Probing the Intramolecular Folding of Nucleic Acids with Native Ion Mobility Mass Spectrometry: Strategies and Caveats

Sanae Benabou[1], Anna de Juan[2] and Valérie Gabelica[1,3]*

1. Univ. Bordeaux, CNRS, INSERM, ARNA, UMR5320, U1212, IECB, F-33600 Pessac, France.
2. Chemometrics group. Universitat de Barcelona. Martí i Franquès, 1. 08028 Barcelona.
3. Section of Pharmaceutical Sciences, University of Geneva, CMU Bat. B09, 1 rue Michel-Servet, 1211 Geneva 4, Switzerland.

ABSTRACT: The goal of native mass spectrometry is to obtain information on non-covalent interactions in solution through mass spectrometry measurements in the gas phase. Characterizing intramolecular folding requires using structural probing techniques such as ion mobility spectrometry. However, inferring solution structures of nucleic acids is difficult because the low-charge state ions produced from aqueous solutions at physiological ionic strength get compacted during electrospray. Here we explored whether native supercharging could produce higher charge states that would better reflect solution folding, and whether the voltage required for collision-induced unfolding (CIU) could reflect preserved intramolecular hydrogen bonds. We studied pH-responsive i-motif structures with different loops, and unstructured controls. We also implemented a multivariate curve resolution procedure to extract physically meaningful pure components from the CIU data and reconstruct unfolding curves. We found that the relative unfolding voltages reflect to some extent, but not always unambiguously, the number of intramolecular hydrogen bonds that were present in solution. Reaching phosphate charging densities over 0.25 makes it easier to discriminate between structures, and the use of native supercharging agents is thus essential. We also uncovered several caveats in data interpretation: (1) when different structures (for example the i-motif with and without hairpin) unfold via different pathways, the unfolding voltages do not necessarily reflect the number of hydrogen bonds, (2) unstructured controls also undergo unfolding, and the base composition influences the unfolding voltage, (3) changing the solution pH also unexpectedly changed the unfolding voltage, and (4) the ion mobility patterns become more complicated when two structures are present simultaneously, such as an i-motif and a harpin, because of opposite effects on the collision cross section upon activation.

INTRODUCTION

Native mass spectrometry (native MS) involves recording mass spectra in solution conditions that mimic the physiological conditions, and with instrumental conditions that perturb the non-covalent interactions as little as possible (electrospray ionization and low-energy collisions in the instrument).[1] When the intermolecular non-covalent assemblies are detected intact, the stoichiometries can be read directly from the detected masses. In nucleic acids, our past work focused mostly on DNA duplexes and G-quadruplexes, where either the number of strands or incorporating specific cations provides a mass signature for the secondary structure formation.[2] However, the cases where intramolecular folding is not accompanied by a mass change are more difficult to assess with mass spectrometry.

With proteins, a change of folding is usually indicated by a change in the electrospray charge state distribution: the folded fraction of the population is ionized by the charged residue mechanism, and thus adopt charge states limited by the Rayleigh limit,[3, 4] while the unfolded fraction can be ionized at least partly by chain extrusion,[5] and thus adopts higher charge states.[6-9] The gas-phase compactness can be probed by ion mobility spectrometry, confirming the correlation between charge state and structure.[10] Even if the gas-phase mobility of low charge states at low internal energy cannot differentiate the molecular interactions involved in two molecular systems, one can increase the internal energy before the ion mobility analyzer and investigate at which voltage the structure unfolds.[11] This type of experiment is often called “collision-induced unfolding” (CIU).

In contrast with proteins, nucleic acids electrosprayed in the negative mode from aqueous solutions at physiological ionic strength (150 mM NH_4OAc) induce predominantly low charge states below the Rayleigh limit, independently of whether the strand was folded or not in solution.[12] Ion mobility measurements and comparisons with calculated collision cross sections revealed that DNA as well as RNA single strands, duplex and hairpins present low charge states often more compact in the gas phase than in solution.[13-15] Finally, upon collisional activation, ions at those charge states do not extend in the gas phase, but get more compact.[13, 14] Our problem is thus: how can we analyze nucleic acid structures from aqueous solutions at physiological ionic strength, produce ions that preserve the non-covalent interactions that were present in solution, yet with charge states high enough to undergo unfolding instead of compaction upon internal energy increase?

Here we studied sequences containing i-motif structures, which are secondary structures mediated by hemi-protonated cytosine base pairs (see Figure 1).[16-18] The *in vitro* formation of i-motifs in solution has been demonstrated by different classical spectroscopic techniques,[19] and their *in vivo* presence in human cells and control regulatory functions has been demonstrated.[20] In the gas phase, intramolecular i-motif folding is hard to prove because the charge state represents the global protonation/deprotonation count, but does not tell whether protons are located in-between base pairs. Ion mobility or ion spectroscopy has to be used.[12, 21, 22] The solution stability can be modulated by the sequence (e.g., a hairpin in the loop increases the stability of the entire structure, replacing cytosines by thymines disrupts the i-motif entirely) or by pH changes (the i-motif requires a slightly acidic pH to form), and thus this molecular system is ideally suited to conceive control experiments. We studied here a 34mer C-rich sequence related to the promoter region of the *n-myc* oncogene (Figure 1, Table 1).[23] The wild type sequence shows an i-motif and an unusually long loop containing two complementary TGCA sequences which promote the formation of a stable hairpin. This sequence will be called **IH** (i-motif+hairpin). In addition, we studied 34-mer control sequences, which can form only the i-motif (**IC**), only the hairpin (**CH**) and the last control, which can form none of the ordered structures (**CC**).

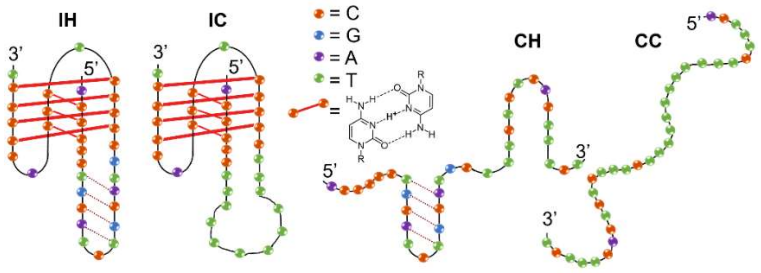

Figure 1. Secondary structure of our 34-mer DNA models at pH = 5.6 (wild type **IH** = i-motif + hairpin; **IC** = i-motif + coil; **CH** = coil + hairpin; **CC** = all coil). At pH = 7.1, the C—H^+—C base pairs and i-motifs are disrupted.

Our goal was to establish whether CIU can be used to assess the preserved intramolecular folding of nucleic acids in electrospray conditions compatible with native-MS, with i-motifs and controls as model systems. To overcome the limitation of compaction at low charge states, we used supercharging agents in low amounts to promote native supercharging.[24] We studied the resulting charge state distributions (CSD)

and collision cross section distribution (CCSD) at low internal energy, and the CIU profiles of all charge states. We varied both the sequence (Figure 1, Table 1) and the pH (at pH = 5.6 the i-motif is formed and at pH = 7.1 it is disrupted). A problem to be surmounted is that current software to describe unfolding processes in the gas phase, such as CUIsuite2[25], allowed to distinguish automatically only two conformations in our molecular systems. To lift this limitation, we applied the method multivariate curve resolution-alternating least-squares to extract two or more physically meaningful pure components from the CIU data, according to the complexity of the unfolding procedure.[26]

Although we found evidence of a memory of the solution hydrogen bonds in the CSD, CCSD or CIU, we also uncovered several caveats in the conception of control experiments, and pitfalls in data interpretation. We hope that their description and discussion will help readers who would face similarly ambiguous situations in their native ion mobility MS data interpretations.

