Native Ion Mobility–MS Reveals the Bead Ejection Mechanism in Multi-Domain Nucleic Acids

Debasmita Ghosh,[1] Frédéric Rosu[2,3] and Valérie Gabelica[1,3]*

[1] Univ. Bordeaux, INSERM, CNRS, ARNA, UMR 5320, U1212, F-33000 Bordeaux, France

[2] Univ. Bordeaux, CNRS, INSERM, IECB, UAR3033, US01, F-33600 Pessac, France

[3] School of Pharmaceutical Sciences, University of Geneva, Geneva, Switzerland (valerie.gabelica@unige.ch)

Abstract:

Elucidating how biomolecules acquire charge and retain their solution conformation during electrospray ionization (ESI) is crucial for native MS analysis. Here, we examine the charging and gas phase conformation of nucleic acid constructs comprising folded G-quadruplex "beads" linked by unstructured polythymine regions. Under physiological ionic strength, these oligonucleotides exhibit a multimodal charge-state and collision cross section distribution, revealing multiple conformational ensembles, in contrast to the unimodal profiles typically observed for shorter oligonucleotides. The ion mobility results for intermediate charge states are compatible with ion production via the recently proposed bead ejection mechanism (BEM), in addition to the charge residue model (CRM) for low charge states and chain ejection model (CEM) for the highest charge states or for sequences with thymine overhangs on both ends. The position of the G-quadruplex domain and ionic strength governs the charging and retention of G-quadruplex folded regions. Our findings broaden the existing mechanistic framework underpinning nucleic acid ionization.

Introduction

Native electrospray ionization (ESI) mass spectrometry (MS) preserves the covalent and non-covalent interactions of biomolecules during their transition from solution to the gas phase.[1,2] Although ESI was introduced over three decades ago,[3] the precise mechanisms that govern the final stages of ion charging and desolvation remain a matter of debate. Three major models are commonly discussed: the ion evaporation model (IEM), which primarily applies to small ions[4] and was also modeled for folded proteins,[5] the charge residue model (CRM)[6] for large globular analytes, and the chain ejection model (CEM)[7] for disordered (bio)polymers. Intermediate mechanisms were also proposed.[8] For example, multi-domain therapeutic proteins can ionize via a hybrid CRM-CEM route called the bead ejection mechanism (BEM), wherein folded domains (beads) remain compact while disordered linkers extend.[9]

The charge state distribution (CSD) of nucleic acids during ESI often reflects the ionization process more than the solution folding state.[10] Studies on oligonucleotides containing up to 34 bases,[11,12] reveal that high ammonium acetate concentrations (100-150 mM, physiological ionic strength) produce low charge states and compact conformations in the gas phase.[13] High charge states and extended conformations are observed either from solutions of low ionic strength, or in the presence of supercharging agents.[11] Under high ionic strength, the presence of more acetate anions (charge carriers) on the droplet surface preserves the nucleic acids in the droplet center and favors CRM, whereas fewer acetate ions (low ionic strength) promote CEM.[10]

Here we wanted to establish whether the BEM also applies to longer nucleic acids in the negative ion mode. We used sequences containing G-quadruplexes as model systems. Folded G-quadruplex nucleic acid structures are driven by the specific coordination of cations (NH_4^+, K^+) within guanine quartets. Both the nature and concentration of these cations determine the structure and stability of G-quadruplexes.[14,15] Ammonium acetate is commonly used in native MS as a volatile electrolyte, with its concentration chosen based on the desired ionic strength.[11,16] Inspired by the findings for multi-domain proteins, we designed mixed nucleic acid structures containing both G-quadruplex (G4) and unstructured polythymine (Tn) sequences, and placed the G-quadruplex at various positions along a long thymine chain (Table 1). As a control, we examined a non-folding

single strand sequence (NG) which lacks enough contiguous G-tracts to form a G-quadruplex and thus does not exhibit specifically bound cations. Each G4 domain contains 3 G-quartets and thus 2 specifically bound NH_4^+ ions. Thus, the mass measurement informs us on how many G4 domains are preserved in the gas phase. We also used native ion mobility MS[17–21] to deduce the gas-phase compactness of the gas-phase structures and deduce the possible ESI mechanism of the different conformers.

Table 1. 60-mer DNA sequences studied herein.

Name	Sequence	Structure
G4Tn	$[dT(TG_3)_4T_{43}\bullet(NH_4^+)_2]$	1G4
TnG4	$[dT_{43}T(TG_3)_4\bullet(NH_4^+)_2]$	1G4
TnG4Tn	$[dT_{21}T(TG_3)_4T_{22}\bullet(NH_4^+)_2]$	1G4
G4TnG4	$[dT(TG_3)_4T_{26}T(TG_3)_4\bullet(NH_4^+)_4]$	2G4
NG	$[dT_4GT_5TGT_4(GT_3)_5(GT_2)_2GT_4TGT_4GT_4T_2]$	non-structured

Experimental Section:

Sample Preparation: All single strands were purchased from Eurogentec (Seraing, Belgium, with RP cartridge purification) and dissolved in water from Biosolve. Desalting is crucial for our sample preparation, particularly since some experiments are conducted at low ionic strength. We used Amicron ultracel-3 centrifugal filters (Merck Millipore Ltd.) with a 3-kDa molecular weight cutoff. The samples underwent several centrifugation passes with a 200 mM aqueous NH_4OAc solution, followed by several passes with a 100 mM aqueous NH_4OAc solution, and finally with water to achieve satisfactory desalting.

All G-quadruplexes were formed in 150 mM ammonium acetate. Intramolecular G-quadruplexes were formed from 200 μM single strand, incubated for at least 24 h in the folding buffer (150 mM aqueous NH_4OAc). Single strands were annealed at 85°C in water for 5 min to ensure the removal of any preformed structures. Control experiments were done using non-structured oligonucleotides of the same length. The supercharging agent propylene carbonate (PC) (99.7% purity) was purchased from Sigma-Aldrich. The DNA

structures were diluted and injected at a final concentration of 15 μM final concentration for IM-MS analysis.

UV spectroscopy:

The absorbance was recorded at 260 nm using a Uvikon XS spectrophotometer and the stock sample concentrations were determined using the Beer−Lambert law. Molar extinction coefficients were estimated using the nearest-neighbor model and Cantor's parameters.[22] The stability of G4 structures in solution was examined with a UVmc2 double-beam spectrophotometer (SAFAS, Monte Carlo, Monaco) equipped with a temperature-controlled 10-cell holder and a high-performance Peltier temperature controller. [11]

Thermal denaturation was monitored by measuring the changes in the absorbance at 295 nm as a function of temperature. Samples were heated to 90 °C, then the absorbance was recorded at 260, 295, and 350 nm in a series contains a cooling down to 4 °C at a rate of 0.2 °C min^{-1}, followed by reheating to 90 °C at the same rate.

