and the influence of the G•U localization on the stability of the kissing complex. Molecular dynamics simulations are traditionally used to grasp the atomistic details of the kissing complex structure and dynamics,<sup>[12]</sup> and the role of monovalent or divalent cations.<sup>[13]</sup> To estimate the energetics, we used molecular mechanics Poisson–Boltzmann surface area (MMPBSA)<sup>[14]</sup>. Even though applying this method to single trajectories on the complex neglects conformational rearrangement upon binding,<sup>[15]</sup> we found that the method correctly accounts for the *differences* due to G•U mutations.

As a model, we used the kissing complex K1/K1' previously obtained by SELEX.<sup>[10b]</sup> The aptamer K1' (Figure 1a and 1e) contains a U at position 7 (5'-end of the loop) forming a G•U wobble base pair with its target K1, and a C at position 12 (3'-end of the loop) forming a canonical G•C base pair on the other end of the loop. K1/K1' and its variants with U or C at positions 7 and 12 of K1' (Fig. 1e–f) are therefore good models to understand the impact of the G•C→G•U modification on the stability of RNA-RNA kissing complexes. Our molecular modeling investigation revealed that the electrostatic environment of the entire loop dictates the effect of non-canonical base pairing on the kissing complex.

## Results and Discussion

### 1. G•U base pairs favor kissing complexes only at specific loop locations

We used surface plasmon resonance (SPR), UV-melting, and native mass spectrometry to assay the relative stability of the four variants of the kissing complex K1/K1', without and with magnesium. The analysis of SPR data was problematic because the data could not be fit by a simple 1:1 binding model (see Supporting Information Figures S1-S2). UV-melting analyzes the dimer stability as a function of temperature, as a proxy for the relative stability at a given temperature. With or without magnesium, the stability ranking is CC > UC > CU > UU (Supporting Information Fig. S3).

Native mass spectrometry is the most direct assay to distinguish heterodimers from monomers or homodimers. To evaluate the relative stability of the four kissing complexes CC, CU, UC and UU, we measured the peak area ratio between the complex and the free target. The association equilibrium is:



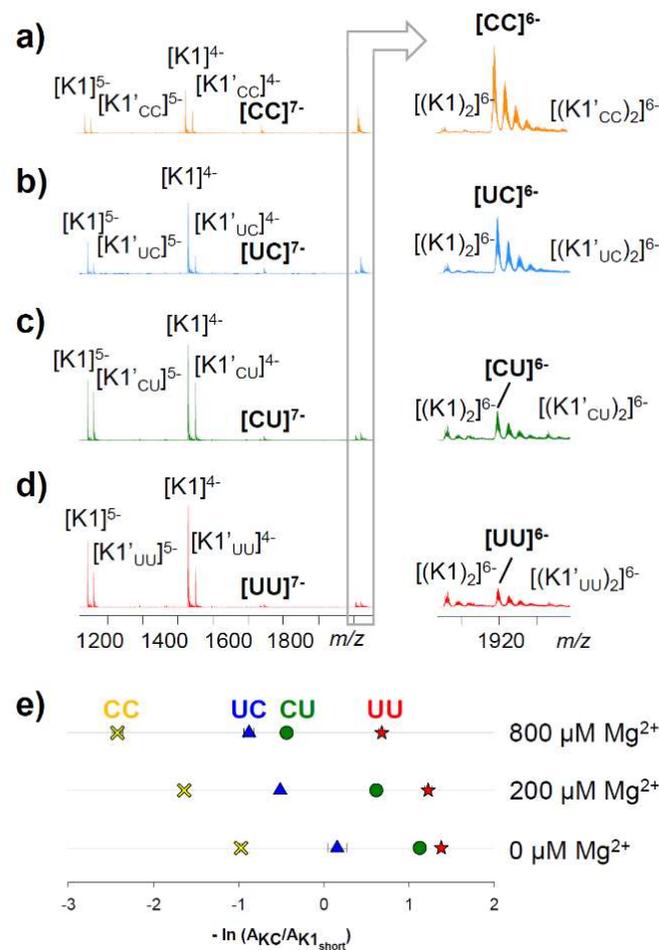
Hence the  $\Delta G_{assoc}$  is given by:

$$\Delta G_{assoc} = -RT \cdot \ln \left( \frac{[KC]}{[target][ligand]} \right) \quad (2)$$

Here the target is the same within each series, and the ligand varies. For a given total concentration of target and ligand, assuming that all KCs of a series respond similarly in electrospray ionization, the  $\Delta G_{assoc}$  is proportional to the ratio between the mass spectral peaks areas of KC ( $A_{KC}$ ) and target ( $A_{target}$ ), as follows:

$$\Delta G_{assoc} \propto -\ln \left( \frac{A_{KC}}{A_{target}} \right) \quad (3)$$

More negative values for Eq. (3) therefore indicate more stable kissing complexes. All observed charge states were summed up to calculate the peak areas (6- and 7- for the KC, 4- and 5- for the target).



**Figure 2. Effect of C→U mutation on kissing complex stability: experiments. a–d)** ESI-MS spectra of the equimolar mixture (10  $\mu\text{M}$  each) of K1 target and K1' ligands (see Figure 1 for sequences) in 150 mM

NH<sub>4</sub>OAc. The right hand side shows a zoom on the kissing complexes at charge state 6-, and the close signals of homodimers of K1 and K1'. **e)** Relative quantification of the peak areas of kissing complex, relative to the free target K1<sub>short</sub> (K1 with terminal base pair truncated, see Figure 1). The most stable complexes are on the left, and the least stable are on the right.

Experiments were first carried out with target K1 and the four variants of ligand K1' (Figure 1). The results are shown in Figure 2a–d in 150 mM NH<sub>4</sub>OAc and no magnesium. The heterodimers (kissing complexes) were detected at charge states 6- (major) and 7- (minor). However, signals of homodimers of K1 and of the ligand are also detected at the charge state 6-. Immobilization of both monomers and homodimers on the SPR chip could explain the complexity of our SPR analysis. When magnesium ions were added to the solution, these signals become difficult to integrate separately (see Supplementary Figure S4a-c).

We therefore truncated K1 (the terminal base pair is deleted) to allow better mass separation of all species (Supplementary Figure S4d-f). The experiments were carried out in pure NH<sub>4</sub>OAc, and in the presence of magnesium (200 μM or 800 μM). The calculations of  $-\ln(A_{KC}/A_{target})$  are shown in Figure 2e. The kissing complexes can form without magnesium. In all conditions, the most stable kissing complex is **CC**, followed by **UC**, then **CU**. **UU** is the least stable. In presence of magnesium, the kissing complexes are more stable, but the ordering is the same. In all cases, a G•U base pair is stabilizing the K1.K1' complex better when placed on the 5'-side of the K1' loop than when placed on the 3'-side. We then set to understand the influence of the environment of the G•U base pair on its stabilizing/destabilizing effect, using molecular modelling and dynamic simulations.

## 2. G•U stabilization, with or without magnesium

Based on RMSDs for the different MD simulations, (Table 1 and Supplementary Figure S5), the structure of all four KCs is stable in magnesium and in sodium. The kissing loop motif forms a tunnel through the loop-loop where ions can diffuse,<sup>[16]</sup> and be trapped in pockets. We observe two main pockets of high affinity towards the sodium ions in all the kissing complexes (Supplementary Figure S6). The first pocket is located in the loop-loop region, in the vicinity of the G-tract of the K1' ligands. The second pocket is observed near the exit of the tunnel of the K1 hairpin. MD simulations with magnesium ions were conducted by initially placing the two Mg<sup>2+</sup> ions close to the pockets found in sodium, and upon unrestrained MD the magnesium ions indeed stayed in the vicinity (Supplementary Figure S6).

In **UC** and **UU** MD in sodium, we observe that in the  $U_7 \bullet G_{30}$  base pair,  $U_7$  oscillates between open and closed states, whereas magnesium favors the closed state. However, if a cytosine is present in position 7, the base pairing is fully maintained. The residues at the 5'-end of the loop are always more flexible than those at the 3'-end, so a uracil at position 7 makes a weaker base pair in the context of a  $G \bullet U$  wobble base pair. In the MD simulations, we could observe a density of  $Na^+$  around the  $U_7 \bullet G_{30}$  pair, and switching to  $Mg^{2+}$  leads to a more stable base pairing (Supplementary Figure S7). Magnesium ions help forming the non-canonic base pair. To better reflect cellular conditions, we will therefore mostly focus on the MD results in the presence of magnesium.

**Table 1. Overview of the MD simulations.** The four kissing complexes studied are shown with their respective nomenclature (see Figure 1): each complex is denoted by its 5' and 3' end residue of the loop of the K1' ligand. For the MD simulations, the root mean square deviation (RMSD in Å) calculated on all the atoms is reported with its standard deviation for the total time of the MD using the average structure as reference.

Kissing complex	Ions	MD time (ns)	RMSD (Å)
CC	34 Na <sup>+</sup>	200	1.5 ± 0.4
	30 Na <sup>+</sup> , 2 Mg <sup>2+</sup>	300	1.4 ± 0.3
UC	34 Na <sup>+</sup>	200	1.7 ± 0.5
	30 Na <sup>+</sup> , 2 Mg <sup>2+</sup>	300	1.5 ± 0.4
CU	34 Na <sup>+</sup>	200	1.3 ± 0.4
	30 Na <sup>+</sup> , 2 Mg <sup>2+</sup>	300	1.4 ± 0.4
UU	34 Na <sup>+</sup>	200	1.8 ± 0.5
	30 Na <sup>+</sup> , 2 Mg <sup>2+</sup>	300	1.8 ± 0.5

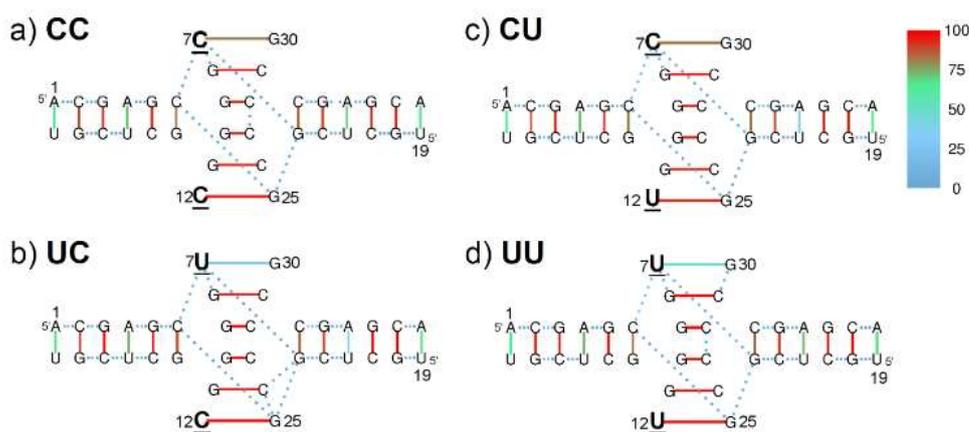
### 3. Effect of G•U on the structure of the kissing complexes

Substituting residues at the 5' or 3' end loop residues of the K1' ligand will have different impact on the whole KC structure. We first describe the effect of substituting G•C for G•U on the hydrogen bond network and on the helicoidal parameters.

#### 3.1. Hydrogen bond network

We first calculated the hydrogen bond network on a window frame from 50 to 300 ns of the MD. The four systems show subtle differences in their hydrogen bond networks in magnesium and sodium (Figure 3 and supplementary Figure S8, respectively).

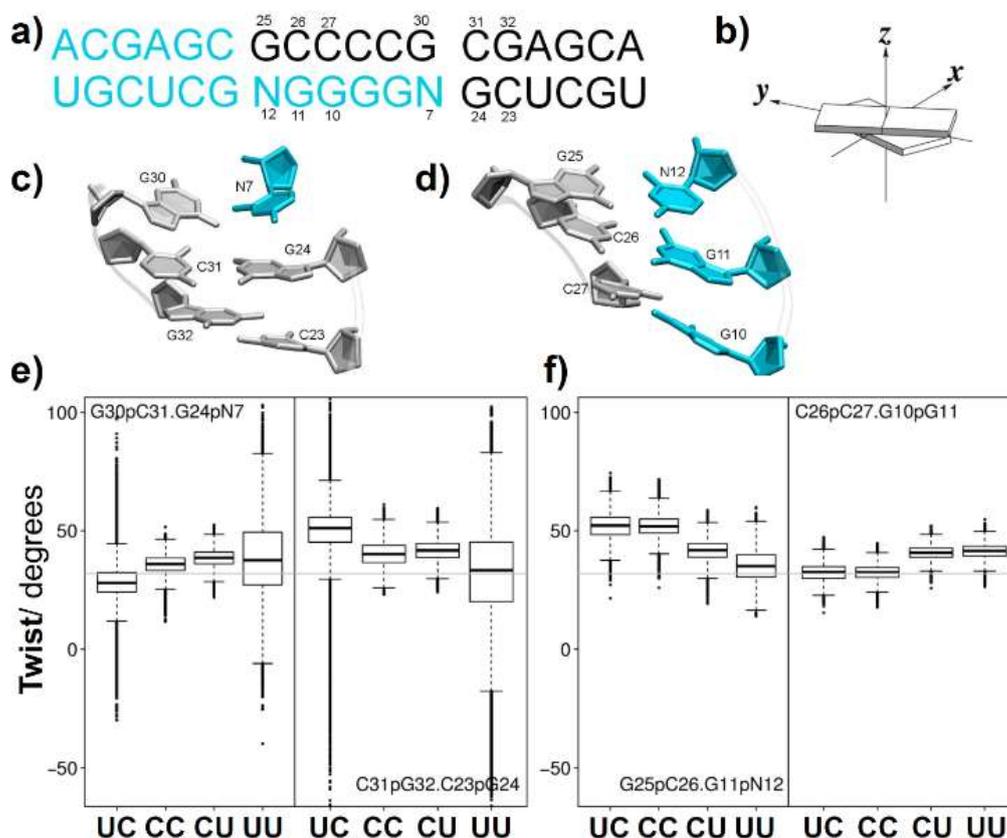
First, the strength of hydrogen bonding (indicated by the percentage of occupancy in the trajectory) differs according to the base pair location. The 3'-end base of the ligand loop (position 12) is strongly paired with 78–80% in all four systems, independently of whether the base pair is C<sub>12</sub>•G<sub>25</sub> or U<sub>12</sub>•G<sub>25</sub>. In contrast, on the 5'-end of K1' loop, the hydrogen base pairing occupancy is ~71% for CC and CU, whereas it is only 54% and 8% in UC and UU, respectively. Introducing a uracil at the 5' end therefore leads to a weaker wobble base pair than at the 3' end. Second, a transient hydrogen bonding network forms between the ribose hydroxyl group of residues G<sub>25</sub> and C<sub>6</sub> and the phosphate OP1 or OP2 atoms of the residues G<sub>24</sub> and U<sub>7</sub>/C<sub>7</sub> (except in UU in Mg<sup>2+</sup>). This network is the only interaction between the two hairpins (besides the base pairing of the loops), and is more pronounced in the MD in sodium than in magnesium in the time window chosen for the calculations (Supplementary Figure S9). In summary, the hydrogen bonding network alone does not allow to understand why substituting C for U at position 7 is more favorable than at position 12.



