



**HAL**  
open science

# DNA and RNA telomeric G-quadruplexes: what topology features can be inferred from ion mobility mass spectrometry?

Valentina d'Atri, Valérie Gabelica

## ► To cite this version:

Valentina d'Atri, Valérie Gabelica. DNA and RNA telomeric G-quadruplexes: what topology features can be inferred from ion mobility mass spectrometry?. *Analyst*, 2019, 144 (20), pp.6074-6088. <10.1039/C9AN01216H>. <hal-02358205>

**HAL Id: hal-02358205**

**<https://hal.science/hal-02358205v1>**

Submitted on 11 Nov 2019

**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.



HAL Authorization

# DNA and RNA Telomeric G-Quadruplexes: What Topology Features Can be Inferred from Ion Mobility Mass Spectrometry?

Valentina D'Atri<sup>‡</sup>, Valérie Gabelica\*

*University of Bordeaux, INSERM and CNRS, ARNA Laboratory, IECB site, 2 rue Robert Escarpit, 33600 Pessac, France.*

\* Corresponding author: [v.gabelica@iecb.u-bordeaux.fr](mailto:v.gabelica@iecb.u-bordeaux.fr)

<sup>(1)</sup> <sup>‡</sup> Current address: *School of Pharmaceutical Sciences, University of Geneva, University of Lausanne, CMU, Rue Michel Servet 1, 1211, Geneva 4, Switzerland*

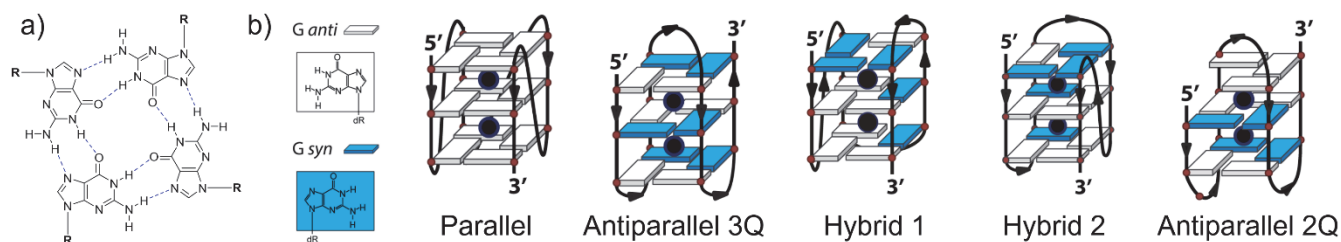
**Abstract:** Maintenance of the telomeres is key to chromosome integrity and cell proliferation. The G-quadruplex structures formed by telomeric DNA and RNA (TTAGGG and UUAGGG repeats, respectively) are key to this process. However, because these sequences are particularly polymorphic, solving high-resolution structures is not always possible, and there is a need for new methodologies to characterize the multiple structures coexisting in solution. In this context, we evaluated whether ion mobility spectrometry coupled to native mass spectrometry could help separate and assign the G-quadruplex topologies. We explored the circular dichroism spectra, multimer formation, cation binding, and ion mobility spectra of several 4-repeat and 8-repeat telomeric DNA and RNA sequences, both in  $\text{NH}_4^+$  and in  $\text{K}^+$ . In 1 mM  $\text{K}^+$  and 100 mM trimethylammonium acetate, all RNAs fold intramolecularly (no multimer). In 8-repeat sequences, the subunits are not independent: in DNA the first subunit disfavors the folding of the second one, whereas in RNA the two subunits fold cooperatively via cation-mediated stacking. Ion mobility spectrometry shows that gas-phase structures keep a memory of the solution ones, but not identical. At the native charge states, the loops can rearrange in a variety of ways (unless they are constrained by pre-formed hydrogen bonds), thereby wrapping the core and masking the strand arrangements. Our study highlights that, to progress towards structural assignment from IM-MS experiments, deeper understanding of the solution-to-gas-phase rearrangement mechanisms is warranted.

## INTRODUCTION

Guanine-rich nucleic acids can form G-quadruplex secondary structures by a special hydrogen bonding pattern between four guanines called the G-quartet (Figure 1a). Quartets stack on top of one another via cation-mediated binding.<sup>1</sup> Strands that contain at least four stretches of guanines can fold intramolecularly, and form higher-order multimers as well in vitro. Bioinformatic studies revealed a large amount of sequences filling the requirements to potentially form G-quadruplexes,<sup>2-6</sup> and evidence is accumulating that G-quadruplexes do actually form at specific functional locations in cells.<sup>7-10</sup>

One of these locations is the telomere, which consists of guanine-rich repeats (TTAGGG in mammals) with a G-rich 3' overhang. Telomeric G-quadruplexes were proposed early on as an anticancer drug target, on the premise that folding of the 3'-overhang into G-quadruplexes would hamper the action of telomerase (an enzyme that adds new repeats to the telomeres and is re-activated in 85% of cancers).<sup>11</sup> Now, a more complex picture emerges, wherein both G-quadruplexes and single-stranded forms could be recognized (although differently) by telomerase.<sup>12, 13</sup> Also, it was found that telomeric DNA is transcribed into telomeric RNA (TERRA, with UUAGGG repeats),<sup>14, 15</sup> which co-localize at the telomeres and affect telomere maintenance, both in telomerase-active and ALT (Alternative Lengthening of Telomeres) cancer cells, the other 15% of cancer cells that have alternative mechanisms of telomere elongation.<sup>16-18</sup>

Telomeric DNA and RNA still pose challenges in structural biology and drug design. Years of structural investigations by NMR and X-ray crystallography have revealed a wide variety of folding topologies (part of which are shown in Figure 1b) for DNA repeats.<sup>19-34</sup> Topologies (overall folding) are characterized by the number of quartets, the strand orientation in the guanine-rich stem, the loop arrangement, and the base stacking. These factors are inter-dependent.<sup>35, 36</sup> The strand arrangement is the most often used topology descriptor for G-quadruplex structure classification, and antiparallel, parallel and hybrid arrangements are illustrated in Figure 1b. G-quadruplex RNA topologies are often thought to be simpler, because the ribose favors the *anti* orientation for guanines, which in turn favors the formation of parallel-stranded structures (Fig. 1b).<sup>37-39</sup> Yet, telomeric RNA of four or more repeats have evaded direct high-resolution structural characterization (2-repeat bimolecular structures were solved<sup>40, 41</sup>), because despite the *anti* preference they are still polymorphic, and the underlying reasons are not yet clear.<sup>42-45</sup>



**Figure 1.** a) G-quartet hydrogen bonding mode. b) Main topologies encountered in telomeric G-quadruplexes with four GGG repeats.

Given that (1) predicting the structure that a given sequence will adopt in a given solution environment is still elusive (although reasons that forbid certain topologies emerge<sup>46</sup>), and that (2) not all sequences are amenable to high-resolution structural characterization, it is important to have alternative methods that can quickly inform us on any aspect of the topology. *Spectroscopic* techniques prove useful in that respect. The most popular method is circular dichroism (CD) spectroscopy, which informs on the predominant stacking arrangement.<sup>47, 48</sup> In right-handed G-quadruplexes, all-*anti* homo-stacking (which predominates in parallel strand arrangements) gives a positive CD peak at 264 nm and a negative peak at 245 nm, heterostacking (alternating *syn/anti* stacks, which predominate in antiparallel strand arrangements) gives a positive peak at 290 nm, a negative peak at 264 nm and a positive peak at 245 nm. Hybrid structures or mixtures of structures with both types of stacking typically give two maxima at 295 and 270 nm, and a negative peak at 245 nm.<sup>47, 48</sup>

*Spectrometric* techniques can also be used to characterize G-quadruplexes.<sup>49, 50</sup> Mass spectrometry (MS) informs us on strand stoichiometry and cation binding stoichiometry. The latter was long confined to studying G-quadruplexes with  $\text{NH}_4^+$  ions bound<sup>51</sup> (ammonium acetate being a volatile buffer frequently used in mass spectrometry of biological complexes), until we found that by using trimethylammonium acetate doped with potassium chloride, we can mass analyze G-quadruplexes with  $\text{K}^+$  ions trapped in-between the G-quartets.<sup>52</sup> The cation binding stoichiometry informs us on the number of G-quartets in the topology:  $n$  cations mean there are  $n+1$  stacked G-quartets.<sup>51-55</sup>

To obtain further information on each species separated by mass, ion mobility (IM) spectrometry can be coupled to MS (IM-MS). The ion mobility spectrometer, in a drift tube configuration, measures the ion transit time in a buffer gas when dragged by a uniform electric field. It provides an information on the ion

shape: the mobility is related to the ion's momentum transfer collision integral, often called "collision cross section" (CCS, in Å<sup>2</sup>), which measures how ions are slowed down by collisions.<sup>56-58</sup> By comparing experimental CCS values with those calculated for three-dimensional models, one can test structural hypotheses for their compatibility with the measurements.<sup>59</sup>

Early studies on a home-made ion mobility spectrometer in helium under a constant low electric field showed promising results: a memory of the solution structure of telomeric G-quadruplexes seemed preserved from the solution to the gas phase, provided that soft pre-IMS conditions were used and that inner ammonium cations were preserved.<sup>60, 61</sup> Since that time, commercial instruments with improved mobility resolving power became available<sup>62</sup> and molecular modeling force fields better adapted to nucleic acids have been released,<sup>63</sup> re-parameterization of models to calculate the CCS values have been issued,<sup>64</sup> the same experiments can be performed in K<sup>+</sup>,<sup>52</sup> and the telomeric RNA became of interest.<sup>14</sup> Here, we report the ion mobility mass spectrometry of telomeric DNA and RNA sequences (4 and 8 GGG repeats) in ammonium and in potassium, and discuss the results in light of molecular modeling. Our study provides an update on the prospects and challenges in assigning G-quadruplex structures using ion mobility spectrometry. We highlight that, although a memory of the solution structure is preserved in the gas-phase ions, the solution and gas-phase structures are not identical. Specifically, although the stacking between G-quartet and thus the strand arrangement is preserved, the shape as measured by ion mobility changes because the loops can rearrange.

## MATERIALS AND METHODS

**Sample preparation.** Lyophilized oligonucleotide sequences were purchased from Eurogentec (Seraing, Belgium) with RP-cartridge-Gold purification (DNA) or RP-HPLC purification (RNA) and used as received. The sequences used in this work are summarized in Table 1.

**Table 1.** Sequences studied herein.

Name	DNA Sequence	Name	RNA Sequence
<b>24TTA</b>	d(TTAGGG) <sub>4</sub>	<b>24UUA</b>	r(UUAGGG) <sub>4</sub>
<b>24TTG</b>	dTTGGG(TTAGGG) <sub>3</sub> A	<b>24UUG</b>	rUUGGG(UUAGGG) <sub>3</sub> A
<b>26TTA</b>	d(TTAGGG) <sub>4</sub> TT	<b>26UUA</b>	r(UUAGGG) <sub>4</sub> UU
<b>48TTA</b>	d(TTAGGG) <sub>8</sub>	<b>48UUA</b>	r(UUAGGG) <sub>8</sub>
<b>48TTG</b>	dTTGGG(TTAGGG) <sub>7</sub> A	<b>48UUG</b>	rUUGGG(UUAGGG) <sub>7</sub> A
<b>50TTA</b>	d(TTAGGG) <sub>8</sub> TT	<b>50UUA</b>	r(UUAGGG) <sub>8</sub> UU

Nuclease-free water was from Ambion (Fischer Scientific, Illkirch, France). Ammonium acetate (NH<sub>4</sub>OAc, BioUltra ~ 5M, Fluka), trimethylammonium acetate (TMAA, Ultra for UPLC ~ 1M, Fluka), and potassium (KCl, > 99.999%, Sigma) were purchased from Sigma-Aldrich (Saint-Quentin Fallavier, France). Depending on the starting material, single strand stock solution were prepared at 800 μM (all DNA sequences), 600 μM (48UUA), 450 μM (48UUG), 400 μM (50UUA), and 350 μM (24UUG, 24UUA, and 26UUA) in water, and stored at -20°C. The stock concentrations were determined by UV absorbance at 260 nm, with extinction coefficients provided by the manufacturer (RNA sequences) or calculated using Cavaluzzi-Borer Correction<sup>65</sup> (DNA sequences). The samples were prepared at 10 μM single strand in either 100 mM NH<sub>4</sub>OAc or 100 mM TMAA + 1 mM KCl, and the corresponding quadruplexes were formed by annealing at either 90°C (DNA sequences) or 80°C (RNA sequences) for 5 min and cooling down to room temperature over ~3 hours. Samples were stored at 4°C for at least 24 h before measurements.