The resulting absorbance versus temperature data were converted into a fraction folded (θ) versus temperature plot.[23] We fitted the upper and lower baselines, then calculated the fraction folded at each temperature, θ_T (Eq. 1), where $L0_T$ and $L1_T$ are baseline values of the unfolded and folded species. The temperature at which θ = 0.5 is known as the melting temperature.[23]

$$\theta_T = \frac{(L0_T - A_T)}{(L0_T - L1_T)} \quad (1)$$

Native ion mobility mass spectrometry (IM-MS):

IM-MS and collision-induced unfolding (CIU) experiments were conducted at 24 °C using a 6560 DTIMS-Q-TOF instrument (Agilent Technologies, Santa Clara, CA) adapted to run with helium in the drift tube.[20] All experiments were performed in negative ion mode with the standard electrospray ionization (ESI) source. A syringe pump flow rate of 3 μL/min was used. The parameters used were fragmentor = 320 V (unless stated otherwise for CIU experiments), nebulizing gas = 4 psi, drying gas = 1 L/min, trap fill time = 1000 μs, trap

release time = 150 µs, and trap entrance grid delta (TEGD) = 2 V. For collision cross-section (CCS) quality control, $[(dTG_4T)_4(NH_4)_3]^{5-}$ was injected prior to the analysis, ensuring a $^{DT}CCS_{He}$ within 785−791 Å² of its previously determined value (788 Å²).[24] The data extraction was performed using the IM-MS Browser software version B.08.00 (Agilent Technologies).

The conversion of arrival time distributions to CCS distributions is performed as follows.[25] First, the CCS of the centroid of one of the peaks is measured using the step-field method: five IM-MS spectra (segments) are recorded with varying drift tube entrance voltages (650, 700, 800, 900, 1000 V), and the arrival time of the centroid of the peak (t_A) is measured as a function of the voltage difference between the entrance and exit of the drift tube (ΔV). A linear regression with Eq. (2) provides CCS from the slope, knowing that ($\mu = m_{gas}m_{ion}/(m_{gas} + m_{ion})$), p_0 is the standard pressure (760 Torr), p is the pressure in the drift tube (3.89 ± 0.01 Torr), T_0 is the standard temperature (273.15 K), T is the temperature in the tube (296.15 ± 1 K), N_0 is the buffer gas number density at T_0 and p_0 (2.687 10^{25} m^{-3}), L is the physical length of the drift tube (78.1 ± 0.2 cm), k_B is the Boltzmann constant, z is the absolute value of the ion nominal charge, and e is the charge of the electron.

$$t_A = t_0 + \left(CCS.\frac{\sqrt{\mu}}{z}\right)\frac{16L^2}{3e}\sqrt{\frac{k_BT}{2\pi}}\frac{N_0T_0p}{Tp_0} \times \left(\frac{1}{\Delta V}\right) \quad (2)$$

To reconstruct CCS distributions from the arrival time distributions, the CCS determined with the step-field method and t_A determined for the 650 V segment are used to calculate the factor a using Equation (3), which is then used to change the axes from t_A (recorded at 650 V) to CCS for all other peaks.

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

Collision-Induced Unfolding (CIU) experiments were conducted by increasing the fragmentor voltage from 300V to 500V. 20 V steps at 2-minute intervals were acquired. The arrival time distributions were plotted at different fragmentor voltages to compare the gas phase stabilities of 60-mers.

CCS calculations for gas-phase structural models

Molecular dynamics simulations, carried out using Hyperchem 8.0.10 (Hypercube, Inc.), were used to generate structures of the G_4T_n in vacuo to calculate theoretical collision cross sections. The propeller-type parallel-stranded G-quadruplex 5'-TTGGGTGGGTGGGTGGGT-3' is taken from the NMR structure (PDB entry: 2LK7)[26] and modified by coupling a 42-mer strand of thymine. The initial structure of the single strand dT_{42} was generated as a B-DNA type helix. The charge states (8- and 29-) were generated manually using localized charges (protons). Using a localized or delocalized charges model has only a small impact on the generated structure when using force fields (amber[27], parmbsc0[28]). The position of the charge has a important effect on the structure only when using higher level of calculation like semi-empirical PM7[29] or DFT. Given the size of the oligonucleotide (60-mer), only force field is used on the complete sequence. PM7 will be used only on portions of the structure (the G-quadruplex structure, and a single strand truncated to dT_{21}). The three parts are then concatenated to calculate the collision cross section for low charge states.

To generate a structure of $(G4Tn)^{8-}$ with a more folded tail, the initial structure of the homothymine strand dT_{42} has been generated using the curvature of the DNA on the histone H2 (PDB entry: 7LV9)[30]. A 42-mer was extracted and the bases were mutated to thymine. The structure was cut in two 21-mers and optimized using PM7. Each dT_{21} carries three negative charges arbitrarily located every 10 bases. The G-quadruplex with 2- charge is also optimized and concatenated to the thymine strands. EHSSrot[31] is used for CCS calculations, with the helium revised parameters from Siu *et al*[32].

For the extended forms of G4Tn, we generated a 42-mer of homothymine with regular B-helix and connected it to the quadruplex structure. Starting from the 3'end, 29 charges were located on the phosphate groups (i.e., one phosphate deprotonated follow by a neutralized one). Amber force field (parmbsc1)[33] was used and molecular dynamics was performed for 40 ns. The structure elongates rapidly (after less than 1 ns) to reach an extended structure. TnG4Tn is generated with the concatenation of the homothymine strands at each extremities, and the same procedure at for G4Tn is applied.