**Figure 3. Hydrogen bond network of the kissing complexes.** Hairpin K1 (right hand side) contains residues 19 to 36 and ligand hairpin variants K1' (left hand side) contains the residues 1 to 18. The hydrogen bond network obtained by MD in magnesium, on a window frame from 50 to 300 ns, is shown for kissing complex CC (a), UC (b), CU (c) and UU (d). The colors indicate the fractional population of hydrogen bonds, from blue 0% (absence) to red 100% (present). H-bonds between bases are represented by solid lines, and H-bonds between the ribose and the phosphate groups are represented as dashed lines.

### 3.2. Helicoidal parameters

However, in wobble base pairs, the shifting of G and U towards the minor (narrowest) and major (widest) groove respectively could affect the helicoidal parameters compared to a canonical base pair. We can write the sequences on two lines highlighting the base pairs (Figure 4a), then calculate various inter-base parameters, indicating the position and orientation of a base pair relative to the helical axis (defined here by the repetition of a two base-pair unit). Although the kissing loop is not a canonical double-helix motif, the helicoidal parameters serve here to highlight the structural differences between the four systems. The most significant changes when mutating C to U were found in the base pairs twist and slide. The twist is a rotation about the helical axis Z-axis (Figure 4b), and the slide is a translation around the Y-axis (Supplementary Figure S10e). The canonical helical twist for an A-helix is  $32^\circ$  and the slide is  $-1.5^\circ$ .<sup>[17]</sup> Here we found that the values of the twist (Figure 4), and to a lesser extent of the slide (Supplementary Figure S10), depends on the identity of the bases closing the loop-loop helix.



**Figure 4. Effect of G•U substitution on inter-base pair twist.** The helicoidal parameters are calculated on the MD in magnesium on a window frame (50-300 ns) on **a)** the linearized sequences (K1 hairpin in black, K1' hairpin in blue, and base pairs of the kissing complex facing one another). N stands for cytosine or uracil. **b)** Scheme of the twist of two base pairs around the z-axis. **c-d)** The steps which yield to a change across the four kissing complexes: **c)** the steps G<sub>30</sub>pC<sub>31</sub>•G<sub>24</sub>pN<sub>7</sub> and the step C<sub>31</sub>pG<sub>32</sub>•C<sub>23</sub>pG<sub>24</sub>, and **d)** the steps G<sub>25</sub>pC<sub>26</sub>•G<sub>11</sub>pN<sub>12</sub> and the step C<sub>26</sub>pC<sub>27</sub>•G<sub>10</sub>pG<sub>11</sub>. **e-f)** corresponding twist value (in degrees) as a boxplot. The gray line in the boxplots indicates the canonical value of the twist parameter (32°) in an A-helix.

### 3.2.1. Effect of C → U mutation at the 5'-end of K1' (position 7):

When a cytosine is present at the 7<sup>th</sup> position (**CC** and **CU**), the twist for the step G<sub>30</sub>pC<sub>31</sub>•G<sub>24</sub>pC<sub>7</sub> is ~38°, followed by a twist of ~41° for the step C<sub>31</sub>pG<sub>32</sub>•C<sub>23</sub>pG<sub>24</sub>. However, for **UU**, the twist for the step G<sub>30</sub>pC<sub>31</sub>•G<sub>24</sub>pU<sub>7</sub> is lower (29±9°), and that of the next step C<sub>31</sub>pG<sub>32</sub>•C<sub>23</sub>pG<sub>24</sub> is high (48±4°). For **UU**, the order is reversed: we observe first high twist (38±16°) for the step G<sub>30</sub>pC<sub>31</sub>•G<sub>24</sub>pU<sub>7</sub>, followed by a lower

twist ( $28 \pm 27^\circ$ ) for the step  $C_{31}pG_{32} \bullet C_{23}pG_{24}$  (Figure 4 b,d). For the slide parameter, the differences are more subtle (Supplementary Figure S10 b,d). The changes are detected on the step  $C_{31}pG_{32} \bullet C_{23}pG_{24}$ : a more negative slide is observed for **CC** and **CU**, ( $-1.2 \pm 0.5^\circ$  and  $-0.9 \pm 0.4^\circ$ , respectively) than for **UC** and **UU** ( $0 \pm 1.2^\circ$  and  $0.7 \pm 1.7^\circ$ ). The twist and slide values for **UC** and **UU** are much more widely spread than for **CC** and **CU**, and this reflects structural changes occurring when the  $G_{30} \bullet U_7$  base pair opens.

### 3.2.2. Effect of $C \rightarrow U$ mutation at the 3'-end of $K1'$ (position 12):

When a cytosine is present at the 12<sup>th</sup> position (**UC** and **CC**), the twist for the step  $G_{25}pC_{26} \bullet G_{11}pC_{12}$  is high ( $52 \pm 6^\circ$ ), followed by a lower twist for the step  $C_{26}pC_{27} \bullet G_{10}pG_{11}$  ( $33 \pm 4^\circ$ ). When an uracil is in the 12<sup>th</sup> position in the systems **CU** and **UU** (Figure 4 c,e) the twist for the step  $G_{25}pC_{26} \bullet G_{11}pU_{12}$  is  $42 \pm 5^\circ$  and  $35 \pm 6^\circ$  respectively, followed by a similar twist for the neighboring step  $C_{26}pC_{27} \bullet G_{10}pG_{11}$  ( $41 \pm 3^\circ$ ). The effect on the slide, however, is more pronounced at the step  $G_{25}pC_{26} \bullet G_{11}pN_{12}$  (Supplementary Figure S10 c,e). In presence of  $U_{12}$  we observed more negative values for **CU** and **UU** ( $-2 \pm 0.4^\circ$  and  $-1.8 \pm 0.4^\circ$ , respectively) than in the presence of  $C_{12}$  ( $-1.2 \pm 0.4^\circ$  for **CC** and  $-1.1 \pm 0.5^\circ$  for **CU**).

In summary, the effect of  $C \rightarrow U$  mutation on the helicoidal parameters is centered around the  $G \bullet U$  wobble base pairs, but is also context-dependent. Each complex has a distinct helicoidal signature, because mutations at the 5'-end or 3'-end of the  $K1'$  loops are not equivalent. Although the mutations occurred only in  $K1'$ , the effects on helicoidal parameters propagate towards the target hairpin  $K1$ , suggesting an induced fit mechanism. The same tendencies were also found in the MD in sodium (Supplementary Figures S11-S12). Still, the structural description alone does not yet reveal why a  $G \bullet U$  wobble base pair is stabilizing at the 5'-end and destabilizing at the 3'-end of the  $K1'$  loop. We therefore turned to energy calculations on the structures generated by MD.

## 4. Structure-thermodynamics relationships unravel why the $G \bullet U$ stabilizing effect is position-dependent

### 4.1. MMPBSA calculations

First, we computed the binding energies of the four kissing complexes. We let each system equilibrate for 50 ns and we calculated a global molecular mechanics Poisson–Boltzmann surface area (MMPBSA), from 50 to 300 ns. We also divided this time frame into 5 windows (50-100 ns, 100-150 ns, 150-200 ns, 200-

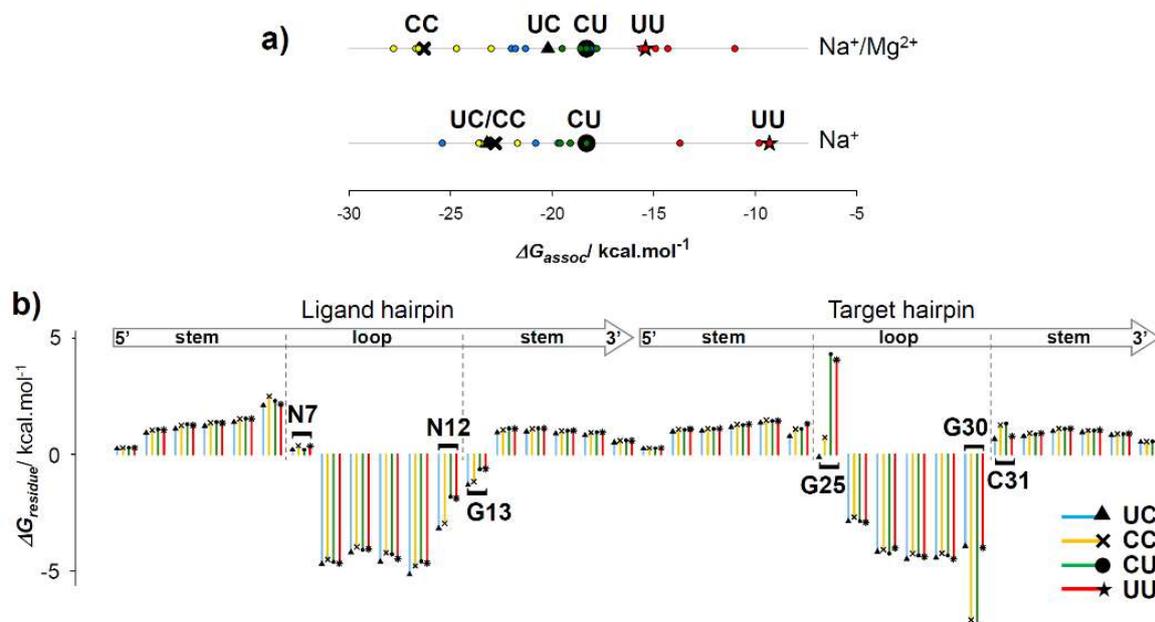
250 ns and 250-300 ns) to grasp the variability due to structural rearrangements during the MD simulation. The association free energy of the hairpin-hairpin binding ( $\Delta G_{assoc}$ ) is the difference between the free energies of the kissing complex ( $G_{KC}$ ) and the unbound target ( $G_{target}$ ) and ligand ( $G_{ligand}$ ):

$$\Delta G_{assoc} = G_{KC} - G_{target} - G_{ligand} \quad (4)$$

Note that for UC and UU in sodium, on those time scales the  $U_7 \bullet G_{30}$  base pair can be either closed or opened (Supplementary Figure S4), and this will influence the  $\Delta G$  values and their variations. Although the magnitude binding constants are by far overestimated (a notorious problem of the MMPBSA method, which neglects target and ligand rearrangement<sup>[18]</sup>), the stability *ranking* of  $\Delta G_{assoc}$  values (Figure 5a; supporting Table S1) agrees with the experimental trends. The per-residue dissection of MMPBSA calculations (Figure 5b in presence of  $Mg^{2+}$ , supporting information Figure S13 in  $Na^+$  alone) can then help us understand the position-dependent effect of  $G \bullet U$  base pair on KC stability. This decomposition allows exploring the respective contributions of the modified base pair, of its nearest neighbors, and of more distant residues. The free energy contribution of a residue is obtained by summing up its interaction with all residues in the system and calculating the difference of interaction in the kissing complex versus the free hairpins.

The  $C_{12} \rightarrow U_{12}$  mutation is unfavorable, especially for its base-pairing residue  $G_{25}$  which greatly destabilizes the complex CU (Figure 5b). There is only a small effect of  $C_{12} \rightarrow U_{12}$  on the neighboring bases ( $C_6$ ,  $G_{13}$ ,  $G_{24}$  and  $G_{11}$ ), with the most marked effect being on  $G_{13}$ . This contrasts with the effect of the  $C_7 \rightarrow U_7$  mutation.

**Figure 5. Total free energy, and contribution each residue.** a) Computed  $\Delta G_{\text{assoc}}$  (most stable complexes on the left) in presence and absence of  $\text{Mg}^{2+}$ . Black symbols: average over the entire trajectory. Smaller



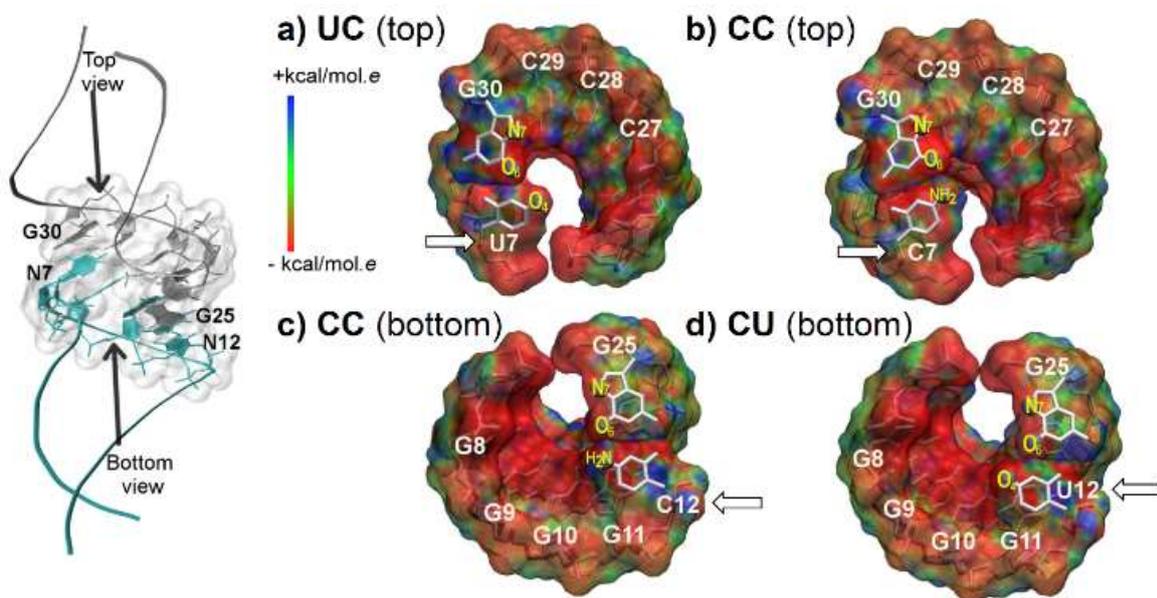
symbols: the average values over each 50-ns chunk of the trajectory, showing that the differences are significant. The stability ordering in the presence of  $\text{Mg}^{2+}$  is  $\text{CC} > \text{UC} > \text{CU} > \text{UU}$ , in agreement with the experiment. The stabilities of UC and CC are very close in the absence of  $\text{Mg}^{2+}$ . The values are given in Supplementary Table S1. **b)** Decomposition of  $\Delta G$  per residue (see text) in presence of  $\text{Mg}^{2+}$ , highlighting which residues are most influenced by each mutation: the entire loop plus the closing base pair (residues 6 to 13 and 24 to 31) are influenced by a single  $\text{C} \rightarrow \text{U}$  mutation.

