Linking molecular models with ion mobility experiments. Illustration with a rigid nucleic acid structure

Valentina D 'Atri, Massimiliano Porrini, Frédéric Rosu, Valérie Gabelica

Linking molecular models with ion mobility experiments. Illustration with a rigid nucleic acid structure

Valentina D'Atri,[a,b] Massimiliano Porrini,[a,b] Frédéric Rosu[c] and Valérie Gabelica[a,b]*

Ion mobility spectrometry experiments allow the mass spectrometrist to determine an ion's rotationally averaged collision cross section Ω_{EXP}. Molecular modelling is used to visualize what ion three-dimensional structure(s) is(are) compatible with the experiment. The collision cross sections of candidate molecular models have to be calculated, and the resulting Ω_{CALC} are compared with the experimental data. Researchers who want to apply this strategy to a new type of molecule face many questions: (1) What experimental error is associated with Ω_{EXP} determination, and how to estimate it (in particular when using a calibration for traveling wave ion guides)? (2) How to generate plausible 3D models in the gas phase? (3) Different collision cross section calculation models exist, which have been developed for other analytes than mine. Which one(s) can I apply to my systems? To apply ion mobility spectrometry to nucleic acid structural characterization, we explored each of these questions using a rigid structure which we know is preserved in the gas phase: the tetramolecular G-quadruplex [dTGGGGT]$_4$, and we will present these detailed investigation in this tutorial. © 2015 The Authors. *Journal of Mass Spectrometry* published by John Wiley & Sons Ltd.

Additional supporting information may be found in the online version of this article at the publisher's web site.

Introduction

Electrospray mass spectrometry (ESI-MS) in native conditions can preserve the structures of biomolecules and separate them according to their mass-to-charge ratios.[1,2] Ion mobility spectrometry (IMS) separates ions according to their size-to-charge ratios in the gas-phase.[3–6] Size is related to the mass (or number of atoms) and to the three-dimensional shape of a molecule. Therefore, by hyphenating IMS to mass spectrometry (IM-MS), one can sort ions according to both mass and shape. The challenge is to decipher structural information from ion mobility experiments.[7–9]

The physical quantity characterizing the shape is the collision cross section (CCS),[10] which will be introduced in more detail in the Section on Ion mobility and collision cross sections. The present tutorial clarifies how to interpret CCS measurements in terms of three-dimensional structure for ions larger than 100 atoms extracted from the solution by electrospray ionization (ESI). The gold standard is to match CCSs experimentally obtained from IMS with CCSs calculated for modelling three-dimensional structures of the ion of interest. We will discuss the factors that can affect the precision (reproducibility) and accuracy in both the measurements and the calculations. Understanding each of these factors is crucial to interpret quantitative matches confidently and assess the meaningfulness of structural assignments.

On the experimental side, we will describe the determination of CCS from traveling wave ion mobility spectrometers and from drift tube ion mobility spectrometers, with example data recorded on commercially available instruments of each type. For structural interpretation, CCS experimental data are compared with CCS calculated data generated independently. We will therefore discuss

how to estimate the error and provide recommendations on how experimental data should be documented.

On the calculation side, we will discuss the choice of the starting model(s), the methods for three-dimensional structure calculation in the gas phase, the sampling of structures in the gas phase and the different methods available to calculate CCSs. To assess the errors and biases pertaining to modelling, we needed a system with limited conformational diversity in the gas phase. We chose the parallel-stranded [dTGGGGT]$_4$ G-quadruplex DNA structure (Fig. 1) because (1) starting atomic coordinates of [dTGGGGT]$_4$ are available from X-ray crystallography,[11–14] (2) the same molecular assembly was studied by nuclear magnetic resonance (NMR) as well and presents the same G-quadruplex core arrangement, although the thymine placement differs, with evidence for T–T interactions,[15–17] and (3) the central guanine-rich core remains very rigid in the gas phase,[18–21] thanks to the preservation of the G-quartet hydrogen bonds and to the coordination of three
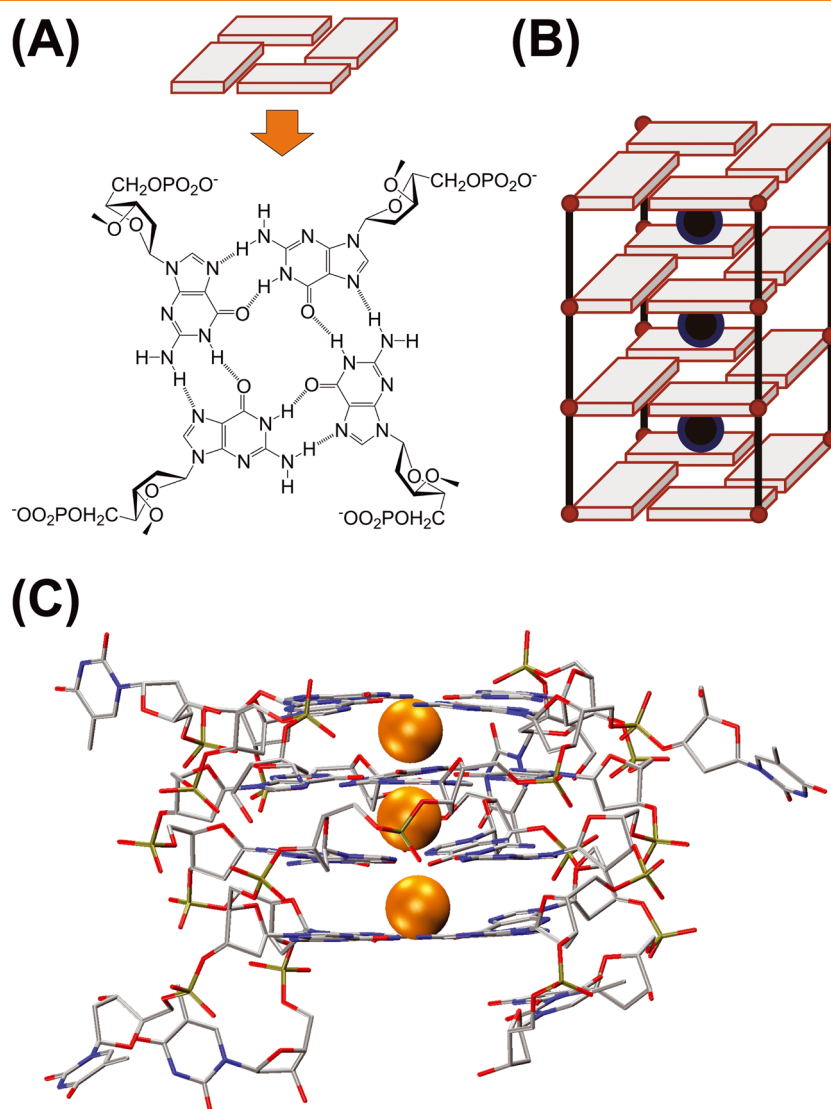
---

* Correspondence to: Valérie Gabelica, INSERM, U869, ARNA laboratory, F-33000 Bordeaux, France. E-mail: valerie.gabelica@inserm.fr

a Univ. Bordeaux IECB, ARNA laboratory Pessac F-33600, France

b INSERM U869, ARNA laboratory Bordeaux F-33000, France

c CNRS UMS3033, IECB Pessac F-33600, France

Figure 1. What is a G-quadruplex? (A) Four guanines forming a G-quartet via Hoogsteen H-bonds, involving N1-O6 and N2-N7 atoms. (B) Sequences containing consecutive guanines form G-quartets core stabilized by the coordination of monovalent cations in-between G-quartets. (C) 3-D crystal structure of the DNA G-quadruplex [dTGGGGT]$_4$; cations shown as orange spheres; PDB reference 352D.[11]

ammonium cations in-between the four guanine G-quartets. Only the position of thymines in the gas phase is less certain.

This ammonium-bound, 24-nucleobase structure (795 atoms and 7504 Da) therefore constitutes a rigid model system in which to assess the structure determination from CCS matching. Two important questions for using IMS-MS in the field of nucleic acid structural biophysics are (1) which model and parameterization of CCS calculation work for nucleic acids of this size, although none of the models was parameterized for nucleic acids specifically, and (2) if ion mobility approaches linked to molecular modelling methodologies are able to detect minor structural changes such as the different positioning of one or two nucleobases from a 24-nucleotide structure.

Ion mobility and collision cross sections

Ions moving in a chamber filled with gas under the influence of an external electric field (E) will reach an apparent steady state drift velocity (v_d) if the collisions are frequent enough to reach a regime where friction forces compensate the electric drag forces. The drift velocity is linearly proportional to the electric field (E), and the proportionality factor is termed the mobility (K):

$$v_d = K \cdot E \tag{1}$$

Because the velocities (v_d) of the ions are influenced by the collisions with the drift gas molecules, K depends on the number density of the drift gas (N). The reported value is the reduced mobility (K_0), normalized to standard pressure and temperature ($p_0 = 760$ Torr and $T_0 = 273.16$ K):

$$K_0 = K \cdot \frac{p}{p_0} \cdot \frac{T_0}{T} \tag{2}$$

In the low field limit (E/N ratio $\leq 2 \times 10^{-21}$ V m^2, or 2 Td), K_0 is independent on the electric field and, based on the gas kinetic theory, can be expressed as follows: [22]

$$K_0 = \frac{3ze}{16N_0} \cdot \sqrt{\frac{2\pi}{\mu k_B T}} \cdot \frac{1}{\Omega_{avg}^{(1,1)}} \tag{3}$$

where z is the nominal ion charge, e is the absolute charge of

an electron, N_0 is the gas number density at p_0 and at T_0 ($N_0 = 2.687 \times 10^{25}$ m^{-3}), μ is the reduced mass of the collision partners ($1/\mu = 1/M + 1/m$, where M is the ion mass and m is the mass of the collision gas), k_B is the Boltzmann's constant, T is the temperature of the gas and $\Omega_{avg}^{(1,1)}$ is the orientationally averaged momentum transfer collision integral. $\Omega_{avg}^{(1,1)}$ is also commonly called the 'collision cross section' and noted Ω or 'CCS' for brevity.

The CCS influences the ion mobility because it represents the effective area for the interaction between the ions and the gas responsible for slowing down the ions by friction.

Section on experimental collision cross section determination explains how to deduce K_0 and therefore the experimental CCS (Ω_{EXP}) from measured drift times. The section also clarifies how measurements are done directly in drift tube IMS cells and how to obtain a $\Omega_{EXP,He}$ from measurements carried out in a nitrogen-filled travelling wave ion guide IMS cell, following proper calibration. On the theoretical side, to determine theoretically CCS (Ω_{CALC}) values, the most representative gas-phase conformations of the ion have to be generated first by molecular modelling (we will discuss this procedure in Section on relating ion mobility spectrometry measurements to molecular structure), and then some approximate methods can be applied on these potential geometries, with the ultimate goal to fit at the best within the Ω_{EXP} values. Most of the approximate methods have been parameterized using helium as collision gas because helium is most likely to behave as a hard sphere upon collision with an ion, but efforts are now made to predict CCS in other gases. A good knowledge of the theory behind the calculation methods guides the proper method choice and is crucial for the interpretation of the matching between Ω_{EXP} and Ω_{CALC} values, as will be discussed in Section on how to obtain theoretical collision cross section from atomistic models.

Experimental collision cross section determination

Introduction: hardware considerations, electric field regimes and collision gas

Knowledge of the instrument hardware is essential to obtain best quality results or to assess the quality of published results. Although we will focus here on the IMS cell, other parts of the mass spectrometer are also important. In particular, all parameters affecting the ion transfer from the solution to the IMS cell can potentially affect the ion structure and the overall result (Ω_{EXP}). For structural biology applications in which the goal is to trace back the solution structure from the gas-phase measurement, soft ionization from native solutions (usually by electrospray) is usually preferred. Besides the experiment, the conception of the modelling workflow also benefits from deeper knowledge of the ionization mechanism[23–25] and of ion activation prior to entrance in the IMS cell (effective temperature of the ions and possible ensuing rearrangements).[26,27] Here, with the aim of assessing the determination of experimental and theoretical CCS, we chose a conformationally rigid molecule that is little affected by instrument tuning. However, the reader should beware that the conformations of floppy molecules can be affected in the gas phase by parameters such as desolvation gas temperature, direct current (DC) and radio frequency (RF) voltages and even molecular charge states.