Results and Discussion

60-mer oligonucleotides in physiological ionic strength produce multimodal charge state distributions

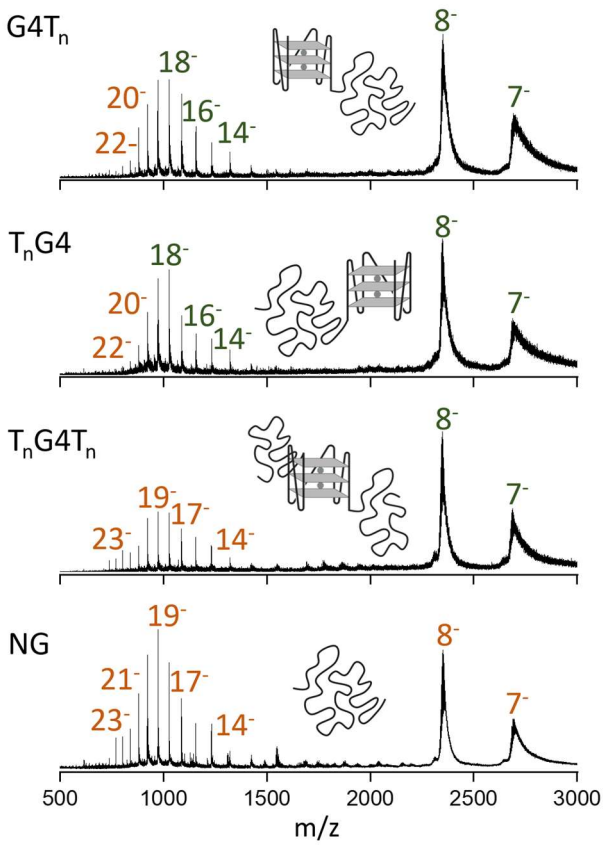

Figure 1. ESI-MS spectra of four 60-mer oligonucleotide structures: G4Tn, TnG4, TnG4Tn, and NG (top to bottom, 15 μM each) in 150 mM aqueous NH_4OAc. Peaks annotated in green contain 2 NH_4^+ ions, and peaks annotated in orange contain no NH_4^+ adducts.

To investigate the ESI mechanisms in mixed nucleic acid structures, we examined four 60-mer oligonucleotides: G4Tn, TnG4, TnG4Tn, and G4TnG4. Each variant contains one or two parallel-stranded intramolecular G-quadruplexes, each stabilized by two NH_4^+ ions per G-quadruplex subunit, along with a long thymine chain, and where we varied the position of the G4 domains. We confirmed that all G4-containing constructs are folded at room

temperature, with melting temperatures of 62°C for G4Tn and TnG4, 58°C for TnG4Tn, and 70°C for G4TnG4 (Figure S1). In this section, we discuss the results of G4Tn, TnG4, TnG4Tn (which contain one G-quadruplex) and the NG control, which cannot form a G-quadruplex.

Native MS spectra of four 60-mers obtained from 150 mM aqueous NH_4OAc show low charge states such as 7- and 8-, along with higher charges ranging from 14- to 22-, or even up to 26-, depending on the constructs. This finding contrasts with oligonucleotides up to 24-mers, which typically only exhibit low charges (such as 4- or 5-) at comparable ionic strength. All four 60-mers, $G4T_n$, T_nG4, T_nG4T_n, and NG, show distinct low and high charge populations through multimodal distributions (Figure 1). Thus, although these constructs are folded in solution, native MS data reveal multimodal CSDs for all 60-mers at physiological ionic strength, irrespective of their solution folding state.

Here, we identify four types of charge states: i) low charge states (7- and 8-); ii) missing charge states (9- to 12-); iii) high charge states that still contain 2 NH_4^+ ions (dark green), centered on 16- for TnG4 and G4Tn; iv) and high charge states that do not contain NH_4^+ ions (orange), centered on 18- for G4Tn and TnG4 and 19- for TnG4Tn and NG. Additionally, TnG4Tn and NG display an extra distribution at even higher charge states. This effect is much more pronounced at 50 mM aqueous NH_4OAc (Figure S2). The extent of high charges depends on the solution structure, and NG forms more of the high charge states (Figure S2, and Figure 1).

Lowering the ionic strength to 1 mM NH_4OAc (Figure S3) drastically favors higher charges. Both TnG4Tn and NG populate the highest charge states. Thus, when a polythymine overhang is present at both ends, the CSD is similar to the NG sequence. MS data are filtered in the IM-MS browser to exclude background signals with low charges (1- to 4-), while the unfiltered data are shown in Figure S4. Throughout all experiments, a fragmentor voltage was kept at 320 V to preserve the integrity of folded structures; however, fully desolvating low-charge ions (e.g., 7- or 8-) at high ionic strength is often challenging, resulting mass shifts to higher values. This is due to the presence of multiple ammoniums adducts peaks. Increasing the fragmentor voltages to 400 V or 450 V could

fully desolvate ammonium adducts on the 8- ions but also led to the loss of internally bound cations (Figures S5 and S6).

Ion mobility measurements suggest beads-on-a-string gas-phase structures for G-quadruplex containing sequences

Next, we examined the collision cross section of each charge state, for structures with intact (2 NH_4^+ ions bound) or disrupted (no NH_4^+ ions bound) G4s. The average CCS increases with the number of charges, but groups also emerge. For instance, the CCSDs of G4Tn show three distinct groups at 150 mM (Figure 2A). The first group comprises low charge states (7- and 8-), that retains two NH_4^+ ions, exhibits CCS of less than 1500 Å². Charge states 9- to 13- are missing. The second group shows CCS values from 2500 Å² to 3200 Å² while maintaining intact two-cation binding, and the third group, lacking internally bound NH_4^+, exhibits higher CCS (3200 Å² and higher). The relative abundance of these groups changes with lowering the ionic strength. The first group dominates at 150 mM NH_4OAc, the third group dominates at 50 mM NH_4OAc (Figure 2B), the fourth group dominates at 10 mM NH_4OAc (Figure 2C). Although all four groups are present upto 10 mM ionic strength, the fourth group takes over at 1 mM, along with a few additional high charges (Figure 2D).