The pairing base  $\text{G}_{30}$  has a more favorable energy contribution in the case of  $\text{C}_7$  (-7.2 kcal/mol for CC and CU) than in the case of  $\text{U}_7$  (-3.4 kcal/mol for UC and -1.8 kcal/mol for UU), reflecting the number of hydrogen bonds in the wobble base pair. But importantly, the contribution of nearest neighbor base pairs changes as well. For the base pair  $\text{C}_6 \bullet \text{G}_{13}$ , and for the G-tract in the loop of UC, the interaction energy is slightly more favorable for UC than for CC, while the direct neighbors of  $\text{G}_{30}$  ( $\text{C}_{29}$  and  $\text{G}_{31}$ ) are little affected. The stem residues, apart from the closing base pair, are not affected, but the loop residues are each a little affected by the mutations. As a result, the addition of all these contributions nearly counterbalances the loss of free energy upon  $\text{C}_7 \rightarrow \text{U}_7$  mutation, whereas the  $\text{C}_{12} \rightarrow \text{U}_{12}$  mutation is more unfavorable. An

important take-home message is that the whole kissing loop and the closing base pairs of the hairpin must be taken into account to understand the effect of the mutation.

#### 4.2. PBSA of the loop-loop interaction

Going back to the global structure is required to understand the interplay between the mutated residue and the other residues of the kissing loop motif. In order to visualize the position-dependent effect of the G•U and G•C base pairs, we computed the electrostatic potential of the loop-loop interaction with the Poisson Boltzmann surface area (PBSA). The tunnel of the kissing complex has on one side a stretch of cytosines (C<sub>26</sub>–C<sub>29</sub> from the target K1) and on the other side a stretch of guanines (G<sub>8</sub>–G<sub>11</sub> from the ligands K1'). The cytosines display positive charges in the major groove, due to the NH<sub>2</sub> groups. Accordingly, Figure 6 shows a surface highlighted in blue and defining an electropositive surface. In contrast, the guanines possess two electronegative atoms (N7 and O6) in the major groove, giving rise to an electronegative surface, in red in Figure 6.



**Figure 6. The tunnel electrostatic potential surfaces by PBSA.** The general structure of the kissing complex is shown on the left, where K1 is in gray and the ligand K1' in cyan with the position in 3D of the

closing base pairs N<sub>7</sub>•G<sub>30</sub> and N<sub>12</sub>•G<sub>25</sub>. The potential surface is calculated via non-linear Poisson Boltzmann surface area on an average structure from the window frame 50 to 80 ns for the systems <sub>5</sub>UC<sub>3</sub>, <sub>5</sub>CC<sub>3</sub> and <sub>5</sub>CU<sub>3</sub>. The red surfaces represent a negative electrostatic potential, which is observed along the backbone, along the stretches of guanines (O6 and N7 atoms of the residues G<sub>9</sub>, G<sub>10</sub>, G<sub>11</sub>) in the bottom view and on the atom O4 of the residues U<sub>7</sub> and U<sub>12</sub> towards the interior of the tunnel. The blue surface represents a positively charged area, which is observed along the stretches of cytosines (NH<sub>2</sub> atoms of the residues C<sub>27</sub>, C<sub>28</sub>, C<sub>29</sub>) in the top view and on the NH<sub>2</sub> atoms of the residues C<sub>7</sub> and C<sub>12</sub> towards the interior of the tunnel.

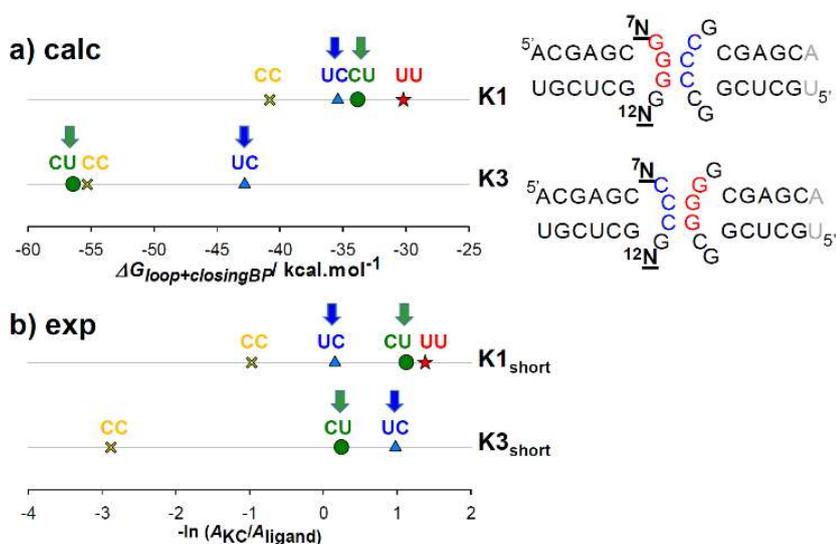
The presence of a G•U wobble base pair provides a distinctive electrostatic feature in the major groove, with the protrusion of the electronegative O4 of the uracil into the tunnel. Consequently, when the G•U is positioned at the 5' end of the loop (UC), the protrusion of the electronegative O4 of the uracil is facing the stretch of electropositive cytosines of the receptor K1, and this is energetically favorable. In contrast, when the G•U is positioned in the 3' end of the loop (CU and UU), the protruding electronegative O4 of the Uracil is now facing the stretch of electronegative guanine of the ligand, which is less favorable (Figure 6). Thus, the description of the electrostatic surface of the entire loop/loop region is necessary to understand why the effect of the C→U mutation is position-dependent.

##### **5. Towards predicting when a G•U mutation will stabilize a kissing complex**

In order to test our hypothesis, we constructed another series of kissing complexes, named K3, with a C-tract in the loop of the target hairpin instead of the G-tract of K1 (Figure 7). With the change of electrostatics in the tunnel, the effect of the G•U wobble base position should be the opposite as in K1. This is what is observed experimentally (Figure 7b). In this complex, the surface of the K1 hairpin switched from positively charged to a negatively charged surface. Therefore, the position of the uracil from the G<sub>30</sub>•U<sub>7</sub> base pair does not benefit anymore for a favorable environment, however the G<sub>25</sub>•U<sub>12</sub> is more favorable which explains the switch in affinity.

In this example, owing to the drastic change in electrostatics occurring in the tunnel surface, a simple reasoning enables one to predict on which side the U would be more favorable. However in the presence of more subtle electrostatic changes, calculating an affinity score is necessary. The three systems were modeled, built and equilibrated as described in the materials and methods, but here no MD production was run. In order to evaluate an affinity score, we summed the per-residue contributions of the all loop

residues, plus the first base pair (Figure 7a). On the four original complexes K1, the simplified affinity score correctly accounts for the stability order. In the case of the K3 series, the affinity score correctly predicts that a U at position 12 is now more favorable than at position 7. Although crude (for example, CU is predicted to be as stable as CC, but experimentally it is less stable), the simplicity and speed of the calculation makes it promising to make rough predictions and design new sequence variants and estimate their change of affinity. Experimental validation remains necessary.



**Figure 7. Effect of loop electrostatics on the stability of mutated complexes.** The K1 series was mutated in its loop-loop region to switch cytosines and guanines, to create the K3 series. **a)** Computed affinity score (no Mg<sup>2+</sup>): the per-residue contributions of the loop and the first stem base pair of each hairpin were summed. **b)** Experimental results on the relative stability of the KC in 150 mM NH<sub>4</sub>OAc. The most stable complexes appear on the left. The experimental stability ordering of CU and UC is correctly predicted by the calculation.

## Conclusions

Here we investigated the sequence effect in the formation of a RNA kissing complex. We show, that a G•U wobble base pair provides a favorable environment for the kissing complex to form due to its specific geometry. We also underlined the crucial role of the identity of the base pairing in the whole loop-loop

motif. Thus a strategically placed non-canonic base pair might affect the interaction towards a better affinity. However, predicting the stabilizing effect of a G•U base pair requires considering all residues close in space, beyond the nearest neighbors. In the case of kissing loop complexes, this means that all residues of the loop influence the global stability of the complex. In a case of a G•U wobble base pair, the protrusion of the electronegative O4 atom of the uracil towards the major groove and into the tunnel of the loop-loop interaction allows one to understand that this is favorable only when this protrusion is made in the direction of an electropositive pocket, here provided by the cytosine residues.

These conclusions might lead to the understanding of the electrostatic impact of the introduction of a G•U wobble base pair and may be a lead for a promising approach for the rational engineering of affine kissing complexes. We can envision that by strategically placing non-canonical base pairs such as G•U and G•A with defined geometry and with unique profiles of donor and acceptors in the major groove, we can take advantage of this structural characteristics to rationally design new aptamers with higher affinity, for therapeutic, nanotechnology or analytical applications. One should however not forget that RNA aptamer structure, ligand specificity and ligand affinity depend on the ionic environment,<sup>[19]</sup> and that exploring the effect of the divalent ions is necessary if different specific binding sites are involved. Also, affinity is not the sole selection criteria. Kinetics might also be crucial, especially in therapeutics. In that respect, investigating the effects of mutations and of magnesium concentration on the reaction dynamics is worthy of further exploration.

## Experimental Section

**Oligonucleotides.** RNA oligonucleotides were purchased in lyophilized form from Integrated DNA Technology (Leuven, Belgium) with standard desalting. They were dissolved in RNase free water (Ambion, Fisher Scientific, Illkirch, France), then diluted to a stock solution at 100  $\mu$ M of each hairpin in 150 mM Ammonium Acetate NH<sub>4</sub>OAc (prepared from BioUltra  $\approx$  5 M stock solution, Fluka, Sigma-Aldrich, Saint-Quentin Fallavier, France). This solution was desalted using Amicon Ultra-0.5 3K ultrafiltration devices (Merck Millipore, Saint-Quentin en Yvelines, France). The sequences were then heated at 90°C during 90 s, then immediately placed on ice and allowed to cool for 10 min, then stored at 4°C. All the injected solutions were prepared at least one day prior to the experiment. The injected solutions were 10  $\mu$ M of each hairpin in 150 mM NH<sub>4</sub>OAc. 2  $\mu$ M DNA oligonucleotide dT<sub>6</sub> was added as an internal standard. For

the Mg-dependent experiments, an  $\text{Mg}(\text{OAc})_2$  solution (Sigma-Aldrich, Saint-Quentin Fallavier, France) is added to obtain the desired concentrations.

**Native Electrospray Ionization Mass Spectrometry (ESI-MS).** Experiments were performed on an Agilent 6560 DTIMS-Q-TOF instrument (Agilent Technologies, Santa Clara, CA), with the dual-ESI source operated in negative ion mode. The syringe pump flow rate was 180  $\mu\text{L}/\text{h}$ . The drift tube was filled with helium and the pressure in the drift tube was  $3.89 \pm 0.01$  Torr, and the pressure in the trapping funnel is  $3.67 \pm 0.01$  Torr (this ensures that only helium is present in the drift tube). The source temperature and fragmentor voltage were set at  $200^\circ\text{C}$  and 350 V, and the Trap Entrance Grid Delta fixed at 12 V (Trap entrance Grid low 105 V, Trap Entrance Grid High 117 V). All these parameters ensured a good desolvation/declustering and transmission of ions, without causing any dissociation of the complexes.

**Generation of the structural models.** The RNA models were constructed using the webserver of ModeRNA,<sup>[20]</sup> with the NMR structure of TAR/TAR\* (PDB ID=2RN1)<sup>[21]</sup> as template. The complexes were neutralized with either only  $\text{Na}^+$  ions or with a mixture of 30  $\text{Na}^+$  and 2  $\text{Mg}^{2+}$  (magnesium parameters from reference<sup>[22]</sup>) and hydrated with TIP3P water molecules<sup>[23]</sup> in a truncated octahedron. The sodium-only simulations were run first. Then the initial placement of  $\text{Mg}^{2+}$  was decided based on the sodium simulations, wherein we found two high-density regions (In the loop-loop region and at the exit of the tunnel, see supplementary Figure S6) for all four systems. Thus magnesium ions do not have to explore the whole octahedron to find their most affine binding sites. The obtained structures were then minimized and equilibrated using the module pmemd.MPI of Amber12<sup>[24]</sup> suite of packages, using the parmbsc0+ $\chi\text{OL3}$  force field.<sup>[25]</sup> We applied periodic boundary conditions at constant temperature (300 K) and pressure (1 bar) using the Berendsen algorithm.<sup>[26]</sup> Covalent bonds involving hydrogens were restrained using SHAKE,<sup>[27]</sup> allowing for a 2-fs integration time-step. Long-range electrostatic interactions were treated using the particle mesh Ewald (PME)<sup>[24]</sup> algorithm with a radial cutoff of 9 Å; the same cutoff was used for the van der Waals interactions. The non-bonded pair-list was updated heuristically and the center-of-mass motion removed every 10 ps. Water molecules and counter-ions were energy-minimized (steepest descent followed by conjugate gradient), and equilibrated at 100 K for 100 ps at constant volume and temperature while RNA molecules were positionally restrained. The whole system was then heated from 100 K to 300 K in 10 ps by 5-K increments with harmonic positional restraints on solute atoms (5.0 kcal/mol/Å<sup>2</sup> force constant). The simulation was continued at constant pressure and temperature, with positional restraints gradually removed over 250 ps.