Ion mobility separation hardware can be of three types:

(1) In *tubes* filled with static gas, the ions move under the influence of electric fields, and the ions transit time in the tube is related to their mobility. IMS tubes are the mobility spectrometry analogues of time-of-flight (TOF) mass spectrometers. We will distinguish constant and low-field drift tube ion mobility spectrometry (DTIMS, Fig. 2A) from travelling-wave ion mobility spectrometry (TWIMS, Fig. 2B). In the latter case, the field is not constant and the regime is not in the low field limit at all times [26,28] and, therefore, calibration with ions of known collision cross section is always required.

(2) *Filters* are filled with gas, and only ions with a given mobility can reach the exit slit of the filter. They are the mobility analogues of quadrupole or sector-type mass analyzers. Filters can operate in the low field limit (like differential mobility analyzers or 'DMA') or involve high fields (like field asymmetric waveform ion mobility spectrometry or 'FAIMS'). In the low field limit, DMA can potentially provide collision cross sections as defined in Eqn (3), [29,30] whereas high-field phenomena in FAIMS are much more complicated.[31]

(3) In *trapping* IMS devices (TIMS), ions face a counter-current gas flow and are trapped, depending on the electric field. Scanning the electric field provides a mobility-selective instability scan, the IMS analogue of mass-selective instability scans in quadrupole ion trap mass spectrometers.[32]
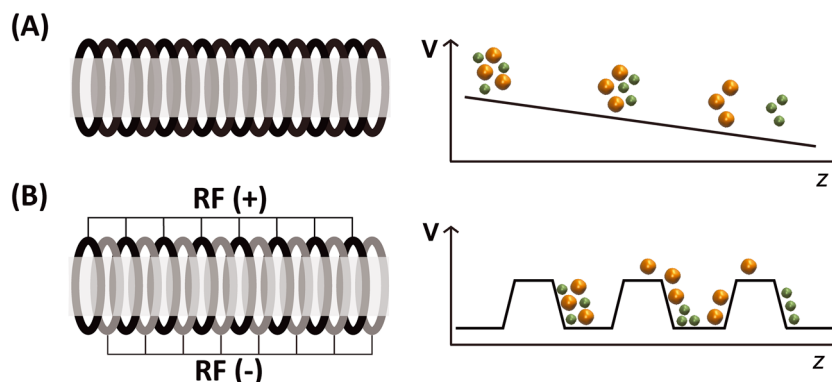
Figure 2. Main difference between (A) drift tube ion mobility spectrometry (DTIMS) and (B) traveling wave IMS (TWIMS). High mobility ions are in green and low mobility ions are in orange. (**A**) In DTIMS, a constant and homogeneous potential gradient is applied along the tube. (**B**) In TWIMS, the ions are confined by a radio frequency (RF) applied to a stacked ring ion guide. In addition, a direct current voltage wave is traveling to the exit (T-wave). Ions of higher mobility are picked up more easily by the waves, whereas larger ions are subjected to larger friction with the gas and slip more often behind the waves and therefore take longer to exit the mobility cell.

Any configuration in the low field limit can provide collision cross sections from first principles (without calibration), provided that the gas flows, pressure and purity are carefully controlled and accurately known and that the electric field is known exactly in all parts of the IMS device. If not, calibration with compounds of known collision cross section remains possible. Here, we describe the two tube-type IMS cells (DTIMS and TWIMS) that are the most commonly encountered in the literature and are commercially available, coupled with high resolution TOF mass analyzers. Commercial instruments are usually specified to work best with nitrogen in the mobility cells, whereas collision cross sections obtained in helium are preferable for matching with calculations. We will therefore also discuss the adaptation of the commercial DTIMS hardware to enable measurements in helium and calibration of TWIMS measurements performed in nitrogen using ions of known collision cross sections in helium.

Drift tube ion mobility spectrometry

Principle of drift tube ion mobility spectrometry

In DTIMS, a static uniform electric field propels the ions in the drift tube (Fig. 2a), and there is (usually) no RF confinement. The electric field is $E = V/L$ (V is the voltage applied across the cell length L). The drift tube is filled with gas, typically helium or nitrogen, kept at constant and known pressure. The drift velocity (v_d) of the ions is determined from the drift time, knowing that $v_d = L/t_{drift}$, where L is the drift cell length, and t_{drift} is the time spent inside the tube. The ions exiting the drift tube are detected as a function of time, yielding an arrival time distribution. All previous equations can be combined with Eqns (1) and (2) to express the arrival time t_A,

$$t_A = \frac{L^2}{K_0} \cdot \frac{T_0}{p_0 T} \cdot \frac{p}{V} + t_0 \qquad (4)$$

which is the sum of t_{drift} (the time spent in the tube) and t_0 (the time spent outside the drift cell and before detection). t_0 is also ion-dependent and must be determined in each experiment through voltage-dependent experiments: a graph of measured t_A as a function of $1/V$ (or of p/V if the pressure changes significantly for each voltage) is linear and the slope is proportional to $1/K_0$. The accuracy of the determined K_0 depends directly on the accuracy with which the cell length L, voltage V, pressure p and temperature T are known. The pressure determination is the most delicate, for several reasons. (1) The pressure gauge can not be placed directly inside the tube, but as close as possible to it so as to reflect the inner pressure. (2) Capacitance diaphragm gauges must be used instead of Pirani gauges for all gases except nitrogen. This is because Pirani gauges are based on a thermal conductivity measurement, which conversion to pressure is calibrated for nitrogen and air. The calibration with other gases greatly diverges above 1 mbar. Capacitance diaphragm gauges measure the pressure on a membrane, and the reading is gas-independent. (3) The gas must be static on average and homogeneously distributed throughout the tube, and care must be given to the design of flow controllers and pumping system for gas regulation inside the tube.

The experimental CCS is then determined from K_0 using Eqn (3). For known analytes, the mass and charge are known accurately. The temperature is determined using a thermocouple plunged in the tube exit. An isolating blanket around the whole tube helps in homogenizing the temperature and minimizing temporal fluctuations. Our instrument is placed in a room with robust air conditioning and the inner temperatures fluctuate by $<0.5\,°C$

day-to-day, $<0.1\,°C$ during a CCS measurement. We however draw the attention on the gas purity. Operating drift tubes in 100% nitrogen is not particularly difficult with ambient sources (for electrospray, the source nebulizing gas is nitrogen, and most commercial instrument operate with counter-current nitrogen flows in their electrospray interface to reduce contamination). However, achieving 100% helium purity in the tube (even if using high purity helium as a supply), while having nitrogen up front, is more difficult and requires some precautions. If helium is contaminated with nitrogen, the ions would be slowed down more than with pure helium and the Ω_{EXP} would be biased and too high.

Materials and methods

Sample preparation

Lyophilized dTGGGGT and dT$_6$ oligodeoxynucleotides were purchased from Eurogentec (Belgium) with RP-cartridge-Gold purification grade. Ammonium acetate (BioUltra ~ 5M) was from Fluka (Sigma-Aldrich, Saint-Quentin Fallavier, France), and nuclease-free water from Ambion (Fisher Scientific, Illkirch, France). Single strand stock solutions were prepared at 1 mM in water. dTGGGGT was annealed at 200 µM single strand (50 µM tetramolecular G-quadruplex [dTGGGGT]$_4$ when fully formed; beware the formation rates are slow)[33] in 100 mM aqueous NH$_4$OAc. Before injection, the samples were prepared at room temperature and injected at a final concentration of ~5 µM quadruplex [dTGGGGT]$_4$ and 5 µM single strand dT$_6$ (added as internal standard).

Drift tube ion mobility spectrometry

Experiments were performed on an Agilent 6560 DTIMS-Q-TOF instrument (Agilent Technologies, Santa Clara, CA), with the dual-ESI source operated in the negative ion mode. The syringe pump flow rate was 180 µl/h. The source and Q-TOF parts were optimized, so as to minimize the fragmentation of a reference oligonucleotide complex $[(dG_4T_4G_4)_2(NH_4^+)_3]^{5-}$.[27] The drying gas was 100 °C or 200 °C (the CCS results were not affected by this temperature). There is no ion injection in this instrument, but a relatively low field (~17 V/cm) between the trap funnel exit and the drift tube entrance. The trap funnel is the region where ion fragmentation is most prominent because of RF heating, and diminishing the trapping funnel RF amplitude from 180 to 90 V improves softness. Fortunately, the abundance and collision cross section of our ion $[(dTGGGGT)_4(NH_4^+)_3-8H]^{5-}$ is insensitive over that tuning range, and it is another reason why we chose this G-quadruplex nucleic acid as a test system.

Experiments were first carried out using the standard configuration of the Agilent 6560 IMS-Q-TOF (Agilent Technologies). The drift tube length is 78.1 cm. For nitrogen experiments, the collision gas pressure was around 4.2 Torr and monitored throughout the experiment with the pirani gauge. The drift tube pressure was 0.14 Torr higher than the trapping funnel pressure. The drift cell exit voltage was 210 V, and the drift cell entrance voltage was ramped from 1000 to 1700 V. For helium experiments, the pumping system was modified by connecting an E2M80 oil pump (Edwards SAS, Gennevilliers, France) to the source region, whereas the original pump was connected to the Q-TOF region. Helium (Alphagaz 1 grade from Air Liquide, H$_2$O <3 ppm, C$_n$H$_m$ <0.5 ppm, O$_2$ <2 ppm) was introduced in the drift tube through the same tubings as in the standard configuration. The pressures in the drift tube and in the front funnel were monitored using capacitance diaphragm gauges (CDG-500, Agilent Technologies). The drift tube pressure was set around 3.70 Torr (measured accurately by the CDG), and

0.23 Torr higher than the trapping funnel pressure, to ensure no nitrogen enters the drift tube. Lower pressure differential resulted in higher collision cross sections, indicating nitrogen contamination.

Experiments were also repeated with the 'Alternate Gas Option' added to the standard configuration. Here, CDGs are connected to the trapping funnel and to the drift tube for all measurements. An additional flow controller admits gas in the trapping funnel, and the flow controller is regulated by a feedback reading of the pressure in the drift tube. We connected a second Tri-scroll 800 pump (Agilent Technologies) 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. This configuration stabilizes the drift tube pressure. For nitrogen and helium measurements, the pressure in the drift tube was 3.95 ± 0.01 Torr for helium and 3.65 ± 0.01 Torr for nitrogen. The actual pressure can vary and the results would not change, but the pressure differential between the drift tube and the trapping funnel must always be at least 0.18 Torr.

The data were analysed using the IM-MS Browser software version B.06.01 (Agilent Technologies), which measures the centroid arrival time of a selected peak as a function of the drift voltage and performs all calculations described in Eqns (1–4). The automatic calculations were checked by manually integrating the peaks using PeakFit v4.11 (Systat Softwares, San Jose, CA). The CCS obtained in the two different instrument configurations were the same, but the second configuration makes it easier to reach the same pressure differential every day.