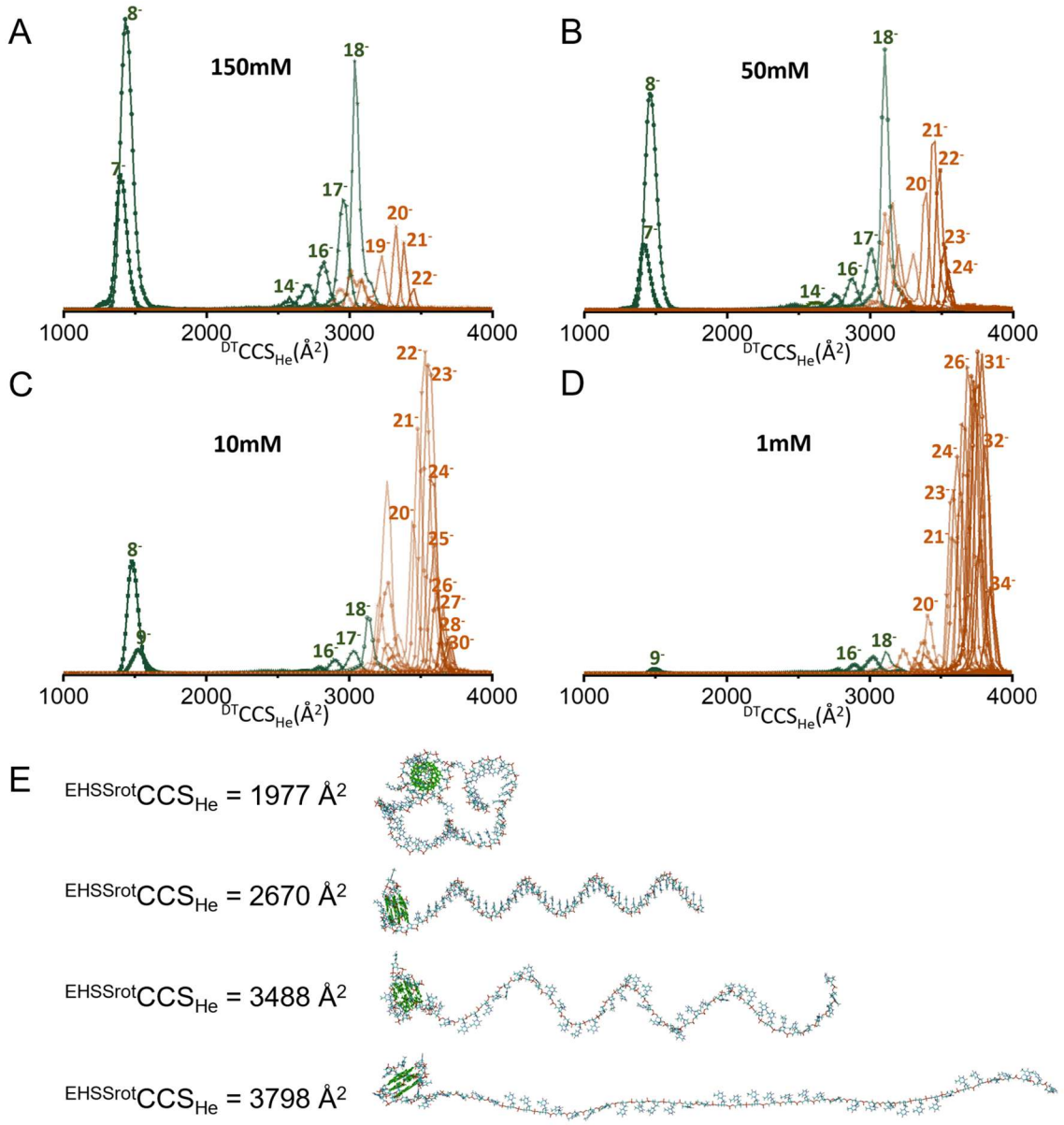

Figure 2. CCSD for the G4Tn oligonucleotide in various ionic strength, 150 (A), 50 (B), 10 (C), and 1 (D) mM aqueous NH_4OAc. Ions with two specifically bound NH_4^+ are in green, and ions without these bound cations are in orange. Some charge states appear both with and without two specific NH_4^+. Panel E shows four different modelled conformations of G4Tn in vacuo, along with their calculated CCS values, demonstrating that a wide range of CCS values are compatible with a preserved G-quadruplex structure.

To rationalize how these distributions can arise from solutions in which the G-quadruplexes are folded, we generated candidate structures for G4Tn: the compact conformations were generated thanks to semi-empirical PM7 calculations, while extended structures were generated using force fields. At physiological ionic strength, G4Tn exhibits four distributions of CCS (Figure 2A). We attribute the low charge state (7- to 9-) conformations to globular structures, wherein the G-quadruplex core and thymine tail remain closely packed. This full compaction may be due to the preservation of the quadruplex structure through native intramolecular interactions and the folded thymines by non-native intramolecular interactions, consistent with the CRM. The compact structure modeled here (8-) has a CCS of 1977 Å², which corresponds to the value of the 10- native state obtained using propylene carbonate (Figure S10). We could not generate a structure compact enough to match a CCS of 1500 Å², and this suggests a additional structural compaction of the polythymine tail onto the G-quadruplex in the gas phase.

After the missing intermediate charge states, the next group of ions is attributed to preserved G-quadruplex regions with extended thymine overhangs. Their CCS value match well with the structure calculated at 2670 Å² (Figure 2E). This kind of bead-and-string structure is compatible with the BEM. Here the extension of the long thymine chain at the end is mainly responsible for the high CCS values. The next CCS distributions correspond to more extended structures without G-quadruplex. These fully extended structures were formed either directly through the CEM, or initially via the BEM but they underwent Coulomb unfolding in the gas phase, leading to the loss of inner NH_4^+ cations. At physiological ionic strength, given that the folded fraction is close to 100% (Figure S1), we think the latter scenario is plausible, and thus that the multimodal distributions arise from a combination of CRM and BEM. Figure 2E also shows a structure with overhang extension yet still containing a G-quadruplex, and its calculated CCS is 3488 Å².

Results are similar at 50 mM aqueous NH_4OAc but the proportion of intact G4s diminishes when [NH_4OAc] is further decreased. The CEM becomes more likely to occur at lower ionic strength and for nonfolded structures. At very low ionic strength (1 mM), the thymine tail is more extended, inner cations are lost, indicating the disruption of the G-quadruplex

core. The gas-phase structure with a overstretched overhang is compatible with the ionization by the CEM. Indeed, IM-MS results showed most of the ions have CCSD ranging from 3500-4000 $Å^2$ at 1 mM ionic strength, confirming the maximum elongation of thymine with G-quadruplex unfolding.

In contrast, in purely aqueous ammonium acetate, TnG4Tn does not display a high-charge state distribution with preserved ammonium adducts (Figure S7). However, experiments under native supercharging conditions (see below), favoring the CRM,[11] enabled the production of 14- and 15- charge states with preserved ammonium ions. The behavior without supercharging agent could be explained by two scenarios. (i) Ions are initially formed via the BEM in a similar fashion but then undergo gas phase unfolding and lose the ammonium ions. It is possible that Coulomb force pulling from both thymine overhangs disrupts the G-quadruplex structure in the center at a comparative low charge like 13- or 14-. (ii) Ion production must start from an overhang and thus by CEM, and once the process is started it keeps unfolding the central G-quadruplex during ion formation. Note that the sequence TnG4Tn has the lowest solution thermal stability (Figure S1), but it should still be almost fully folded at room temperature. Molecular modeling (Figure S8) indicates that unfolding of the centrally located G-quadruplex can also occur before the thymine chains fully extend. Our data does not allow us to choose one scenario over the other, but the ion production mechanism clearly depends on the bead and overhang positions.

Extended structures are prominent in the control NG, even at physiological ionic strength. CCSD produced by high charges indicates the presence of extended structures at 150 mM and 50 mM NH_4OAc (Figure S9). At 1 mM aqueous NH_4OAc, NG exhibits multimodal CCSD centered on 23-, 29-, and 32-, consistent with a fully extended conformation (modeled CCSD: ~3798 $Å^2$). This multimodal distribution might indicate that two mechanisms are at stake: the CEM for the highest charge states and a BEM (although the beads have no set localization) at slightly lower charge states.