**Molecular dynamics simulations.** MD production runs (listed in Table 1) started from the end point of the equilibration. The cpptraj<sup>[28]</sup> module of AmberTools14<sup>[29]</sup> was used to calculate the hydrogen bond occupancies, by taking snapshots from 50 ns to 300 ns with a step of 5ps. The statistical analysis was carried out using the R package.<sup>[30]</sup> Image rendering was performed with tachyon in VMD,<sup>[31]</sup> and the helicoidal parameters were calculated with *curves+*.<sup>[32]</sup>

**Thermodynamics calculations. MMPBSA analyses:** The binding energies of the hairpins were computed using the molecular mechanics Poisson–Boltzmann surface area (MMPBSA) approaches,<sup>[14]</sup> implementing the MMPBSA.py script.<sup>[33]</sup> The  $\Delta G_{\text{assoc}}$  is composed of the changes in the molecular mechanical gas phase energy ( $\Delta E_{\text{MM}}$ ), entropic contribution, and solvation free energy:

$$\Delta G_{\text{assoc}} = \Delta E_{\text{MM}} - T\Delta S + \Delta G_{\text{solv}} \quad (5)$$

$\Delta G_{\text{solv}}$  is estimated by solving the linearised Poisson Boltzmann equation for each of the three states ( $\Delta G_{\text{polar}}$ ) and adding an empirical term for hydrophobic contributions to it ( $\Delta G_{\text{nonpolar}}$ ). The hydrophobic contribution is calculated from the solvent accessible surface area (SASA). We assume that the entropic contribution of the four systems will be in the same order because of only slight differences in atomic composition, therefore the entropic contribution is neglected from the calculations that aim at comparing the mutants. **PBSA analyses:** We used the non-linear Poisson-Boltzmann solver (PBSA)<sup>[34]</sup> in AmberTools14<sup>[29]</sup> for the calculation of the electrostatic potential surface. The interior (solute/molecule) and exterior (solvent) dielectric constants were set to 2 and 80, respectively. The temperature was set to 300 K, a solvent probe radius of 1.4 Å was used to define the dielectric boundary, a 150 mM ionic strength and a PB grid resolution of 0.5 Å were chosen.

## Acknowledgements

This work was supported by the European Research Council (ERC DNAFOLDIMS to VG) and the CNRS (post-doctoral fellowship to JAG).

## References

- [1] M. Mascini, I. Palchetti, S. Tombelli, *Angew. Chem. Int. Ed.* **2012**, *51*, 1316-1332.
- [2] M. Mandal, R. R. Breaker, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 451-463.
- [3] M. Darmostuk, S. Rimpelova, H. Gbelcova, T. Ruml, *Biotechnol. Adv.* **2015**, *33*, 1141-1161.
- [4] a) C. Boiziau, E. Dausse, L. Yurchenko, J. J. Toulme, *J Biol Chem* **1999**, *274*, 12730-12737; b) F. Ducongé, C. Di Primo, J. J. Toulmé, *J. Biol. Chem.* **2000**, *275*, 21287-21294; c) L. Aldaz-Carroll, B. Tallet, E. Dausse, L. Yurchenko, J. J. Toulme, *Biochemistry* **2002**, *41*, 5883-5893; d) K. Kikuchi, T. Umehara, K. Fukuda, J. Hwang, A. Kuno, T. Hasegawa, S. Nishikawa, *J. Biochem.* **2003**, *133*, 263-270.
- [5] A. Wittmann, B. Suess, *FEBS Lett* **2012**, *586*, 2076-2083.
- [6] a) J. C. Paillart, E. Skripkin, B. Ehresmann, C. Ehresmann, R. Marquet, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 5572-5577; b) J. Tomizawa, *Cell* **1986**, *47*, 89-97.
- [7] a) A. M. Pyle, S. Moran, S. A. Strobel, T. Chapman, D. H. Turner, T. R. Cech, *Biochemistry* **1994**, *33*, 13856-13863; b) J. Chen, A. Ganguly, Z. Miswan, S. Hammes-Schiffer, P. C. Bevilacqua, B. L. Golden, *Biochemistry* **2013**, *52*, 557-567.
- [8] W. H. McClain, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 4570-4575.
- [9] a) F. H. Allain, G. Varani, *Nucleic Acids Res* **1995**, *23*, 341-350; b) L. R. Stefan, R. Zhang, A. G. Levitan, D. K. Hendrix, S. E. Brenner, S. R. Holbrook, *Nucleic Acids Res* **2006**, *34*, D131-134; c) G. Varani, W. H. McClain, *EMBO Rep* **2000**, *1*, 18-23.
- [10] a) E. Dausse, A. Barre, A. Aime, A. Groppi, A. Rico, C. Ainali, G. Salgado, W. Palau, E. Daguerre, M. Nikolski, J. J. Toulme, C. Di Primo, *Biosens. Bioelectron.* **2016**, *80*, 418-425; b) G. Durand, E. Dausse, E. Goux, E. Fiore, E. Peyrin, C. Ravelet, J. J. Toulme, *Nucleic Acids Res.* **2016**, *44*, 4450-4459.
- [11] D. Xu, T. Landon, N. L. Greenbaum, M. O. Fenley, *Nucleic Acids Res.* **2007**, *35*, 3836-3847.
- [12] a) F. Beaurain, C. Di Primo, J.-J. Toulmé, M. Laguerre, *Nucleic Acids Res.* **2003**, *31*, 4275-4284; b) K. Reblova, N. Spackova, J. E. Sponer, J. Koca, J. Sponer, *Nucleic Acids Res.* **2003**, *31*, 6492-6952; c) J. Golebiowski, S. Antonczak, J. Fernandez-Carmona, R. Condom, D. Cabrol-Bass, *J. Mol. Model.* **2004**, *10*, 408-417.
- [13] a) A. A. Chen, D. E. Draper, R. V. Pappu, *J. Mol. Biol.* **2009**, *390*, 805-819; b) A. Singh, L. Sethaphong, Y. G. Yingling, *Biophys. J.* **2011**, *101*, 727-735.
- [14] a) P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, T. E. Cheatham, 3rd, *Acc. Chem. Res.* **2000**, *33*,

- 889-897; b) J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman, D. A. Case, *J. Am. Chem. Soc.* **1998**, 120, 9401-9409; c) J. Srinivasan, J. Miller, P. A. Kollman, D. A. Case, *J. Biomol. Struct. Dyn.* **1998**, 16, 671-682.
- [15] S. Genheden, U. Ryde, *Exp. Opin. Drug Discov.* **2015**, 10, 449-461.
- [16] A. Singh, L. Sethaphong, Y. G. Yingling, *Biophys. J.* **2011**, 101, 727-735.
- [17] A. Perez, A. Noy, F. Lankas, F. J. Luque, M. Orozco, *Nucleic Acids Res.* **2004**, 32, 6144-6151.
- [18] a) X. Du, Y. Li, Y. L. Xia, S. M. Ai, J. Liang, P. Sang, X. L. Ji, S. Q. Liu, *Int. J. Mol. Sci.* **2016**, 17; b) U. Ryde, P. Soderhjelm, *Chem. Rev.* **2016**, 116, 5520-5566.
- [19] M. Ilgu, T. Wang, M. H. Lamm, M. Nilsen-Hamilton, *Methods* **2013**, 63, 178-187.
- [20] a) M. Rother, K. Rother, T. Puton, J. M. Bujnicki, *Brief. Bioinform.* **2011**, 12, 601-613; b) M. Rother, K. Rother, T. Puton, J. M. Bujnicki, *Nucleic Acids Res.* **2011**, 39, 4007-4022.
- [21] H. Van Melckebeke, M. Devany, C. Di Primo, F. Beaurain, J. J. Toulmé, D. L. Bryce, J. Boisbouvier, *Proc. Natl. Acad. Sci. USA* **2008**, 105, 9210-9215.
- [22] O. Allner, L. Nilsson, A. Villa, *J. Chem. Theor. Comput.* **2012**, 8, 1493-1502.
- [23] a) W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, 79, 926-935; b) M. Mahoney, W. Jorgensen, *J. Chem. Phys.* **2000**, 112, 8910-8922.
- [24] AMBER 12. D. A. Case, T. A. Darden, T.E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B.Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F.Wong, F. Paesani, J. Vanicek, X.Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, P. A. Kollman, University of California, San Francisco, **2008**.
- [25] a) P. Banáš, D. Hollas, M. Zgarbová, P. Jurečka, M. Orozco, T. E. Cheatham, J. Šponer, M. Otyepka, *J. Chem. Theor. Comput.* **2010**, 6, 3836-3849; b) M. Zgarbova, M. Otyepka, J. Sponer, A. Mladek, P. Banas, T. E. Cheatham, 3rd, P. Jurecka, *J. Chem. Theor. Comput.* **2011**, 7, 2886-2902.
- [26] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Dinola, J. R. Haak, *J. Chem. Phys.* **1984**, 81, 3684-3690.
- [27] W. F. van Gunsteren, H. J. C. Berendsen, *Mol. Phys.* **1977**, 34, 1311-1327.
- [28] D. R. Roe, T. E. Cheatham, *J. Chem. Theor. Comput.* **2013**, 9, 3084-3095.
- [29] J. T. B. D.A. Case, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko,

T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, G. Monard, P. Needham, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, R. Salomon-Ferrer, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, D.M. York and P.A. Kollman, *University of California, San Francisco*. **2015**.

[30] R core team, R Foundation for Statistical Computing, Vienna, Austria, **2015**.

[31] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.* **1996**, *14*, 33-38, 27-38.

[32] R. Lavery, M. Moakher, J. H. Maddocks, D. Petkeviciute, K. Zakrzewska, *Nucleic Acids Res.* **2009**, *37*, 5917-5929.

[33] B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke, A. E. Roitberg, *J. Chem. Theor. Comput.* **2012**, *8*, 3314-3321.

[34] Q. Cai, M.-J. Hsieh, J. Wang, R. Luo, *J. Chem. Theor. Comput.* **2010**, *6*, 203-211.



III. A3. Article 2: [Compaction of duplex nucleic acids upon native electrospray mass spectrometry](#), Porrini et al, et al, *ACS Central Science*, 2017

This document is the Accepted Manuscript version (acceptance date 19-Apr-2017) of a Published Work that will appear in final form in *ACS Central Science*, copyright © American Chemical Society after technical editing by the publisher.

## Compaction of Duplex Nucleic Acids upon Native Electrospray Mass Spectrometry

Massimiliano Porrini,<sup>1</sup> Frédéric Rosu,<sup>2</sup> Clémence Rabin,<sup>1</sup> Leonardo Darré,<sup>3,4</sup> Hansel Gómez,<sup>3,4</sup> Modesto Orozco,<sup>3,4,5</sup> and Valérie Gabelica\*<sup>1</sup>

<sup>1</sup> Univ. Bordeaux, INSERM, CNRS, Laboratoire Acides Nucléiques Régulations Naturelle et Artificielle (ARNA, U1212, UMR5320), IECB, 2 rue Robert Escarpit, 33607 Pessac, France.

<sup>2</sup> CNRS, INSERM, Univ. Bordeaux, Institut Européen de Chimie et Biologie (IECB, UMS3033, US001), 2 rue Robert Escarpit, 33607 Pessac, France.

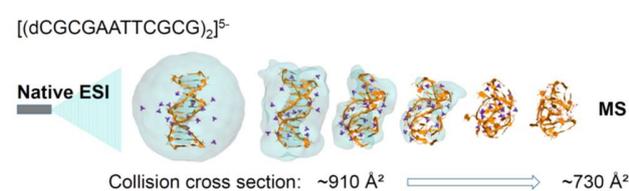
<sup>3</sup> Institute for Research in Biomedicine (IRB) Barcelona, The Barcelona Institute of Science and Technology, Baldiri Reixac 10, 08028 Barcelona, Spain.

<sup>4</sup> Joint BSC-CRG-IRB Research Program in Computational Biology, IRB Barcelona, Barcelona, Spain.

<sup>5</sup> Department of Biochemistry and Biomedicine, University of Barcelona, Avda Diagonal 647, 08028 Barcelona, Spain.

\* Corresponding Author: Valérie Gabelica, IECB, 2 rue Robert Escarpit, 33607 Pessac, France. Tel: +33 (0)5 4000 2940. E-mail: [v.gabelica@iecb.u-bordeaux.fr](mailto:v.gabelica@iecb.u-bordeaux.fr)

### SYNOPSIS



solution.

Electrospray ionization is widely used in mass spectrometry. But to what extent is the solution structure conserved in the gas phase? Using ion mobility, we demonstrate that DNA duplexes are much more compact in the gas phase than in

## ABSTRACT

We report on the fate of nucleic acids conformation in the gas phase as sampled using native mass spectrometry coupled to ion mobility spectrometry. Based on several successful reports for proteins and their complexes, the technique becomes popular in structural biology, and the conformation survival becomes more and more taken for granted. Surprisingly, we found that DNA and RNA duplexes, at the electrospray charge states naturally obtained from native solution conditions ( $\geq 100$  mM aqueous  $\text{NH}_4\text{OAc}$ ) are significantly more compact in the gas phase compared to the canonical solution structures. The compaction is observed for all duplex sizes (gas-phase structures are more compact than canonical B-helices by  $\sim 20\%$  for 12-bp, and by up to  $\sim 30\%$  for 36-bp duplexes), and for DNA and RNA alike. Molecular modeling (density functional calculations on small helices, semi-empirical calculations on up to 12-bp, and molecular dynamics on up to 36-bp duplexes) demonstrates that the compaction is due to phosphate group self-solvation prevailing over Coulomb repulsion. Molecular dynamics simulations starting from solution structures do not reproduce the experimental compaction. To be experimentally relevant, molecular dynamics sampling should reflect the progressive structural rearrangements occurring during desolvation. For nucleic acid duplexes, the compaction observed for low charge states results from novel phosphate-phosphate hydrogen bonds formed across both grooves at the very late stages of electrospray.