Results

Figure 3A shows the mass spectrum of $[dTGGGGT]_4$, acquired in negative ESI and ion mobility mode, with helium drift gas. The m/z confirms that the tetramolecular complex contains three ammonium ions at all charge states. The most abundant is $5-$ ($m/z = 1499.8$), meaning that eight negative charges sit presumably on phosphate groups and three positive charges on the inner cations specifically trapped in-between G-quartets. Figure 3B shows the arrival time distribution of $[d(TGGGGT)_4 + 3NH_4]^{5-}$ at a drift cell entrance voltage of 800 V. Figure 3C shows the linear regression of the average arrival time t_A as a function of $1/V$, which was used to determine Ω_{EXP} and t_0. The error from a single regression is on the order of 1 A^2, and manual peak fitting to obtain the average t_A was helpful in reducing this source of error. CCS distributions can be reconstructed from the arrival time distributions, after determining t_0, using Eqn (5), with t_A and t_0 in seconds, L in centimeter, μ in Daltons:

$$\Omega = (t_A - t_0) \cdot \frac{V}{L^2} \cdot \frac{T}{T_0} \cdot \frac{p_0}{p} \cdot \frac{z \cdot 18500}{\sqrt{\mu T}} \quad (5)$$

In helium, we measured $\Omega_{EXP,He} = 787 \pm 3$ A^2 (average value $\pm$ standard deviation on 10 measurements) without the alternate gas option, and $\Omega_{EXP,He} = 789.1 \pm 1.6$ A^2 (7) with the alternate gas option. Overall average is 788.0 ± 2.5 A^2, and a typical CCS distribution is shown in Fig. 5D. The CCS peak resolution in helium is ~55. For the same nucleic acid complex, Gidden et al. had reported an experimental CCS of 775 ± 15 A^2,[20] measured with a home-made, 5-cm-long drift tube.[34] The average helium experimental values differ by 1.5%. In nitrogen, we obtain $\Omega_{EXP,N2} = 1010 \pm 2$ A^2 (six experiments, see Fig. 5E for a typical distribution) on both instrument configurations, and the CCS peak resolution is ~90.
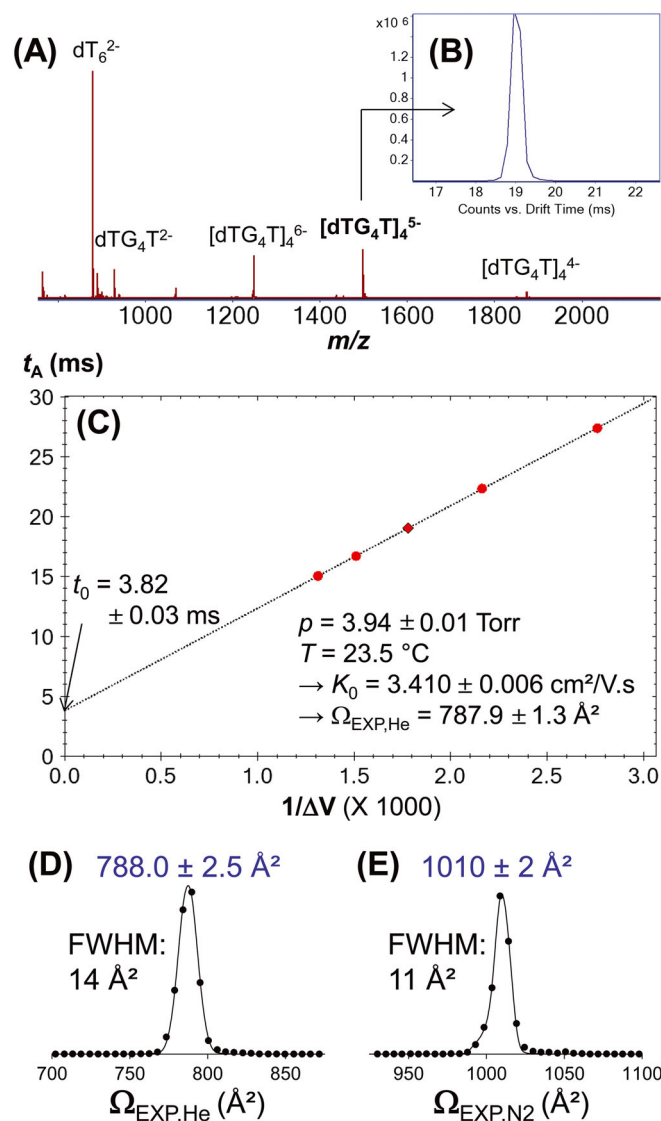
Figure 3. (A) Electrospray ionization mass spectrum of d(TGGGGT) quadruplex in ammonium acetate, sprayed in negative ion mode. dT_6 was an internal standard. (B) Arrival time distribution of $[d(TGGGGT)_4 + 3NH_4]^{5-}$ in helium, with a drift cell entrance of 800 V (drift cell exit $= 210$ V, hence $\Delta V = 590$ V). (C) Linear regression of the average arrival time t_A as a function of $1/\Delta V$ to determine the reduced mobility K_0 and the collision cross section Ω from the slope. (D) Reconstructed collision cross section distributions in helium. (E) Reconstructed collision cross section distributions in nitrogen.

Travelling-wave ion mobility spectrometry

Principle of travelling wave ion mobility spectrometry

Travelling wave ion guides are stacked ring ion guides (SRIG) enabling RF confinement, and DC potentials are applied on some of the rings (Fig. 2B). The DC potentials travel towards the exit of the SRIG. The experimental variables are the wave height and the wave speed, the nature of the gas and its pressure. If the pressure in the SRIG is low (few or no collisions), the device functions as an ion guide. However, if the pressure is sufficiently high that the ions undergo numerous collisions and therefore undergo friction, the time spent in the traveling wave ion guide depends on the ion mobility. Ions are separated according to the transit time in the TWIMS cell because of the fact that collisions slow the larger ions

more than the smaller ions, causing them to exit later in time than the smaller ions as the waves sweep them through the cell.[35,36]

Like in DTIMS, ions with higher mobility exit the TWIMS earlier than ions with lower mobility. However, unlike the DTIMS, analytical equations describing ion movement are complicated to derive because the field is not linear and varies both in time and space. Schvartsburg *et al.*[28] approximated the ion motion using triangular or half-sinusoidal waves and demonstrated that the drift time is related to approximately the Ω^2. In practice, this exponent and the proportionality constant have to be determined by calibration using molecules of known CCS. Eqn (6) is expressed here in a way it can be compared with Eqn (5).

$$\Omega \cdot \frac{\sqrt{\mu}}{z} = A \cdot (t_A - t_0)^B \qquad (6)$$

Factors A and B include the effect of all instrumental conditions (wave height, velocity and profile, cell length, gas pressure and temperature). The manufacturer provides information to calculate t_0 for each ion mass and charge as a function of the instrumental parameters,[37] and there is no mobility-dependent contribution to t_0 by instrumental design (low pressure outside the TWIMS). The calibration involves prior knowledge of CCSs of a range of ions, which cover the transit time range of the ions of interest. The calibration curve to find A and B is usually represented as Ω' (Ω corrected for ion reduced mass and charge) as a function of t'_d (the time spent in the TWIMS):

$$\Omega' = \Omega_{literature} \cdot \frac{\sqrt{\mu}}{z} = A \cdot \left(t'_d\right)^B \qquad (7)$$

If the literature values were obtained in helium, the calibration will provide predicted Ω_{EXP} in helium, even if other gases (typically nitrogen) may have been used in the TWIMS. However, the relationship between nitrogen and helium CCS depends on the molecule size, charge and chemical nature.[38,39] Therefore, calibrating a TWIMS operated in nitrogen using helium CCSs is more reliable using molecules of similar size, charge and chemical nature as the analytes.

Materials and methods

Experiments were carried out both on a first-generation Synapt HDMS (commonly called 'Synapt G1') and a Synapt G2 HDMS. The source and desolvation gas temperatures were 40 °C and 60 °C, respectively. The instruments differ in their TWIMS cell length, pressure, wave profile and optimal heights and velocities. On the Synapt G1, the TWIMS gas was nitrogen at pressure $p = 0.532 \pm 0.002$ mbar, and we used 8-V waves traveling at 300 m/s. Soft ion injection conditions correspond to a bias voltage of 10 V and harsher conditions to a bias voltage up to 30 V. On the Synapt G2, the TWIMS cell is longer, has a helium cell in front (inlet flow maintained at 180 ml/min) and is filled with nitrogen flow regulated so as to read a pressure of 2.760 ± 0.005 mbar on a capacitance gauge (CERAVAC CTR 90, Oerlikon Leybold Vacuum, London, UK) linked to a vacuum controller (Center two, Oerlikon). Soft ion injection conditions were obtained using a bias voltage of 30 V and harsher conditions using a bias voltage of 45 V. The wave profile also differs from the Synapt G1. Optimal separation was obtained using 40 V wave height, travelling at 1000 m/s.

The quadruplex [dTGGGGT]$_4$ was prepared as described in the Section on Materials and methods. Nucleic acid calibrants included dT$_{10}$, injected at 2 μM in 50:50 (v:v) water/methanol ($\Omega_{literature}$ = 446 Å^2 (3−), 537 Å^2 (4−), 627 Å^2 (5−) and 641 Å^2

(6−),[40] and several G-quadruplexes injected at 5 μM strand annealed in 100 mM NH$_4$OAc, and injected in 20% methanol: d (TTAGGG)$_6$ ($\Omega_{literature}$ = 989 Å^2 (5−), 1010 Å^2 (6−), 1074 Å^2 (7−, peak favored at low injection energy and with ammonium ions preserved), 1232 Å^2 (7−, peak favored at high injection energy and without ammonium ion adducts) and 1405 Å^2 (8−, always denatured),[41] d(TTAGGG)$_4$ ($\Omega_{literature}$ = 789 Å^2 (5−, to record only on the species with no ammonium adducts obtained in relatively harsh injection conditions) and 805 Å^2 (6−),[41] dGGG(TTAGGG)$_3$ ($\Omega_{literature}$ = 688 Å^2 (4−), 718 Å^2 (5−),[42] dGAGGGTGGGGAGGG TGGGGAAG ($\Omega_{literature}$ = 701 Å^2 (4−), 757 Å^2 (5−, to record on the complex with two ammonium ions and only in soft injection conditions), 696 (5−, to record on the complex without ammonium adduct and in harsh injection conditions) and 900 Å^2 (6−, denatured),[42] and the dimer of the latter sequence $\Omega_{literature}$ = 1098 Å^2 (7−) and 1113 Å^2 (8−).[43] All literature values are in helium. The calibration curves were generated using Sigmaplot 12.5 (Systat software).

Results and discussion

Figure 4A shows the calibration for the Synapt G2 mass spectrometer (the calibration for the Synapt G1 can be found in the supplementary material of reference[44]). The black line shows the curve fitting using Eqn (7). The blue lines delimit the 95% confidence band: the interval in which there is 95% confidence that the true value lies, assuming that the calibrants literature measurements were not biased. Here, we used calibrant literature data coming from two different labs, and all data fit the same trend, increasing our confidence that no bias is introduced. The red lines delimit the 95% prediction band: the interval in which there is 95% chance to make one or more future observations if the measurement is repeated in a similar way as the given data were sampled (here, with a drift tube similar to those used for measuring the calibrants).

Figure 4B and 4C shows the CCS distributions reconstructed from the arrival time distributions obtained on the Synapt G1 and Synapt G2, using their respective calibration factors. The position of the peak average, obtained by Gaussian peak fitting is virtually the same (within the 95% confidence interval of 5 Å^2), although the experiments were completely independent. This peak position also matches with the DTIMS measurements.

The prediction interval of the TWIMS calibration curve is particularly useful to assess the comparison between TWIMS measurements and measurements carried out on other instruments in other laboratories: measurements carried out independently have 95% chance to fall in this prediction interval. With the current set of calibrants for oligonucleotides, the best prediction interval (in the middle of the curve where there is least extrapolation) is 3% of the average CCS. The previously published literature value (775 ± 15 Å^2) [20] falls in this interval.