EXPERIMENTAL SECTION

DNA strands

All strands were purchased from Eurogentec (Seraing, Belgium, with RP cartridge – Gold™ purification), dissolved in nuclease-free water from Ambion (Applied Biosystems, Lennik, Belgium), and used without further purification. The stock concentrations were determined using the Beer-Lambert law. The absorbance was recorded at 260 nm on a Uvikon XS, and molar extinction coefficients calculated using the IDT website by applying the Cavaluzzi-Borer Correction.[27]

The 34-mer oligodeoxynucleotides used for the native IMS-MS study are listed in Table 1. dT_6 and dTG_4T sequences were used to verify the instrument each day for the determination of collision cross section (CCS) values ($^{DT}CCS_{He}$ = 306 Å² for dT_6^{2-} (see supporting information of ref. [28]), and 788 Å² for $[(dTG_4T)_4(NH_4)_3]^{5-}$).[29] The $[(dTG_4T)_4(NH_4)_3]$ G-quadruplex was formed in 150 mM ammonium acetate from 200 mM single strand, incubated overnight at 4°C. Final solution used in IMS-MS analysis contained 150 mM NH_4OAc, 2.5 µM dT_6 or 5 µM $[(dTG_4T)_4(NH_4)_3]$. ESI-MS samples were prepared at 10 or 20 µM single strand concentration in 15-, 100- and 150-mM ammonium acetate at pH = 5.6 or 7.1 and let fold overnight at 4 °C. Acetic acid or ammonium hydroxide were used to adjust the pH. The pH was measured using a Multi9420 digital pH meter (WTW, Weilheim, Germany) calibrated daily.

Table 1. 34-mer DNA sequences studied. In the control sequences, some bases were replaced by thymines to disrupt secondary structures.

Name	Sequence (5'→ 3')	Structure at pH = 5.6	Structure at pH = 7.1	# thymines
IH	ACCCCCTGCATCTGCATGCCCCCTCCCACCCCCT	i-motif + hairpin	hairpin	6
IC	ACCCCCTTTTTTTTTTTTCCCCCTCCCACCCCCT	i-motif	coil	14
CH	ACTTTCTGCATCTGCATGCTTTCTCTCACTTTCT	hairpin	hairpin	16
CC	ACTTTCTTTTTTTTTTTTCTTTCTCTCACTTTCT	coil	coil	24

Ion mobility spectrometry-mass spectrometry

Experiments were performed on an Agilent 6560 DTIMS-Q-TOF instrument (Agilent Technologies, Santa Clara, CA), equipped with the dual-ESI source operated in the negative ion mode. The pre-IMS introduction conditions were optimized for softness as described elsewhere.[30] Note that the fragmentor is different for the present work: an electrode was introduced between the capillary and high-pressure funnel with small distance compared to the previous configuration, so the field generated at a given "fragmentor" value is greater than in reference [30]. The drift tube was filled with helium and the pressure was fixed at 3.89 ± 0.01 Torr, measured accurately by a capacitance diaphragm gauge (CDG-500, Agilent Technologies). The following default IMS parameters were used: fragmentor voltage 320 V, trap fill time 1000 µs, trap release time 100 µs, trap entrance grid delta (TEGD) 2 V. For the CCS determination, the step-field experiment included 5 segments (1 min each) where drift tube entrance voltage was –650 V, –700 V, –800 V, –900, V, –1000 V. For CIU experiments, the drift tube entrance voltage was fixed at –650 V, and the activation was performed by varying the fragmentor voltage from 320 to 450 V.

Different supercharging additives (*m*-nitrobenzyl alcohol *m*-NBA, imidazole, propylene carbonate and ethylene carbonate) have been tested at different source temperatures (to reduce the build-up of SC agents). The optimized conditions where no shift of ion mobility profile is observed were 0.25% *m*-NBA at 280 °C. On the course of this work, we noticed that, although the trends reported here were consistently observed, the actual values of CCS and unfolding voltages of extracted components (see MCR-ALS procedure below) were actually very sensitive to the ambient conditions. All results reported here were recorded on final experiments carried out in a single day.

The IMS data were extracted using the IM-MS Browser software version B.06.01 (Agilent Technologies). The arrival time and ion counts were extracted for the *m/z* range of interest (encompassing the isotopic distribution of the non-adducted species) using in-house RStudio script.

To estimate the CCS values, the ion mobility peaks were fitted by Gaussian functions using PeakFit v4.11 (Systat Softwares, San Jose, CA). The arrival time for the center of each peak was determined for each segment and plotted as a function of the inverse of the drift voltage ΔV, to extract the CCS value from the slope according to Equation (1).

$$t_a = t_0 + CCS \cdot \left(\frac{L^2 N_0 \frac{T_0 p}{T p_0}}{\frac{3ze}{16}\sqrt{\frac{2\pi}{\mu k_B T}}} \right) \cdot \frac{1}{\Delta V} \qquad (1)$$

Where t_a is the measured arrival time, t_0 is the time spent outside the drift tube and is deduced from the intercept, L length of the mobility cell (L = 78.1 cm), μ is the reduced mass of the analyte/gas couple, k_B the Boltzmann constant, T the temperature (T = 296.5 ± 1 K), p the pressure in the drift tube (p = 3.89 ± 0.01 Torr), N_0 = 2.687 x $10^{25} m^{-3}$, p_0 = 760 $Torr$ and T_0 = 273.15 K.

Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) extraction of pure components

The ion mobility data as a function of fragmentor voltage were extracted using CIUSuite2.[25] The arrival time and signal intensity were extracted for the *m/z* range of interest using batch files. For our CIU data processing, the intensities in each IMS segment were normalized by peak area for each charge state in a given segment, converted to CCS distributions, and then analyzed by using MCR-ALS to extract the pure components information (pure concentration profiles and pure signal species). MCR-ALS is an iterative self-modelling approach that optimizes **C** (concentration profiles) and $\mathbf{S^T}$ (here CCS distributions) under constraints, which has been successfully applied in numerous fields.[31-33] MCR-ALS has been widely applied for process analysis since it provides pure component spectra and process profiles by using the information coming from the mixture spectra recorded during the evolution of the chemical system.[34, 35] The main advantages of this method are that no prior or little information is needed about the nature and composition of the species involved in the process of interest, the possibility to include complementary and physically meaningful information (such as non-negativity of ion abundances) as a mathematical

constraint, and the possibility to conduct a simultaneous analysis of data recorded along complementary experiments. MCR-ALS aims at resolving the bilinear model $\mathbf{D} = \mathbf{CS^T}$ by using the sole information contained in the raw data set **D**.