## INTRODUCTION

Besides genetic information storage, nucleic acids perform pivotal regulatory functions in living organisms.<sup>1</sup> Conformational changes are key to these regulation mechanisms, whether at the DNA level (e.g., particular structures in gene promoters)<sup>2</sup> or at the RNA level (e.g., riboswitches).<sup>3</sup> There is thus a need for new experimental approaches enabling one to detect and characterize the different ensembles of conformation simultaneously present in solution. “Native” electrospray ionization mass spectrometry (ESI-MS) helps deciphering the complexation equilibria.<sup>4-7</sup> “Native MS” means that the solution conditions must be compatible with electrospray ionization, yet allow the system to fold in the same way as it would in physiological conditions.<sup>8</sup> The typical solution condition for native MS is aqueous ammonium acetate (neutral pH, no organic co-solvent). Also, the mass spectrometer must be operated in “soft” conditions, meaning that the internal energy imparted to the ions must be just enough to desolvate the biomolecules and strip nonspecific ionic adducts, but not too high so as not to disrupt the complexes before they reach the mass analyzer.

Ion mobility spectrometry (IMS) further enables to characterize the shape of each complex separated by mass, based on the electrophoretic mobility of the ions in a buffer gas.<sup>9</sup> Coupled to ion mobility spectrometry (IMS), native mass spectrometry thus makes it possible to reveal the topology of multi-protein complexes,<sup>10</sup> conformational changes upon ligand binding,<sup>11-13</sup> or the architecture of synthetic supramolecular assemblies.<sup>14</sup> But before native ESI-IMS-MS can be applied to characterize nucleic acids conformations in solution, it is essential to understand to what extent the different types of DNA/RNA secondary structures are preserved, or affected, by the transition from the solution to the gas phase.

Tandem mass spectrometry<sup>15-18</sup> experiments have shown that the kinetic stability of gas-phase DNA duplexes was correlated with the fraction of guanine-cytosine (GC) base pairs and with the base pair ordering. This suggested that *Watson-Crick* (*WC*) hydrogen bonding and base stacking were at least partially maintained in the gas phase. Infrared multiphoton dissociation spectroscopy on duplexes and single strands suggested that GC bases are engaged in hydrogen bonds in the gas-phase duplexes, although no such evidence was found for AT base pairs.<sup>19</sup> Early molecular dynamics (MD) validated this view: although (12-mer)<sub>2</sub><sup>6-</sup> and (16-mer)<sub>2</sub><sup>8-</sup> duplexes were distorted in the gas phase, most *WC* H-bonding and stacking interactions were preserved, particularly in GC-rich regions.<sup>20</sup>

The Bowers group further explored the gas-phase structure of DNA duplexes by IMS.<sup>21,22</sup> By comparing the experimental collision cross section (CCS) of d(GC)<sub>n</sub> duplexes with those obtained on duplexes relaxed by short (5-ns) MD, they showed that the gas-phase structures resemble an A-helix for the short duplexes (8—16-mer) and a B-helix for longer ones (>18-mer). However, those experiments had been performed on

the relatively high charge states (1 negative charge per 2 base pairs) obtained from solution conditions (49:49:2 mixture of H<sub>2</sub>O:MeOH:NH<sub>4</sub>OH) differing markedly from those typically used in native MS. Here we show that, at the lower charge states typically obtained from native MS conditions (100—150 mM NH<sub>4</sub>OAc in 100% H<sub>2</sub>O, pH=7), the gas-phase conformations are significantly more compact than expected for B- or A-helices. We investigated the physical origin of this gas-phase compaction by molecular modeling.

## RESULTS AND DISCUSSION

### **DNA duplexes in their predominant native charge state are more compact than a canonical B-helix.**

The sequences studied here were designed based on the Dickerson-Drew dodecamer,<sup>23</sup> here noted 12-d<sub>66</sub> indicating the number of base pairs (12), hyphen the nucleotide type (i.e. *d* for DNA and *r* for RNA), and followed by the GC base pair percentage content (66%) as a subscript. This notation was used throughout the text. DNA or RNA duplexes were prepared by annealing in aqueous ammonium acetate and analyzed using drift tube ion mobility in helium (see methods and supporting information section S1). In aqueous NH<sub>4</sub>OAc, RNA sequences fold into the A-form, and DNA sequences fold into the B-form (according to the circular dichroism<sup>24</sup> spectra in supporting information Figure S2).

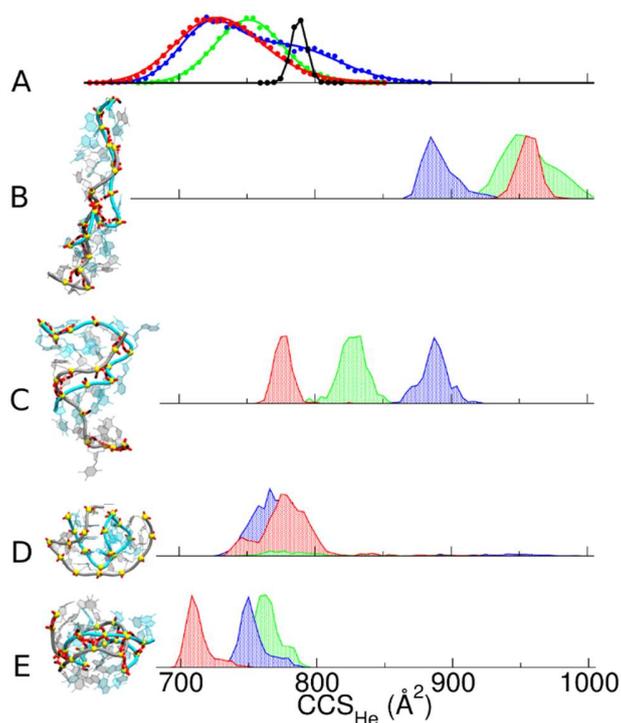
When sprayed from aqueous 100 mM NH<sub>4</sub>OAc, the major charge state of 12-bp duplexes is 5<sup>-</sup> (Figure S3). This may seem at odds with the solution charge of the nucleic acid alone, which is 22<sup>-</sup>. One should however remind that an atmosphere of counter-ions surrounds nucleic acids in solution.<sup>25</sup> If counter-ions were not present, the electrostatic repulsion between strands would amount to hundreds of kT, and double helices would not exist. The effective charge state of nucleic acids in solution is hard to define: bulk solutions being electrically neutral, the net charge of a duplex depends at which distance from the nucleic acid atoms one places the boundary. Molecular dynamics simulations on 18-bp B-DNA duplexes<sup>26</sup> showed that neutrality is reached at an average distance of 24 Å from the center of the helix, and that 76% neutralization (a threshold defined by Manning's counterion condensation model<sup>27</sup>) is reached at 18 Å. Electrospray droplets are however not neutral, and the final charge states could reflect the thickness of the layer of solvent and counterions surrounding the nucleic acid in the final droplets.

The <sup>DT</sup>CCS<sub>He</sub> distributions for the 5<sup>-</sup> duplexes without ammonium ions bound and in the softest ion transfer conditions are shown in Figure 1A (full results in supporting Figure S4). These distributions suggest more compact structures than those of canonical B- or A- forms (Table 2). Upon pre-IMS activation, the CCS distribution of 12-d<sub>33</sub> is unchanged, 12-d<sub>66</sub> is losing the high-CCS peak to the profit of the low-CCS peak, and the entire distribution of 12-d<sub>100</sub> is shifting towards lower CCS (from 730 Å<sup>2</sup> down to ~705 Å<sup>2</sup>, see supporting information Figure S5). At low pre-IMS activation, duplexes<sup>5-</sup> with ammonium adducts are also detected. The CCS of these ions is similar to that of the bare duplexes, but for 12-d<sub>66</sub> the higher-CCS peak is more abundant—see supporting Figure S6). The peak center values obtained for soft and harsh conditions are listed in Table 2. Charge states 4<sup>-</sup> and 6<sup>-</sup> are also detected in aqueous NH<sub>4</sub>OAc. The duplex CCS distributions are highly charge-state dependent (see supplementary Figure S4): charge states 4<sup>-</sup> and 5<sup>-</sup> are similarly compact, whereas charge state 6<sup>-</sup> has a ~20% larger CCS. Charge state 7<sup>-</sup>, obtained for 12-d<sub>66</sub> and 12-d<sub>100</sub> by adding the “supercharging” agent sulfolane,<sup>28</sup> has a >30% larger CCS. The duplex CCS

distributions are significantly broader than those of the tetramolecular G-quadruplex [dTG<sub>4</sub>T]<sub>4</sub> (in black in Figure 1A, and supplementary Figure S4), a rigid structure with the same number of bases. This indicates that a greater conformational space is explored in the gas phase by nucleic acid duplexes compared to the G-quadruplex,<sup>18</sup> and that gas-phase duplexes consist of an ensemble of conformations not fully interconverting on the time scale of the mobility separation (10-30 ms).

**Table 1.** Size, name and sequences of the duplexes under study, and outline of calculations.

Size	Name	Sequence	Theoretical levels of study
10 bp	10-bp	dCGCGGGCCCG•dCGGGCCCGCG	DFT (M06-2X - 4-31G(d)) semi-empirical (PM7)
12 bp	12-d <sub>33</sub>	(dCGTAAATTTACG) <sub>2</sub>	semi-empirical (PM7)
	12-d <sub>66</sub>	(dCGCGAATTCGCG) <sub>2</sub>	MD from different starting structures (1 μs each)
	12-d <sub>100</sub>	(dCGCGGGCCCGCG) <sub>2</sub>	
	12-r <sub>33</sub>	(rCGUAAAUUUACG) <sub>2</sub>	semi-empirical (PM7)
	12-r <sub>66</sub>	(rCGCGAAUUCGCG) <sub>2</sub>	T-REMD (1 μs × 18 replicas)
	12-r <sub>100</sub>	(rCGCGGGCCCGCG) <sub>2</sub>	
24 bp and 36 bp	Concatenations of the 12-bp duplexes above. See supporting information Section S9.		MD on B-helix (0.25 μs) MD on zipped helix (0.5 μs)



**Figure 1.** (A) Experimental  $^{DT}CCS_{He}$  distribution measured under soft native conditions for the 12-bp DNA duplexes (green:  $(12-d_{33})^{5-}$ , blue:  $(12-d_{66})^{5-}$ , red:  $(12-d_{100})^{5-}$ ; see Table 1 for full sequences) and the rigid G-quadruplex  $([d(TG_4T)]_4)^{5-}$  (black). (B-E) Calculated CCS distributions for molecular models generated by (B) gas-phase MD of the B-helices, (C) gas-phase MD of the A-helices, (D) T-REMD simulations on B-helices using distributed charges (note that most of the population of  $(12-d_{33})^{5-}$  duplex dissociated during simulations), and (E) MD following a restrained minimization forcing H-bond formation between the phosphate groups across both grooves of a B-helix. The final MD structures of each duplex model, created with VMD software,<sup>29</sup> are shown for  $(12-d_{66})^{5-}$  on the same scale (see supporting Figure S11 for a magnification).

**Table 2.** Collision cross section values in helium ( $CCS_{He}$ ) of 12-bp duplexes, experimental and calculated prior and subsequent to PM7 optimization, unbiased MD, T-REMD (DC model; results with LC model are not listed because they depend on the chosen locations—see text), and unbiased MD on the helix zipped by restrained minimization. Errors reported for MD are the standard deviation on the different structures along the trajectory.

	$^{DT}CCS_{He} (\text{Å}^2)$	$^{CALC}CCS_{He} (\text{Å}^2)$								
		B-helix				A-helix				Zipped helix
		Canonical	PM7	MD	T-REMD (DC)	Canonical	PM7	MD	T-REMD (DC)	MD
$[12-d_{33}]^{5-}$	Soft/harsh: 760	918	914	$954 \pm 18$	$780 \pm 24$	900	861	$826 \pm 11$		$759 \pm 12$
$[12-d_{66}]^{5-}$	Soft: 735/803 Harsh : 735	903	831	$881 \pm 6$	$770 \pm 15$	893	942	$886 \pm 10$		$752 \pm 10$
$[12-d_{100}]^{5-}$	Soft: 730 Harsh: 705	908	838	$953 \pm 8$	$771 \pm 14$	892	869	$774 \pm 6$		$711 \pm 9$
$[12-r_{33}]^{5-}$	-					944	938		$741 \pm 21$	
$[12-r_{66}]^{5-}$	Soft/harsh: 725					945	902		$747 \pm 18$	
$[12-r_{100}]^{5-}$	Soft: 735 Harsh: 695					940	896		$743 \pm 11$	

**MD trajectories, DFT and semi-empirical optimizations reveal phosphate-phosphate hydrogen bond formation.** To find out which three-dimensional structures are compatible with the experimental CCS values of the 12-bp<sup>5-</sup> duplexes, we first carried out unbiased MD simulations directly from B- and A-helix structures, stripped of the solvent. The two possibilities to reduce the total charge to -5 (major charge state) are the localized charges (LC) and distributed charges (DC) models.<sup>20</sup> With the DC model, the net charge of each phosphate group is reduced so that the total charge of the duplex is -5. With the LC model, protons are added on 17 out of the 22 phosphate groups. LC and DC gave similar results upon unbiased MD of duplexes,<sup>20</sup> so here only the LC model was tested extensively.