The accuracy of CCS data obtained from TWIMS crucially depends on the accuracy of the calibrant data. For oligonucleotides, the number of calibrants selected for calibration is limited. CCS data on other oligonucleotides exist in the literature, and we tried to include them all. However, we could not always reproduce the literature data and in many cases, obtained calibration points that were clear outliers. The reader may have noticed the level of detail in which the peak attribution is described in the Section on Materials and methods. The non-trivial point when calibrating using literature data is that the gas-phase structure and hence the CCS depends on gas-phase activation and possible rearrangements.
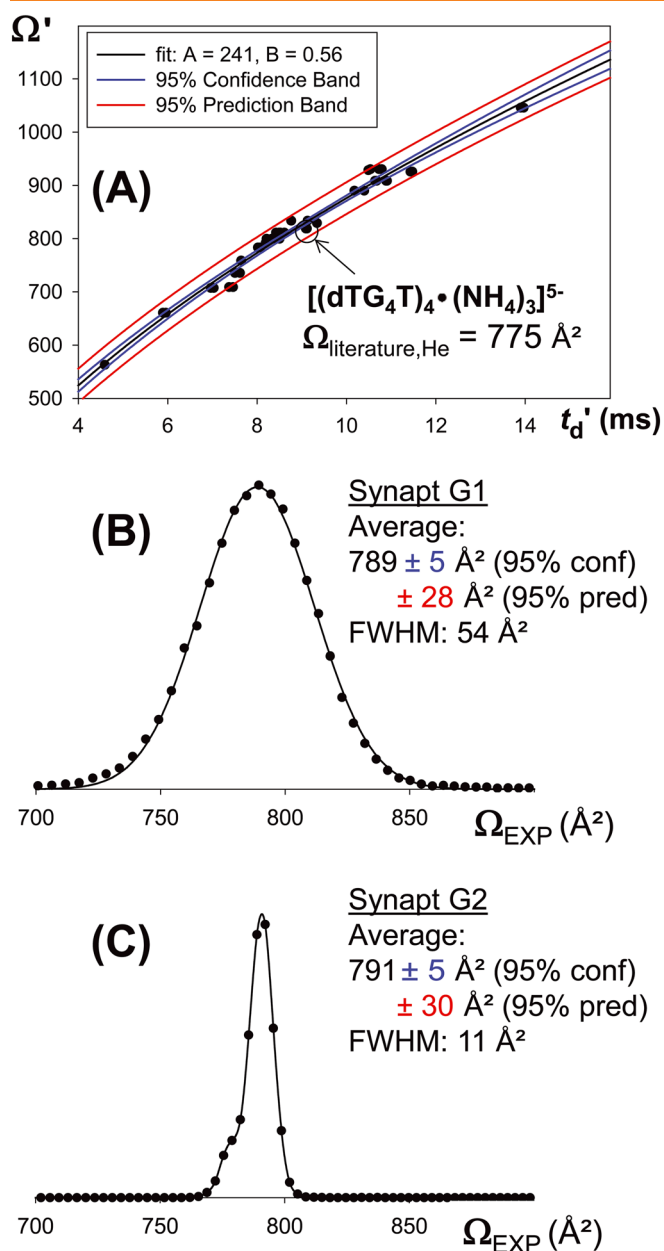
Figure 4. (A): Oligonucleotide-based calibration curve for the Synapt G2 HDMS The datapoint of the structure of interest is circled: it falls outside the 95% confidence band and inside the 95% prediction band. $\Omega_{literature}$ refers to data of reference [20] (B): Reconstructed collision cross section distributions for the ion $[(dTG_4T)_4 \bullet (NH_4)_3]^{5-}$ on the Synapt G1 and (C) Reconstructed collision cross section distributions for the ion $[(dTG_4T)_4 \bullet (NH_4)_3]^{5-}$ on the Synapt G2. Summary collision cross section data of the main peak.

Some of the calibrant structures used here undergo such rearrangement. Prior experience in the drift tube measurement with some structures that served for calibration made it easier to assign the peaks in the TWIMS experiments. Peak assignment is also more secure if exactly the same instrument source and ion transfer geometries are used for the calibration and the experiment. For this reason, a modified Waters Synapt in which the SRIG is operated in a drift tube mode (without travelling waves)[38,45–47] was built and calibration data were recorded. In the future, an inter-laboratory calibration would be useful to identify the most reliable and robust calibrants.

Relating ion mobility spectrometry measurements to molecular structure: how to choose the simulation strategy?

Introduction

The measured CCS distributions reflect the conformational space that molecules explore in the gas phase under the experimental conditions, but IMS data alone do not offer structural interpretation. IMS data can however be interpreted using molecular modelling techniques. The gold standard for structural biology interpretation of IMS data is a satisfactory match between the experimental CCSs and those calculated on a pool of candidate structures generated via computational chemistry techniques. This section explains how to choose or develop a proper simulation strategy to generate structures, and the comparison between calculated and experimental data will be discussed in Section on how to obtain theoretical collision cross section from atomistic models.

Computational approaches

Protein Data Bank (PDB) entries of biomolecules resolved at atomistic level, coming from either X-ray crystallography or NMR techniques, represent a valuable starting point. They can be used either as initial coordinates for the conformational sampling or for first guess of structural assignment after a gas phase energy minimization, or quantum optimization if size-wise applicable. If no atomic-level structure is available, software to manually build initial atom coordinates include, but are not limited to Hyperchem,[48] Discovery Studio,[49] Maestro,[50] NAB[51] and Avogadro.[52]

Computational chemistry methods include quantum mechanical (QM) methods and classical methods. QM methods are capable of describing the electronic structure, thus, they achieve a high level of accuracy, but are typically viable only for small systems. QM methods include *ab initio* geometry optimization methods, such as Hartree-Fock (HF), post-HF like coupled cluster (CC) and perturbation theory methods, density functional theory (DFT) and *semi-empirical* methods (useful for systems ranging from hundreds up to thousands of atoms in size).[53] Then, *ab initio* molecular dynamics (MD) methods, such as Car–Parrinello MD and Born–Oppenheimer MD, are to be implemented in case one is interested into electrons quantum effects on nuclei classical motion.

Conversely, classical methods describe large biosystems properties by a parameterized potential energy surface, and they include the molecular mechanics for optimization procedures, Monte Carlo (MC) and MD techniques for sampling. There are pros and cons in applying MD or MC, and one technique is not better than the other, as both work out ensemble properties. The pro of MD, a deterministic method, is that time-dependent properties can be studied (indeed, the time variable is absent in MC), whereas the pro of MC, a stochastic method, lies in its higher flexibility in sampling regions of the configuration space and, hence, the user can in principle 'decide' to explore certain conformational areas of interest.[54]

A first method selection criterion is the size of the molecular system under investigation. MD using classical methods are preferred for biomolecules, such as proteins, protein/protein complexes and large oligonucleotides, where a satisfactory conformational sampling is more desirable than accounting for electronic structure effects.[41,55,56] MD simulations can however be hampered by multiple local energy minima, where the biomolecule can be trapped. Several enhanced sampling techniques allow the species to overcome energetic barriers and explore other conformational

wileyonlinelibrary.com/journal/jms

basins, for example simulated annealing molecular dynamics,[57–60] replica-exchange molecular dynamics [61–63] and adaptively biased molecular dynamics. [64,65] On the other hand, when the molecular system taken into account is too big, as for instance large multiprotein complexes (>500 kDa), all the aforementioned atomistic methods are still too computationally expensive, but some qualitative models can be built instead, such as (bead-type) coarse grained models.[37,66–69]

Force fields

A possible pitfall intrinsic in classical MD simulations in gas phase is related to the use of force fields (FFs) that are presently parameterized based on high-level quantum calculations, but eventually tested and tuned in aqueous phase (with either explicit or implicit water models). For example, in the development of FFs, such as AMBER[70] and CHARMM[71] (two of the most popular in simulations of biosystems), the charges are derived and/or optimized to be compatible with interactions with the highly polarizing water molecules, and then all the parameters of the remaining contributions (stretching, bending, torsion, etc.) to the FF are generated subsequently. In principle, to run simulations of biomolecules in solvent-free environments, one should recalculate the charges in the gas phase and re-optimize all the FFs parameters, but unfortunately, structural biology experimental data in gas phase against which the comparison should be performed are not available. Therefore, the sampling in gas phase using the available FFs has to always be considered with caution.

Notwithstanding that, Orozco *et al.* have carried out control gas phase MD simulations for DNAs in which they scaled all the AMBER charges by a factor of 0.8 (so that to take into account the depolarization that all the bonds would undergo during the dehydration process), showing that no major differences arise in the final average properties.[30]

Charge location

To generate the charge states detected from the IM-MS spectra, the neutralization of the titrable groups of the macromolecule under investigation is commonly accepted. In the case of proteins, the basic and acidic side chains plus the N- and C-terminus represent the titrable groups that can be neutralized by protons to achieve the desirable net charge state. Therefore, many ways of distributing the charges among these titrable groups are compatible with a specific net charge shown in the ESI spectra. To tackle this issue, different methodologies have been developed, which either take into account how the acidic strength (the pK_a) of the (conjugated) acid (e.g.—$COOH$ and—NH_3^+), varies depending on the surrounding environment, going from residues deeply buried in the protein (very low pK_a) up to residues surrounded nearly by only water (with a pK_a practically equal to that of the lone residue in bulk water),[55,72] or consider energy states resulting from different protonation schemes. [73,74]

In the case of nucleic acids, assuming that all the phosphates can be equally titrable, a different strategy must be used to obtain the net charge state, and two main methods are available in literature: the distributed charge (DC) and the localized charge (LC) method.[75] With the former, the partial charges of all the phosphate groups present in the oligonucleotide are scaled by a proper factor in order to achieve the aimed net charge; with the latter, one neutralizes as many phosphate groups as sufficient to reach the needed total charge by adding one proton to the available oxygen atom. The LC method requires firstly the generation

of all the possible substates (equal in number to $\binom{n}{k}$, where n is the number of phosphate groups and k is the number of protons to be added), and then the selection of the most representative one according to a chosen criterion (usually the lowest in energy). Both DC and LC methodologies applied on DNA double helix [75] and G-quadruplex cores[76] did not give remarkable different results in terms of structures and energetics; thus, in presence of relatively large systems, if the explicit presence of protons is not needed, the DC model is preferred because it is more easily implemented than the LC method. Nevertheless, the LC method is imperative if one is interested in the investigation of possible intramolecular proton transfer reactions.[77]

Materials and methods for [dTGGGGT]$_4$ molecular dynamics simulations

The [dTGGGGT]$_4$ G-quadruplex DNA structure represents a suitable system to assess the structure determination by matching experimental and theoretical CCS. Even though the stability of the G-core was already assessed in the gas phase,[76] we chose to investigate the robustness of the MD trajectories of this system by using diverse starting conformations, differing for the initial positioning of thymine residues. We wanted (i) to define the possible convergence of the simulations to a common ensemble, despite the inclusion of the thymine residues in the MD simulation runs; and (ii) to better evaluate the experimental results, we wanted to appreciate whether and how the positioning of the thymines influences or not the theoretical CCS calculations. To avoid subjective (user-dependent) definition of the starting points, we used an MD simulation trajectory in solution to generate different structures with various thymine positions.

All simulations have been carried out using the AMBER12 suite of programs [78] and applying the *parmbsc0* force field.[79] The *parmbsc0* force field is the improved version of Cornell *et al.* force field [80] and, together with the CHARMM force fields,[81] is the best and most popular force field for nucleic acids.[82] Moreover, *parmbsc0* was already successfully used by Orozco *et al.* for computational studies of diverse nucleic acids structural motifs in gas phase;[44,75–77] therefore, we opted to make use of this force field, which can be readily utilized within the Amber12 suite of programs.