For all 60-mers, charge states 9- to 13- are notably absent under native conditions. In our previous work, we reported that adding 0.4% of PC in 150 mM aqueous NH_4OAc solution can induce native SC via the CRM, boosting ion charge states without destroying G-quadruplex structures.[11] Here, adding 0.4% PC likewise reveals otherwise hidden charges

from 9- to 12- for instance for G4Tn (Figure S10), and their CCS increase gradually with the charge state. Furthermore, for TnG4Tn, a low-abundance population of 13- to 15- ions retaining two bound NH_4^+ ions emerges under native SC conditions, as shown in Figure S11. This supports Coulomb-induced unfolding of the overhangs away from the G4 core. This also suggests that the high-charge state distribution observed in native conditions without supercharging agents were produced via a different mechanism.

In summary, IM-MS and molecular modeling confirm that 60-mers containing a single G-quadruplex can adopt conformations consistent with both CRM and BEM. However, the gas phase stability of TnG4Tn is lower than G4Tn, presumably due to the Coulomb force pulling by the two poly(T), which can happen alongside with ion production by the BEM or after ion production by the CRM, as shown in the CIU results of Figure S12).

Pre-folded beads at both termini: G4TnG4 shows unimodal CSD and CCSD in physiological ionic strength, and multimodal distributions at low ionic strength

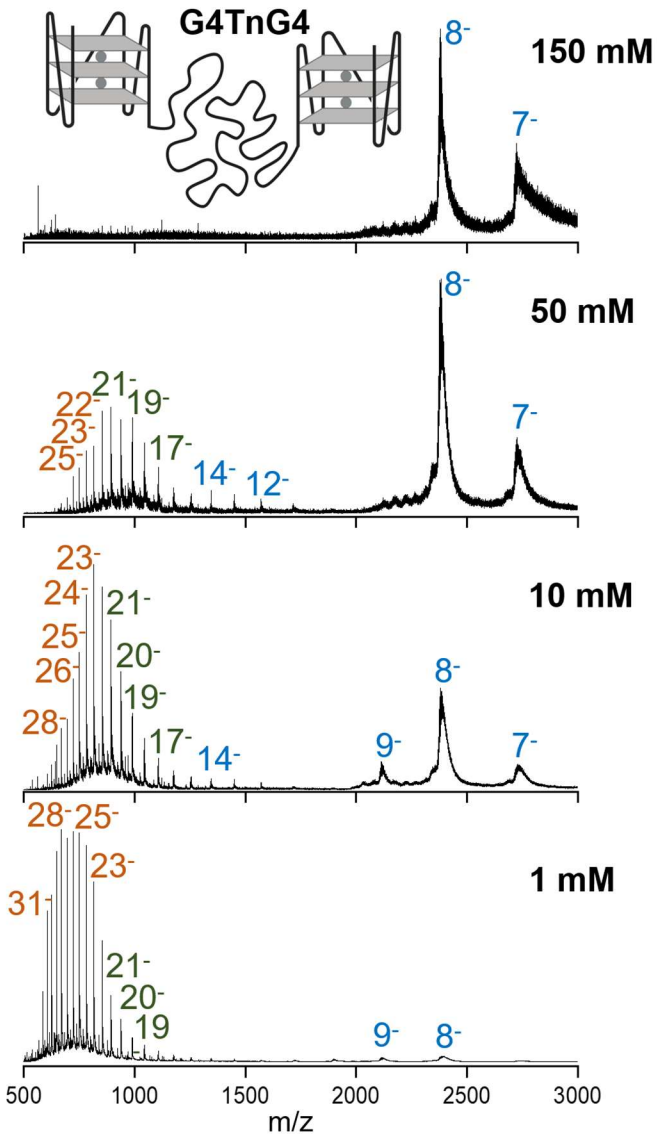

Figure 3. ESI-MS spectra of the G4TnG4 oligonucleotidem comprising one quadruplex at each end, measured at four ionic strengths (150, 50, 10, and 1 mM, top to bottom, 15 µM each). Peaks annotated in blue contain 4 NH_4^+ adducts, in dark green: 2 NH_4^+ adducts, and in orange: no NH_4^+ adducts. Unfiltered spectra are shown in Figure S13.

Unlike single-G4 sequences, G4TnG4, which contains two G-quadruplexes at opposite termini, forms only low charge states 7- and 8- ions at physiological ionic strength (Figure 3). Its unimodal CCSD (around 1500 Å², Figure 4A) corresponds to the complete compaction of the structure. The existence of two quadruplexes at the termini renders the extension of central long thymine chains unfavorable. Even under native supercharging, only 7- to 9- charges were detected in 150 mM NH_4OAc (Figure S14). Thus, the CRM predominates in all cases resulting in a fully compact structure during charging and native super charging conditions at physiological ionic strength. Stacking between the two G-quadruplex structures is possible,[34,35] keeping the overall fold compact in solution and preventing supercharging.

At 50 mM NH_4OAc, we begin to observe both normal compact and partially extended ions. These are even more evident at 10 mM. The first CCSD again corresponds to the fully compact structure and similarly here the second distribution is absent. The next three distributions include conformers that: i) retain both G-quadruplexes with less extended thymine (blue), ii) retain only one G-quadruplex (green), and iii) lack both G-quadruplexes and display extended thymines (brown). The relative abundances of these conformers change between different ionic strengths. Conformations with CCS values of 2500–3000 Å² represent compact structures compatible with formation via the BEM, retaining both G-quadruplex domains, internal cations, and a partially extended thymine linker. A fourth group between CCS values of 3000 Å² and 3200 Å² involves one terminus unfolded while the other remains compact, possibly arising from BEM followed by gas phase unfolding. In some intermediate charge states such as, 18- to 20-, two or three conformations appear, each bearing four, two, or zero bound NH_4^+, illustrating a bead ejection pathway, where one or both G-quadruplex domains remain intact during ion release. The fifth set of conformations (3200 Å² and higher) with both unfolded G-quadruplexes may be formed by the CEM. Overall, the BEM emerges from 50 mM ionic strength and dominates at 10 mM, producing ions that preserve both, only one, or no G-quadruplexes intact in the gas-phase. At 1 mM aqueous NH_4OAc, the fifth conformation shifts to a range of 3400 to 3700 $Å^2$, centered on 29-, suggesting that CEM partially take over.