**Circular Dichroism.** Circular dichroism (CD) spectra were recorded on a Jasco J-815 spectropolarimeter equipped with a Peltier Jasco PFD-425S/15, using a 2-mm path length quartz cuvette. The annealed samples had a single strand concentration of 10 μM for the DNA sequences and 5 μM for

the RNA sequences. The CD profiles were obtained by summing three scans, acquired at 20°C with a scan speed of 50 nm/min and a 2 nm bandwidth. The CD data were normalized to molar circular-dichroic absorption ( $\Delta\epsilon$ ), based on DNA/RNA concentration, after subtraction of the background buffer spectrum.

**Native Electrospray Ion Mobility Mass Spectrometry.** Native ESI IM-MS experiments were performed on an Agilent 6560 drift tube IMS-Q-TOF spectrometer (Agilent Technologies, Santa Clara, CA). Data analysis was performed by using the IM-MS Browser software version B.06.01. The instrument was used with the dual-ESI source operated in the negative ion mode and the syringe pump flow rate was 180  $\mu\text{L/h}$ . Experiments were carried out in helium drift gas using the “Alternate Gas Option”. The drift tube length is 78.1 cm and the collision gas pressure was 3.89 Torr. An in-house modification to the pumping system allows faster equilibration of the pressures: a second Tri-scroll 800 pump (Agilent) was connected to the source region (with an Edwards SP16K connected to the front pumping line), while the original Tri-scroll 800 pump is connected to the Q-TOF region. To avoid nitrogen contaminations in the drift tube, the trapping funnel pressure was always set  $>0.20$  Torr (typically, 0.22 Torr) lower than that of the drift tube. The full list of instrument settings is provided in supporting information Table S1. In a previous paper we analyzed the effect of all pre-IMS tuning parameters on the collisional activation of the ions.<sup>66</sup> Here, each experiment was acquired with two different fragmentor voltages (350V and 600V), to investigate how samples were affected by different pre-IMS activation conditions. Furthermore, the 8-repeat RNAs required higher pre-IMS activation conditions than the other samples, notably with higher funnel RF values (see Table S1). The trapping time was set at 600  $\mu\text{s}$ .

As the ions exit the drift tube, they are analyzed by TOFMS and sorted as a function of their arrival time on the TOF pusher, yielding an arrival time distribution (ATD). The arrival time ( $t_A$ ) of the ions is the sum of the time ions spent in the drift tube ( $t_d$ ) and the time ions spent outside the drift tube and before detection ( $t_0$ ). Knowing that  $v_d = L/t_d$  (where  $L$  is the drift tube length), and that the electric field is  $E = \Delta V/L$  (where  $\Delta V$  is the voltage difference applied across the drift tube), the arrival time ( $t_A$ ) can be expressed as follows:

$$t_A = t_0 + \frac{L^2 T_0 p}{K_0 p_0 T} \cdot \left( \frac{1}{\Delta V} \right) \quad (1)$$

with  $p_0 = 760$  Torr and  $T_0 = 273.15$  K. In our experiments,  $p = 3.89$  Torr,  $T = 296$  K, and  $\Delta V$  was varied between 390 V and 790 V. By acquiring a voltage-dependent experiment (five different drift voltages

were applied, ramping from 390 V to 790 V), a linear plot of the  $t_A$  of each peak center versus  $1/\Delta V$  is obtained. The reduced mobility  $K_0$ , obtained from the slope, is then converted to CCS using Equation (2):

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

with  $z$  the ion net charge,  $\mu$  is the reduced mass of the ion-gas pair (the gas being helium),  $e$  is the electron charge and  $k_B$  is the Boltzmann constant. To reconstruct CCS distributions from arrival time distributions, the  $t_A$  of the peak center of the most intense peak recorded at  $\Delta V = 390$  V and its corresponding cross section CCS were used to determine the conversion factor  $a$ , according to Equation (3). Then, this conversion factor was used to transform the  $t_A$  values into CCS values for the entire arrival time distribution recorded at 390 V.<sup>67</sup>

$$CCS = a \cdot \frac{z}{\sqrt{\mu}} \times t_A \quad (3)$$

**Molecular Modelling.** The structures of hybrid-1 and hybrid-2 DNA quadruplex units were generated by using the coordinates of the reported NMR structures (PDB code 2GKU and 2JPZ, for 24TTG<sup>68</sup> and 26TTA<sup>28</sup>, respectively). The reported 26mer dAAAGGG(TTAGGG)<sub>3</sub>AA (hybrid-1, PDB code 2HY9<sup>27</sup>) was modified to obtain a starting structure for 24TTA. Coordinate files for DNA higher-order DNA G-quadruplex structures<sup>69</sup> containing two quadruplex units were provided by John O. Trent (University of Louisville, Kentucky).

The structures of parallel RNA quadruplex units were generated by modifying the coordinates of the reported 12mer r(UAGGGUUAGGGU)<sub>2</sub> X-ray structure (PDB code 3IBK).<sup>41</sup> Coordinate files for RNA higher-order G-quadruplex structures containing two parallel quadruplex units were generated by manually linking the 3'-end of one quadruplex unit with the 5'-end of the other quadruplex unit, and by adding a type-8 UUA propeller loop (from PDB structure 3MIJ).<sup>70</sup> Accelrys Discovery Studio was used to modify the NMR and X-ray structures, and to intercalate one K<sup>+</sup> cation in-between each pair of G-tetrads in the models.

Except for the DNA higher-order G-quadruplex structures which had already been subjected to solution MD,<sup>69</sup> all the structures were relaxed independently by performing a 100 ns MD run in explicit water. The molecular dynamics simulations were carried out using the AMBER12 suite of programs<sup>71</sup> and applying the *parmbsc0* force field<sup>63</sup> for the DNA strands and the *parmbsc0* with the *parmχOL3* modification<sup>72</sup> for the RNA strands. The potassium cation parameters were those derived by Cheatham III

and co-workers.<sup>73</sup> The models were solvated in a 10 Å truncated octahedron box of TIP3P water molecules,<sup>74</sup> and additional potassium counter-ions were added to neutralize the negative charge of the quadruplexes. The systems were optimized and equilibrated for 300 ps through a gradual heating and the removal of the positional restraints by applying a multistep protocol. First, a minimization step was performed in five rounds: in the first one, 500 steps of steepest descent followed by 1000 steps of conjugate gradient were performed, applying an harmonic restraints of 5 kcal/mol.Å<sup>2</sup> on the solute; in the following rounds the harmonic restraints were decreased to 2.5, 1.0, and 0.5 kcal/mol.Å<sup>2</sup> respectively, and in the final round were totally removed. Then, the equilibration step was performed in five 20-ps rounds of restricted MD simulations, followed by a final round of 200-ps of unrestricted MD simulation. In the first 20-ps round, the system was kept at constant volume and gently heated to 300 K, while the DNA was weakly restrained by a 10 kcal/mol.Å<sup>2</sup> harmonic potential. The following four 20-ps rounds were performed at isothermic isobaric ensemble ( $p = 1$  atm,  $T = 300$  K), with the DNA constraints gradually decreasing from 10 to 5, 2.5, and 1 kcal/mol.Å<sup>2</sup>, until no constraints were applied to the final 200-ps equilibration round. After the equilibration phase, a 100-ns unrestrained MD production was performed with *pmemd* module of AMBER12. Periodic boundary conditions and Particle Mesh Ewald (PME)<sup>75</sup> method were used. Bonds involving hydrogens were constrained by the SHAKE<sup>76</sup> algorithm, while the radial cut-off for the non-bonded interaction was set to 8 Å in both equilibration and production runs. An integration time step of 1 fs and 2 fs was used in the equilibration and production runs, respectively.

From the MD simulations in aqueous solution, the structures with the lowest potential energy were selected to create the starting structures for the gas phase MD runs. After the removal of water molecules and external counter-ions, the “distributed charge model” described by Rueda et al.<sup>77</sup> was applied to equate the charge state of the models to the one identified for the major charge state species in the corresponding mass spectra. Previous studies on duplexes<sup>77</sup> and G-quadruplex cores<sup>78</sup> showed that localized charges and distributed charge models gave similar structures and energetics. Here we chose the distributed charge model for its simplicity. All structures were simulated independently, with a 500-ns MD run in gas phase. The structures were relaxed with a minimization step performed in two rounds: in the first one, 500 steps of steepest descent followed by 2000 steps of conjugate gradient were performed and an harmonic restraints of 1 kcal/mol.Å<sup>2</sup> was applied on the guanines involved in the G-tetrads formation during this step; in the second round, the harmonic restraints were totally removed. A thermalization step was then performed in two 200 ps rounds of restricted MD simulations, followed by a final round of 200 ps of unrestricted MD simulation. In the first 200 ps round, the system was gently heated to 300 K, while the

guanines involved in the tetrads formation were weakly restrained by a 1 kcal/mol.Å<sup>2</sup> harmonic potential. The second 200 ps round was performed in canonical ensemble (at T = 300 K), with the same constraints on the G-tetrads, until no constraints were applied to the final 200 ps equilibration round. The equilibration step was then followed by a 500-ns unrestrained MD production run, using the *sander* module of Amber12, and applying the *parmbsc0* force field. The temperature was kept constant at 300 K using a weak-coupling algorithm.<sup>79</sup> The integration time step was 1.0 fs and no radial cut-off for the non-bonded forces were applied. SHAKE algorithm was used to constrain all the bonds involving hydrogen atoms. The final trajectories (solution and gas phase), resulting by collecting frames every 2 ps, were analyzed with *cpptraj* module<sup>80</sup> in AmberTools16<sup>81</sup> and visualized using VMD.<sup>82</sup>

From the gas phase MD simulation runs, snapshots were collected each 2.5 ns in order to calculate the theoretical CCS, for a total of 200 candidate geometries per each sequence. The theoretical CCS (<sup>EHSS</sup>CCS<sub>He</sub>) were calculated applying the exact hard sphere scattering (EHSS) method,<sup>83</sup> by using the *EHSSrot* algorithm<sup>84</sup> with the optimized helium parameters proposed by Siu and coworkers.<sup>64</sup> This algorithm was selected as the most suitable for theoretical CCS calculation of nucleic acids,<sup>59</sup> as tested on the [dTGGGGT]<sub>4</sub> parallel G-quadruplex structure of 24 bases.