The crystal structure of the parallel [dTGGGGT]$_4$ G-quadruplex (PDB entry 352D)[11] was relaxed performing a 50 ns MD run in explicit water. The module *pmemd*, which uses the Particle Mesh Ewald (PME)[83] method to compute electrostatic interactions, was used and the radial cut-off for the non-bonded interactions was set to 8 Å. This is a crucial step to be applied when dealing with MD simulations in explicit water and periodic boundary conditions: PME method implemented to calculate the pair-wise electrostatic forces increases both computation speed and accuracy with respect to the direct Coulomb summation and, as we shall discuss in the succeeding section, the approach is totally different when the same electrostatic interactions are to be computed in gas phase.

Four snapshots (specifically those at 1, 5, 10 and 50 ns) of this solution simulation were selected to create the starting structures needed for four independent gas phase MD runs, the detailed procedure of which is given in the Supporting Information. Water molecules and counter ions external to the core were removed; then, the charge state of the models was adapted with the predominant observed in ESI-MS experiments (corresponding to—5) by applying the DC model.[76]

Gas phase MD simulations were run with the *sander* module of Amber12, in canonical ensemble (at T = 298 K) using a weak coupling thermostat,[84] implementing a direct Coulomb summation for evaluating the electrostatic interactions, using an integration time-step of 1 fs and no radial cut-off for the non-bonded forces. Because the screening effect of the solvent is missing *in vacuo*, (1) we set no radial cut-off for both Coulomb and van der Waals interactions and (2) the PME algorithm can not be applied. In order to calculate the Ω_{CALC} from the four gas phase MD simulations, snapshots were collected at each 2.5 ns, for a total of 800 candidate geometries (see the last paragraph of the section on how to obtain theoretical collision cross section from atomistic models for the results).

For all the simulations (solution and gas phase), frames were collected every 2 ps and the resulting trajectories analysed with *cpptraj* module [85] in Ambertools13.[86] Pictures were created with VMD software.[87]

Results and discussion

Solution molecular dynamics simulation

The MD simulation in solution yields a stable trajectory, as demonstrated from the low variation of the structural descriptors of the system (Fig S1). The radius of gyration (R_g) accounts for an initial compaction of the structure, which take takes place in the first few nanosecond of the simulation, then remains constant at a value of ~10 Å for the rest of the dynamics (Fig S1, panel **a**). The average backbone root mean square deviation (RMSd) of the G-tetrads (1.2 ±0.2 Å) is significantly lower than that of all the backbone atoms (4.6 ±0.4 Å), denoting the great stability of the stacked G-quartets core. Wobbling effects are experienced by the terminal thymines with an average RMSd of 2.9 ± 0.1 Å (Fig S1, panel **b**). The rigidity of the G-quadruplex core is also highlighted by the root mean square fluctuations (RMSF) calculation: none of the guanines is subjected to fluctuations exceeding 1.2 Å, while thymines are more mobile, with fluctuations ranging from 1.5 to 5.5 Å (Fig S1, panel **c**). To further ascertain the maintenance of the Hoogsteen H-bonding geometry, we also analysed the interatomic distances between guanine N1-O6 and N2-N7 atoms (directly involved in the characteristic G-quartets H-bonds set), which are maintained throughout all the simulation run, (Fig S1**d**). All results suggest that only one major conformational transition occurs during the simulation run: the rearrangement of the thymines nucleobases, which fold above the rigid quadruplex core, in agreement with solution NMR measurements.

The snapshot structures obtained after 1, 5, 10 and 50 ns simulation in solution are shown on the left hand side of Fig. 5. The G-quartet core is basically unchanged, whereas the thymines, located on the outside in the starting point (X-ray crystal structure) close on the top and bottom tetrads within 50 ns. Interaction
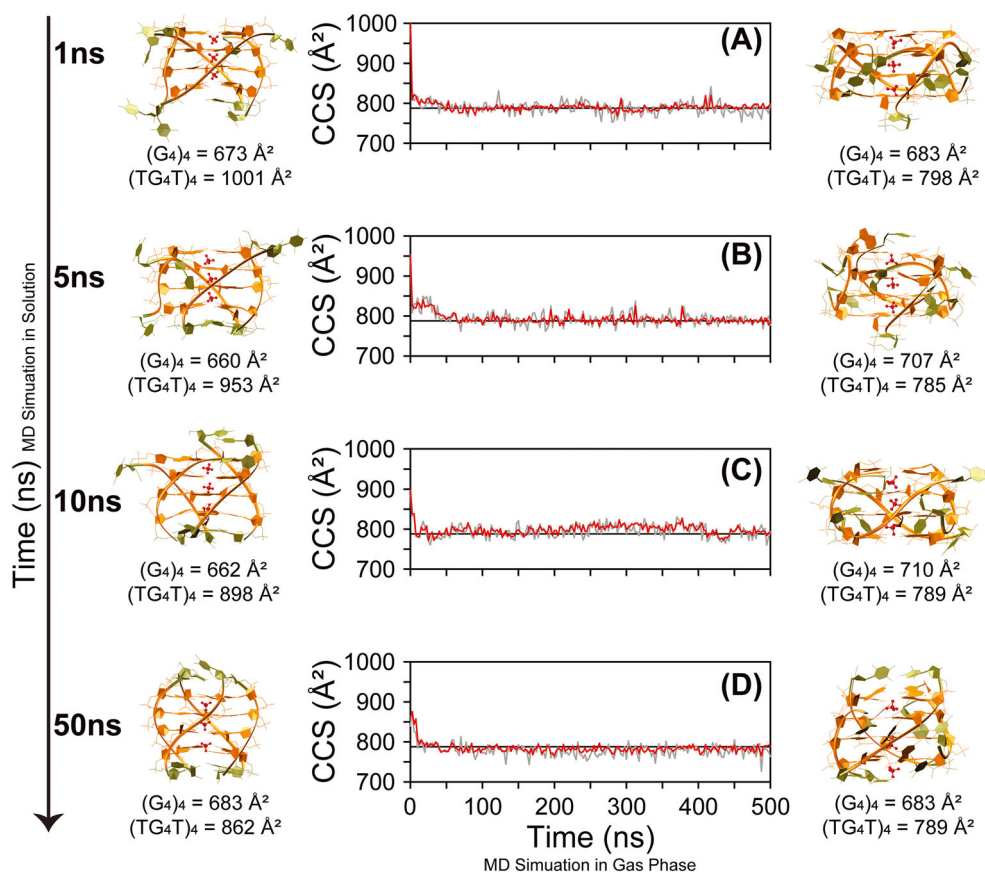
Figure 5. Monitoring of Ω_{CALC} of [(dTGGGGT)$_4$ + 3 NH$_4$]$^{5-}$ for four gas phase 500 ns MD replicas started from different time points of the solution molecular dynamics (MD): after (A) 1 ns, (B) 5 ns, (C) 10 ns, and (D) 50 ns of MD in solution. Red and grey line are to Ω_{CALC} obtained by using *EHSSrot-Siu* and *mobcal-He25*, respectively. The black horizontal line indicates the average experimental value. The starting structures taken from the solution MD and the final structures of the MD simulations in gas phase are shown on the left and on the right, respectively. Furthermore, the Ω_{CALC} of the starting structures and the final structures of the MD simulations in gas phase are indicated below each structure. (G$_4$)$_4$ refers to the Ω_{CALC} calculated for the sole G-core (in orange on the molecular models), while (TG$_4$T)$_4$ refers to the Ω_{CALC} calculated for the whole structure.

between thymines agrees with NMR data. This illustrates that relaxing crystal structures (that could be hampered by lattice effects) in water prior the running of the MD simulations in gas phase is recommended, particularly when the question is whether the solution phase structure is preserved in the gas phase.

Gas phase molecular dynamics simulation

The four snapshots mentioned earlier were selected as starting structures for gas phase simulations. We will refer to the four ensuing gas phase MDs as replicas 1 to 4. Each [d(TGGGGT)$_4$ + 3NH$_4$]$^{5-}$ replica was simulated for 0.5 µs.

The results, summarized in Figs S2–S5, show that all the performed simulations lead to stable trajectories, sharing a common compaction that occurs during the first nanoseconds of the MD simulations, as depicted from the variation of the R$_g$s (Figs S2–S5, panels a). In agreement with the experiments, all the G-quadruplex-ammonium ions complexes are retained; in particular, their positions in the central channel of the structures do not experience fluctuations exceeding 0.7 Å for replicas 1 to 3, and reaching the 1.4 Å for replica 4 (Figs S2–S5, panels c, depicted in grey). The four replicas of the gas phase MD simulation show that (i) memory of the G-tetrads geometry from the starting conformations, as shown from the average RMSDs of the guanines that quickly plateaus at ca. 1.4 Å, and (ii) rearrangements of the thymine residues that occur in the first few nanoseconds of the simulations and are maintained during the entire trajectories, as shown from the average RMSDs of the thymines plateau at ca. 4 Å for replicas 1 to 3, and ca. 6 Å for replica 4 (Figs S2–S5, panels b). Thymine bases interact mostly with the grooves of the quadruplex.

Representative gas-phase structures for each replica are shown on the right hand side of Fig. 5. The [dTGGGGT]$_4$ G-quadruplex

DNA structure is clearly able to maintain its G-core structuring in the gas phase, even though the flipping of the terminal thymines can lead to slightly diverse final conformers.

How to obtain theoretical collision cross section from atomistic models?

Introduction to the different collision cross section calculation models

The experimental CCS measured by IM-MS is the momentum transfer collision integral $\Omega_{avg}^{(1,1)}$ averaged over all possible ion collision geometries defined by the angles θ, ϕ and γ. $\Omega_{avg}^{(1,1)}$ reflects the effect of the collisions on slowing down the ion and therefore depends also on the scattering angle χ (the angle through which the gas is deflected after the collision with the ion). $\Omega_{avg}^{(1,1)}$ is obtained by integrating the momentum transfer cross section over the impact parameter b and relative velocities g:

$$\Omega_{avg}^{(1,1)} = \frac{1}{8\pi^2}\int_0^{2\pi}d\theta\int_0^{\pi}d\phi\,\sin\,\phi\int_0^{2\pi}d\gamma\,\frac{\pi}{8}\left(\frac{\mu}{k_BT}\right)^3\int_0^{\infty}dg\,e^{-\mu g^2/(2k_BT)}g^5$$
$$\times\int_0^{\infty}db\,2b(1-\cos\,\chi(\theta,\phi,\gamma,g,b)) \tag{8}$$

Several computational models (schematized in Fig. 6) have been developed to approximate $\Omega_{avg}^{(1,1)}$, and those computed values will be denoted Ω_{CALC}. The most widely used CCS calculation methods are the trajectory method (TM),[88,89] the exact hard sphere scattering (EHSS)[90] and the projection approximation (PA).[88,90] All were initially parameterized to evaluate Ω_{CALC} with helium collision gas. Several optimizations either of the calculation algorithm or of the
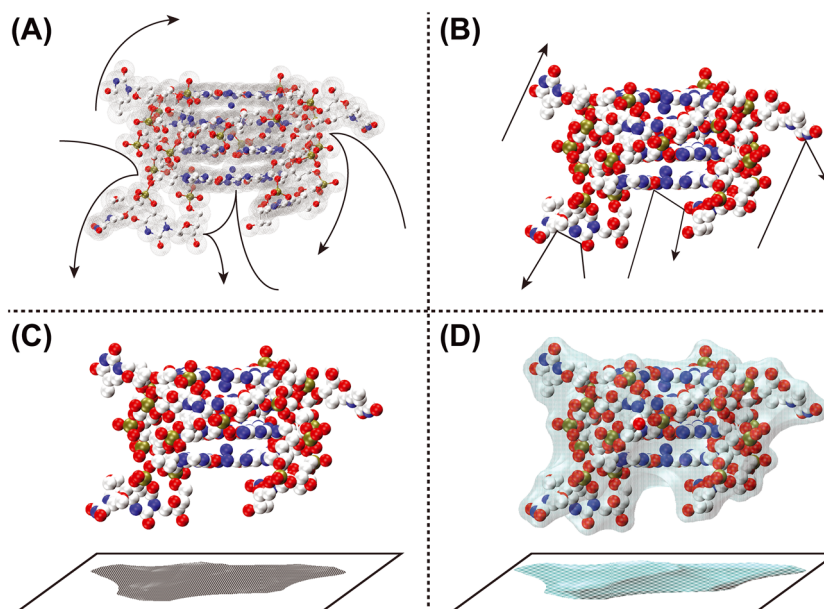
Figure 6. Schematic representation of how the different models calculate Ω_{CALC}. (A): Trajectory model: the long-range interactions between the drift gas and the polyatomic ion are taken into account to determine gas deflection trajectories (shown as arrows) and the scattering angle. (B): Exact hard sphere scattering model: the ion's atoms are considered as hard spheres and the scattering angles between the incoming and departing collision gas atom trajectories (shown as arrows) resulting from the collision with the polyatomic ion are calculated. (C): Projection approximation (PA): the ion's atoms are considered as hard spheres and the average Ω_{CALC} is defined by calculating the projected 'shadow' of the polyatomic ion, averaged over all possible orientations. (D): Projection superimposition approximation: similar to PA, but the ion's atoms are considered as 'soft' spheres, accounting for the collective size effect by superimposition of all atomic contributions, and the resulting value is scaled by a shape factor ρ, which takes into account the shape effects by assessing the molecule's concavities.