The MCR-ALS *modus operandi* can be summarized in the following steps:

1. Determination of the number of components in the raw data set **D**;

2. Generation of initial estimates of either the evolving concentration profiles, **C**, or spectra, $\mathbf{S^T}$;

3. Iterative least-squares calculation of the spectra matrix, $\mathbf{S^T}$, and the matrix of concentration profiles, **C,** under constraints until convergence is achieved (i.e. the LOF difference between two consecutive iterations is below a threshold or a predefined number of iterations is reached).

The percentage of lack of fit (LOF) and explained variance (R2) are used to determine the quality of the resolution model and are calculated by using Eqs. (2) and (3), respectively:

$$LOF = \sqrt{\frac{\sum_i \sum_j (d_{i,j} - \hat{d}i,j)^2}{\sum_i \sum_j d_{i,j}^2}} \qquad (2)$$

$$R^2 = \frac{\sum_i \sum_j \hat{d}_{i,j}^2}{\sum_i \sum_j d_{i,j}^2} \qquad (3)$$

$d_{i,j}$ and $\hat{d}_{i,j}$ are the ij^{th} element of **D** and the ij^{th} element of the reconstructed matrix by the MCR-ALS model, respectively.

More detailed description of the algorithm can be found elsewhere.[34, 35] Briefly, the number of components can be estimated from singular value decomposition (SVD).[36] The procedure needs an initial estimate of either **C** or $\mathbf{S^T}$. The initial **C** matrix can be generated using Evolving Factor Analysis (EFA),[37] which is a local rank analysis method, very appropriate to obtain initial estimates of concentration profiles from sequential evolutionary processes. The resolved profiles in **C** and $\mathbf{S^T}$ are not unique, being subject to rotational ambiguity. Constraints are therefore used to limit the span of feasible solutions and to provide chemically meaningful concentration profiles and resolved spectra.[38, 39] All figures were prepared using Matlab, Excel, SigmaPlot or OriginPro softwares.

Solution spectroscopic tests of solution folding

The solution structure and stability of the different sequences used in this study were confirmed by CD melting and NMR spectroscopies as described previously.[23] CD experiments were performed on solutions prepared following the same protocol as for ESI-IM-MS. CD spectra were recorded on spectrophotometer Jasco J-1500 at 20°C. The measured CD ellipticity (θ, in milidegrees) was transformed to molar circular dichroic absorption ($\Delta\varepsilon$) using Equation (4):

$$\Delta\varepsilon = \frac{\theta}{32980 \times C \times l} \qquad (4)$$

where C is the DNA concentration in mol/L ($C = 2 \times 10^{-5}$ mol/L) and l is the path length in cm ($l = 0.2$ cm). The CD ellipticity of DNA sequences was recorded as a function of temperature using a Jasco J-1500 spectrophotometer. The DNA concentration in the ammonium acetate solutions was 5 or 10 μM. The temperature ramp was 1 °C/min from 4 to 90°C and back to 4 °C. The CD ellipticity was monitored from 220 to 320 nm.

Melting experiments monitored by circular dichroism spectroscopy (supplementary Figure S1 and accompanying text) confirmed that IC and IH are folded at room temperature and pH 5.6, unfolded at pH 7.1, and CC and CH are unfolded at both pH values. In addition, NMR spectroscopy was used to investigate the effect of NBA on the i-motif formation in solutions with ammonium concentrations at 150 mM. The results confirm the formation of i-motif and hairpin in these conditions (See Supplementary Figure S2 and accompanying text).

RESULTS AND DISCUSSION

Charge state distributions (CSDs)

Typical charge state distributions are shown in Figure 2, and further mass spectra are shown in Supplementary Figures S3–S4. In the absence of supercharging agents, charge states 7– to 9– are the least intense, and their abundance is too low for conveniently recording CIU curves. The three sequences having a secondary structure (**IH** with i-motif + secondary structure, **IC** with only i-motif and **CH** with only hairpin) do not ionize at charge states above 7–, but the fully unstructured sequence **CC** shows a second, high-charge distribution, indicating that a fraction of the population underwent chain ejection from electrospray droplets. Upon addition of *m*-NBA at pH 5.6 (the i-motif in **IC** is folded), the charge state distribution shifts and the 7– and 8– charge states are now the most intense. **IC** and **IH** (folded at pH 5.6) charge up to 13-, while **CH** and **CC** charge up to 16-. The fraction of the high-charge states is larger in the fully unstructured **CC** compared to the hairpin-containing **CH**. When the pH is increased to 7.1, the i-motif unfolds, and all sequences can become ionized with charge states up to 16–. IH and IC at pH 5.6 are the only two molecular systems that could not reach charging up to 16–. In that way, the charge state distribution indicates a memory of the solution folding into an i-motif. Note that if ions are collisionally activated before the IMS region (high fragmentor voltage; 420 V), the high charge states fragment and all charge state distributions become more similar (Supplementary Figure S5).

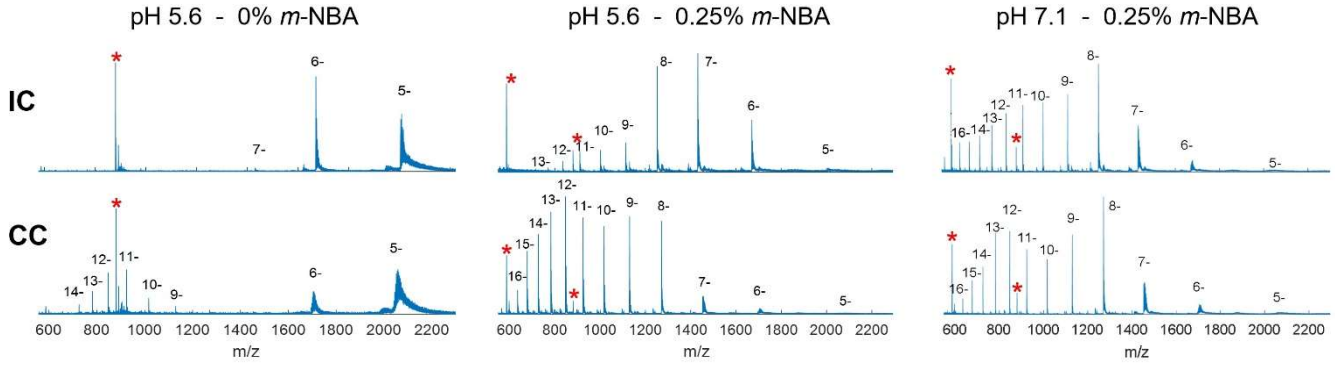

Figure 2. Native electrospray mass spectra and charge state distribution of **IC** and **CC** in 150 mM NH_4OAc adjusted at pH = 5.6 (i-motif formed in IC), without (left) and with (middle) 0.25% vol. meta-nitrobenzyl alcohol (*m*-NBA) or adjusted at pH = 7.1 with 0.25% vol. *m*-NBA. The asterisks indicate dT_6^{z-} internal standard peaks.