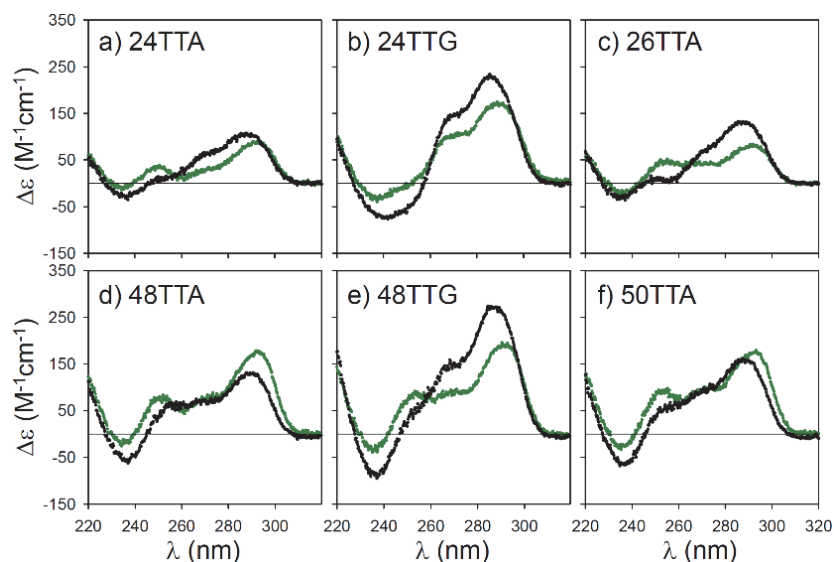
## RESULTS AND DISCUSSION

### Circular Dichroism

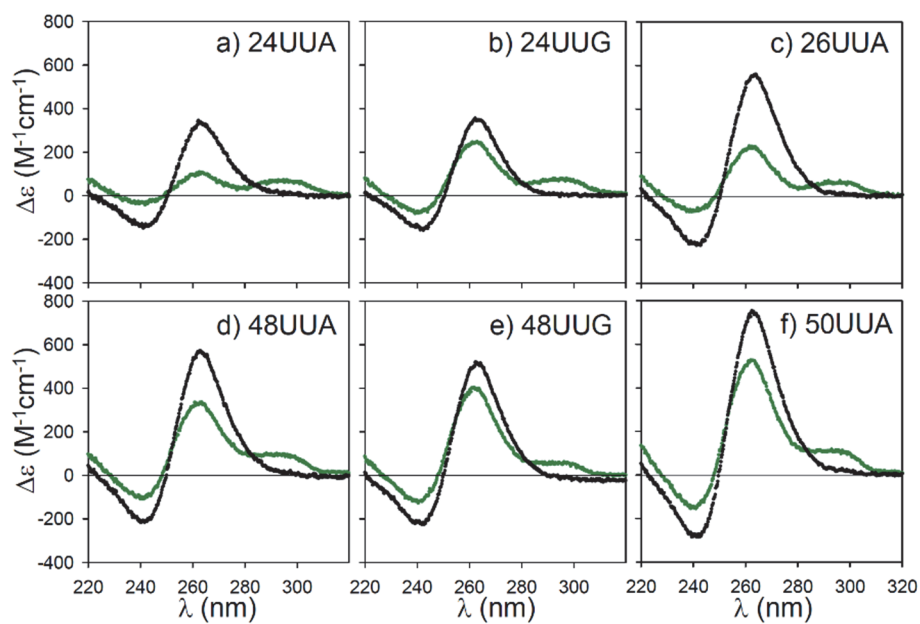
CD spectra of all DNA sequences indicate the formation of hybrid G-quadruplex structures or mixtures, whether they were annealed in 100 mM NH<sub>4</sub>OAc (Figure 2, green lines) or in 100 mM TMAA/1 mM KCl (Figure 2, black lines). The sequence 24TTG has the highest CD signal per nucleotide, suggesting it is the best folded. The spectral shape is similar in both buffers. For 24TTA and 26TTA, the spectral shapes differ more significantly between ammonium and potassium conditions. Among the 8-repeat sequences, 48TTG has the highest CD signal.

Concerning the RNA sequences, the CD profiles show that there is only all-*anti* homostacking in potassium (Figure 3, black lines), and mainly homostacking in ammonium (Fig. 3, green lines). In ammonium acetate, however, another band appears at ~300 nm. This could be due either to a fraction of

G-quartets flipping to *syn*, a phenomenon that can happen at the 5' end of purely parallel structures,<sup>44, 85</sup> or to another type of stacking involved in a dimerization interface between two parallel-stranded G-quadruplex subunits,<sup>86, 87</sup> as observed already previously on telomeric RNA sequences.<sup>88</sup>



**Figure 2.** CD spectra of annealed (a) 24TTA, (b) 24TTG, (c) 26TTA, (d) 48TTA, (e) 48TTG, and (f) 50TTA sequences in 100 mM ammonium acetate (green lines) and 100 mM tri-methyl-ammonium acetate + 1 mM potassium chloride (black lines).



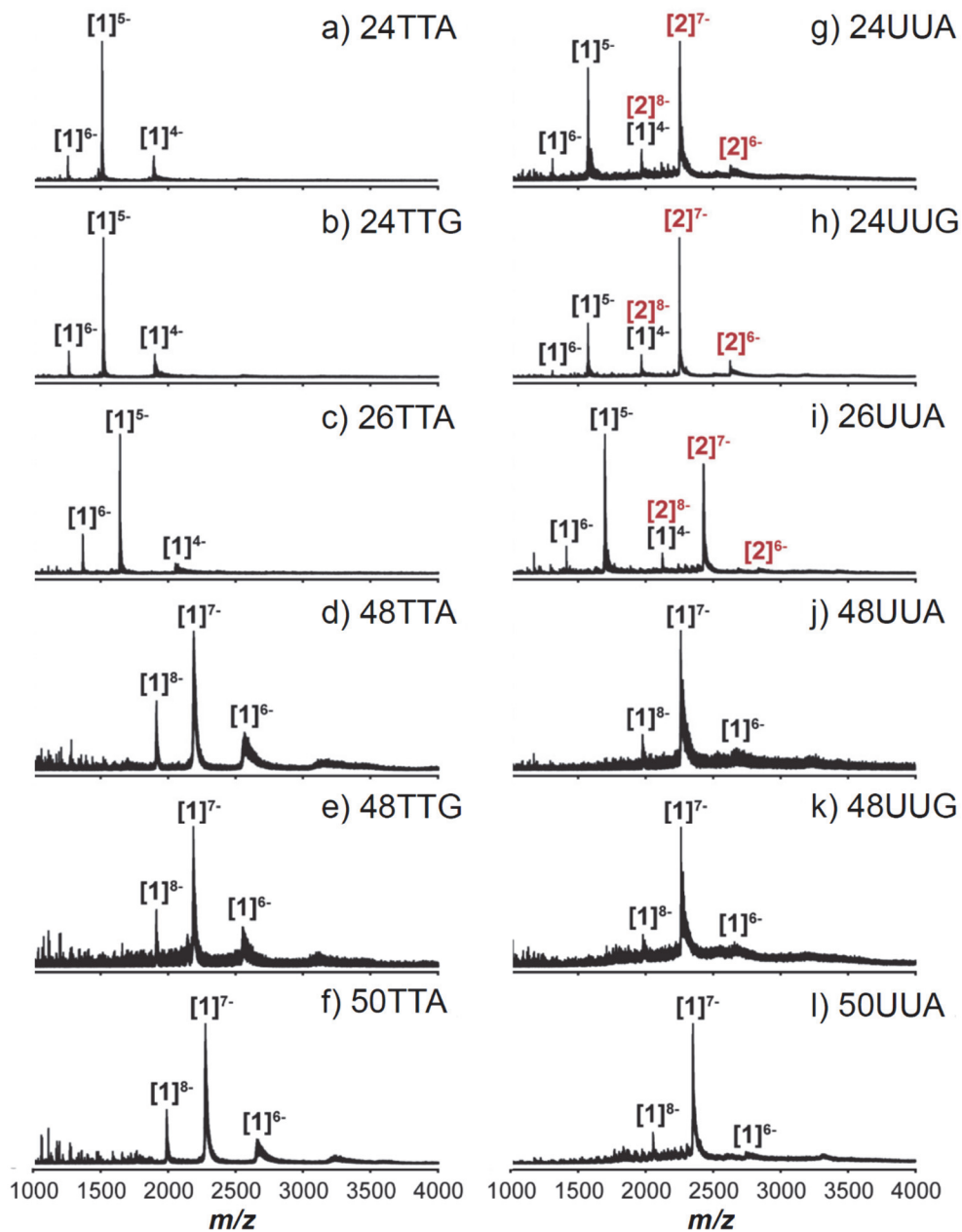
**Figure 3.** CD spectra of annealed (a) 24UUA, (b) 24UUG, (c) 26UUA, (d) 48UUA, (e) 48UUG, and (f) 50UUA sequences in 100 mM ammonium acetate (green lines) and 100 mM tri-methyl-ammonium acetate + 1 mM potassium chloride (black lines).

## Native Mass Spectrometry

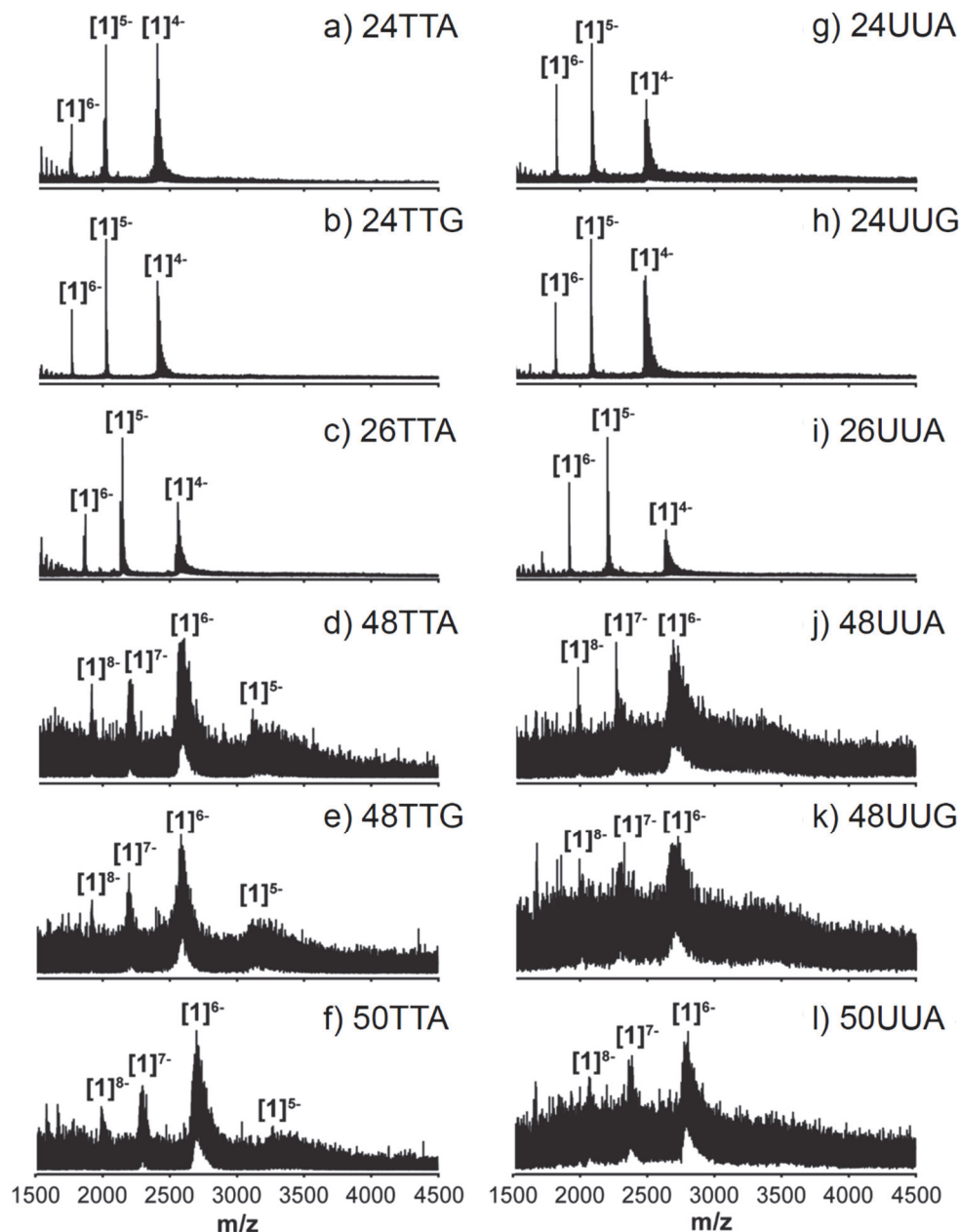
### 1) *Strand stoichiometry: intramolecular vs. intermolecular folds*

In NH<sub>4</sub>OAc, all DNA sequences (Figure 4, a–f) and the 8-repeat RNA sequences (Figure 4, panels j–l) were detected as monomers. In contrast, 4-repeat RNA sequences formed both monomeric and dimeric G-quadruplex structures (Figure 4, panels g–i), in line with previous findings.<sup>88</sup> In TMAA/KCl, all DNA and RNA sequences were detected as monomers (Figure 5). Dimerization depends both on cation concentration and cation size. The ionic strength is the same in our two conditions; but NH<sub>4</sub>OAc provides 100 mM of small cations (NH<sub>4</sub><sup>+</sup>) whereas the TMAA/KCl conditions provide only 1 mM of K<sup>+</sup>, TMAA being too bulky and less apt to screen repulsion between the subunits. Supplementary experiments on the sequence 24UUA in TMAA and [NH<sub>4</sub>OAc] at either 10 mM or 100 mM (supplemental Figure S1) show that the concentration of NH<sub>4</sub><sup>+</sup> influences the monomer/dimer equilibrium, without altering the cation uptake of each form. We deduce that high concentrations of small monovalent cations assist the stacking between subunits, due mainly to electrostatic effects.<sup>89, 90</sup> A previous study showed that stacking of two subunits was fast, whereas subunit rearrangement with cation-mediated stacking was slow.<sup>88</sup> Also, the dimers were not significantly converted to monomers under harsher pre-IMS conditions, suggesting that they are stabilized by numerous attractive electrostatic interactions.