parameterization,[45,91,92] including for nitrogen collision gas,[93,94] have been proposed. This section aims at clarifying these variations. Besides, new computational methods are still being developed to provide Ω_{CALC} in a faster and/or more accurate way and in different gases. Among these methods are the projection superposition approximation (PSA) described below,[95–98] the diffuse hard spheres scattering and diffuse trajectory models [99] and the scattering on electron density isosurfaces.[100]

Trajectory model

The TM simulates the trajectories in a realistic potential that accounts for the long-range interactions between the drift gas and the polyatomic ion by evaluating the effective potential of the ion (using Lennard-Jones potentials):

$$\Phi(\theta, \phi, \gamma, b, r) = 4\epsilon \sum_{i}^{n} \left[\left(\frac{\sigma}{r_i} \right)^{12} - \left(\frac{\sigma}{r_i} \right)^{6} \right] - \frac{\alpha}{2} \left(\frac{ze}{n} \right)^2 \left[\left(\sum_{i}^{n} \frac{x_i}{r_i^3} \right)^2 + \left(\sum_{i}^{n} \frac{y_i}{r_i^3} \right)^2 + \left(\sum_{i}^{n} \frac{z_i}{r_i^3} \right)^2 \right]$$

(9)

The first term is a sum of Lennard-Jones two-body 6–12 short-range van der Waals interactions between the collision gas and individual atom i in the polyatomic ion. The second term is the charge-induced dipole interaction. ϵ and σ are the Lennard-Jones parameters (ϵ is the well depth and σ is the distance where the potential become positive), α is the polarizability of the IM collision gas and n represents the number of atoms in the polyatomic ion. This potential is then used to determine the scattering angle of the incoming and departing collision gas atom trajectories.

The evaluation of the Ω_{CALC} with TM is not only the most physically appealing but also the most computationally expensive type of calculation, especially if applied to large polyatomic ions (>1000 atoms). It requires information about the intermolecular potential that should be obtained by fitting measured mobilities as a function of temperature. A TM algorithm is available in the *mobcal* program,[101] and the original parameters were obtained based on ion mobility measurements of carbon clusters in helium collision gas.[88,89]

In 2008, Kim *et al.* proposed a modified TM algorithm and parameterization to calculate the Ω_{CALC} with nitrogen collision gas.[94] They modified the potential of Eqn (9) by taking the Lennard-Jones parameters described in the universal force field [102] and setting the polarizability of neutral gas for N_2, without neglecting the ion–quadrupole interaction and molecular orientation due to the linear geometry of N_2. Based on Kim's modified TM, Campuzano and Bush parameterized both the N_2-based and the He-based trajectory method algorithms by applying a scaling factor to the Lennard-Jones parameters, based on IM measurements on drug-like molecules. The optimized N_2 and He trajectory method source codes, *mobcal_He* and *mobcal_N2*, are provided upon request to the corresponding authors .[45]

Exact hard spheres scattering

The EHSS model assumes that both the individual atoms of the polyatomic ion and the collision gas behave as hard spheres.[10] EHSS computations determine the scattering angles from the trajectories resulting from the gas collision with the polyatomic ion. Therefore, Eqn (8) is approximated by the orientationally averaged hard-sphere CCS Ω,

$$\Omega = \frac{1}{4\pi} \int_0^\pi d\phi \sin \phi \int_0^{2\pi} d\gamma \, \Omega_{dir}(\phi, \gamma)$$

(10)

where φ, and γ are the spatial angles defining the ion orientation and Ω_{dir} is the corresponding directional cross section, which depends on how the interaction between the ion and the collision gas is evaluated. In EHSS, the scattering angle χ of the collision gas is taken into account as follows:

$$\Omega_{dir}(\phi, \gamma) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} [1 - \cos \chi(\phi, \gamma, y, z)] dy dz$$

(11)

An EHSS algorithm is available in the original *mobcal* program,[101] where original atom size parameters are provided again based on IMS measurements on carbon clusters. In 2007, A. A. Shvartsburg proposed *EHSSrot*, a more computationally effective algorithm in which the EHSS calculations are accelerated by reducing the number of ion rotation steps through evaluation of multiple ion/gas collisions per orientation. The *EHSSrot* source code is provided for academic use upon request to the corresponding author. [92] Regarding the atom size parameterization, Siu *et al.* optimized the EHSS parameters by varying the collision radii of He-H and He–X atoms (where X is C, N or O) in order to match IM experiments from several peptide ions, which structures could be computed with high accuracy DFT calculations.[91] Siu's atom parameters can be included either in the *mobcal* algorithm or in the *EHSSrot* algorithm.

Projection approximation

The PA approach calculates Ω_{CALC} by finding the average 'shadow' of the polyatomic ion (onto a defined plane) by rotating the structure through all possible orientations. The projection of the polyatomic ion onto the yz-plane is computed as follows:

$$\Omega_{dir}(\phi, \gamma) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} M(\phi, \gamma, y, z) \, dy \, dz$$

(12)

where M is unity when a collision occurs for the configuration defined by φ, γ, y and z, and null otherwise. PA algorithms are the most simple and computationally effective and can be found in the *sigma* program by Bower *et al.* [103] and in *mobcal*.[101] It is now known that the PA greatly underestimates $\Omega_{avg}^{(1,1)}$ because the gas slow-down effect is greater in cavities than on convex surfaces, and the projection masks the cavities effect. Consequently, PA correctly estimates $\Omega_{avg}^{(1,1)}$ only for small convex molecules (10–200 atoms).

Projection superposition approximation

The computational speed of the PA method inspired further refinements to render the method more accurate for very large molecules, including those containing cavities. Bower *et al.* developed the PSA model.[95–98] PSA is based on the PA concept, improved by taking into account (1) the cooperative size effect by superposition of diffuse atomic contributions and (2) the presence of cavities using a shape factor ρ. This algorithm, whose atomic parameters were obtained for H, C, N, O, Na and K based on experimental cross section measurements, is made available to the community as a web service[104] to compute Ω_{CALC} in both He and N_2.

Materials and methods

The calculations with all the listed algorithms (except for PSA that is available as web server) were carried out on an Intel

wileyonlinelibrary.com/journal/jms

Xeon Processor E5-2650 v2 server, equipped with two CPUs consisting of 16 hyper-threading cores each. One thread per computation was used, because none of the previous listed codes is parallelized.

For TM, the following programs have been tested: *mobcal*,[88] *mobcal-He*,[45] *mobcal-N2*[45] and their related optimizations. In *mobcal* and *mobcal-He* source codes, atom masses and sizes are defined starting on lines 586 and 2616. We added there parameters for phosphorus, by setting the atomic energy and the van der Waals distance parameters (the variables eolj and rolj in the source codes) equal to those of silicon. Furthermore, we also tested a slightly modified version of *mobcal-He* and *mobcal-N2*, in which we decreased the impact parameter (reported as *imp* in the source codes and representing the number of points in the Monte Carlo integration of the impact factor and orientation) from 1000 to 25, to check the influence of this parameter on the balance between calculation speed and accuracy. We will refer to these modified versions as *mobcal-He25* and *mobcal-N25*. Besides, because of the size of our system (795 atoms), the *mobcal-N2* and *mobcal-N25* algorithms were not able to compute the CCS in a reasonable computer time. We therefore tested modified input files excluding the partial atomic charges (keyword 'none' in line 5 of the input file). These modified algorithms are referred to as *mobcal-N2-mod* and *mobcal-N25-mod*.

For EHSS, we used all *mobcal* programs (which are equivalent for the EHSS part) and *EHSSrot*. In the latter, the collision radii of the atoms in He as collision gas are defined starting on line 52; here, we added the optimized parameters proposed by Siu *et al.*,[91] and we will refer to this optimized version of the algorithm as *EHSSrot-Siu*. For PA, we used *mobcal* with the original parameters, with the addition of phosphorus parameters. For PSA, theoretical CCS data were obtained by participating to the beta-testing of WebPSA, the web interface provided by the University of California, Santa Barbara.[101]

Results and discussion: matching between theoretical and experimental collision cross section for the [dTGGGGT]$_4$ quadruplex nucleic acid

The Ω_{CALC} values were calculated on 200 structures collected along the gas-phase MD simulation of replica 1. After discarding those coming from the first 25 ns of the MD simulation in the gas phase, the CCS values vary by no more than 1.5% (Fig. 5A). The averages and standard deviation are listed in Table 1, together with the percentage differences between experimental and theoretical values ($<\%\mathrm{diff}> = <[(\Omega_{CALC}-\Omega_{EXP})/\Omega_{EXP}] \times 100\%>$) and the computer time required per structure calculation.

In helium, Ω_{CALC} and Ω_{EXP} match within 4%, except for the PA method that underestimates the CCS by 12.8%. This underestimation was expected because our molecule is much larger (795 atoms) than typically appropriate for PA (10–200 atoms). The PSA method has correctly taken into account the effect of the cavities. Both for TM and EHSS, the original atom parameterizations give rise to a Ω_{CALC} about 3–4% higher than Ω_{EXP} (with TM calculations converged with a precision within 1.4%), and more recent parameters (implemented in *mobcal-He* for TM and Siu's parameters for EHSS) lead to a very good match (within 1%, with TM calculations converged with a precision within 1.4%). The *EHSSrot-Siu* calculations were 1000 times faster than using *mobcal-He*. Reducing the impact parameter from 1000 to 25 in *mobcal-He* reduced the computation time by a factor of 30 (from 18 h to 35 min per structure) and led to the same average value and twice the standard deviation (with calculations converged with a precision within 7.2%).