Collision cross section distributions (CCSDs) at low collision energy

Let us now examine if the collision cross section distribution of specific charge states is correlated with the solution folding state (Figure 3). From now on we will use the supercharging conditions (the CCS distributions without supercharging are shown in Supplementary Figure S6) because, as we will see below, significant differences in CCS or CIU are only observable at charge states higher than 6–. The CCS ranges of ions from 6– to 8– are similar for the 4 sequences, and similar whatever the pH. Bimodal CCSDs at low collision energy start to appear at charge state 9–.

i-motif structures (**IH** and **IC** at pH 5.6) show monomodal CCSDs at low pH, but when the pH is increased and thus i-motifs are disrupted, the CCSDs become bimodal, with a higher-CCS contribution which increases with the charge state. A bimodal distribution can mean two things: the charge state can be produced both by a charged residue and a chain ejection mechanism,[12] or the charge state is produced by the charged residue mechanism but is already in the middle of a collision-induced unfolding transition.

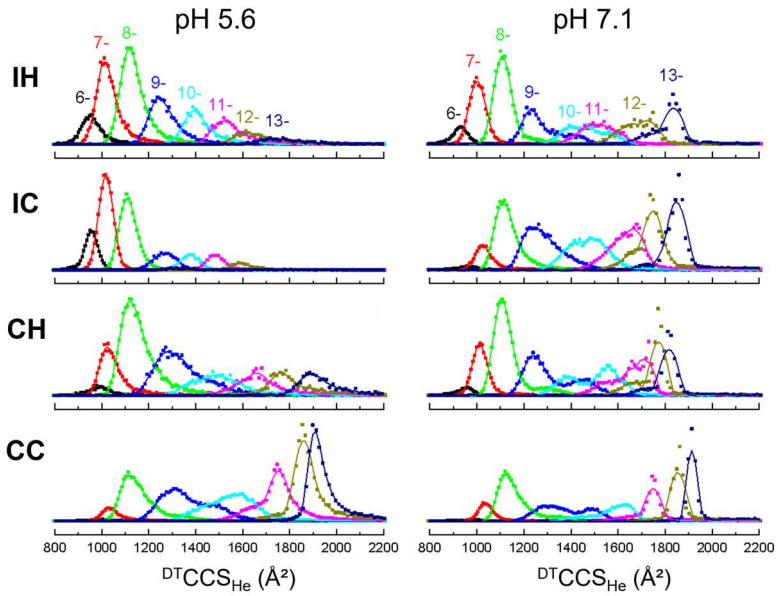

Figure 3. CCS distributions for different charge states (produced from 150 mM NH_4OAc + 0.25% *m*-NBA) of the four sequences pH 5.6 (left, i-motifs formed in **IH** and **IC**) and pH 7.1 (i-motif structures disrupted). All spectra are recorded at the lowest fragmentor voltage (320 V). CCS distributions normalized by height, which better show the shifts for the low-abundant charge states, are shown in Supplementary Figure S7.

However, the monomodal and low-CCS distribution for the folded i-motifs can only mean that they were produced exclusively by the charged residue mechanism. Thus, with 0.25% m-NBA, charge states up to 13– can be produced via the CRM, i.e., by native supercharging. However, if the structure is too unstable either in solution or in the gas phase (e.g., in **CH**, where the hairpin is quite short compared to the disordered region), this does not hold anymore. In summary the CCSD of charge states 9– to 13– show a memory of the solution structure already at low in-source activation voltages, with higher charge states showing a more clear-cut difference. For charge states 7– to 10–, collision-induced unfolding experiments better reveal the differences.

Collision-induced unfolding (CIU)

Let us now examine the entire range of pre-IMS activation voltages. Figure 4 shows the data (collision cross section distributions for different voltages) for the most clear-cut example (sequence **IC**, at pH 5.6 where the i-motif is folded and pH 7.1 where it is unfolded), and the other sequences are shown in Supplementary Figures S8—S10. A visual examination of the data is important to ensure that the mathematical deconvolution of the populations is meaningful. On Figure 4 we can notice: (1) differences as a function of the pH, indicating a memory of the solution i-motif folding; (2) clearer differences at some voltages for the higher charge states. At charge states 11– and 12–, the CCS distribution at low collision energy even indicates full folding at pH=5.6 and full unfolding at pH 7.1, the most faithful reflection of the solution folding state; (3) several CCS distribution are tri-modal; (4) even the sequences that are unstructured in solution show a collision-induced unfolding behavior.

Next, we determined the V_{50} values, i.e., the voltage at which half of the DNA structure is unfolded assuming a two-step conformation, for each condition and charge state using CIUsuite2. A typical result is shown in Figure 5B, and the trends in V_{50} values are shown in Figure 5D. We see the expected trend that **IH** and **IC** are more stable at pH 5.6 than 7.1, and more stable than the **CH** and **CC** controls. However, (1) CIUsuite2 could only distinguish two populations each time, and (2) each CIU curve is treated independently, while in our dataset we postulate that the data for the same sequence in two pH conditions should be processed together, as we suspect that the same gas-phase ensembles should be involved, just in different proportions. As there was more to do with the data, we turned to the MCR-ALS deconvolution procedure.