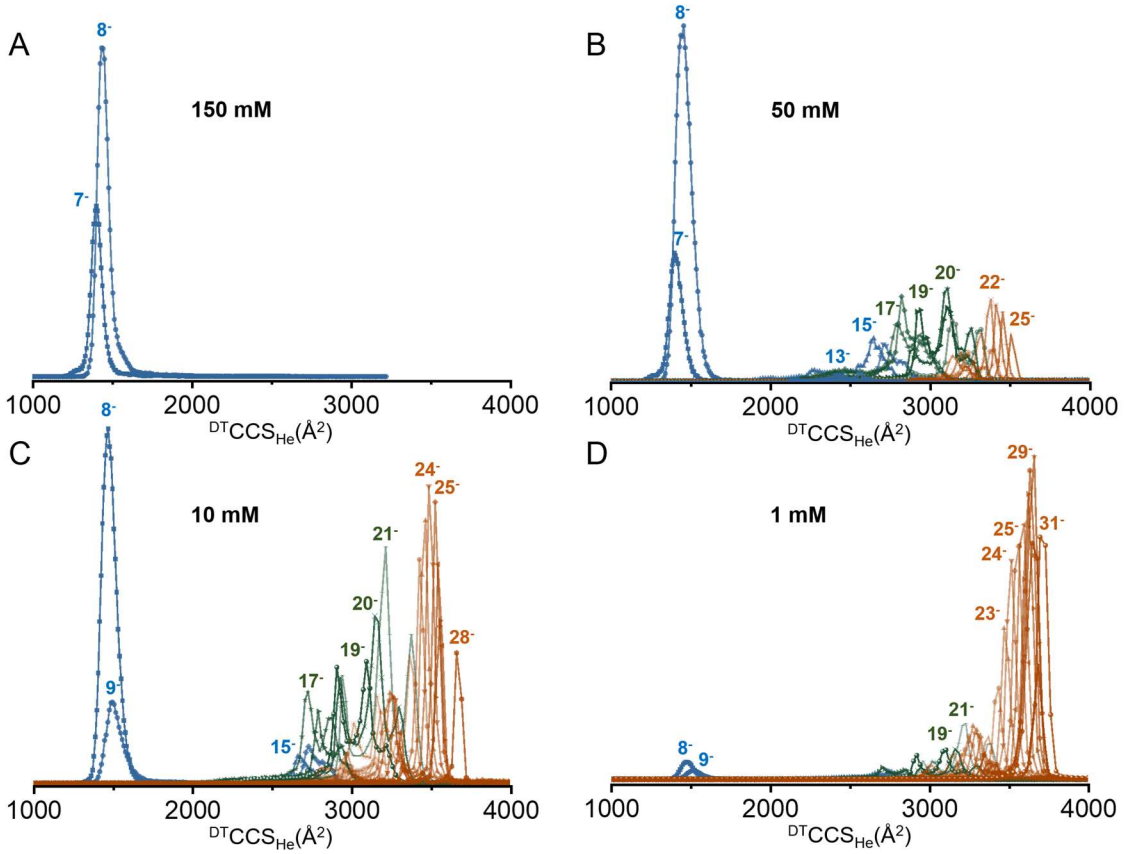

Figure 4. CCSD of G4TnG4 oligonucleotide in various ionic strengths, 150 (A), 50 (B), 10 (C), and 1 (D) mM aqueous NH_4OAc. Ions with four specifically bound NH_4^+ are in blue, two bound cations are in green, and ions without bound cations are in brown.

The competition between CRM, BEM and CEM depends on the position of beads and overhangs

Up to now, ESI mechanisms of oligonucleotide have been explained by a competition between CRM and CEM.[10] The present study with longer sequences containing folded domains and unstructured linkers suggests that the BEM process, which is hybrid between CRM and CEM, can also happen. Figure 5 conceptualizes this continuum.

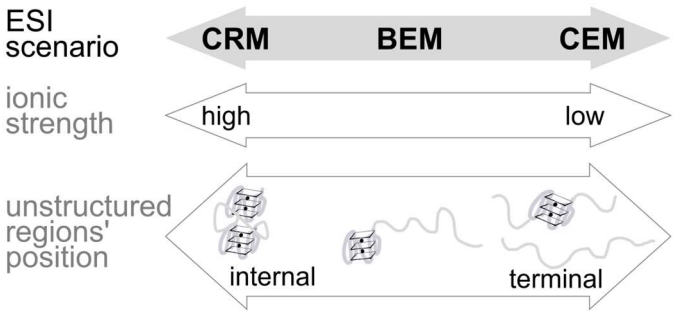

Figure 5. Ionic strength and position of unstructured regions are the major factors influencing the nucleic acids ion production mechanism in negative ion mode. The bead ejection mechanism (BEM) is intermediate and hybrid between the charged residue mechanism (CRM) leading to globular structures and the chain ejection mechanism (CEM) leading to fully extended structures.

High ionic strength favors the CRM while low ionic strength favors the CEM. Unstructured terminal regions favor the CEM, while folded domains on the termini favor the CRM. Folded G-quadruplex domains at both termini favor the CRM, which is the only scenario at stake at physiological (high) ionic strength. However, a higher level of charging and extension becomes possible by lowering the ionic strength, and beads-on-a-string structures appear, which can only be produced by the BEM. When one terminus consists of a G-quadruplex domain and the other by an unstructured region, the BEM is observed even at physiological ionic strength.

Surprisingly, for TnG4Tn, the G-quadruplex domain remains folded only in the low charge states produced by the CRM, which is observed in presence of supercharging agent. None of the peaks at charge states 13- or higher contained inner ammonium ions indicative of a preserved G-quadruplex. Also, the charge states were on average higher with this sequence: CSD and CCSDs were similar to the unstructured control. We reason that non-CRM ionization mechanisms must start from one of the ends, and that once the CEM process has started, it could induce the unfolding during ionization. Alternatively, the BEM could be at stake, but due to greater Coulombic pull from dual poly(T) tails flanking the G-quadruplex core. Given that the CRM can also occur for part of the nonstructured sequences, and that intermediate charge states have intermediate CCS values between globular and fully unstructured, we still conclude that BEM can occur for a fraction of the ions also when unstructured sequences are present at both termini.

Charge-to-phosphate ratios between 0.25 and 0.33 are optimal to discriminate folded domains preserved from solution

Another interesting question is: what charge states are the most useful to obtain information on the solution structures? As reported before, the lowest charge states produced by the CRM are not particularly informative because the ion structures are equally compact and globular. In our 60-mers, these charge states correspond to a charge density (charge-to-phosphate ratio, z/P) of 0.11 to 0.14. In duplexes, such charge density leads to a compaction of the double helix into an almost globular shape (see Figure 51 of reference 2).[2] The highest charge states (21- and more, i.e. $z/P > 0.35$), are equally uninformative, this time because the solution structure is disrupted in the gas phase, as indicated by the loss of inner NH_4^+ cations.