The perfect match (within 1%) may have been obtained just by luck with one single MD trajectory. To test the robustness of our conclusion on the best matching computational method in helium, the same procedure was followed for *EHSSrot-Siu* and *mobcal-He25* on the three other replica, that is, with different starting structures coming from different points of the solution MD. The results are summarized in Fig. 5 and Table 2. The data show that the solution

Table 1. Summary of theoretical collision cross section of $[(dTGGGGT)_4 + 3NH_4]^{5-}$ replica 1: average values with standard deviations (Ω_{CALC}, reported in Å^2); average percent difference between experimental and theoretical CCS ($<\%\mathrm{diff}> = <[(\Omega_{CALC}- \Omega_{EXP})/\Omega_{EXP}]\times100\%>$); and computer time required per structure calculation (time, expressed in minutes); n. a. stands for not acquired

Method	Algorithm	Ω_{CALC}	$<\%\mathrm{diff}>$	Time (min)
Trajectory method	*mobcal*	810 ± 12	2.8%	~10
	Mobcal-He	788 ± 7	0%	~1080
	Mobcal-He25	788 ± 14	0%	~35
	Mobcal-N2	n. a.	n. a.	~40320
	Mobcal-N25	n. a.	n. a.	~25920
	Mobcal-N2-mod	n. a.	n. a.	~11520
	Mobcal-N25-mod	1004 ± 15	−0.6%	~360
Exact hard sphere scattering	*mobcal*	813 ± 8	3.2%	~10
	Mobcal-He	813 ± 8	3.2%	~1080
	Mobcal-He25	813 ± 8	3.2%	~35
	EHSSrot	816 ± 8	3.5%	~1
	EHSSrot-Siu	789 ± 7	0.1%	~1
Projection approximation	*mobcal*	687 ± 6	−12.8%	~10
	Mobcal-He	687 ± 6	−12.8%	~1080
	Mobcal-He25	687 ± 6	−12.8%	~35
Projection superposition approximation	*He*	781 ± 14	−0.9%	~15
	N$_2$	941 ± 17	−6.8%	~15

Table 2. Summary of Ω_{CALC} obtained for each $[(dTGGGGT)_4 + 3NH_4]^{5-}$ replica through *EHSSrot-Siu* and *mobcal-He25* algorithms: average values with standard deviations reported in Å^2

Replica	$\Omega_{EHSSrot\text{-}Siu}$	$\Omega_{mobcal\text{-}He25}$
1	789 ± 7	788 ± 14
2	791 ± 9	793 ± 13
3	798 ± 10	795 ± 15
4	782 ± 6	778 ± 13
Av.	790 ± 10	788 ± 15

phase structures experience a significant decrease of the CCSs while in the gas phase. By measuring the CCS of the starting and final conformations of the four replicas with and without the thymine residues, we find that the CCS of the G-quadruplex core is not undergoing significant changes, denoting that the decrease of the CCS is mainly due to the rearrangements of the thymine residues, as anticipated from the variations of the structural descriptors of the replicas during the MD simulations in gas phase (see also Figs S2–S5).

Although the structures differ with respect to the thymines positions relative to the G-quadruplex core, the average Ω_{CALC} of each trajectory matches remarkably well with Ω_{EXP}. This gives us confidence to ascertain that the modified atom parameters, although optimized on molecules of a different chemical nature, are transferrable to nucleic acids (at least in the size range around 24 bases). The results also underline, however, that several different structures (in the present case with different thymine positions) can give the same Ω_{CALC}. It is also possible that all structures with all the different thymine placements are contributing to the observed population. Note also that a narrow Ω_{EXP} distribution indicates either that all structures present in the population have a narrow distribution of $\Omega_{avg}^{(1,1)}$ or that structures that could have a broader distribution of $\Omega_{avg}^{(1,1)}$ interconvert rapidly compared with the time scale of the experiment.

In nitrogen, the only TM calculation that can be carried out within affordable computational time (yet still 6 h per structure) is using *mobcal-N25-mod*, with a decreased number of impact trajectories and without considering partial charges. The agreement between the calculation and the experiment is again notable (within 1%, with calculations converged with a precision within 7.7%). The PSA calculations in N$_2$ are 25 times faster than with this modified TM, but unfortunately, PSA in N$_2$ underestimates the CCS by 6.8%. It is important to note that, contrary to helium parameters, nitrogen parameters for PSA are still preliminary.

The quest for a universally applicable and computationally affordable method to compute CCSs is however far from finished, particularly for other gases than helium. The [dTGGGGT]$_4$ quadruplex structure is a good candidate to benchmark the applicability of novel methods to negatively charged nucleic acids. Meanwhile, for structural interpretation applications, it is advisable to keep to helium CCS measurements and calculations. Based on our results, for future high throughput nucleic acid CCS determination during MD trajectories, we will choose the *EHSSrot-Siu* combination, which is the most convenient because it is the fastest. For rigorous confirmation, the Ω_{CALC} of representative structures will also be checked with TM and PSA.

Conclusions and perspectives

In summary, relative CCS values for a system of given size inform us on whether the ion is more compact or more extended. Molecular modelling's added value is to provide a spatial representation of ion structures that are compatible with the experiment. Quantum mechanics or molecular dynamics simulations describe structures with atomistic detail. However, the information from the CCS measurement of each separated species comes only from the average value (entailed with error) and from the peak width (if larger than the instrumental function). Even for the particular rigid DNA molecule presented here, it is important to realize that several structurally and energetically realistic structures are compatible with a given experiment. In other words, the solution is not unique, although often a single representative image is chosen to illustrate the discussion of IMS results. These images should be interpreted as representative structure(s) from an ensemble compatible with the experiment. Extensive modelling and computational resources are required to ascertain how large this ensemble can be.

One criterion for delimiting the ensemble of compatible molecular models is matching values of Ω_{EXP} and Ω_{CALC}. This tutorial has highlighted the different sources of error pertaining to the three key steps: the CCS measurement, the generation of the molecular models, and the calculation of the CCS from the models.

The CCS measurements are subject to experimental errors. The prediction intervals of the Synapt calibrations tell us that for nucleic acids, there is a 95% chance that any repeated measurement by any lab would fall within 3% of the average value. The reproducibility of CCS measurements with current commercial instruments is actually much better (standard deviation between average values is less than 1%). Improvement must therefore come from reducing systematic biases. An inter-laboratory study of CCS measurements, with different instruments (commercial and non-commercial) would be needed to approach true values for a set of compounds that could be then used by the community for quality control, calibration or internal standard correction of the CCS. Such an incentive is currently under way in Europe.

On the molecular modelling side, the biggest challenge for biomolecules larger than 100 atoms is to properly explore realistic conformational ensembles. In the particular case where large systems were measured using soft ion desolvation and transfer in order to preserve as many structural features as possible from the solution to the gas phase, the phase space exploration can be biased voluntarily. This approach was chosen here: a plausible structure is modelled first in solution, then solvent and counter-ions not observed in the mass spectra are removed, charges are modified to mimic the experimental charge state and the structure is left to evolve in the gas phase. The confrontation of these modelling results with the experiments can teach us what structural features are maintained from the solution to the gas phase (and for how long, depending on the temperature). They can also potentially teach us something on the ionization mechanism and specifically what happens to ion structure upon transfer to the gas phase. Modelling the ESI process to generate solvent-free polyatomic ions and the subsequent ion structural changes in the gas phase however encounters three major challenges to be addressed in the future as follows: (1) realistic description of solvent evaporation effects on the species conformation, (2) partial charges location for a given charge state, and (3) the time scale (milliseconds for an IMS experiment). The first two issues have not yet been clarified for all biological molecules, although several

attempts been made in this direction, [55,72,74,75,105–107] whereas the latter point would require a rather nowadays unaffordable computer time, above all for large systems like biopolymers (to the best of our knowledge, the sub-milliseconds time scale in a MD simulation in gas phase has been achieved only for a 7-mer oligonucleotide).[77]

Finally, although there is still room for improvement in CCS calculation approaches, algorithms and parameterization (especially for gases other than helium), we were pleased to find out that several of the currently available methods, although developed and parameterized for totally different molecules, transpose well to the 24-nucleotide nucleic acid structure studied here. This is an important stepping stone for future studies of other nucleic acid structures, for which IMS-MS can bring important insight on the populations coexisting in solution. It will be important for future work to further how robust each CCS calculation method can be when applied to ions of diverse chemical nature and different sizes, shapes and degrees of flexibility. For that, reference molecules for which the gas-phase structure is predicted robustly, in a similar way to the (dTGGGGT)$_4$ quadruplex nucleic acid, should be preferred. Ideally, calibrants envisaged for an inter-laboratory experimental calibration study should also ideally have robust predicted gas-phase structure. Meanwhile, for all readers interested in using IMS-MS to investigate ion structures in the gas phase, all three steps of this tutorial will have to be carried out carefully for the system of interest.

References

[1] A. J. R. Heck, R. H. H. van Den Heuvel. Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom. Rev.* **2004**, *23*, 368–389.

[2] M. Sharon, C. V. Robinson. The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu. Rev. Biochem.* **2007**, *76*, 167–193.

[3] D. E. Clemmer, M. F. Jarrold. Ion mobility measurements and their applications to clusters and biomolecules. *J. Mass Spectrom.* **1997**, *32*, 577–592.

[4] E. A. Mason, E. W. McDaniel. *Transport Properties of Ions in Gases.* Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, FRG, **1988**, 0–471.

[5] B. C. Bohrer, S. I. Merenbloom, S. L. Koeniger, A. E. Hilderbrand, D. E. Clemmer. Biomolecule analysis by ion mobility spectrometry. *Annu. Rev. Anal. Chem. (Palo Alto. Calif).* **2008**, *1*, 293–327.

[6] A. B. Kanu, P. Dwivedi, M. Tam, L. Matz, H. H. Hill. Ion mobility-mass spectrometry. *J. Mass Spectrom.* **2008**, *43*, 1–22.

[7] E. Jurneczko, P. E. Barran. How useful is ion mobility mass spectrometry for structural biology? The relationship between protein crystal structures and their collision cross sections in the gas phase. *Analyst* **2011**, *136*, 20–28.

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

[9] M. T. Bowers. Ion mobility spectrometry: a personal view of its development at UCSB. *Int. J. Mass Spectrom.* **2014**, *370*, 75–95.

[10] T. Wyttenbach, C. Bleiholder, M. T. Bowers. Factors contributing to the collision cross section of polyatomic ions in the kilodalton to gigadalton range: application to ion mobility measurements. *Anal. Chem.* **2013**, *85*, 2191–2199.

[11] K. Phillips, Z. Dauter, A. I. Murchie, D. M. Lilley, B. Luisi. The crystal structure of a parallel-stranded guanine tetraplex at 0.95 A resolution. *J. Mol. Biol.* **1997**, *273*, 171–182.

[12] C. Creze, B. Rinaldi, R. Haser, P. Bouvet, P. Gouet. Structure of a d (TGGGGT) quadruplex crystallized in the presence of Li+ ions. *Acta Crystallogr. D Biol. Crystallogr.* **2007**, *63*, 682–688.

[13] C. Cáceres, G. Wright, C. Gouyette, G. Parkinson. J. A subirana. A thymine tetrad in d(TGGGGT) quadruplexes stabilized with Tl+/Na+ ions. *Nucleic Acids Res.* **2004**, *32*, 1097–1102.

[14] M. P. H. Lee, G. N. Parkinson, P. Hazel, S. Neidle. Observation of the coexistence of sodium and calcium ions in a DNA G-quadruplex ion channel. *J. Am. Chem. Soc.* **2007**, *129*, 10106–10107.

[15] S. Cosconati, L. Marinelli, R. Trotta, A. Virno, S. De Tito, R. Romagnoli, B. Pagano, V. Limongelli, C. Giancola, P. G. Baraldi, L. Mayol, E. Novellino, A. Randazzo. Structural and conformational requisites in DNA quadruplex groove binding: another piece to the puzzle. *J. Am. Chem. Soc.* **2010**, *132*, 6425–6433.

[16] G. R. Clark, P. D. Pytel, C. J. Squire. The high-resolution crystal structure of a parallel intermolecular DNA G-4 quadruplex/drug complex employing Syn glycosyl linkages. *Nucleic Acids Res.* **2012**, *40*, 5731–5738.

[17] L. Martino, A. Virno, B. Pagano, A. Virgilio, S. Di Micco, A. Galeone, C. Giancola, G. Bifulco, L. Mayol, A. Randazzo. Structural and thermodynamic studies of the interaction of distamycin a with the parallel quadruplex structure [d(TGGGGT)]4. *J. Am. Chem. Soc.* **2007**, *129*, 16048–16056.