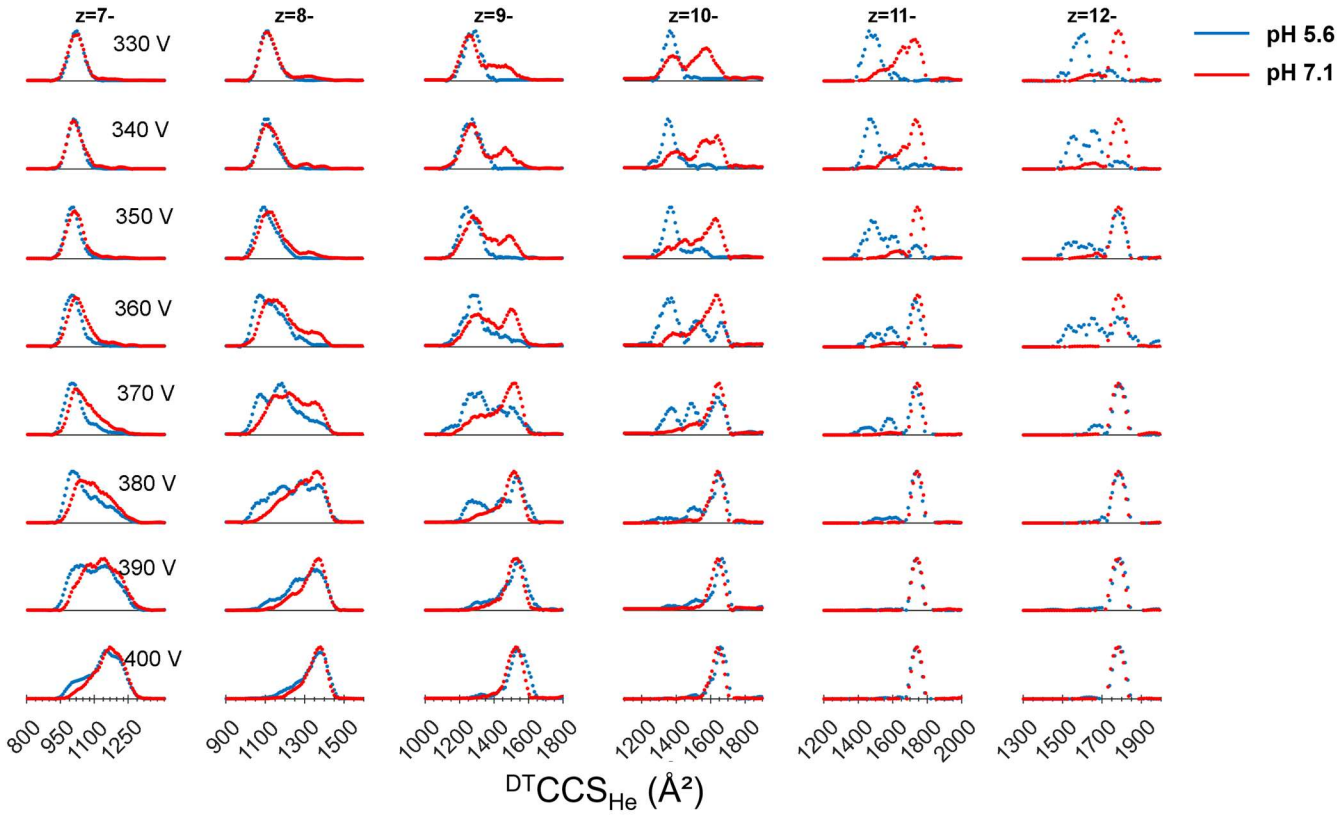

Figure 4. Collision-induced unfolding data: arrival time distribution as a function of the pre-IMS fragmentor voltage, for different charge states of sequence **IC**, prepared at pH 5.6 (blue) or pH 7.1 (red) in 150 mM NH_4OAc and sprayed in presence of 0.25% *m*-NBA.

For each sequence and each charge state, we processed together the data coming from both pH values. We applied to the **C** profiles non-negativity and closure to make the sum of the component fractions for each dataset (at each pH), and we forced the CCS distribution of components ($\mathbf{S}^T$ profiles) to obey a Gaussian shape. The number of components used to perform the MCR model was first suggested by the SVD analysis and was most often two or three. However, in all cases, MCR models with 2 components, 3 components, and, sometimes, 4 components, were tested. The final number of components was adopted as that providing a good model fit and meaningful resolved profiles. MCR models with a larger number of components were discarded when they provided meaningless profiles and were not improving significantly the fit. The resolved component profiles by MCR in this context were the collision cross section distributions of the pure components (normalized by area) ($\mathbf{S}^T$ profiles), and the component fraction as a

function of the fragmentor voltage (i.e., our CIU curves) (**C** profiles) for each pH. Typical outputs are shown in Figure 6, and all of them are shown in Supplementary Figure S11.

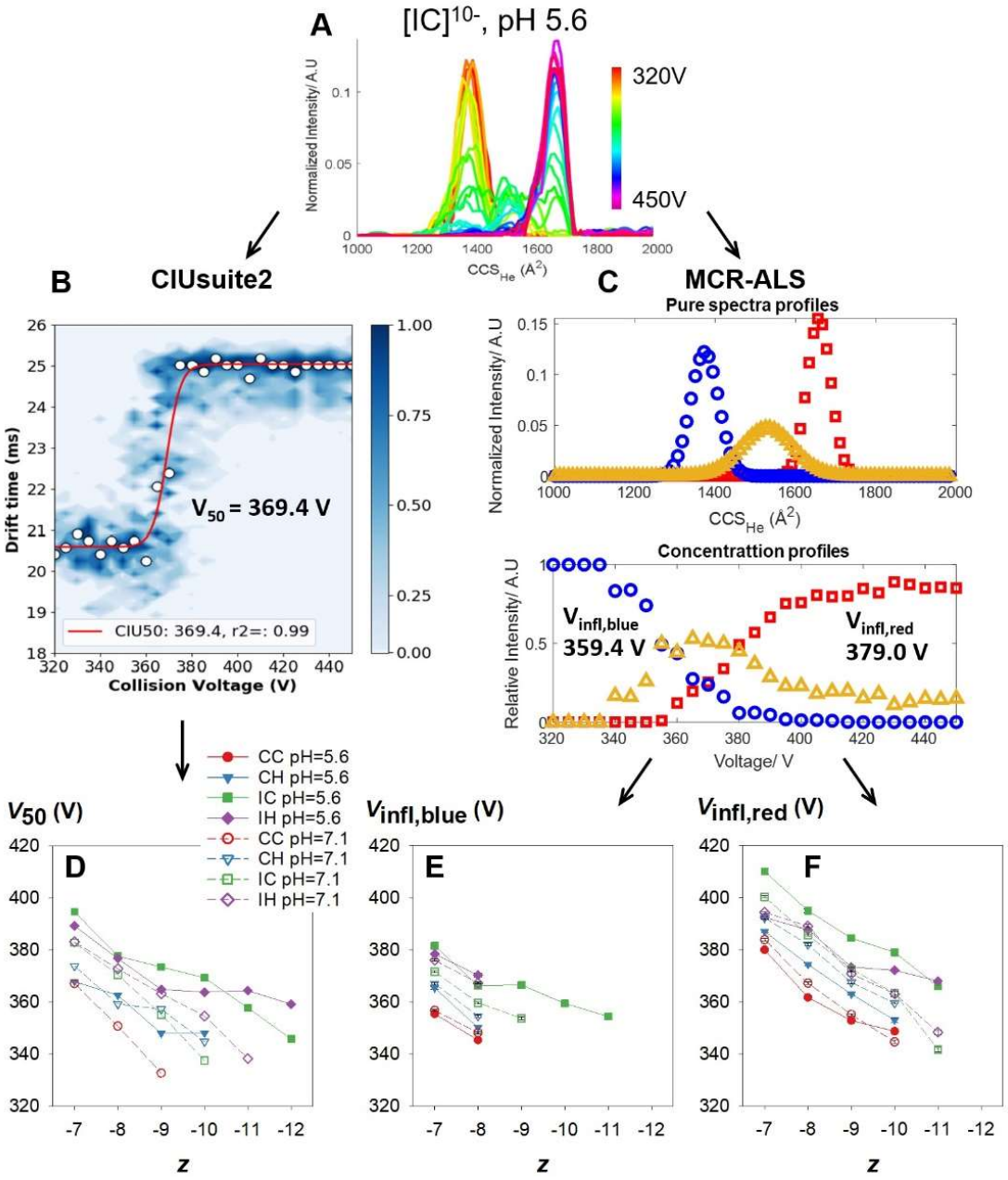

Figure 5. Comparison of MCR-ALS with CIUsuite2 for deconvoluting components of $[IC]^{10-}$ and determining V_{50} values. (A) Original data; (B) CUIsuite2 deconvolution finds two components and the V_{50}; (C) MCR-ALS finds three components and provides their relative abundance as a function of voltage; two V_{50} values were defined as the inflexion point of the abundance of the low-CCS component (blue) and of the high-CCS component (red). Lower panels: V_{50} values as a function of the charge state for the different molecular systems, determined by (D) CIUsuite2 and (E-F) MCR-ALS for the inflexion points of the blue and the red components.