TMAA/KCl conditions are thus particularly favorable for studying *intramolecular* G-quadruplex folding by mass spectrometry, avoiding the higher-order multimers that often form in NH<sub>4</sub>OAc.<sup>91</sup> However, the signal-to-noise ratio is particularly low for the longer 8-repeat intramolecular RNAs, and declustering is incomplete (lots of solvent/cation adducts remain even at 600 V). The difference with DNA sequences of the same length is significant, thus RNA G-quadruplexes seem to form tighter interactions with solvent molecules and counter-ions than the corresponding DNA, as suggested previously.<sup>45</sup>



**Figure 4.** Negative mode ESI mass spectra of annealed telomeric sequences in 100 mM ammonium acetate: (a) 24TTA, (b) 24TTG, (c) 26TTA, (d) 48TTA, (e) 48TTG, (f) 50TTA, (g) 24UUA, (h) 24UUG, (i) 26UUA, (j) 48UUA, (k) 48UUG, and (l) 50UUA. Final strand concentration is 10  $\mu$ M. In peak annotation  $[n]^{z-}$ ,  $n$  stands the number of strands and  $z-$  is the charge state.



**Figure 5.** Negative mode ESI mass spectra of annealed telomeric sequences in 100 mM tri-methyl-ammonium acetate doped with 1 mM potassium chloride: (a) 24TTA, (b) 24TTG, (c) 26TTA, (d) 48TTA, (e) 48TTG, (f) 50TTA, (g) 24UUA, (h) 24UUG, (i) 26UUA, (j) 48UUA, (k) 48UUG, and (l) 50UUA. Final strand concentration is 10  $\mu$ M. In peak annotation  $[n]^{z-}$ ,  $n$  stands the number of strands and  $z-$  is the charge state.

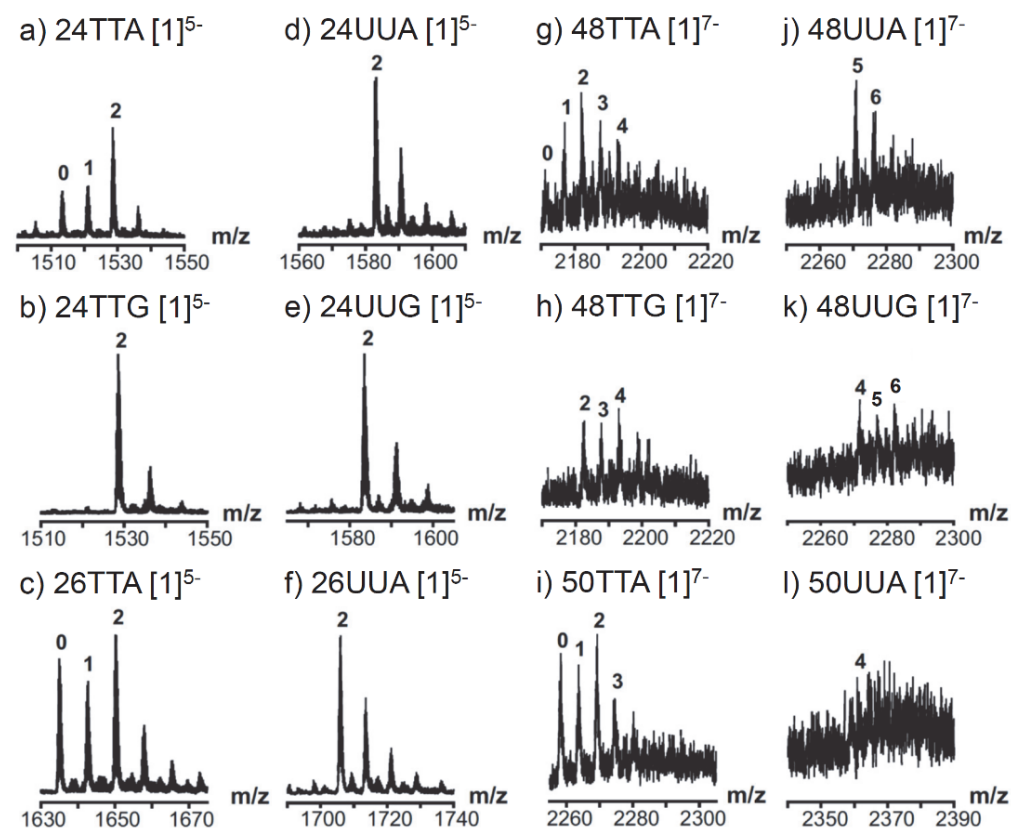
## 2) Cation binding stoichiometry and number of tetrads

4-repeat human telomeric sequences can fold into G-quadruplex units containing maximum three G-quartets, and for such structure two cations should be trapped in-between the quartets. 8-repeat sequences can form up to two G-quadruplex subunits of three G-quartets each, and at least four cations are expected to be trapped in the structure if both subunits are folded. If the subunits are stacked in a cation-mediated manner, then five cations would be observed.<sup>88</sup> The number of cations preserved by the G-quadruplex units is determined from the mass of each detected ion. In addition, two different pre-IM-MS energetic collisions (fragmentor at 350V or 600V) were tested for how they affect the global declustering from solvent and external counterions and the preservation of labile specific cations. Upon collisional activation,  $\text{NH}_4^+$  adducts can be lost in the form of  $\text{NH}_3$  (involving a proton exchange with the backbone, the precise mechanism being unknown) and trimethylammonium can be lost as trimethylamine, whereas  $\text{K}^+$  ions cannot be lost from the polyanions.<sup>51</sup> Thus, in TMAA/KCl, we don't expect a stoichiometry change in terms of bound  $\text{K}^+$ ,<sup>52</sup> but upon activation, the signal/noise ratio improves thanks to losses of neutral solvent molecules and/or losses of neutral ionic clusters. In the main text we will thus show the TMAA/KCl spectra at 600 V and the  $\text{NH}_4\text{OAc}$  spectra at 350 V. The other conditions are shown in supporting information Figures S2-S3.

Figure 6 shows the  $\text{K}^+$  adduct distributions on the 4-repeat DNAs (a–c), 4-repeat RNAs (d–f), 8-repeat DNAs (g–i) and 8-repeat RNAs (j–l). As validated previously for 4-repeat telomeric DNAs,<sup>54</sup> the  $\text{K}^+$  adducts distribution indicates the equilibrium concentrations of unfolded form (0  $\text{K}^+$  adducts), 2-tetrad forms (1  $\text{K}^+$  adducts) and 3-tetrad forms (2  $\text{K}^+$  adducts plus nonspecific adducts). The 24TTG sequence is fully folded into a 3-tetrad G-quadruplex, hybrid according to CD (Figure 3) and NMR (supporting information of reference<sup>92</sup>). However, 24TTA and 26TTA are not fully folded, and contain some 2-tetrad G-quadruplexes. In contrast, all 4-repeat RNA sequences are fully folded into exclusively 3-tetrad (2  $\text{K}^+$ ) G-quadruplex structures, which according to CD have homostacking and thus parallel-strand arrangements.

We then compare the  $\text{K}^+$  adduct distributions of the 8-repeat sequences with their 4-repeat counterparts in TMAA/KCl. In DNA 8-repeat sequences, the two subunits are not fully formed. The adducts distribution starts at 2  $\text{K}^+$  for 48TTA, and at 0  $\text{K}^+$  for 48TAA and 50TAA, like in their 4-repeat counterparts. The results of 48TTG vs. 24TTG clearly show that, while the first subunit can fully form in those conditions, the second subunit has a lower equilibrium constant of formation. The folding of

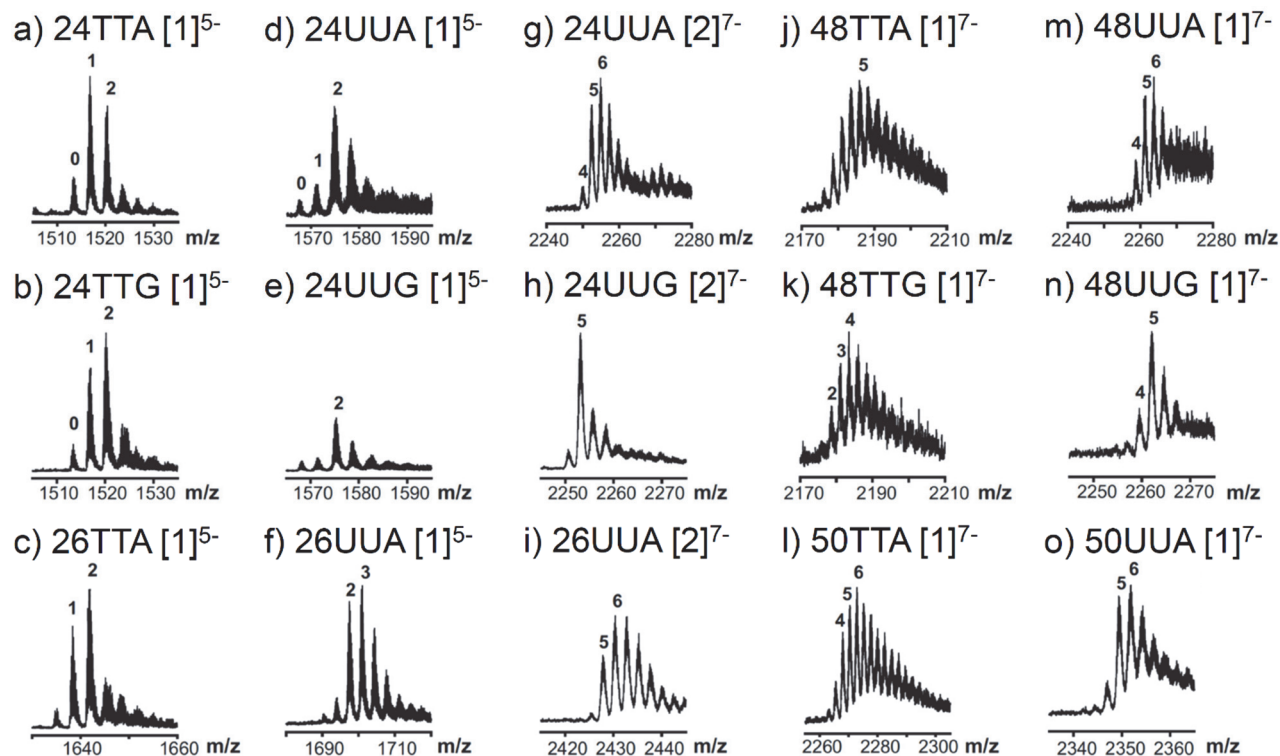
consecutive DNA G-quadruplex subunits is either non-cooperative or anti-cooperative. For RNAs, the number of specific adducts is harder to distinguish because of the poor signal-to-noise ratio. In the clearest example (48UUA, Figure 6j) the major peak contains 5  $K^+$ . Thus, two G-quadruplex subunits are formed, and they interact by sandwiching an extra cation in-between them. This involves the stacking of tetrads of the different subunits, as reported before for 12-mer telomeric RNAs.<sup>93</sup> For 48UUG, the first peak of the adduct distribution has only 4  $K^+$ , indicating two subunits that are not necessarily stacked.