The most discriminative region for our 60-mers consists of charge states 15- to 20- ($0.25 < z/P < 0.35$). Figure 6 shows how the collision cross section of the peak maxima depends on the sequence at those charge states: the structures containing two preserved G-quadruplexes are significantly more compact than those that contain only one, and those that contain none (NG control). This discrimination based on CCS is made possible by the BEM ionization scenario. Interestingly, when G4TnG4 contains only two ammonium ions

and not four at charge states 17- to 20-, it is more compact than G4Tn or TnG4 with two ammoniums, so despite the partial loss of ammonium ions there is a memory of the presence of two folded domains.

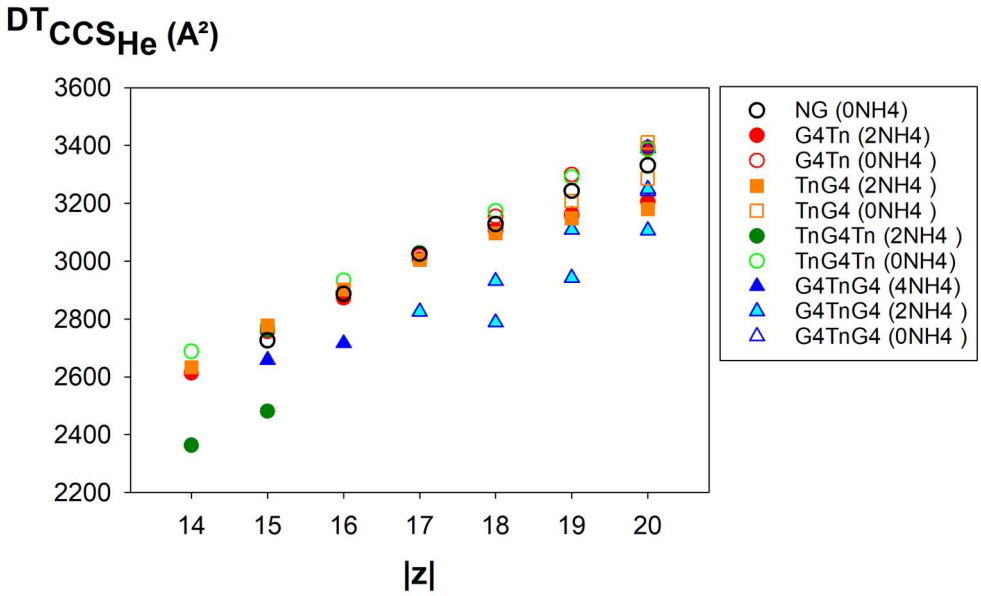

Figure 6. Helium collision cross sections of the 60-mer DNAs, recorded from aqueous 50 mM NH_4OAc, except for TnG4Tn with 2 NH_4^+ 14- and 15-, which was recorded in 150 mM NH_4OAc and 0.4% propylene carbonate. Filled symbols correspond to sequences with all ammonium ions and thus with preserved G-quadruplex structures.

However, those charge states are not always present when infusing the samples from purely aqueous NH_4OAc solutions at physiological ionic strength. Supercharging agents do not always succeed at charging to the optimum level. It can be useful in specific cases, such as TnG4Tn here, to produce some charge states with the CRM and have a chance to observe folded domains in low-energy conditions. In contrast, it is always possible to lower the ionic strength in aqueous solution to generate the most informative charge states for ion mobility analysis, while favoring the BEM.

Conclusions

We presented here several gas-phase structures (elongated but with inner ammoniums indicating preserved G4 folded domains) which can only have a bead on a string structure. Our study confirms that the bead ejection mechanism (BEM), which should be seen as a charged residue scenario for the folded regions and a chain ejection scenario for some of the unstructured regions, can also apply to nucleic acids. Unstructured regions located at the 5' or the 3'-end favor the CEM, whereas the CRM is favored when the strand has folded domains on both termini. In future work, it would be interesting to investigate whether the sequence of the unstructured regions plays a role, and if the model applies to other kinds of folds than G-quadruplexes.

Favoring the BEM scenario is particularly interesting for native MS studies of nucleic acid in solution, because it leads to higher levels of charging. We found that a charge-to-phosphate ratio $0.25 < z/P < 0.35$ is optimal to distinguish the solution folds based on collision cross section studies. The CRM produces too low charge states, wherein folded domains and unstructured regions rather form extra non-native hydrogen bonds with one another,[21] resulting in indistinguishably compact structures. This contrasts with the usual tenet according to which the lowest charge states are always the best choice for native MS.

Charge state tuning can be achieved either through the addition of supercharging agents or through lowering the ionic strength. The latter approach worked better. Adding supercharging agents did not always succeed at producing higher charge states, and the gradual charge state increase observed is more in line with the CRM. However, the results on TnG4Tn, where the G-quadruplex domain is flanked by two unstructured thymine overhangs, show that the BEM sometimes fails at preserving the folded domains. Thus, caution must be exercised when interpreting the results, and having a good set of control sequences is advisable.

Acknowledgements

We acknowledge funding for this project by the ANR (ANR-18-CE29-0013-POLYnESI grant to VG) and the Conseil Régional Nouvelle Aquitaine (PROMIS project to VG).

References

(1) Leney, A. C.; Heck, A. J. R. Native Mass Spectrometry: What Is in the Name? *J. Am. Soc. Mass Spectrom.* **2017**, *28* (1), 5–13.

(2) Largy, E.; König, A.; Ghosh, A.; Ghosh, D.; Benabou, S.; Rosu, F.; Gabelica, V. Mass Spectrometry of Nucleic Acid Noncovalent Complexes. *Chem. Rev.* **2022**, *122* (8), 7720–7839.

(3) Yamashita, M.; Fenn, J. B. Electrospray Ion Source. Another Variation on the Free-Jet Theme. *J. Phys. Chem.* **1984**, *88* (20), 4451–4459.

(4) Iribarne, J. V. On the Evaporation of Small Ions from Charged Droplets. *J. Chem. Phys.* **1976**, *64* (6), 2287–2294.

(5) Aliyari, E.; Konermann, L. Formation of Gaseous Proteins via the Ion Evaporation Model (IEM) in Electrospray Mass Spectrometry. *Anal. Chem.* **2020**, *92* (15), 10807–10814.

(6) Fernandez de la Mora, J. Electrospray Ionization of Large Multiply Charged Species Proceeds via Dole's Charged Residue Mechanism. *Analytica Chimica Acta* **2000**, *406* (1), 93–104.

(7) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013**, *85* (1), 2–9.