For charge states 7– and 8–, three components are necessary to describe the gas-phase populations, but in the case of **IH**, four components were necessary. Indeed, the low-CCS population undergoes first a slight but detectable compaction, before unfolding to the intermediate-CCS and then to the high-CCS population. This compaction is reminiscent of what happens to (low-charge state) duplexes and kissing loops[13, 14]. For the other sequences, this effect is not visible, and thus to compare all sequences we chose to sum up the two low-CCS components. When there are three components, the low, medium and high-CCS components are colored blue, yellow and red, respectively, a color coding indicating which component predominates in cold or hot ion activation conditions.

For charge states 9– and 10–, except for **IC**, two components sufficed to describe the data (a broad ensemble at lower CCS converting to a sharper ensemble at high CCS). We then wondered to which of the three components previously described they were similar with. We examined their average CCS values (Supplementary Figure S12), and the analysis is as follows. We first plotted the CCS as a function of the base composition (ranked by thymine content) for the three- or four-component decompositions. For each charge state, the average CCS increases when the thymine content increases, in line with the behavior of short homo-base strands.[40] The narrow, high-CCS, high-voltage component is present for all sequences and all charge states, and is thus assigned the red color. The blue and yellow components are present for all sequences at charge states 7– and 8–, and for **IC** at charge states 9– and 10–, providing anchor points to place the unknown broad component. Compared to the anchor points and based on the trend with thymine content (see next section), the average CCS of the unknown component seems intermediate between that of the blue and yellow component. It thus contains contributions from both ensembles and is assigned an intermediate color (green). The ensembles may be undistinguishable because both are broad and close in CCS, because they partially exchange on the time scale of the ion mobility separation, or because the transition between blue and yellow involves both compaction of some parts and extension of other parts of the molecule.

When the nature of the structural ensembles are not exactly the same for different sequences and charge states, one cannot compare V_{50} values defined as the crossing point (50/50 abundance) between curves. We therefore use as a proxy the inflexion point of the red curve in all cases, and the inflexion point of the blue curve (or sum of blue curves) when distinguishable (see Figure 5C). The unfolding curve analysis is therefore more nuanced with MCR-ALS (Figure 5D-E) than with CIUsuite2 (Figure 5D). CIUsuite3,[41] which was released after this work and data analysis was completed, was not tested herein.

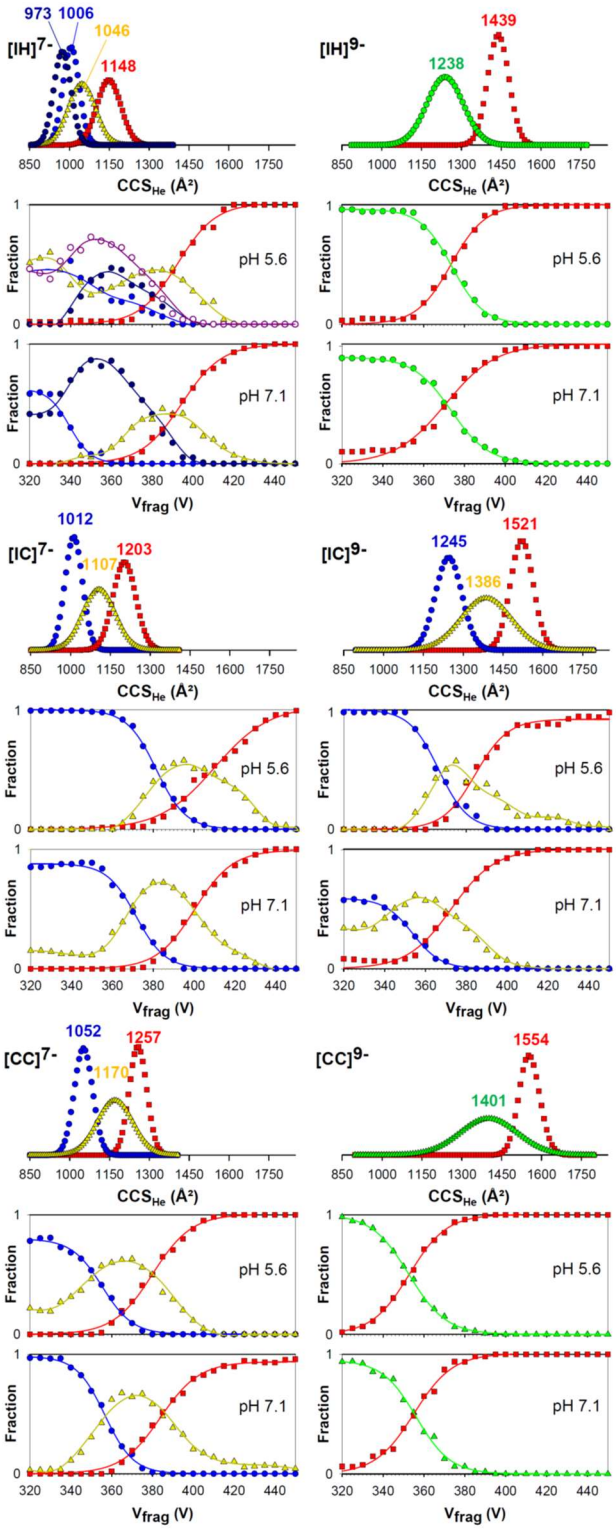

Figure 6. Typical components and unfolding transitions discovered with MCR-ALS. The full set of results is shown in Supplementary Figure S12.

Interpreting the unfolding voltages based on control experiments

Here we have two conditions wherein the i-motif is folded in solution (**IH** and **IC** at pH 5.6), and we have two types of control experiments (mutating the sequence by placing T's instead of bases forming base pairs and changing the pH). Let us start by comparing the unfolding curves as a function of the pH, for each sequence. Here we will compare the inflexion points of the lowest-CCS component (V_{blue}, Figure 7A) and highest-CCS component (V_{red}, Figure 7B), by plotting the difference between the two pH values. If the pH does not influence the structure (**CH** and **CC**), we would expect no difference, yet we found that the V_{infl} are slightly lower at pH 7.1 than pH 5.6. It is a surprising result, indicating that the change in solution composition (addition of acetic acid) slightly changes the droplet desolvation/declustering or ion activation conditions in the source collisional activation region. Reassuringly, the difference for IH and IC is reversed, with the i-motif containing sequences being more stable in the gas phase when produced from acidic solution conditions. But this illustrates a first caveat if one would just examine if there is a voltage difference as a function of pH, without looking at the control sequences without i-motif.