**Figure 6.** Zoom of the ESI-IM-MS spectra (fragmentor: 600 V) showing the distribution of the number of potassium ions preserved in the intramolecular structures of (a) 24TTA, (b) 24TTG, (c) 26TTA, (d) 24UUA, (e) 24UUG, (f) 26UUA, (g) 48TTA, (h) 48TTG, (i) 50TTA, (j) 48UUA, (k) 48UUG, and (l) 50UUA, where  $[n]^{z-}$  refers to the number of strands and the charge state ( $n$  and  $z$ -, respectively), while the number of potassium ions is indicated on each peak. Distributions at 350 V are shown in supplemental Figure S2.

Figure 7 shows the ammonium ion distributions obtained in soft pre-IMS conditions (350 V) for the major charge states of the 4-repeats DNAs (a–c), 4-repeats RNA (d–f), bimolecular assemblies of 4-repeat DNA (g–i), and intramolecular assemblies of 8-repeat DNA (j–l) and RNA (m–o). For all DNAs, the harsher conditions led to the loss of all inner  $\text{NH}_4^+$  ions, but for RNAs, some of the inner cations are still preserved in those conditions (supplemental Figure S3). To interpret the ammonium adducts distribution in terms of G-quartet forms in solution, we will thus examine only the distributions observed in the soft conditions (Fig. 7). For the structures containing four repeats, in the case of DNA the main peaks correspond to 1 and 2  $\text{NH}_4^+$  adducts, whereas for RNA the main peak corresponds to 2  $\text{NH}_4^+$  adducts. However, in contrast to the TMAA/KCl conditions, with ammonium acetate it is impossible to deduce whether the 1- $\text{NH}_4^+$  adduct peak is due to significant amounts of 2-tetrad species in solution, or to loss of one  $\text{NH}_3$  from a 3-tetrad species.<sup>51</sup> We observed before that antiparallel-stranded G-quadruplexes are more prone to lose ammonia in the gas phase than parallel ones.<sup>94</sup> A hypothesis is that this may be due primarily to the availability of diagonal or lateral loops that can provide proton exchange sites to the departing ammonium ions. This mechanism can be at stake in hybrid-stranded DNA structures as well.

Let's examine now the structures containing 8-repeats in ammonium acetate, either for strands containing 8 repeats, or for dimers of the 4-repeat RNA sequences. For all RNAs, the first peak of the adducts distribution is the 4- $\text{NH}_4^+$  adduct, but the 5- $\text{NH}_4^+$  adduct peak has even greater intensity. As reported previously, all annealed sequences in 100 mM  $\text{NH}_4\text{OAc}$  form 5-cation species, where the two subunits are likely stacked on one another in a cation-mediated fashion. In the case of 8-repeat DNAs, the  $\text{NH}_4^+$  binding stoichiometries are less clear-cut than for RNAs: they resemble a hump starting at 2  $\text{NH}_4^+$ . The 4- $\text{NH}_4^+$  or 5- $\text{NH}_4^+$  stoichiometries do not particularly stand out. We interpret this as indicating that one subunit per 8-repeat is certainly formed in all cases, but the two G-quadruplex subunits are not formed in all cases. Similar results were obtained in TMAA/KCl. Also, this is in line with the fact that the CD signals (molar ellipticities) of 8-repeat DNAs are less than twice those of the 4-repeat DNAs.



**Figure 7.** Zoom of the ESI-IM-MS (fragmentor: 350 V) showing the distribution of the number of ammonium ions preserved in the structure of (a) 24TTA, (b) 24TTG, (c) 26TTA, (d) intramolecular 24UUA, (e) intramolecular 24UUG, (f) intramolecular 26UUA, (g) bimolecular 24UUA, (h) bimolecular 24UUG, (i) bimolecular 26UUA, (j) 48TTA, (k) 48TTG, (l) 50TTA, (m) 48UUA, (n) 48UUG, and (o) 50UUA. Annotation  $[n]^{z-}$  stands for species of molecularity  $n$  at the  $z$ - charge state, and the number of ammonium ions are indicated on each peak. Distributions at 600 V are shown in supplemental Figure S3.

## **Ion Mobility Mass Spectrometry: experimental collision cross section distributions and comparison with calculated values**

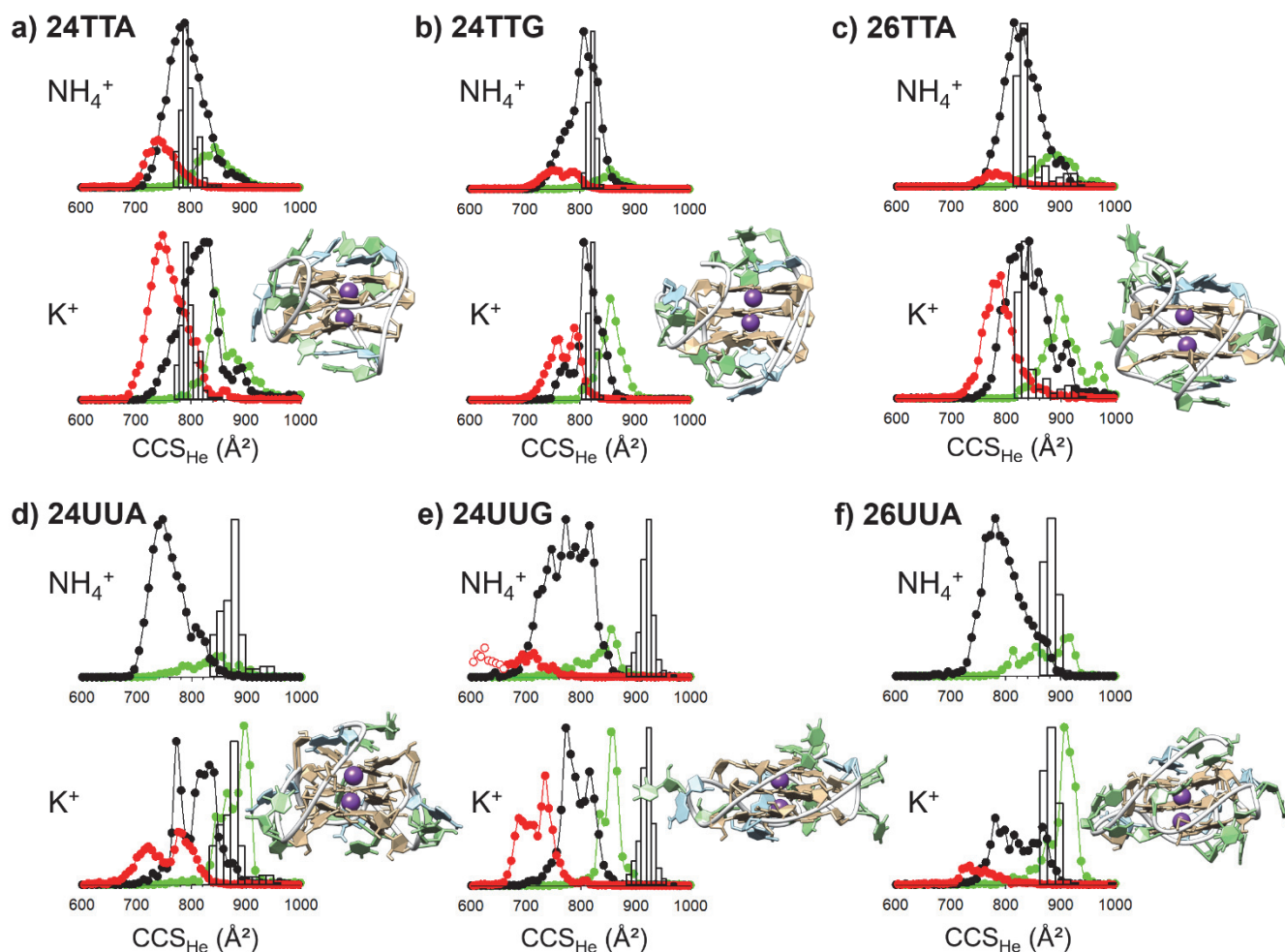
### ***1) 4-repeat DNA sequences***

Figure 8 shows the reconstructed collision cross section distributions (CCSD) of the 1-subunit G-quadruplexes formed in  $\text{NH}_4\text{OAc}$  ( $\text{NH}_4^+$ ) and in  $\text{TMAA/KCl}$  ( $\text{K}^+$ ). We selected the  $m/z$  of the complexes with two cations bound. The main charge state is 5- in all cases (black), and the CCSDs for the 4- charge state (red) and 6- charge state (green) are also shown. The average CCS decreases when the charge state decreases. The Coulomb repulsion between like charges is responsible for this phenomenon, but two scenarios are possible: 1) the lowest charge state is the native state (closest to the solution) and higher charge states are more extended due to Coulomb repulsion causing rupture of native non-covalent interactions, or 2) the structures at lower charge states are deformed compared to the native structures, because insufficient Coulomb repulsion gives rise to new, non-native non-covalent interactions that render the structure more compact. In the case of DNA and RNA double helices, the second scenario applies:<sup>95</sup> Watson-Crick double helices become more compact in the gas phase, at (low) charge states such as produced by spraying from 100 mM  $\text{NH}_4\text{OAc}$ , due to the formation of non-native hydrogen bonds between charged and protonated (neutralized) phosphate groups.

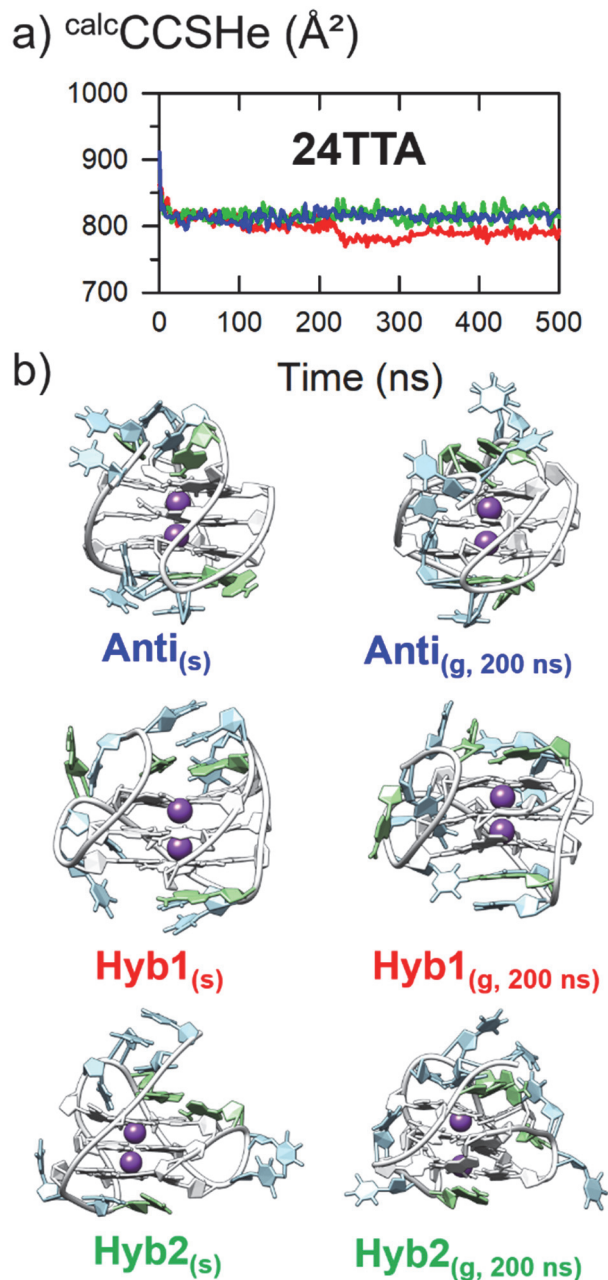
We thus need calculations of starting structures and gas-phase structures to assess whether this phenomenon is also at stake in G-quadruplexes. The gas-phase molecular dynamics simulations were run on the 5- charge states. The structures in the gas phase get more compact than the initial structures in solution, in most cases within the first 20 ns of simulation (Figure 9 for 24TTA and supplemental Figure S4 for 24TTG and 26TTA). For the DNA sequences, the CCS values of the resulting structures match with the experiments, suggesting that the 5- ions are more compact in the gas phase compared to the solution. When examining the 3D structures, we can see that mainly the loops and overhangs rearrange, while the G-quartet core structure is maintained. However, the gas-phase collision cross sections do not allow us to assign the structures to antiparallel, hybrid 1 or hybrid 2 (Figure 9). All structures have their core G-quartets intact, their loops compacting around it, and the resulting structures are equally compact. These results clearly prompt us to moderate the claims made a decade ago regarding the potential use of ion mobility spectrometry to *assign* solution structures.<sup>60</sup> We also note that the cation effects are not particularly remarkable, despite the fact that CD signals differ. The rearrangements thus probably mask most of the cation effects on the solution conformational ensembles.