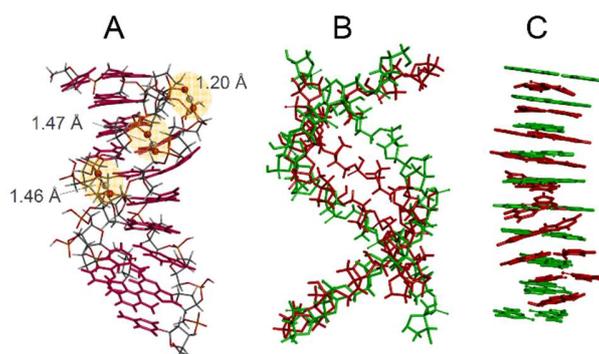
Figure 1B,C shows representative CCS distributions. All such simulations give CCS values significantly larger than the experimental ones. Independent trajectories, started from different conformations and with different sets of localized charges, confirmed this result (supporting Figure S8). The experimental CCS values of the duplexes<sup>6-</sup> are nonetheless compatible with the simulated A- and B-helices, and duplexes<sup>7-</sup> are compatible with B-helices. This does not necessarily mean that the gas-phase conformations of 6- and 7-charge states *are* A- and B-helices, but helps to understand conclusions previously derived solely based on more densely charged duplexes.<sup>21,22</sup>

When starting from the B-form, MD simulations always show spontaneous hydrogen bonds formation between phosphate groups situated on each side of the minor groove (supporting information Movie S1 and Figure S9). This causes the “zipping” of the minor groove. The structures are stable up to 1  $\mu$ s (Figure S9), but too elongated compared to the duplexes<sup>5-</sup> experimental data (compare Fig. 1B with 1A). In simulations starting from the A-helix, zipping occurs as well, but this time across the major groove (supporting Figure S10). In the gas phase, the closest protonated phosphate groups therefore tend to form hydrogen bonds that did not exist in solution. However, this kind of simulation does not reproduce the experiments.

To check for possible artifacts due to using classical force fields to represent macromolecules in the gas phase, we also studied B-duplexes at the density functional theory (DFT) and semi-empirical (SE) levels. Upon DFT optimization of 7-bp to 10-bp duplexes, phosphate groups form new H-bonds and close the minor groove as well (Figure 2 and Supporting Figure S12). *WC* H-bonds and base stacking are well preserved,<sup>30,31</sup> and the helix compresses along its longitudinal axis (Figure 2C) and the <sup>CALC</sup>CCS<sub>He</sub> of the DFT optimized structure is smaller than that of the canonical helix.

For SE calculations, we first validated that the PM7 method best reproduces the DFT results (supporting Figure S13). The CCS values obtained after PM7 optimization are summarized in Table 2 for all [12-bp duplexes]<sup>5-</sup> (A- and B-form for DNA and A-form for RNA). Upon PM7 optimization, the duplex [12-d<sub>66</sub>]<sup>5-</sup> undergoes minor groove zipping, while *WC* H-bonds and base pair stacking interactions are preserved (Figure S14). The compaction compared to the solution structure is only ~8%, thus still far from the experimental value.

Thus, starting from naked canonical structures, neither geometry optimization nor unbiased MD trajectories lead the system towards the experimentally observed conformational ensemble. So, either the sampling is incomplete (simulated and experimental time scales differ), or the starting structures, obtained by desolvating and charging the duplexes all at once, inadequately reflect the electrospray droplet desolvation and declustering.



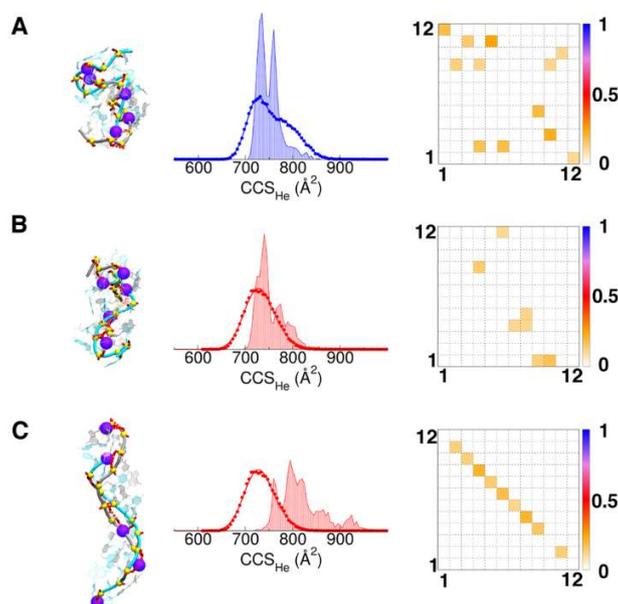
**Figure 2.** DFT optimization of the 10-bp duplex<sup>4</sup>: (A) new hydrogen bonds formed between phosphate groups across the minor groove; (B) superposition of the sugar-phosphate backbones of the canonical B-helix (green) and optimized duplex (red); (C) superposition of the base pairs of the canonical B-helix (green) and optimized duplex (red).

**Temperature replica exchange MD (T-REMD) exploration of gas-phase conformational landscapes do not converge to the experimental structure.** To enhance sampling, we carried out T-REMD simulations on the 12-bp duplexes<sup>5-</sup> starting from their canonical structures. Because temperature replica exchange MD (T-REMD) simulations on gas-phase duplexes have never been attempted before, LC and DC models were both tested for T-REMD. DNA and RNA results, including representative snapshots, total hydrogen bonding, *WC* hydrogen bonding and stacking<sup>32</sup> occupancies, are summarized in supporting information Section S7 (Figures S15—S25). For RNA, the theoretical CCS distributions, both with the DC and LC models, match with the experimental ones (Figures S16 and S19). However, the hydrogen bonding and stacking patterns (Figures S20—S22) always become scrambled in the gas phase. For DNA, the T-REMD results depend more critically on the charge location model: <sup>CALC</sup>CCS<sub>He</sub> values agree with the experiments for the DC model (Figure 1D) and for *some* LC models (Figure 3).

With the DC model, the CCS values are closer to the experimental ones, yet still significantly too large for 12-ds<sub>100</sub><sup>5-</sup>. The GC base pairs are mostly preserved (Figure S37), but the duplex 12-d<sub>33</sub><sup>5-</sup> is mostly melted (separated into single strands; see Figure S15). DC models lack the explicit protons and therefore cannot form phosphate-phosphate hydrogen bonds. As a result, the strands can dissociate upon T-REMD. The rate of strand dissociation occurrence ranks 12-d<sub>33</sub> > 12-d<sub>66</sub> > 12-d<sub>100</sub> (supporting Figure S15), in line with the relative gas-phase kinetic stabilities in tandem mass spectrometry.<sup>15-18</sup>

When LC models are used, the T-REMD final structures depend significantly on the choice of charge location (Figure S18, S20—S22). For example, for [12-d<sub>100</sub>]<sup>5-</sup>, a first model (Figure 3B) gives CCS values matching well with the experiment, but has lost most *WC* H-bonds, whereas a second model (Figure 3C) preserves *WC* H-bonds but its CCS values are much larger than the experimental ones (the representative structure resembles those obtained with unbiased simulations starting from the B-form).

In summary, the T-REMD results do not account simultaneously for the preservation of GC base pairs and the experimental collision cross sections. Yet they teach us that the duplex<sup>5-</sup> conformations are closer to a compact globular shape than to the helices obtained by unbiased MD trajectories, and suggest that sampling problems are at least partially responsible for the lack of agreement between simulation and experimental CCSs.



**Figure 3.** T-REMD on 12-bp DNA duplexes<sup>5-</sup> with localized charges (LC). (A) [12-d<sub>66</sub>]<sup>5-</sup>, (B-C) [12-d<sub>100</sub>]<sup>5-</sup> with different LC models. The location of negatively charged phosphates is shown by the beads on the representative snapshot structure. The experimental <sup>DT</sup>CCS<sub>He</sub> distribution and calculated one (shade) are overlaid. The fractional *WC* H-bond occupancies are shown on the right: the bases of each strand are numbered from 5' to 3', and perfect base pair matching is therefore indicated by a diagonal from top left to bottom right.

**Progressive duplex desolvation leads to experimentally relevant structures.** All simulations presented to this point, and nearly all previous simulations are assuming an instantaneous transfer of the duplex from solution to the gas phase. But in practice, desolvation and declustering proceeds gradually during the

experiment. Consta and co-workers<sup>33</sup> modeled the desolvation process of duplex dA<sub>11</sub>•dT<sub>11</sub> at atomistic level in a water droplet containing Na<sup>+</sup> and Cl<sup>-</sup>, and found that the duplex collapses inside the droplet when the Na<sup>+</sup> cations are numerous enough to interact with the phosphates and reduce the size of both grooves. Interestingly, the resulting charge density is similar to those observed experimentally from native conditions. It is therefore likely that the DNA duplexes are desolvated via the charged residue model<sup>34,35</sup> and that the compaction results from the association of NH<sub>4</sub><sup>+</sup> cations to both minor and major groove before full desolvation. As a result, the starting structure for gas phase simulations might be quite different to the canonical A- or B- helices. Because electrostatic interactions prevail, the gas phase conformational space is very stiff, and an incorrect starting structure can significantly bias the entire trajectory.

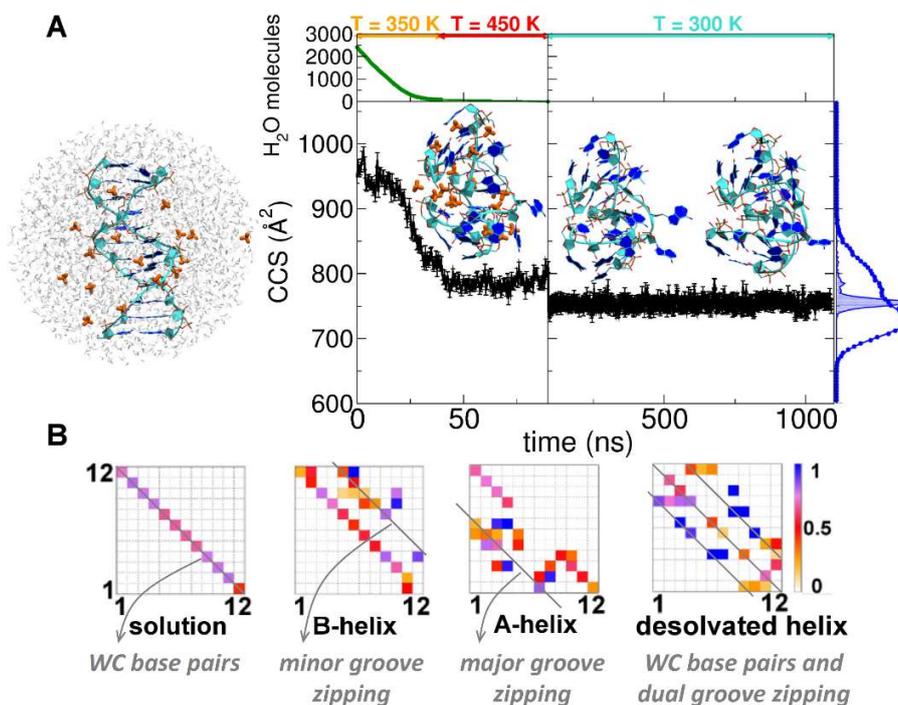
Here we simulated the desolvation of duplex 12-d<sub>66</sub> placed in a droplet containing ~2400 water molecules and 17 NH<sub>4</sub><sup>+</sup> (net charge = -5). When the simulation at 350 K reaches 39.5 ns, 87 water molecules stick on the duplex together with the 17 NH<sub>4</sub><sup>+</sup> cations. The simulation was pursued for 50 ns at 450 K, to allow this ultra-stable inner solvation shell of water molecules to evaporate. The ammonium ions sticking to the duplex are mainly located close to phosphate groups, often in-between them. Several trajectories (see 4-, 5- and 6- charge states in supporting information Section S8; Figures S26—S31) lead to a chain of ammonium ions in the minor groove. For the 5- charge state, ammonium ions are more distributed across both grooves, and thus upon evaporation both grooves get narrower to enable phosphate-ammonium-phosphate salt bridges to form. Accordingly, the CCS value diminishes (Figure 4).

Classical MD cannot model proton transfer, but if water evaporation is almost complete before the proton transfers start, then the structures generated by gradual desolvation are good candidate for modeling the electrosprayed structures. Also, ammonium ion positioning upon desolvation could predict which phosphate groups will share a proton and form hydrogen bonds after complete declustering.

To simulate the eventual ammonia loss, we arbitrarily transferred a proton from each ammonium ion to its closest phosphate oxygen. The resulting desolvated and declustered duplex (12-d<sub>66</sub>)<sup>5-</sup> is stable over 1- $\mu$ s MD at 300 K (Figure 4). The total hydrogen bond occupancy reveals additional contacts across both grooves. Remarkably, the CCS value now matches the experimental one, and at the same time, the generated structure keeps partial memory of the WC base pairs (at least, preserving mostly the GC ones), in line with CID and IRMPD results.

In summary, progressive desolvation, allowing the duplex to form phosphate-phosphate hydrogen bonds across both grooves can account simultaneously for the experimental compactness and for partial preservation of GC base pairs. In contrast, upon unbiased MD or structure optimization (DFT or SE), only

the phosphate groups that were the closest in the starting structure (across the minor and major groove for B- and A-helices, respectively) could mate.

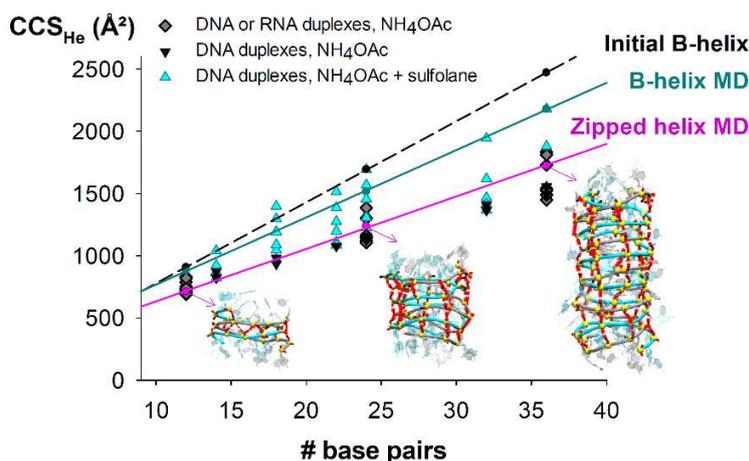


**Figure 4.** (A)  $^{CALC}CCS_{He}$  evolution during the desolvation of the 12-d<sub>66</sub> droplet with -5 net charge. At the top the simulation temperatures are shown along the related trajectory portions. The initial, intermediate and final structures are shown with DNA strands in cyan, bases aromatic rings in blue and NH<sub>4</sub><sup>+</sup> cations in orange. On the right the [12-d<sub>66</sub>]<sup>5-</sup>  $^{DT}CCS_{He}$  distribution (blue circles) is superimposed on the  $^{CALC}CCS_{He}$  (blue area) one. (B) H-bond occupancies of the desolvated helix (right), compared to Watson-Crick H-bond occupancies of a B-helix in solution (left), and to gas-phase B- or A- helices upon unbiased MD in the gas phase (middle). In the B-helix, extra H-bonds form between phosphates across the minor groove; in the A-helix, across the major groove, and in the desolvated helix, across both grooves.