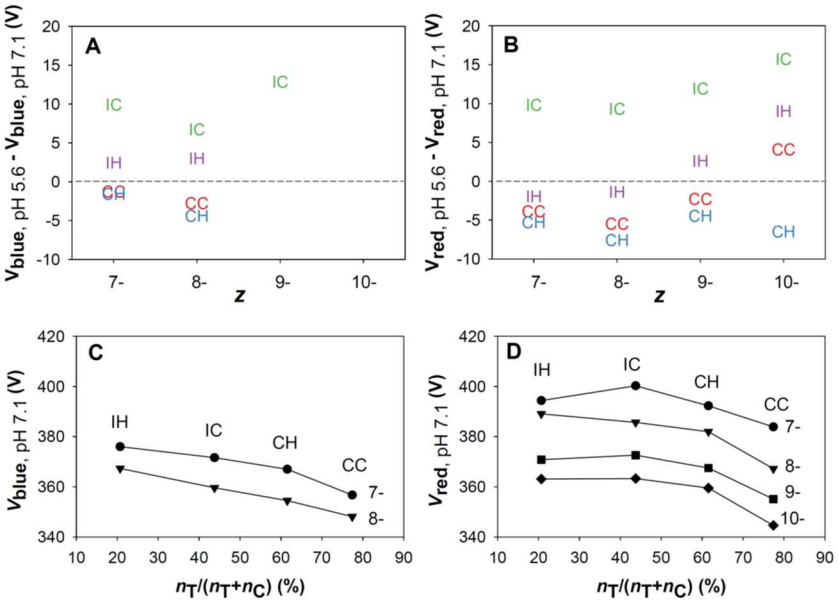

Figure 7. (A-B) Difference of voltage of the inflexion point in disappearance of the blue (low-CCS) component (A) and appearance of the red (high-CCS) component (B). (C-D) Correlation of V_{infl} of each contribution with base composition (expressed in the fraction of thymines compared to total pyrimidines), shown at pH 7.1 where the i-motifs are not formed.

The difference in collisional activation necessary for unfolding is larger for the **IC** (no hairpin) than **IH** (with hairpin). For **IH**, the difference with **CH** becomes higher at high *z*. The fact that IC appears more kinetically stable in the gas phase than **IH** is a surprising result, that seems to contradict the memory of the solution phase structure. However, in the presence of the hairpin, the unfolding path is more complicated: for example, it has 4 components at low charges and the components blend at higher charge. Thus, adding the hairpin secondary structure to the i-motif may change the unfolding pathways. When different pathways are characterized by different activation entropies, one can not necessarily expect that V_{50} or V_{infl} values will reflect the degree of hydrogen bonding, which would translate into a change in activation enthalpy. This is one of the caveats in interpreting CIU data in terms of hydrogen bonding content.

Another interesting observation is that, even among the controls at pH 7.1 (**IH** should have the same structure as **CH,** and **IC** as **CC**), the apparent gas-phase kinetic stability changes; we see a trend when ranking them according to thymine content (lower kinetic stability with higher T content, see Figure 7C-D), and we hypothesize this is linked to the larger CCS in T-rich sequences, itself linked to the fact that there are fewer non-specific intramolecular hydrogen bonds in strands containing thymines.[40] Other unpublished datasets of our lab confirm this trend. In terms of conceiving control experiments to compare CCS, V_{50} or V_{infl} values, this means that replacing bases with only thymines was not such a good idea. The first unfolding event (disappearance of the blue component) is particularly influenced by the thymine content, and we thus suspect that this event is characterized in large proportion by the rupture of nonspecific H-bonds present in the compact forms.

CONCLUSIONS

In summary, the CCSD and CIU data are compatible with the fact that i-motifs are preserved from the solution to the gas phase. Native supercharging was essential to generate higher charge densities at which the differences are clearer. Ions with charge densities of $z = 9$ or 10 per 33 phosphate groups ($z/P = 0.27$ or 0.30) were the most favorable: they are still produced via a charged residue electrospray mechanism and thus preserve the solution interactions in the gas phase, yet the expand upon unfolding at a voltage that reflects the number and strength of preserved intramolecular hydrogen bonds. Ions with z/P of 0.21 or 0.24 underwent unfolding, but the process starts from a more compact state wherein nonspecific hydrogen bonding also contributes (ions with $z/P < 0.2$ did not extend upon collisional activation).

However, our data however revealed several potential pitfalls in the interpretation of CIU curves or unfolding voltages. First, the unfolding process can involve more ensembles than extracted with CIUsuite2, and the MCR-ALS extraction of pure components was key to extract hidden populations. Second, the sole fact of observing unfolding transitions does not indicate that an oligonucleotide was folded in solution (cfr. the **CC** control), only that it was folded in the gas phase, and this can stem from nonspecific hydrogen bonds forming in low charge states upon desolvation/declustering. The *differences* in collision voltages required for a particular transition to occur are relevant, and even here the choice of the control is key. Third, if the unfolding pathways differ between two molecules, one cannot simply assume that the differences in CIU unfolding voltages will simply reflect differences in activation enthalpies, and thus differences in hydrogen bonding content. Hints on different pathways can be gathered from the shape of the CIU curves.

Regarding the choice of the controls, we used control conditions (changing the solution pH to unfold the i-motif) and control sequences (mutating bases to thymines, which disrupts base pair interactions). We found caveats in both approaches, indicating that the difference of unfolding voltage alone was also not sufficient to deduce that extra specific intramolecular bonds were present in solution. In practice, one should further investigate how the nature of the acidifying additive influences the gas-phase unfolding even when the solution structure is unchanged. IMS after mass selection and activation in a collision cell might avoid such problem. When designing control sequences, we would recommend changing the base order rather than replacing many bases with thymines. To cope with this base-dependent effect on CCS and unfolding voltage, future work will be devoted to predicticting the "baseline" CIU curve (starting CCS, end CCS, transition voltages) of unstructured control sequences of any composition, length and

charge state. Deviations from this baseline would indicate additional gas-phase interactions originating from the solution.

ASSOCIATED CONTENT

Supporting Information

Supplementary CD melting, NMR spectroscopy, mass spectra, collision cross section distributions, collision-induced unfolding, and full set of components (CCS distributions) and CIU transitions obtained by MCR-ALS (PDF)

The Supporting Information is available free of charge on the ACS Publications website.

AUTHOR INFORMATION

Corresponding Author

* valerie.gabelica@unige.ch.

Author Contributions

The manuscript was written through contributions of all authors. / All authors have given approval to the final version of the manuscript.

ACKNOWLEDGMENT

This work was funded by the IDEX Bordeaux (project IMOTIMS). We acknowledge the structural biophysical chemistry (BPCS) platforms of the IECB for access to instruments and advice. S.B. and A.J. acknowledge financial support from the Spanish government through the project PID2019-1071586B-IOO.

REFERENCES

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

(2) Largy, E.; Konig, 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) Fernandez de la Mora, J. Electrospray ionization of large multiply charged species via Dole's charged residue mechanism. *Anal. Chim. Acta* **2000**, *406*, 93-104.

(4) Mohimen, A.; Kaltashov, I. A. Estimates of Protein Surface Areas in Solution by Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2005**, *77*, 5370-5379.

(5) Konermann, L.; Rodriguez, A. D.; Liu, J. On the formation of highly charged gaseous ions from unfolded proteins by electrospray ionization. *Anal. Chem.* **2012**, *84* (15), 6798-6804.

(6) Chowdhury, S. K.; Katta, V.; Chait, B. T. Probing conformational changes in proteins by mass spectrometry. *J. Am. Chem. Soc.* **1990**, *112* (24), 9012-9013.

(7) Li, J.; Santambrogio, C.; Brocca, S.; Rossetti, G.; Carloni, P.; Grandori, R. Conformational effects in protein electrospray-ionization mass spectrometry. *Mass Spectrom. Rev.* **2016**, *105*, 111-122.

(8) Beveridge, R.; Covill, S.; Pacholarz, K. J.; Kalapothakis, J. M.; MacPhee, C. E.; Barran, P. E. A mass-spectrometry-based framework to define the extent of disorder in proteins. *Anal. Chem.* **2014**, *86* (22), 10979-10991.