Even though the gas-phase structures are not identical to the solution ones, they keep a memory of the solution. The CCS values of the G-quadruplexes which remain folded in the gas phase differ significantly from those of the non-folded forms of the same sequences. However, the difference goes in a counterintuitive direction for the 5- ions: non-folded forms are more compact than the G-quadruplex ones. This is shown in Figure 10a for 24TTA in  $K^+$ : the structures with two potassium ions bound produced in gentle pre-IMS conditions (350 V) have an average CCS of 815 Å<sup>2</sup>. Those with no potassium bound are centered on 771 Å<sup>2</sup>. In harsh pre-IMS conditions (600 V), all structures transitioned to the compact form. These observations also support the gas-phase compaction upon rearrangements because 5 negative charges over 23 phosphate groups is not providing enough Coulomb repulsion: the molecules prefer to form new non-covalent bonds instead.

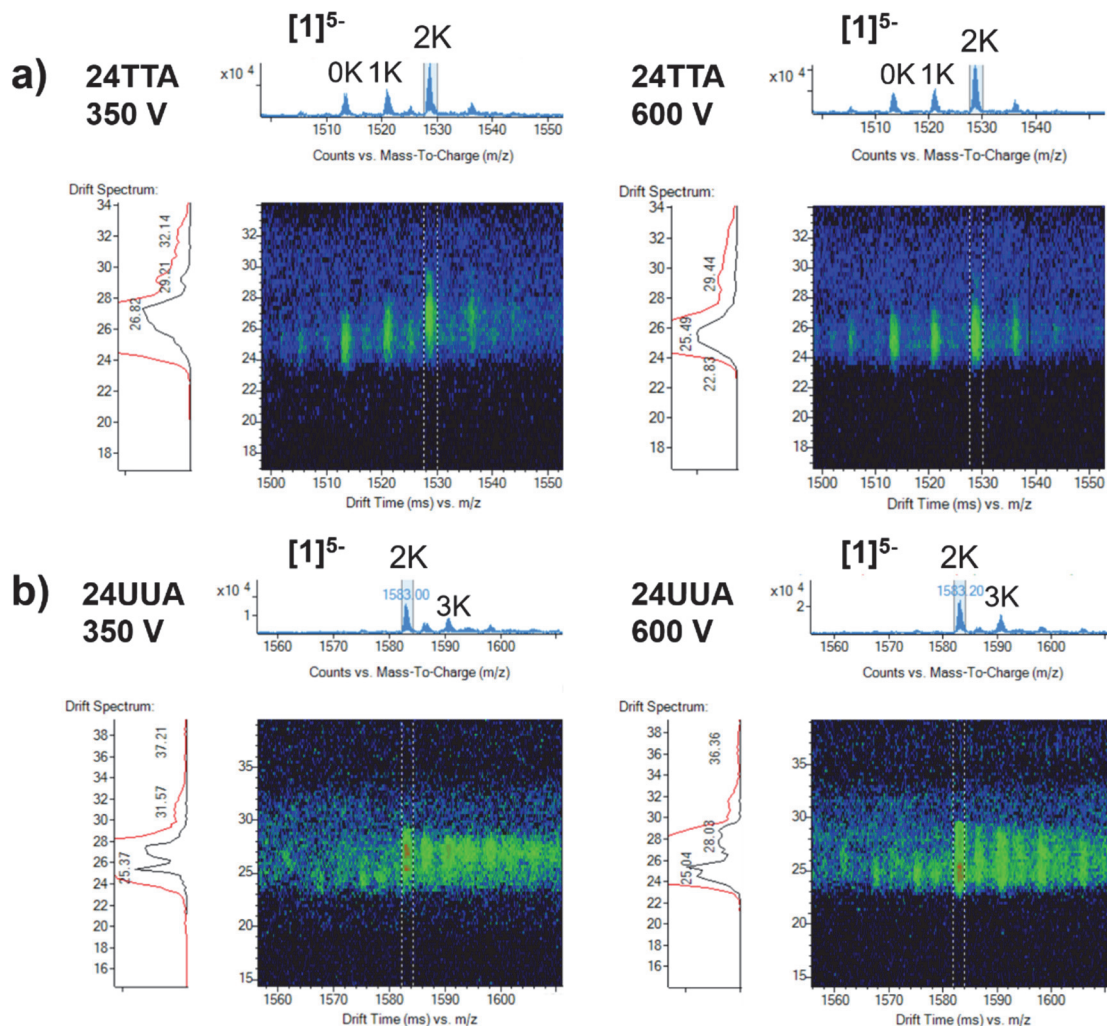
The different sequences also behave differently with regard to peak shape. The CCSDs of 24TTG are clearly narrower than those of 24TTA and 26TTA. This suggests that, although structures rearrange, 24TTG has fewer rearrangement pathways than 24TTA and 26TTA. Figure 11 shows the root mean square fluctuations (RMSF) of the bases on the course of the gas-phase MD simulations started from the expected solution structures. 24TTG has the lowest RMSF. The particularity of this structure is its stabilization in solution by two extra TA base pairs.<sup>26</sup> These are maintained in the gas phase and limit the rearrangement possibilities. Note however that we did not carry out multiple MD simulations in the gas phase from a given starting structure. Multiple trajectories would be required to tentatively account for the peak width. Also, we used a distributed charge model that is incapable of taking into account the specific hydrogen bonds that could be formed between protonated phosphate groups and other chemical functions. Localized charge models, and actually higher-level calculations than force field-based modeling, would be required to explore gas-phase rearrangement pathways in more depth. However, given the similar CCS range for different starting structures, we didn't pursue this endeavor here.



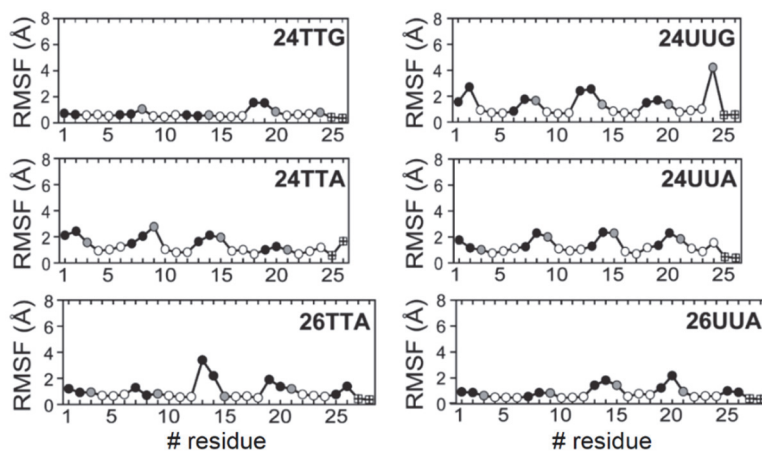
**Figure 8.** Matching between the experimental  $^{\text{DT}}\text{CCS}_{\text{He}}$  (line and scatter plots) and the theoretical  $^{\text{EHSS}}\text{CCS}_{\text{He}}$  (histograms) for the 1-subunit G-quadruplexes DNA (a) 24TTA, (b) 24TTG, (c) 26TTA and RNA (d) 24UUA, (e) 24UUG, (f) 26UUA, with two inner cations. Experimental data are color coded by charge state: red: 4- charge state, black: 5- charge state, green: 6- charge state. All results were obtained in gentle (350 V) pre-IMS activation conditions. On panel e), the empty red circles correspond to dimers of charge 8-. The structures shown are the final structures of the complexes with 2 inner  $\text{K}^+$  ions after gas-phase molecular dynamics modelling. The histograms of calculated structures were calculated over 500 ns.



**Figure 9.** Evolution of the collision cross section as a function of gas-phase molecular dynamics for three models of 24TTA (antiparallel, hybrid 1 and hybrid 2). Structures on the left are solution models extracted from solution MD, which served as starting points for gas-phase force field MD. Structures on the right are gas-phase structures extracted after 200 ns of gas-phase force field MD.



**Figure 10.** Collision cross section distributions as a function of the number of bound potassium ions of a) 24TTA and b) 24UUA in gentle pre-IMS activation conditions (350 V, left) and harsh pre-IMS activation conditions (600 V, right). The 2- $K^+$  complexes in soft conditions should be the best preserved G-quadruplexes, with three G-quartets.



**Figure 11.** Indicators of the stability of the DNA (left) and RNA (right) 4-repeat models during 500 ns MD simulation: root mean square fluctuation (RMSF, panels on the right) of the guanines, the thymines, the adenines (shown as circles in white, black, and grey, respectively), and the  $K^+$  cations (shown as crossed squares).

## 2) 4-repeat RNA sequences

In the 4-repeat RNA sequences (Fig. 8, d–f), the cations have a more significant effect on the CCS profiles. This is in line with the CD results: the structures differ and there is thus a memory of this difference in the electrosprayed ions. According to CD, the structures formed in potassium are more purely parallel, which signifies that the loops are all edgewise (propeller), connecting the strands from the sides of the G-quartet core. The CCSDs in ammonium acetate are broad and the gas-phase structures are more compact than their DNA counterparts. Thus, significant gas-phase rearrangements with compaction occur but, as for DNA, it is illusory to assign a structure based on the CCS distributions. We conclude that several structures coexist in solution for the monomers, adding up to the mixture of monomers and dimers.

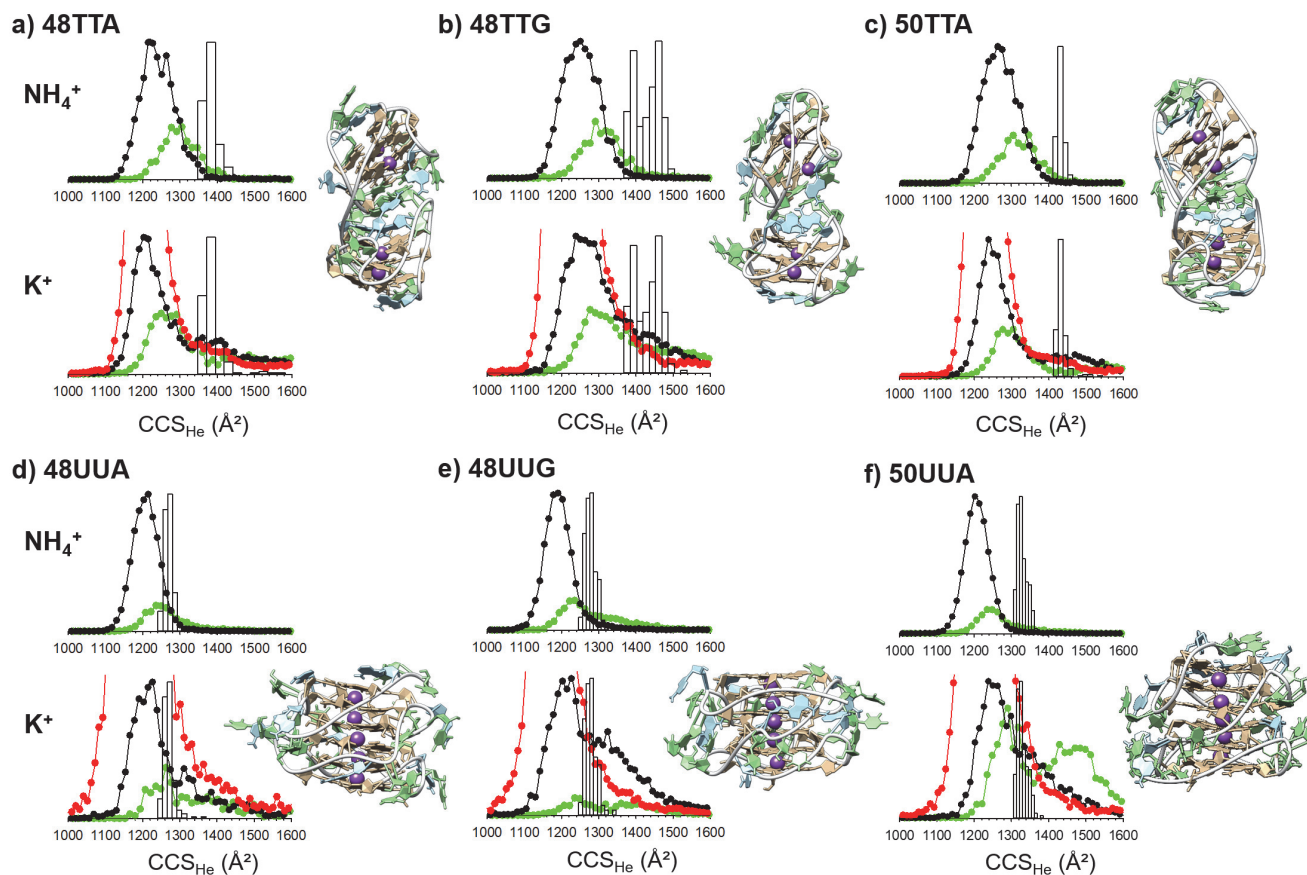
The potassium IM results are intriguing because there are several sharp peaks. Also, the 6- charge state is more abundant than for the DNA counterparts or than for RNA in ammonium. The charges state effect on the collision cross section is also significant (Figure 10b): upon collisional activation of the 5- ions, the structures compact even further, below  $800 \text{ \AA}^2$ . The compaction of the 4- ions is even more dramatic, and upon activation, the most compact peak prevails (not shown). The 6- ions usually just broaden, without change of CCS (not shown). Thus, 1) the peaks above  $800 \text{ \AA}^2$  seem the closest to the solution and 2) gas-phase rearrangements cause compaction. If we examine the three-dimensional models in the gas phase,