**Longer duplexes, DNA and RNA alike, undergo compaction when electrosprayed from aqueous solutions of physiological ionic strength.** Past studies on GC-rich DNA duplexes, with one charge every two base pairs, had suggested that A-helices predominate from 8-bp on, and that B-helices are preserved from 18-bp on.<sup>21,22</sup> To ascertain whether the compaction of lower charge states is a general phenomenon,

we measured 12-bp to 36-bp DNA and RNA duplexes, from native solutions, either 100% aqueous or containing sulfolane. The tested duplexes included multiples of  $d_{66}$ ,  $r_{66}$ ,  $d_{100}$  and  $r_{100}$ , and the DNA duplexes listed as potential CCS calibrants by the Fabris group.<sup>36</sup> The results are shown in Figure 5 (experimental values in supporting Table S3).

We found that DNA and RNA duplexes have very similar gas-phase CCS values, although their helix types differ in solution (B vs. A). The theoretical CCS values obtained with unbiased *in-vacuo* MD directly from the solution B-helix are overlaid (B-helix MD trend line in Fig. 5). They match only with some of the high charge states produced in the presence of sulfolane. However, the low charge states produced from purely aqueous  $\text{NH}_4\text{OAc}$  solutions are significantly more compact.



**Figure 5.** Comparison between experimental and calculated  $\text{CCS}_{\text{He}}$  of 12-bp to 36-bp duplexes. DNA and RNA sequences were either derived from 12-d/ $r_{66/100}$  (diamonds, see Tables 2 and S3) or were identical to the DNA duplexes studied by Lippens et al.<sup>36</sup> (triangles, see Table S3). When measured at the charge states obtained from aqueous 100-mM  $\text{NH}_4\text{OAc}$  (black and dark grey), collision cross section values match with MD on the zipped helix (restrained minimization followed by MD). DNA duplexes at higher charge states produced by sulfolane addition adopt more extended conformations (cyan), which match better with MD simulations on B-helices.

To reproduce the dual-groove closing of longer duplexes while avoiding computationally costly desolvation simulations, we opted for biased exploration. We used restrained minimization and imposed distance constraints based on our knowledge that compact structures can be formed via hydrogen-bond formation between phosphates across both grooves. Distance restraints were imposed using a harmonic potential between the hydrogen atom of neutral phosphates belonging to one strand and the oxygen atom of mating

phosphates belonging to the other strand (Supporting Information Section S10, Figures S32—S34, and Movie S2 for a 12-bp duplex). The systems were minimized *in vacuo*, then all restraints were removed prior to 1- $\mu$ s gas phase MD.

For 12-bp duplexes, the resulting CCS distributions agree with the experiment, with the strongest compaction for 12-d<sub>100</sub> (Figure 1E). Applying the same procedure starting from an A-helix however did not lead to similarly low CCS values (supporting Figure S35). The doubly groove-zipped helices obtained by restrained minimization generally keep *WC* hydrogen bonds less well than those obtained by desolvation (supporting Figure S36—S38), but still reflect the solution trend, with the 12-d<sub>100</sub> preserving the highest fraction of hydrogen bonds. The advantage of the procedure is to reproduce the phosphate-phosphate H-bond pattern of the desolvated helices (two diagonals for zipping across both grooves, plus the central diagonal indicating preserved base pairs, Figure S36). We then applied restrained minimization to the longer (24-bp and 36-bp) helices (supporting Figures S39—S40). Whatever the duplex length, the experimental CCS values obtained for low charge states (Figure 5) match better with these zipped helices than with the canonical structures or with the helices relaxed by long unbiased gas-phase MD.

## CONCLUSIONS

In summary, at the charge states produced by electrospray from  $\geq 100$  mM aqueous ammonium acetate (traditional “native” solution conditions), double-stranded nucleic acids undergo a significant compaction in gas phase compared to the structure in solution. Unbiased molecular dynamics of B-helix or A-helix structures directly transposed from solution to the gas phase fails to reproduce the experimental results. This is due to several reasons: i) only the phosphate groups closest to one another can pair to form hydrogen bonds on the simulation time scale, ii) the starting structure is unrealistic and iii) sampling in unbiased MD simulations is intrinsically limited. In the case of T-REMD, depending on the initial choice of charge location, the final structures either did not have any memory of the solution structure, or resembled those obtained by unbiased MD. T-REMD can help to solve the sampling effect (the question of the maximum internal temperature reached in the experiments remaining open), but not the problem that original charge locations might be incorrect.

Gradual desolvation generates more realistic starting structures for gas phase simulations. Conformational transitions occurring during dehydration cannot be ignored because they guide the entire sampling, within a particularly stiff conformational landscape in the case of nucleic acids. The conformationally restrained duplexes remain stable upon unbiased MD: once formed, they stay locked at room temperature. The

broadness of the experimental CCS distributions therefore indicates a distribution of co-existing—but not interconverting—conformations, wherein each would have a slightly different phosphate-phosphate hydrogen bond network.

Our results highlight a key difference between nucleic acids and proteins native mass spectrometry. Globular proteins can rearrange by relaxing their side chains<sup>37</sup> and undergo minimal salt bridge rearrangement.<sup>38</sup> Briefly optimized structures often have CCS values matching well with the experiments.<sup>35,39-41</sup> Fabris and co-workers have recently underlined the difficulties in transposing to DNA the MD and CCS calculation protocols traditionally used for proteins.<sup>36</sup> As a way out they proposed to calibrate all traveling wave IMS data using short MD simulation results, but our study shows why this approach would lead to a misrepresentation of nucleic acid structures in the gas phase.

DNA and RNA double helices are more compact in the gas phase than in solution, due mostly to new phosphate-phosphate interactions. At the low charge states produced from ammonium acetate, the Coulomb repulsion is not sufficient to keep the phosphate groups apart. They rearrange by self-solvation, cause major rearrangements of the backbone, and lead to a significant compaction (>20%) compared to the starting structure. Yet, they are metastable conformations keeping some memory of the solution structure.

## METHODS

**Electrospray ion mobility spectrometry.** DNA and RNA duplexes were prepared by annealing their corresponding single-strands (purchased from Eurogentec, Seraing, Belgium, with RPl cartridge-Gold purification) in aqueous 100 mM NH<sub>4</sub>OAc. When sprayed at 10 μM duplex, the major charge states are 4- for the 10-bp, 5- for the 12-bp, 7- for the 24-bp, 8- and 9- for the 36-bp duplexes. Higher charge states were generated by adding 0.2% to 0.75% sulfolane to the solution. ESI-IMS-MS experiments were recorded on an Agilent 6560 IMS-Q-TOF, with the drift tube operated in helium (supporting information Section S1). The arrival time distributions were fitted by Gaussian peaks and the CCS values of the center of each peak were determined by the stepped-field method. For visualization, we converted the arrival time distributions into CCS distributions (see supporting information).

**Gas-phase simulations.** The starting structures of the duplexes were built with the Nucleic Acid Builder (NAB) software,<sup>42</sup> both for the A-form (DNA and RNA) and the B-form (DNA). Table 1 lists the main sequences and levels of theory used here. The numbers (33, 66, and 100) in duplex names reflect the GC content (in %) of each 12-bp unit. For 12-d<sub>33</sub>, 12-d<sub>66</sub>, 12-d<sub>100</sub> (B-helix) and 12-r<sub>66</sub> (A-helix), we first carried out MD simulations in water. Then all water molecules and counterions were removed at once, before each

1- $\mu$ s gas-phase simulation. The two possibilities to reduce the total charge to -5 (major charge state) are the localized charges (LC) and distributed charges (DC) models.<sup>20</sup> With the LC model, protons are added on 17 out of the 22 phosphate groups. Among the 26334 possible protonation schemes, we selected a few low-energy ones based on single point molecular mechanics calculations. With the DC model, the net charge of each phosphate group is reduced so that the total charge of the duplex is -5. Because temperature replica exchange MD (T-REMD) simulations on gas-phase duplexes had never been attempted before, LC and DC models were both tested for T-REMD.

Solution and gas phase MD simulations were carried out with the MPI-versions of modules *pmemd* and *sander*, respectively, of the Amber12 suite of programs,<sup>43</sup> implementing parmBSC1 force field<sup>44</sup> for DNA and parmBSC0 force field +  $\chi_{OL3}$  correction<sup>45,46</sup> for RNA. The electrostatic interactions were calculated with the particle mesh Ewald algorithm<sup>47</sup> (real-space cut-off = 10 Å) in solution and direct Coulomb summation (no cut-off) in gas phase. All 12-bp duplexes were subjected gas phase T-REMD<sup>48</sup> (1  $\mu$ s  $\times$  18 replicas) with temperature values from 300.00 to 633.94 K, chosen with predictor from Patriksson *et al.*<sup>49</sup> (average successful exchange rate of ca. 30%). Short duplexes (7–10 bp) were optimized at DFT level,<sup>50</sup> with the M06-2X<sup>51</sup> functional including the dispersion correction GD3.<sup>52</sup> The basis set was 6-31G(d,p) for the 7–9-bp duplexes (see Supporting Information) and 4-31G(d) for the 10-bp duplex. Duplexes up to 12-bp were also studied at the semi-empirical (SE) level with MOPAC,<sup>53</sup> using different methods<sup>54</sup> (further details in Supporting Information). Hydrogen bond and stacking analysis was performed for all simulations as detailed in the Supplementary Information section S4.

**Simulation of desolvation and proton transfer.** Starting from equilibrated MD simulations in solution, we cut droplets of ca. 2400 water molecules (radius  $\sim$  25 Å) containing the duplex 12-d<sub>66</sub>, and 16, 17 and 18 NH<sub>4</sub><sup>+</sup> cations to give a total net charge of -6, -5 and -4 respectively. The droplets were then subjected to gas-phase MD simulations following Konermann's protocol.<sup>55</sup> Briefly, the trajectories were propagated by 500-ps chunks at constant temperature (350 K). To accelerate the evaporation, at the beginning of each chunk the initial velocities were reassigned according to the Boltzmann distribution at T = 350 K. At the end of each chunk, we stripped out all water molecules farther than 60 Å from the N6 atom of the 18<sup>th</sup> residue adenine (which is approximately in the center of the duplex). A further 50-ns chunk at T = 450 K helped the last “sticky” water molecules to evaporate. In total, twelve independent trajectories were obtained (four at each charge state, 4-, 5- and 6-). We then localized the charges (LC model) as follows: on the ultimate conformation of every trajectory a proton from each NH<sub>4</sub><sup>+</sup> cation was transferred to the closest phosphate oxygen atom, and ammonia is removed. The resulting duplexes were then subjected to 1- $\mu$ s unbiased MD.

**CCS calculations.** The collision cross section (CCS) is calculated using the EHSSrot code<sup>56</sup> with the atom parameterization of Siu et al,<sup>57</sup> a combination that is both accurate and efficient for calculating the CCS of nucleic acids in the gas phase.<sup>58</sup> The CCS is calculated for snapshots every 0.5 ns in each MD trajectory.

## SUPPORTING INFORMATION

Experimental procedures including reconstruction of the CCS distributions, detailed computational procedures, ESI-IMS-MS supplementary results, circular dichroism results, full modeling results by *in vacuo* QM optimization, MD, T-REMD, and MD following restrained minimization, structural analysis, supplementary results on longer duplexes, and additional references (PDF).

Movie S1: minor groove zipping upon MD of a 12-bp B-helix (CIF)

Movie S2: dual groove zipping imposed by restrained minimization on a 12-bp B-helix (CIF).

## ACKNOWLEDGMENTS

This work was funded by the European Research Council (ERC DNAFOLDIMS to VG and ERC SimDNA to MO), by the Spanish BIO2015-69802-R grant to MO, and by EU COST action BM1403 (short-term scientific mission to MP). HG is a Juan de la Cierva researcher. LD is a SNI (Sistema Nacional de Investigadores; ANII, Uruguay) researcher, and MO is an ICREA-academia fellow. We are thankful to David E. Condon for donating to us the Perl script used to score the base stacking, and to Adrien Marchand and members of COST action BM1403 for fruitful discussions.

## REFERENCES

- (1) Blackburn, G. M.; Gait, M. J.: *Nucleic acids in chemistry and biology*; Oxford University Press: Oxford, 1996.
- (2) Rhodes, D.; Lipps, H. J.: G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **2015**, *43*, 8627-8637.
- (3) Breaker, R. R.: Prospects for riboswitch discovery and analysis. *Mol. Cell* **2011**, *43*, 867-879.
- (4) Loo, J. A.: Studying non-covalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrom. Rev.* **1997**, *16*, 1-23.