(9) Raab, S. A.; El-Baba, T. J.; Laganowsky, A.; Russell, D. H.; Valentine, S. J.; Clemmer, D. E. Protons Are Fast and Smart; Proteins Are Slow and Dumb: On the Relationship of Electrospray Ionization Charge States and Conformations. *J. Am. Soc. Mass Spectrom.* **2021**, *32* (7), 1553-1561.

(10) Hoaglund-Hyzer, C. S.; Counterman, A. E.; Clemmer, D. E. Anhydrous Protein Ions. *Chem. Rev.* **1999**, *99*, 3037-3079.

(11) Zhong, Y.; Han, L.; Ruotolo, B. T. Collisional and Coulombic Unfolding of Gas-Phase Proteins: High Correlation to Their Domain Structures in Solution. *Angew. Chem. Int. Ed.* **2014**, *53*, 9209-9212.

(12) 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*, 1069-1081.

(13) Abi-Ghanem, J.; Rabin, C.; Porrini, M.; Rosu, F.; Gabelica, V. Compaction of RNA Hairpins and Their Kissing Complexes in Native Electrospray Mass Spectrometry. *J. Am. Soc. Mass Spectrom* **2020**, *31* (10), 2035-2043.

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

(15) Devine, P. W. A.; Fisher, H. C.; Calabrese, A. N.; Whelan, F.; Higazi, D. R.; Potts, J. R.; Lowe, D. C.; Radford, S. E.; Ashcroft, A. E. Investigating the Structural Compaction of Biomolecules Upon Transition to the Gas-Phase Using ESI-TWIMS-MS. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1855-1862.

(16) Irving, K. L.; King, J. J.; Waller, Z. A. E.; Evans, C. W.; Smith, N. M. Stability and context of intercalated motifs (i-motifs) for biological applications. *Biochimie* **2022**, *198*, 33-47.

(17) Abou Assi, H.; Garavis, M.; Gonzalez, C.; Damha, M. J. i-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Res.* **2018**, *46* (16), 8038-8056..

(18) Gueron, M.; Leroy, J. L. The i-motif in nucleic acids. *Curr. Opin. Struct. Biol.* **2000**, *10* (3), 326-331.

(19) Leroy, J. L.; Gueron, M.; Mergny, J. L.; Helene, C. Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an i-motif. *Nucleic Acids Res.* **1994**, *22* (9), 1600-1606.

(20) Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D. I-motif DNA structures are formed in the nuclei of human cells. *Nature Chem.* **2018**, *10* (6), 631-637.

(21) Rosu, F.; Gabelica, V.; Joly, L.; Gregoire, G.; De Pauw, E. Zwitterionic i-motif structures are preserved in DNA negatively charged ions produced by electrospray mass spectrometry. *Phys. Chem. Chem. Phys.* **2010**, *12*, 13448-13454.

(22) Garabedian, A.; Butcher, D.; Lippens, J. L.; Miksovska, J.; Chapagain, P. P.; Fabris, D.; Ridgeway, M. E.; Park, M. A.; Fernandez-Lima, F. Structures of the kinetically trapped i-motif DNA intermediates. *Phys. Chem. Chem. Phys.* **2016**, *18* (38), 26691-26702.

(23) Benabou, S.; Ferreira, R.; Avino, A.; Gonzalez, C.; Lyonnais, S.; Sola, M.; Eritja, R.; Jaumot, J.; Gargallo, R. Solution equilibria of cytosine- and guanine-rich sequences near the promoter region of the n-myc gene that contain stable hairpins within lateral loops. *Biochim. Biophys. Acta* **2014**, *1840* (1), 41-52.

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

(25) Polasky, D. A.; Dixit, S. M.; Fantin, S. M.; Ruotolo, B. T. CIUSuite 2: Next-Generation Software for the Analysis of Gas-Phase Protein Unfolding Data. *Anal. Chem.* **2019**, *91* (4), 3147-3155.

(26) de Juan, A.; Tauler, R. Multivariate Curve Resolution: 50 years addressing the mixture analysis problem - A review. *Anal. Chim. Acta* **2021**, *1145*, 59-78.

(27) Cavaluzzi, M. J.; Borer, P. N. Revised UV extinction coefÆcients for nucleoside-5'-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res.* **2004**, *32* (1), e13.

(28) 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*, 12674-12681.

(29) D'Atri, V.; Porrini, M.; Rosu, F.; Gabelica, V. Linking molecular models with ion mobility experiments. Illustration with a rigid nucleic acid structure. *J. Mass. Spectrom.* **2015**, *50* (5), 711-726.

(30) 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.

(31) Jaumot, J.; Escaja, N.; Gargallo, R.; Gonzalez, C.; Pedroso, E.; Tauler, R. Multivariate curve resolution: a powerful tool for the analysis of conformational transitions in nucleic acids. *Nucleic Acids Res.* **2002**, *30* (17), e92.

(32) Navea, S.; Tauler, R.; de Juan, A. Monitoring and modeling of protein processes using mass spectrometry, circular dichroism, and multivariate curve resolution methods. *Anal. Chem.* **2006**, *78* (14), 4768-4778.

(33) Jaumot, J.; Marchan, V.; Gargallo, R.; Grandas, A.; Tauler, R. Multivariate curve resolution applied to the analysis and resolution of two-dimensional [1H,15N] NMR reaction spectra. *Anal. Chem.* **2004**, *76*, 7094-7101.

(34) Tauler, R.; Maeder, M.; de Juan, A. Multiset Data Analysis: Extended Multivariate Curve Resolution. In *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, Second edition ed.; Vol. 2; Elsevier, 2020; pp 305-336.

(35) de Juan, A.; Rutan, S. C.; Tauler, R. Two-Way Data Analysis: Multivariate Curve Resolution, Iterative Methods. In *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, Second edition ed.; Vol. 2; Elsevier, 2020.

(36) Golub, G. H.; Reinsch, C. Singular value decomposition and least squares solutions. *Numer. Math.* **1970**, *14*, 403-470.

(37) Jaumot, J.; de Juan, A.; Tauler, R. MCR-ALS GUI 2.0: New features and applications. *Chemom. Intell. Lab. Syst.* **2015**, *140*, 1-12.

(38) Maeder, M. Evolving Factor Analysis for the Resolution of Overlapping Chromatographic Peaks. *Anal. Chem.* **1987**, *59*, 527-530.

(39) Tauler, R.; Smilde, A.; Kowalski, B. Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution. *J. Chemometr.* **1995**, *9*, 31-58.

(40) Daly, S.; Porrini, M.; Rosu, F.; Gabelica, V. Electronic spectroscopy of isolated DNA polyanions. *Faraday Discuss.* **2019**, *217* (0), 361-382.

(41) Jeon, C. K.; Rojas Ramirez, C.; Makey, D. M.; Kurulugama, R. T.; Ruotolo, B. T. CIUSuite 3: Next-Generation CCS Calibration and Automated Data Analysis Tools for Gas-Phase Protein Unfolding Data. *J. Am. Soc. Mass Spectrom.* **2024**, *35* (8), 1865-1874.

TOC Graphics