we see that the models of 24UUA and 26UUA match well with the CCSD of the 6- ions and the most extended conformation of the 5- ions. The structures have preserved their G-quartet core but the loops have folded back on the structure (see the comparison between starting and final structure in supplemental Figure S5). For 24UUG, the loops have not compacted so well during the simulations, and the calculated CCS values are too large compared to the experimental ones. All these results indicate that loop rearrangement is a key determinant of gas-phase collision cross sections.

### 3) 8-repeat sequences

Figure 12 shows the collision cross section distributions (experimental and calculated), for all 8-repeat sequences. The starting structures and final structures as well as the evolution of the calculated CCS with MD time are shown in supplemental Figures S6-S7. The collision cross sections do not depend significantly on the cation nature. All peaks are similarly broad. The RNA gas-phase structures are slightly more compact than their DNA counterparts, except for 48TTA/48UUA where there is no difference. All distributions are broad and featureless. Indeed, in all cases where a shoulder or second peak is seen at higher CCS in  $K^+$ , it is actually due to streaks of chemical noise at the same  $m/z$ , and not to oligonucleotide signal, as shown in supporting information Figure S8. Their apparent prominence of noise is due to a low signal to noise ratio (see also mass spectra in Figure 5).

Simple geometric estimations starting from the CCS values of the 4-repeat suggest that the 8-repeat gas-phase structures are mostly spherical. Indeed, if one takes the CCS value of the 4- ions of 4-repeat sequences ( $\sim 755 \text{ \AA}^2$  for 24-mers) as the surface of the cross section of a ball, then estimate the surface of the cross section of a ball with double the volume (48-mers), one finds  $\sim 1200 \text{ \AA}^2$ , in line with observations (Figure 12). The DNA 3D models shown in Figure 12, with the two subunits like beads on a string,<sup>69, 96</sup> are too elongated despite the subunits got closer upon gas-phase MD. If the two DNA subunits are forming independently, as suggested by the specific cation counts, then it means that in the gas phase the compaction is more extensive than captured by our MD simulations.



**Figure 12.** Matching between the experimental  $^{DT}CCS_{He}$  (line and scatter plots) and the theoretical  $^{EHSS}CCS_{He}$  (histograms) for the 2-subunit G-quadruplexes DNA (a) 48TTA, (b) 48TTG, (c) 50TTA and RNA (d) 48UUA, (e) 48UUG, (f) 50UUA, with four to five bound cations (which nature is indicated on the left). All data were obtained using gentle (350 V) pre-IMS activation conditions except RNA sequences in  $K^+$ , which were obtained at 600 V. Experimental data are color coded by charge state: red: 6- charge state, black: 7- charge state, green: 8- charge state. In  $K^+$ , the 6- charge state was reconstructed from the hump of signal, and hence goes off scale compared to the 7- signal. The structures shown are the final structures of complexes with 4 (DNA structures) or 5 (RNA structures) inner  $K^+$  ions, after gas-phase molecular dynamics modelling. The histograms of calculated structures were calculated over 500 ns.

In the case of RNA 8-repeats, the predominance of complexes with 5 cations, combined with the CD spectra indicating all-anti homostacking and thus parallel-stranded structures, suggest that two subunits are stacked. This could give rise to more compact structures than the beads-on-a-string model. We generated two types of stacks: opposite 5'-5', as reported by Phan's group for tetramers of

r(GGGUUAGGGU)<sup>93</sup> and consecutive 5'-3' stacks.<sup>38</sup> The 5'-3' stacking models converged to slightly more compact structures than the 5'-5' stacking models (supporting information Figure S9). Their calculated CCS values are both below 1300 Å<sup>2</sup>, thus matching most closely with the experimental values although the entire compaction was not captured. Yet note that the experimental distributions are broad, and thus several rearrangement pathways or several coexisting structures in solution probably contribute to the peaks. Ion mobility spectrometry could not help assign the structure.

## CONCLUSIONS

Telomeric RNA was studied in TMAA/KCl for the first time, and compared with telomeric DNA. Also, we report the ion mobility spectrometry of 8-repeat sequences (potentially two G-quadruplex subunits) for the first time. The following conclusions can be derived:

- 1) The TMAA/KCl sample preparation conditions, besides the intrinsic advantage of folding the G-quadruplexes in potassium conditions, disfavor intermolecular assemblies compared to ammonium acetate. This is an advantage when the aim is to probe folding of oligonucleotides *in vitro*, where higher-order multimer formation would cause artifacts.
- 2) In the case of 8-repeat DNA sequences, the folding of the second G-quadruplex is less favorable than the folding of the first one. Both subunits fold anticooperatively. In contrast, for the 8-repeat RNA, the presence of a fifth K<sup>+</sup> ion in-between the two folded subunits (2 cations each) indicate the cooperative formation of a higher-order structure.
- 3) RNA shows full incorporation of 2 K<sup>+</sup> ions per subunit, and thus full folding in those conditions. The RNA G-quadruplexes are in that respect more stable than their DNA counterparts. However, the TMAA/KCl conditions present some limitations for the study of longer sequences, due to abundant clustering. Mass analysis of longer RNA G-quadruplex sequences in native conditions would require further work with respect to sample solution conditions.

We also critically assessed the prospects of assigning G-quadruplex structures using ion mobility spectrometry, by comparing experimental and calculated collision cross sections. The results are mixed.

- 4) The promising aspect is that there is a memory of the solution structure in the gas-phase. This is established by three types of controls: sequence variants (e.g. DNA vs. RNA, which changes the structure in solution), control solution conditions in which the sequences are unfolded (as indicated by the absence of trapped cation), and control experiments with pre-IMS collision activation that disrupts pre-existing noncovalent interactions.
- 5) However, despite there is a memory, the gas-phase structures are not identical to the solution ones. Gas-phase rearrangements cause a compaction at the charge states produced from our native MS buffers. Furthermore, the lower the charge state, the greater the compaction. In soft pre-IMS activation conditions, the G-stem prevents full compaction, meaning it is preserved. However, the loops and overhangs rearrange. Subunits folded like beads on a string can also rearrange towards

compaction. The highest charge states produced are closer to the solution structure in terms of compactness. Future work should devote more attention to the charge state dependence of the CCS values, and less to the CCS value of the most abundant charge state. Strategies to manipulate charge states while preserving the solution structure would also be welcome.

- 6) The loop rearrangements mask the strand arrangement topology because they wrap the G-quadruplex core, as shown by gas-phase MD simulations from different starting structures. This hampers the structural assignment based on collision cross section values. Also, we anticipate that higher-level calculation methods might be required to correctly capture these rearrangements.
- 7) Besides the IM peak position, the peak broadness also conveys structural information. Broad peaks can mean either that multiple structures are initially present in solution, or multiple ways for solution structures to rearrange before entering the ion mobility cell. However, a sharp peak indicates a preferred rearrangement pathway, for example due to specific hydrogen bonding interactions involving the loops and overhangs. In the future, we want to explore possible use of IM-MS to screen oligonucleotide sequences in order to identify monomorphic ones with well-defined hydrogen bonding network, which would be more amenable to high-resolution structural studies. The sharp peaks observed for 24TTG, which is reputedly the most monomorphic telomeric DNA variant, support this proposal.
- 8) In addition to 24TTG, sharp peaks were also observed for the 4-repeat RNA sequences 24UUA and 24UUG, yet there were multiple peaks, indicating the coexistence of at least two structures in solution.

To conclude, our ion mobility results indicate a memory of the solution phase structure, but more work is needed to understand the charge-dependent rearrangements occurring upon desolvation and declustering. Understanding these rearrangements required to make them predictable, and realise the full potential of ion mobility spectrometry in assigning the coexisting structures that can be separated.

**Conflicts of interest:** There are no conflicts of interest to declare

## **ACKNOWLEDGEMENTS**

This work was supported by the Inserm (ATIP-Avenir Grant no. R12086GS), the Conseil Régional Aquitaine (Grant no. 20121304005), and the EU (FP7-PEOPLE-2012-CIG-333611 and ERC-2013-CoG-616551-DNAFOLDIMS).

## REFERENCES

1. J. T. Davis, *Angew. Chem. Int. Ed.*, 2004, **43**, 668-698.
2. J. L. Huppert and S. Balasubramanian, *Nucleic Acids Res.*, 2005, **33**, 2908-2916.
3. A. K. Todd, M. Johnston and S. Neidle, *Nucleic Acids Res.*, 2005, **33**, 2901-2907.
4. S. G. Hershman, Q. Chen, J. Y. Lee, M. L. Kozak, P. Yue, L. S. Wang and F. B. Johnson, *Nucleic Acids Res.*, 2008, **36**, 144-156.
5. N. Maizels and L. T. Gray, *PLoS Genet.*, 2013, **9**, e1003468.
6. A. Bedrat, L. Lacroix and J. L. Mergny, *Nucleic Acids Res.*, 2016, **44**, 1746-1759.
7. G. Biffi, D. Tannahill, J. McCafferty and S. Balasubramanian, *Nature Chem.*, 2013, **5**, 182-186.
8. A. Henderson, Y. Wu, Y. C. Huang, E. A. Chavez, J. Platt, F. B. Johnson, R. M. Brosh, Jr., D. Sen and P. M. Lansdorp, *Nucleic Acids Res.*, 2013, **42**, 860-869.
9. A. Henderson, Y. Wu, Y. C. Huang, E. A. Chavez, J. Platt, F. B. Johnson, R. M. Brosh, Jr., D. Sen and P. M. Lansdorp, *Nucleic Acids Res.*, 2017, **45**, 6252.
10. D. Rhodes and H. J. Lipps, *Nucleic Acids Res.*, 2015, **43**, 8627-8637.
11. S. Neidle and G. Parkinson, *Nature Rev. Drug Discov.*, 2002, **1**, 383-393.
12. L. Oganessian, M. E. Graham, P. J. Robinson and T. M. Bryan, *Biochemistry*, 2007, **46**, 11279-11290.
13. A. L. Moye, K. C. Porter, S. B. Cohen, T. Phan, K. G. Zyner, N. Sasaki, G. O. Lovrecz, J. L. Beck and T. M. Bryan, *Nat. Commun.*, 2015, **6**, 7643.
14. C. M. Azzalin, P. Reichenbach, L. Khoriantseva, E. Giulotto and J. Lingner, *Science*, 2007, **318**, 798-801.
15. B. Luke and J. Lingner, *The EMBO journal*, 2009, **28**, 2503-2510.
16. E. Cusanelli, C. A. Romero and P. Chartrand, *Mol. Cell.*, 2013, **51**, 780-791.
17. R. Arora and C. M. Azzalin, *RNA Biol.*, 2015, **12**, 938-941.
18. N. Bettin, C. Oss Pegorar and E. Cusanelli, *Cells*, 2019, **8**.
19. Y. Wang and D. J. Patel, *Structure*, 1993, **1**, 263-282.
20. P. K. Patel, A. S. R. Koti and R. V. Hosur, *Nucleic Acids Res.*, 1999, **29**, 3836-3843.
21. G. N. Parkinson, M. P. H. Lee and S. Neidle, *Nature*, 2002, **417**, 876-880.
22. A. T. Phan and D. J. Patel, *J. Am. Chem. Soc.*, 2003, **125**, 15021-15027.
23. S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle, *Nucleic Acids Res.*, 2006, **34**, 5402-5415.
24. A. Ambrus, D. Chen, J. Dai, T. Bialis, R. A. Jones and D. Yang, *Nucleic Acids Research*, 2006, **34**, 2723-2735.
25. A. T. Phan, V. Kuryavyi, K. N. Luu and D. J. Patel, *Nucleic Acids Res.*, 2007, **35**, 6517-6525.
26. K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix and D. J. Patel, *J Am Chem Soc*, 2006, **128**, 9963-9970.