(8) Beveridge, R.; Phillips, A. S.; Denbigh, L.; Saleem, H. M.; MacPhee, C. E.; Barran, P. E. Relating Gas Phase to Solution Conformations: Lessons from Disordered Proteins. *Proteomics* **2015**, *15* (16), 2872–2883.

(9) Khristenko, N.; Rosu, F.; Largy, E.; Haustant, J.; Mesmin, C.; Gabelica, V. Native Electrospray Ionization of Multi-Domain Proteins via a Bead Ejection Mechanism. *J. Am. Chem. Soc.* **2023**, *145* (1), 498–506.

(10) Khristenko, N.; Amato, J.; Livet, S.; Pagano, B.; Randazzo, A.; Gabelica, V. Native Ion Mobility Mass Spectrometry: When Gas-Phase Ion Structures Depend on the Electrospray Charging Process. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (6), 1069–1081.

(11) Ghosh, D.; Rosu, F.; Gabelica, V. Negative Electrospray Supercharging Mechanisms of Nucleic Acid Structures. *Anal. Chem.* **2022**, *94* (44), 15386–15394.

(12) Benabou, S.; de Juan, A.; Gabelica, V. Probing the Intramolecular Folding of Nucleic Acids with Native Ion Mobility Mass Spectrometry: Strategies and Caveats. *Anal. Chem.* **2024**, *96* (49), 19277–19285.

(13) Xu, N.; Chingin, K.; Chen, H. Ionic Strength of Electrospray Droplets Affects Charging of DNA Oligonucleotides: ESI Charging of DNA Oligonucleotides. *J. Mass Spectrom.* **2014**, *49* (1), 103–107.

(14) Largy, E.; Marchand, A.; Amrane, S.; Gabelica, V.; Mergny, J.-L. Quadruplex Turncoats: Cation-Dependent Folding and Stability of Quadruplex-DNA Double Switches. *J. Am. Chem. Soc.* **2016**, *138* (8), 2780–2792.

(15) Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D. Understanding Nucleic Acid–Ion Interactions. *Annu. Rev. Biochem.* **2014**, *83* (1), 813–841.

(16) Griffeycor, R. H.; Sasmor, H.; Greig, M. J. Oligonucleotide Charge States in Negative Ionization Electrospray-Mass Spectrometry Are a Function of Solution Ammonium Ion Concentration. *J Am Soc Mass Spectrom* **1997**, *8* (2), 155–160.

(17) Bowers, M. T.; Kemper, P. R.; von Helden, G.; van Koppen, P. A. M. Gas-Phase Ion Chromatography: Transition Metal State Selection and Carbon Cluster Formation. *Science* **1993**, *260* (5113), 1446–1451.

(18) Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. Naked Protein Conformations: Cytochrome c in the Gas Phase. *J. Am. Chem. Soc.* **1995**, *117* (40), 10141–10142.

(19) Wyttenbach, T.; von Helden, G.; Bowers, M. T. Gas-Phase Conformation of Biological Molecules: Bradykinin. *J. Am. Chem. Soc.* **1996**, *118* (35), 8355–8364.

(20) Gabelica, V.; Livet, S.; Rosu, F. Optimizing Native Ion Mobility Q-TOF in Helium and Nitrogen for Very Fragile Noncovalent Structures. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (11), 2189–2198.

(21) Porrini, M.; Rosu, F.; Rabin, C.; Darré, L.; Gómez, H.; Orozco, M.; Gabelica, V. Compaction of Duplex Nucleic Acids upon Native Electrospray Mass Spectrometry. *ACS Cent. Sci.* **2017**, *3* (5), 454–461.

(22) Cantor, C. R.; Warshaw, M. M.; Shapiro, H. Oligonucleotide Interactions. III. Circular Dichroism Studies of the Conformation of Deoxyoligonucleolides. *Biopolymers* **1970**, *9* (9), 1059–1077.

(23) Mergny, J.-L.; Lacroix, L. Analysis of Thermal Melting Curves. *Oligonucleotides* **2003**, *13* (6), 515–537.

(24) D'Atri, V.; Porrini, M.; Rosu, F.; Gabelica, V. Linking Molecular Models with Ion Mobility Experiments. Illustration with a Rigid Nucleic Acid Structure: Ion Mobility Calculations and Experiments. *J. Mass Spectrom.* **2015**, *50* (5), 711–726.

(25) Marchand, A.; Livet, S.; Rosu, F.; Gabelica, V. Drift Tube Ion Mobility: How to Reconstruct Collision Cross Section Distributions from Arrival Time Distributions? *Anal. Chem.* **2017**, *89* (23), 12674–12681.

(26) Do, N. Q.; Phan, A. T. Monomer–Dimer Equilibrium for the 5′–5′ Stacking of Propeller-Type Parallel-Stranded G-Quadruplexes: NMR Structural Study. *Chemistry – A European Journal* **2012**, *18* (46), 14752–14759.

(27) Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr., K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *Journal of Computational Chemistry* **2005**, *26* (16), 1668–1688.

(28) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of *α*/*γ* Conformers. *Biophysical Journal* **2007**, *92* (11), 3817–3829.

(29) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods VI: More Modifications to the NDDO Approximations and Re-Optimization of Parameters. *Journal of Molecular Modeling* **2012**, *19* (1), 1.

(30) Valencia-Sánchez, M. I.; Abini-Agbomson, S.; Wang, M.; Lee, R.; Vasilyev, N.; Zhang, J.; De Ioannes, P.; La Scola, B.; Talbert, P.; Henikoff, S.; Nudler, E.; Erives, A.; Armache, K.-J. The Structure of a Virus-Encoded Nucleosome. *Nat Struct Mol Biol* **2021**, *28* (5), 413–417.

(31) 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* (10), 2002–2010.

(32) Siu, C.-K.; Guo, Y.; Saminathan, I. S.; Hopkinson, A. C.; Siu, K. W. M. Optimization of Parameters Used in Algorithms of Ion-Mobility Calculation for Conformational Analyses. *J. Phys. Chem. B* **2010**, *114* (2), 1204–1212.

(33) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpí, J. L.; González, 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* (1), 55–58.

(34) Do, N. Q.; Lim, K. W.; Teo, M. H.; Heddi, B.; Phan, A. T. Stacking of G-Quadruplexes: NMR Structure of a G-Rich Oligonucleotide with Potential Anti-HIV and Anticancer Activity†. *Nucleic Acids Res* **2011**, *39* (21), 9448–9457.

(35) D'Atri, V.; Gabelica, V. DNA and RNA Telomeric G-Quadruplexes: What Topology Features Can Be Inferred from Ion Mobility Mass Spectrometry? *Analyst* **2019**, *144* (20), 6074–6088.