- (5) Breuker, K.: The study of protein–ligand interactions by mass spectrometry—a personal view. *Int. J. Mass Spectrom.* **2004**, *239*, 33-41.
- (6) Rosu, F.; De Pauw, E.; Gabelica, V.: Electrospray mass spectrometry to study drug-nucleic acids interactions. *Biochimie* **2008**, *90*, 1074-1087.
- (7) Kitova, E. N.; El-Hawiet, A.; Schnier, P. D.; Klassen, J. S.: Reliable determinations of protein-ligand interactions by direct ESI-MS measurements. Are we there yet? *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 431-441.
- (8) Leney, A. C.; Heck, A. J. R.: Native Mass Spectrometry: What is in the Name? *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 5-13.
- (9) Bohrer, B. C.; Merenbloom, S. I.; Koeniger, S. L.; Hilderbrand, A. E.; Clemmer, D. E.: Biomolecule analysis by ion mobility spectrometry. *Annu. Rev. Anal. Chem.* **2008**, *1*, 293-327.
- (10) Uetrecht, C.; Rose, R. J.; van, D. E.; Lorenzen, K.; Heck, A. J.: Ion mobility mass spectrometry of proteins and protein assemblies. *Chem. Soc. Rev.* **2010**, *39*, 1633-1655.
- (11) Zheng, X.; Liu, D.; Klamer, F. G.; Schrader, T.; Bitan, G.; Bowers, M. T.: Amyloid beta-protein assembly: The effect of molecular tweezers CLR01 and CLR03. *J. Phys. Chem. B* **2015**, *119*, 4831-4841.
- (12) Canon, F.; Ballivian, R.; Chirot, F.; Antoine, R.; Sarni-Manchado, P.; Lemoine, J.; Dugourd, P.: Folding of a salivary intrinsically disordered protein upon binding to tannins. *J. Am. Chem. Soc.* **2011**, *133*, 7847-7852.
- (13) Hyung, S. J.; Robinson, C. V.; Ruotolo, B. T.: Gas-phase unfolding and disassembly reveals stability differences in ligand-bound multiprotein complexes. *Chem. Biol.* **2009**, *16*, 382-390.
- (14) Ujma, J.; De Cecco, M.; Chepelin, O.; Levene, H.; Moffat, C.; Pike, S. J.; Lusby, P. J.; Barran, P. E.: Shapes of supramolecular cages by ion mobility mass spectrometry. *Chem. Commun.* **2012**, *48*, 4423-4425.
- (15) Schnier, P. D.; Klassen, J. S.; Strittmatter, E. F.; Williams, E. R.: Activation energies for dissociation of double strand oligonucleotide anions: evidence for Watson-Crick base pairing in vacuo. *J. Am. Chem. Soc.* **1998**, *120*, 9605-9613.
- (16) Gabelica, V.; De Pauw, E.: Collision-Induced Dissociation of 16-mer DNA Duplexes with Various Sequences: Evidence for Conservation of the Double Helix Conformation in the Gas Phase. *Int. J. Mass Spectrom.* **2002**, *219*, 151-159.
- (17) Pan, S.; Sun, X. J.; Lee, J. K.: Stability of complementary and mismatched DNA duplexes: Comparison and contrast in gas versus solution phases. *Int. J. Mass Spectrom.* **2006**, *253*, 238-248.
- (18) Burmistrova, A.; Gabelica, V.; Duwez, A. S.; De Pauw, E.: Ion mobility spectrometry reveals duplex DNA dissociation intermediates. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1777-1786.
- (19) Gabelica, V.; Rosu, F.: Gas-Phase Spectroscopy of Nucleic Acids. In *Nucleic Acids in the Gas Phase*; Gabelica, V., Ed.; Springer: Berlin, 2014; pp 103-130.
- (20) Rueda, M.; Kalko, S. G.; Luque, F. J.; Orozco, M.: The structure and dynamics of DNA in the gas phase. *J. Am. Chem. Soc.* **2003**, *125*, 8007-8014.

- (21) Gidden, J.; Ferzoco, A.; Baker, E. S.; Bowers, M. T.: Duplex formation and the onset of helicity in poly d(CG)<sub>n</sub> oligonucleotides in a solvent-free environment. *J. Am. Chem. Soc.* **2004**, *126*, 15132-15140.
- (22) Baker, E. S.; Bowers, M. T.: B-DNA helix stability in a solvent-free environment. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1188-1195.
- (23) Drew, H.; Wing, R.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E.: Structure of B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. USA* **1981**; *78*, 2179-2183.
- (24) Kypr, J.; Kejnovska, I.; Renciuik, D.; Vorlickova, M.: Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* **2009**, *37*, 1713-1725.
- (25) Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D.: Understanding nucleic acid-ion interactions. *Annu. Rev. Biochem.* **2014**, *83*, 813-841.
- (26) Pasi, M.; Maddocks, J. H.; Lavery, R.: Analyzing ion distributions around DNA: sequence-dependence of potassium ion distributions from microsecond molecular dynamics. *Nucleic Acids Res* **2015**, *43*, 2412-2423.
- (27) Manning, G. S.: The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.* **1978**, *11*, 179-246.
- (28) Ogorzalek Loo, R. R.; Lakshmanan, R.; Loo, J. A.: What protein charging (and supercharging) reveal about the mechanism of electrospray ionization. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1675-1693.
- (29) Humphrey, W.; Dalke, A.; Schulten, K.: VMD: visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33-38.
- (30) Zubatiuk, T.; Kukuev, M. A.; Korolyova, A. S.; Gorb, L.; Nyporko, A.; Hovorun, D.; Leszczynski, J.: Structure and Binding Energy of Double-Stranded A-DNA Mini-helices: Quantum-Chemical Study. *J. Phys. Chem. B* **2015**, *119*, 12741-12749.
- (31) Zubatiuk, T. A.; Shishkin, O. V.; Gorb, L.; Hovorun, D. M.; Leszczynski, J.: B-DNA characteristics are preserved in double stranded d(A)<sub>3</sub>.d(T)<sub>3</sub> and d(G)<sub>3</sub>.d(C)<sub>3</sub> mini-helices: conclusions from DFT/M06-2X study. *Phys. Chem. Chem. Phys.* **2013**, *15*, 18155-18166.
- (32) Condon, D. E.; Kennedy, S. D.; Mort, B. C.; Kierzek, R.; Yildirim, I.; Turner, D. H.: Stacking in RNA: NMR of Four Tetramers Benchmark Molecular Dynamics. *J. Chem. Theor. Comput.* **2015**, *11*, 2729-2742.
- (33) Sharawy, M.; Consta, S.: How do non-covalent complexes dissociate in droplets? A case study of the desolvation of dsDNA from a charged aqueous nanodrop. *Phys. Chem. Chem. Phys.* **2015**, *17*, 25550-25562.
- (34) Dole, M.; Mack, L. L.; Hines, R. L.: Molecular beams of macroions. *J. Chem. Phys.* **1968**, *49*, 2240-2249.
- (35) McAllister, R. G.; Metwally, H.; Sun, Y.; Konermann, L.: Release of Native-like Gaseous Proteins from Electrospray Droplets via the Charged Residue Mechanism: Insights from Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2015**, *137*, 12667-12676.

- (36) Lippens, J. L.; Ranganathan, S. V.; D'Esposito, R. J.; Fabris, D.: Modular calibrant sets for the structural analysis of nucleic acids by ion mobility spectrometry mass spectrometry. *Analyst* **2016**, *141*, 4084-4099.
- (37) Breuker, K.; McLafferty, F. W.: Stepwise evolution of protein native structure with electrospray into the gas phase,  $10^{-12}$  to  $10^2$  s. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 18145-18152.
- (38) Loo, R. R.; Loo, J. A.: Salt Bridge Rearrangement (SaBRE) Explains the Dissociation Behavior of Noncovalent Complexes. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 975-990.
- (39) Pacholarz, K. J.; Porrini, M.; Garlish, R. A.; Burnley, R. J.; Taylor, R. J.; Henry, A. J.; Barran, P. E.: Dynamics of intact immunoglobulin G explored by drift-tube ion-mobility mass spectrometry and molecular modeling. *Angew. Chem. Int. Ed.* **2014**, *53*, 7765-7769.
- (40) Meyer, T.; de la Cruz, X.; Orozco, M.: An atomistic view to the gas phase proteome. *Structure* **2009**, *17*, 88-95.
- (41) Harvey, S. R.; Porrini, M.; Konijnenberg, A.; Clarke, D. J.; Tyler, R. C.; Langridge-Smith, P. R.; MacPhee, C. E.; Volkman, B. F.; Barran, P. E.: Dissecting the dynamic conformations of the metamorphic protein lymphotactin. *J. Phys. Chem. B* **2014**, *118*, 12348-12359.
- (42) Macke, T.; Case, D. A.: Modeling unusual nucleic acid structures. In *Molecular Modeling of Nucleic Acids*; Leontes, N. B., SantaLucia, J., Jr., Eds.; American Chemical Society: Washington, DC, 1998; pp 379-393.
- (43) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M. J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A.: AMBER 12. University of California, San Francisco, 2012.
- (44) Ivani, I.; Dans, P. D.; Noy, A.; Perez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goni, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpi, J. L.; Gonzalez, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M.: Parmbsc1: a refined force field for DNA simulations. *Nat. Methods* **2016**, *13*, 55-58.
- (45) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M.: Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **2007**, *92*, 3817-3829.
- (46) Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E., 3rd; Jurecka, P.: Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2886-2902.
- (47) Darden, T.; York, D.; Pedersen, L.: Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089.
- (48) Sugita, Y.; Okamoto, Y.: Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Chem.* **1999**, *314*, 141-151.

(49) Patriksson, A.; van der Spoel, D.: A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2073-2077.

(50) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.: Gaussian 09. Gaussian, Inc.: Wallingford, CT, USA, 2009.

(51) Zhao, Y.; Truhlar, D. G.: The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215-241.

(52) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H.: A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J Chem Phys* **2010**, *132*, 154104.

(53) Stewart, J. J. P.: MOPAC2012. Stewart Computational Chemistry: Colorado Springs, CO, USA, 2012.

(54) Stewart, J. J.: Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **2013**, *19*, 1-32.

(55) Metwally, H.; McAllister, R. G.; Konermann, L.: Exploring the mechanism of salt-induced signal suppression in protein electrospray mass spectrometry using experiments and molecular dynamics simulations. *Anal. Chem.* **2015**, *87*, 2434-2442.

(56) Shvartsburg, A. A.; Mashkevich, S. V.; Baker, E. S.; Smith, R. D.: Optimization of algorithms for ion mobility calculations. *J. Phys. Chem. A* **2007**, *111*, 2002-2010.

(57) Siu, C. K.; Guo, Y.; Saminathan, I. S.; Hopkinson, A. C.; Siu, K. W.: Optimization of parameters used in algorithms of ion-mobility calculation for conformational analyses. *J. Phys. Chem. B* **2010**, *114*, 1204-1212.

(58) D'Atri, V.; Porrini, M.; Rosu, F.; Gabelica, V.: Linking molecular models with ion mobility experiments. Illustration with a rigid nucleic acid structure. *J. Mass. Spectrom.* **2015**, *50*, 711-726.



## **Titre : Etudes de « kissing complexes » d'ARN par spectrométrie de masse native : liaison du magnésium et spectrométrie de mobilité ionique**

**Résumé :** En plus d'être l'intermédiaire entre l'ADN et les protéines, l'ARN est impliqué dans plusieurs processus biologiques : régulation et expression des gènes (riboswitches, ARNm et ARNt) ou encore catalyse (ribozymes). La fonction de chaque ARN est liée à sa structure et à sa dynamique de repliement. Des cations tel que le magnésium se lient à l'ARN et peuvent être essentiels au bon repliement et à la stabilité de ces structures. L'obtention de détails structuraux et thermodynamiques sur l'interaction avec le magnésium a donc une grande importance dans la compréhension de la relation structure-fonction. La première partie de ce travail a consisté en la caractérisation des équilibres de liaison entre le magnésium et des motifs d'ARN modèles, appelés « kissing complexes », par spectrométrie de masse *native* (SM). Grâce à la SM, il est possible de distinguer les stœchiométries de liaison du magnésium. Le travail présenté ici a permis l'élaboration d'une méthode pour quantifier chaque espèce en prenant en compte la distribution d'adduits non-spécifiques. Afin d'aller plus loin dans la localisation du magnésium, nous avons utilisé la spectrométrie de masse en tandem (SM/SM). Nous avons également étudié le comportement des complexes d'ARN en phase gazeuse en utilisant la spectrométrie de mobilité ionique (SMI), avec pour but de détecter des changements de conformation dus à la liaison de cations ou ligands. Contrairement à ce qui était anticipé, nous avons démontré que les duplexes d'ADN et ARN ainsi que les « kissing complexes » subissaient une compaction significative en phase gazeuse aux états de charge initialement obtenus par SM *native*, ce qui pourrait cacher l'effet des cations. Notre travail a montré comment la spectrométrie de masse peut apporter de nouvelles indications sur les stœchiométries et affinités entre ARN et cations, et discute de certaines limitations quant à l'utilisation de techniques en phase gazeuse pour explorer les structures en solution.

**Mots clés :** spectrométrie de masse, mobilité ionique, complexes d'ARN, magnésium, structure en phase gazeuse

---

## **Title: Investigation of RNA kissing complexes by native electrospray mass spectrometry: magnesium binding and ion mobility spectrometry**

**Abstract:** Besides being the molecular intermediate between DNA and proteins, RNA can have many other functions such as gene regulation (riboswitches), gene expression (mRNA and tRNA) or catalysis (ribozymes). RNA function is linked to its structure and its folding dynamics. Cations such as magnesium bind to RNA and are in some instances essential for proper folding and for stability. The need of structural and thermodynamic details about Mg<sup>2+</sup> interactions is then of utmost importance in the study of the structure-function relationships. The first part of our work consists in characterizing the binding equilibria between magnesium and RNA model motifs, called kissing complexes, using *native* mass spectrometry (MS). MS makes it possible to distinguish individual binding stoichiometries, and the present work consisted in developing a method to quantify each species, taking into account the contribution of nonspecific adducts. We also explored how tandem mass spectrometry (MS/MS) could further help localizing magnesium ions. Further, we explored the structures of RNA complexes in the gas phase using ion mobility mass spectrometry (IM-MS), with the aim to detect shape changes upon cation or ligand binding. But in contrast with anticipations, we found that DNA and RNA duplexes as well as RNA kissing complexes undergo a significant compaction at charge states naturally produced by *native* ESI-MS, which may hide the effect of cations. Our work showcases how mass spectrometry can bring novel information on RNA-cation binding stoichiometries and affinities, but also discusses some limitations of a gas-phase method to probe solution structures.

**Keywords :** Mass spectrometry, ion mobility, RNA complexes, magnesium, gas-phase ion structure

---

### **Unité de recherche**

ARNA, INSERM U1212, IECB, 2 rue Robert Escarpit 33607 Pessac