27. J. Dai, C. Punchihewa, A. Ambrus, D. Chen, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2007, **35**, 2440-2450.
28. J. Dai, M. Carver, C. Punchihewa, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2007, **35**, 4927-4940.
29. J. Dai, M. Carver and D. Yang, *Biochimie*, 2008, **90**, 1172-1183.
30. K. W. Lim, S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D. J. Patel, K. N. Luu and A. T. Phan, *J. Am. Chem. Soc.*, 2009, **131**, 4301-4309.
31. D. J. Yue, K. W. Lim and A. T. Phan, *J. Am. Chem. Soc.*, 2011, **133**, 11462-11465.
32. B. Heddi and A. T. Phan, *J. Am. Chem. Soc.*, 2011, **133**, 9824-9833.
33. P. Galer, B. Wang, P. Sket and J. Plavec, *Angew. Chem. Int. Ed.*, 2016, **55**, 1993-1997.
34. C. Liu, B. Zhou, Y. Geng, D. Yan Tam, R. Feng, H. Miao, N. Xu, X. Shi, Y. You, Y. Hong, B. Z. Tang, P. Kwan Lo, V. Kuryavyi and G. Zhu, *Chem. Sci.*, 2018, **10**, 218-226.
35. M. Webba da Silva, *Chemistry*, 2007, **13**, 9738-9745.
36. A. I. Karsisiotis, C. O'Kane and M. Webba da Silva, *Methods*, 2013, **64**, 28-35.
37. A. Arora and S. Maiti, *J. Phys. Chem. B*, 2009, **113**, 10515-10520.
38. H. Martadinata, B. Heddi, K. W. Lim and A. T. Phan, *Biochemistry*, 2011, **50**, 6455-6461.
39. A. Randall and J. D. Griffith, *The Journal of biological chemistry*, 2009, **284**, 13980-13986.
40. H. Martadinata and A. T. Phan, *J. Am. Chem. Soc.*, 2009, **131**, 2570-2578.
41. G. W. Collie, S. M. Haider, S. Neidle and G. N. Parkinson, *Nucleic Acids Res.*, 2010, **38**, 5569-5580.
42. L. Haase, J. Dickerhoff and K. Weisz, *Chemistry*, 2018, **24**, 15365-15371.
43. J. Dickerhoff, B. Appel, S. Muller and K. Weisz, *Angew Chem Int Ed Engl*, 2016, **55**, 15162-15165.
44. J. Zhou, S. Amrane, F. Rosu, G. F. Salgado, Y. Bian, H. Tateishi-Karimata, E. Largy, D. N. Korkut, A. Bourdoncle, D. Miyoshi, J. Zhang, H. Ju, W. Wang, N. Sugimoto, V. Gabelica and J. L. Mergny, *J. Am. Chem. Soc.*, 2017, **139**, 7768-7779.
45. D. H. Zhang, T. Fujimoto, S. Saxena, H. Q. Yu, D. Miyoshi and N. Sugimoto, *Biochemistry*, 2010, **49**, 4554-4563.
46. S. A. Dvorkin, A. I. Karsisiotis and M. Webba da Silva, *Sci. Adv.*, 2018, **4**, eaat3007.
47. A. I. Karsisiotis, N. Ma'ani Hessari, E. Novellino, G. P. Spada, A. Randazzo and M. Webba da Silva, *Angew. Chem. Int. Ed.*, 2011, **50**, 10645-10648.
48. R. Del Villar-Guerra, J. O. Trent and J. B. Chaires, *Angew. Chem. Int. Ed.*, 2018, **57**, 7171-7175.
49. F. Rosu, V. Gabelica, C. Houssier, P. Colson and E. De Pauw, *Rapid Commun. Mass Spectrom.*, 2002, **16**, 1729-1736.
50. G. Yuan, Q. Zhang, J. Zhou and H. Li, *Mass Spectrom. Rev.*, 2011, **30**, 1121-1142.
51. F. Balthasart, J. Plavec and V. Gabelica, *J. Am. Soc. Mass Spectrom.*, 2013, **24**, 1-8.
52. A. Marchand and V. Gabelica, *J. Am. Soc. Mass Spectrom.*, 2014, **25**, 1146-1154.

53. J. Gros, F. Rosu, S. Amrane, A. De Cian, V. Gabelica, L. Lacroix and J. L. Mergny, *Nucleic Acids Res.*, 2007, **35**, 3064-3075.
54. A. Marchand and V. Gabelica, *Nucleic Acids Res.*, 2016, **44**, 10999-11012.
55. A. Marchand, A. Granzhan, K. Iida, Y. Tsushima, Y. Ma, K. Nagasawa, M. P. Teulade-Fichou and V. Gabelica, *J. Am. Chem. Soc.*, 2015, **137**, 750-756.
56. T. Wyttenbach and M. T. Bowers, *Mod. Mass Spectrom.*, 2003, **225**, 207-232.
57. D. E. Clemmer and M. F. Jarrold, *J. Mass Spectrom.*, 1997, **32**, 577-592.
58. V. Gabelica and E. Marklund, *Curr. Opin. Chem. Biol.*, 2018, **42**, 51-59.
59. V. D'Atri, M. Porrini, F. Rosu and V. Gabelica, *J. Mass. Spectrom.*, 2015, **50**, 711-726.
60. V. Gabelica, E. S. Baker, M. P. Teulade-Fichou, E. De Pauw and M. T. Bowers, *J. Am. Chem. Soc.*, 2007, **129**, 895-904.
61. E. S. Baker, S. L. Bernstein, V. Gabelica, E. De Pauw and M. T. Bowers, *Int. J. Mass Spectrom.*, 2006, **253**, 225-237.
62. J. C. May, C. R. Goodwin, N. M. Lareau, K. L. Leaptrot, C. B. Morris, R. T. Kurulugama, A. Mordehai, C. Klein, W. Barry, E. Darland, G. Overney, K. Imatani, G. C. Stafford, J. C. Fjeldsted and J. A. McLean, *Anal. Chem.*, 2014, **86**, 2107-2116.
63. A. Perez, I. Marchan, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton and M. Orozco, *Biophys. J.*, 2007, **92**, 3817-3829.
64. C. K. Siu, Y. Guo, I. S. Saminathan, A. C. Hopkinson and K. W. Siu, *J. Phys. Chem. B*, 2010, **114**, 1204-1212.
65. M. J. Cavaluzzi and P. N. Borer, *Nucleic Acids Res.*, 2004, **32**, e13.
66. V. Gabelica, S. Livet and F. Rosu, *J. Am. Soc. Mass Spectrom.*, 2018, **29**, 2189-2198.
67. A. Marchand, S. Livet, F. Rosu and V. Gabelica, *Anal. Chem.*, 2017, **89**, 12674-12681.
68. K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix and D. J. Patel, *J. Am. Chem. Soc.*, 2006, **128**, 9963-9970.
69. L. Petraccone, C. Spink, J. O. Trent, N. C. Garbett, C. S. Mekmaysy, C. Giancola and J. B. Chaires, *J. Am. Chem. Soc.*, 2011, **133**, 20951-20961.
70. G. W. Collie, N. H. Campbell and S. Neidle, *Nucleic Acids Res.*, 2015, **43**, 4785-4799.
71. D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P. A. Kollman, *Journal*, 2012, DOI: citeulike-article-id:10779586.

72. M. Zgarbova, J. Sponer, M. Otyepka, T. E. Cheatham, 3rd, R. Galindo-Murillo and P. Jurecka, *J Chem Theory Comput*, 2015, **11**, 5723-5736.
73. I. S. Joung and T. E. Cheatham III, *J. Phys. Chem. B*, 2008, **112**, 9020-9041.
74. W. L. Jorgensen, J. Chandrasekhar and J. D. Madura, *J. Chem. Phys.*, 1983, **79**, 926-935.
75. T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089.
76. J.-P. Ryckaert, G. Ciccotti and H. G. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327-341.
77. M. Rueda, S. G. Kalko, F. J. Luque and M. Orozco, *J. Am. Chem. Soc.*, 2003, **125**, 8007-8014.
78. M. Rueda, F. J. Luque and M. Orozco, *J. Am. Chem. Soc.*, 2006, **128**, 3608-3619.
79. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684-3690.
80. D. R. Roe and T. E. Cheatham, 3rd, *J. Chem. Theory Comput.*, 2013, **9**, 3084-3895.
81. D. A. Case, R. M. Betz, W. Botello-Smith, 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. A. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, D. M. York and P. A. Kollman, *Journal*, 2016.
82. W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33-38, 27-38.
83. A. A. Shvartsburg and M. F. Jarrold, *Chem. Phys. Lett.*, 1996, **261**, 86-91.
84. A. A. Shvartsburg, S. V. Mashkevich, E. S. Baker and R. D. Smith, *J. Phys. Chem. A*, 2007, **111**, 2002-2010.
85. P. Sket, A. Virgilio, V. Esposito, A. Galeone and J. Plavec, *Nucleic Acids Res.*, 2012, **40**, 11047-11057.
86. V. D'Atri, N. Borbone, J. Amato, V. Gabelica, S. D'Errico, G. Piccialli, L. Mayol and G. Oliviero, *Biochimie*, 2014, **99**, 119-128.
87. N. Borbone, J. Amato, G. Oliviero, V. D'Atri, V. Gabelica, E. De Pauw, G. Piccialli and L. Mayol, *Nucleic Acids Res.*, 2011, **39**, 7848-7857.
88. G. W. Collie, G. N. Parkinson, S. Neidle, F. Rosu, P. E. De and V. Gabelica, *J. Am. Chem. Soc.*, 2010, **132**, 9328-9334.
89. J. Lipfert, S. Doniach, R. Das and D. Herschlag, *Annu. Rev. Biochem.*, 2014, **83**, 813-841.
90. J. C. Bowman, T. K. Lenz, N. V. Hud and L. D. Williams, *Curr. Opin. Struct. Biol.*, 2012, **22**, 262-272.
91. N. Smargiasso, F. Rosu, W. Hsia, P. Colson, E. S. Baker, M. T. Bowers, E. De Pauw and V. Gabelica, *J. Am. Chem. Soc.*, 2008, **130**, 10208-10216.
92. A. Marchand, F. Rosu, R. Zenobi and V. Gabelica, *J. Am. Chem. Soc.*, 2018, **140**, 12553-12565.
93. H. Martadinata and A. T. Phan, *Biochemistry*, 2013, **52**, 2176-2183.
94. R. Ferreira, A. Marchand and V. Gabelica, *Methods*, 2012, **57**, 56-63.

95. M. Porrini, F. Rosu, C. Rabin, L. Darre, H. Gomez, M. Orozco and V. Gabelica, *ACS Cent. Sci.*, 2017, **3**, 454-461.
96. H. Q. Yu, D. Miyoshi and N. Sugimoto, *J. Am. Chem. Soc.*, 2006, **128**, 15461-15468.