[18] F. Rosu, V. Gabelica, E. De Pauw, R. Antoine, M. Broyer, P. Dugourd. UV spectroscopy of DNA duplex and quadruplex structures in the gas phase. *J. Phys. Chem. A* **2012**, *116*, 5383–5391.

[19] V. Gabelica, F. Rosu, E. De Pauw, J. Lemaire, J. Gillet, J.-C. Poully, F. Lecomte, G. Grégoire, J. Schermann, C. Desfrançois. Infrared signature of DNA G-quadruplexes in the gas phase. *J. Am. Chem. Soc.* **2008**, *130*, 1810–1811.

[20] J. Gidden, E. S. Baker, A. Ferzoco, M. T. Bowers. Structural motifs of DNA complexes in the gas phase. *Int. J. Mass Spectrom.* **2005**, *240*, 183–193.

[21] R. Ferreira, A. Marchand, V. Gabelica. Mass spectrometry and ion mobility spectrometry of G-quadruplexes. A study of solvent effects on dimer formation and structural transitions in the telomeric DNA sequence d(TAGGGTTAGGGT). *Methods* **2012**, *57*, 56–63.

[22] H. E. Revercomb, E. A. Mason. Theory of plasma chromatography/gaseous electrophoresis. *Review. Anal. Chem.* **1975**, *47*, 970–983.

[23] T. Meyer, V. Gabelica, H. Grubmüller, M. Orozco. Proteins in the gas phase. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 408–425.

[24] M. S. Wilm, M. Mann. Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *Int. J. Mass Spectrom. Ion Process.* **1994**, *136*, 167–180.

[25] J. F. de la Mora. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Anal. Chim. Acta* **2000**, *406*, 93–104.

[26] D. Morsa, V. Gabelica, E. De Pauw. Effective temperature of ions in traveling wave ion mobility spectrometry. *Anal. Chem.* **2011**, *83*, 5775–5782.

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

[28] A. A. Shvartsburg, R. D. Smith. Fundamentals of traveling wave ion mobility spectrometry. *Anal. Chem.* **2008**, *80*, 9689–9699.

[29] J. Rus, D. Moro, J. A. Sillero, J. Royuela, A. Casado, F. Estevez-Molinero, J. Fernández de la Mora. IMS–MS studies based on coupling a differential mobility analyzer (DMA) to commercial API–MS systems. *Int. J. Mass Spectrom.* **2010**, *298*, 30–40.

[30] J. F. de la Mora, L. de Juan, T. Eichler, J. Rosell. Differential mobility analysis of molecular ions and nanometer particles. *TrAC Trends Anal. Chem.* **1998**, *17*, 328–339.

[31] A. Shvartsburg. *Differential Ion Mobility Spectrometry: Nonlinear Ion Transport and Fundamentals of FAIMS.* CRC Press: Boca Raton, **2009**.

[32] J. A. Silveira, M. E. Ridgeway, M. A. Park. High resolution trapped ion mobility spectrometry of peptides. *Anal. Chem.* **2014**, *86*, 5624–5627.

[33] J. L. Mergny, J. Gros, A. De Cian, A. Bourdoncle, F. Rosu, B. Sacca, L. Guittat, S. Amrane, M. Mills, P. Alberti, M. Takasugi, L. Lacroix. Energetics, kinetics and dynamics of quadruplex folding. In *Quadruplex Nucleic Acids*, S. Neidle, S. Balasubramanian, (Eds). The Royal Society of Chemistry: Cambridge, **2006**, pp. 31–80.

[34] T. Wyttenbach, P. R. Kemper, M. T. Bowers. Design of a new electrospray ion mobility mass spectrometer. *Int. J. Mass Spectrom.* **2001**, *212*, 13–23.

[35] K. Giles, S. D. Pringle, K. R. Worthington, D. Little, J. L. Wildgoose, R. H. Bateman. Applications of a travelling wave-based radio-frequency-only stacked ring ion guide. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 2401–2414.

[36] K. Thalassinos, S. E. Slade, K. R. Jennings, J. H. Scrivens, K. Giles, J. Wildgoose, J. Hoyes, R. H. Bateman, M. T. Bowers. Ion mobility mass spectrometry of proteins in a modified commercial mass spectrometer. *Int. J. Mass Spectrom.* **2004**, *236*, 55–63.

[37] B. T. Ruotolo, J. L. P. Benesch, A. M. Sandercock, S.-J. Hyung. C. V Robinson. Ion mobility-mass spectrometry analysis of large protein complexes. *Nat. Protoc.* **2008**, *3*, 1139–1152.

[38] M. F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson, B. T. Ruotolo. Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Anal. Chem.* **2010**, *82*, 9557–9565.

[39] 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. A. McLean. Conformational ordering of biomolecules in the gas phase: nitrogen collision cross sections measured on a prototype high resolution drift tube ion mobility-mass spectrometer. *Anal. Chem.* **2014**, *86*, 2107–2116.

[40] K. S. Bose, R. H. Sarma. Delineation of the intimate details of the backbone conformation of pyridine nucleotide coenzymes in aqueous solution. *Biochem. Biophys. Res. Commun.* **1975**, *66*, 1173–1179.

[41] E. S. Baker, S. L. Bernstein, V. Gabelica, E. De Pauw, M. T. Bowers. G-quadruplexes in telomeric repeats are conserved in a solvent-free environment. *Int. J. Mass Spectrom.* **2006**, *253*, 225–237.

[42] V. Gabelica, E. S. Baker, M.-P. Teulade-Fichou, E. De Pauw, M. T. Bowers. Stabilization and structure of telomeric and c-Myc region intramolecular G-quadruplexes: the role of central cations and small planar ligands. *J. Am. Chem. Soc.* **2007**, *129*, 895–904.

[43] N. Smargiasso, F. Rosu, W. Hsia, P. Colson, E. S. Baker, M. T. Bowers, E. De Pauw, V. Gabelica. G-quadruplex DNA assemblies: loop length, cation identity, and multimer formation. *J. Am. Chem. Soc.* **2008**, *130*, 10208–10216.

[44] A. Arcella, G. Portella, M. L. Ruiz, R. Eritja, M. Vilaseca, V. Gabelica, M. Orozco. Structure of triplex DNA in the gas phase. *J. Am. Chem. Soc.* **2012**, *134*, 6596–6606.

[45] I. Campuzano, M. F. Bush, C. V. Robinson, C. Beaumont, K. Richardson, H. Kim, H. I. Kim. Structural characterization of drug-like compounds by ion mobility mass spectrometry: comparison of theoretical and experimentally derived nitrogen collision cross sections. *Anal. Chem.* **2012**, *84*, 1026–1033.

[46] R. Salbo, M. F. Bush, H. Naver, I. Campuzano, C. V. Robinson, I. Pettersson, T. J. D. Jørgensen, K. F. Haselmann. Traveling-wave ion mobility mass spectrometry of protein complexes: accurate calibrated collision cross-sections of human insulin oligomers. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 1181–1193.

[47] The Bush Research Group http://depts.washington.edu/bushlab/ (accessed Dec 17, 2014).

[48] M. Froimowitz. HyperChem : a software package for computational chemistry and molecular modeling. *Biotechniques* **1993**, *14*, 1010–1013.

[49] Accelrys Software Inc.. *Discovery Studio Modeling Environment, Release 4.0.* Accelrys Software Inc.: San Diego, **2013**.

[50] Maestro, Version 10.0, Schrödinger. . LLC: New York, NY, **2014**.

[51] T. J. Macke, D. A. Case. *Molecular Modeling of Nucleic Acids*, Leontis N. B., SantaLucia J. (Eds). ACS Symposium Series; American Chemical Society: Washington, DC, **1997**; Vol. *682*, pp. 379–393.

[52] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, G. R. Hutchison. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **2012**, *4*, 17.

[53] J. E. Šponer, A. Mládek, P. Banáš, P. Jurec. How to understand quantum chemical computations on DNA and RNA systems ? a practical guide for non-specialists. *Methods* **2013**, *64*, 3–11.

[54] M. Orozco. A theoretical view of protein dynamics. *Chem. Soc. Rev.* **2014**, *43*, 5051–5066.

[55] S. R. Harvey, M. Porrini, A. Konijnenberg, D. J. Clarke, R. C. Tyler, P. R. R. Langridge-Smith, C. E. MacPhee, B. F. Volkman, P. E. Barran. Dissecting the dynamic conformations of the metamorphic protein lymphotactin. *J. Phys. Chem. B* **2014**, *118*, 12348–12359.

[56] K. J. Pacholarz, M. Porrini, R. A. Garlish, R. J. Burnley, R. J. Taylor, A. J. Henry, P. E. Barran. Dynamics of intact immunoglobulin G explored by drift-tube ion-mobility mass spectrometry and molecular modeling. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 7765–7769.

[57] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi. Optimization by simulated annealing. *Science* **1983**, *220*, 671–680.

[58] F. A Fernandez-Lima, H. Wei, Y. Q. Gao, D. H. Russell. On the structure elucidation using ion mobility spectrometry and molecular dynamics. *J. Geophys. Res.. Chem. A* **2009**, *113*, 8221–8234.

[59] H. A. Sawyer, J. T. Marini, E. G. Stone, B. T. Ruotolo, K. J. Gillig, D. H. Russell. The structure of gas-phase bradykinin fragment 1-5 (RPPGF) ions: an ion mobility spectrometry and H/D exchange ion-molecule reaction chemistry study. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 893–905.

[60] T. Wyttenbach, G. von Helden, M. T. Bowers. Gas-phase conformation of biological molecules: bradykinin. *J. Am. Chem. Soc.* **1996**, *118*, 8355–8364.

[61] Y. Sugita, Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

[62] F. Albrieux, F. Calvo, F. Chirot, A. Vorobyev, Y. O. Tsybin, V. Lepère, R. Antoine, J. Lemoine, P. Dugourd. Conformation of polyalanine and polyglycine dications in the gas phase: insight from ion mobility spectrometry and replica-exchange molecular dynamics. *J. Phys. Chem. A* **2010**, *114*, 6888–6896.

[63] A. Baumketner, S. L. Bernstein, T. Wyttenbach, G. Bitan, D. B. Teplow, M. T. Bowers, J. Shea. Amyloid beta-protein monomer structure: a computational and experimental study. *Protein Sci.* **2006**, *15*, 420–428.

[64] F. Chirot, F. Calvo, F. Albrieux, J. Lemoine, Y. O. Tsybin, P. Dugourd. Statistical analysis of ion mobility spectrometry. I. Unbiased and guided replica-exchange molecular dynamics. *J. Am. Chem. Soc.* **2012**, *23*, 386–396.

[65] F. Calvo, F. Chirot, F. Albrieux, J. Lemoine, Y. O. Tsybin, P. Pernot, P. Dugourd. Statistical analysis of ion mobility spectrometry. II. Adaptively biased methods and shape correlations. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1279–1288.

[66] F. Canon, R. Ballivian, F. Chirot, R. Antoine, P. Sarni-Manchado, J. Lemoine, P. Dugourd. Folding of a salivary intrinsically disordered protein upon binding to tannins. *J. Am. Chem. Soc.* **2011**, *133*, 7847–7852.

[67] S. L. Bernstein, N. F. Dupuis, N. D. Lazo, T. Wyttenbach, M. M. Condron, G. Bitan, D. B. Teplow, J.-E. Shea, B. T. Ruotolo, C. V. Robinson, M. T. Bowers. Amyloid-B protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nat. Chem.* **2009**, *1*, 326–331.

[68] A. Politis, A. Y. Park, S.-J. Hyung, D. Barsky, B. T. Ruotolo, C. V. Robinson. Integrating ion mobility mass spectrometry with molecular modelling to determine the architecture of multiprotein complexes. *PLoS One* **2010**, *5*, e12080.

[69] B. T. Ruotolo, K. Giles, I. Campuzano, A. M. Sandercock, R. H. Bateman, C. V. Robinson. Evidence for macromolecular protein rings in the absence of bulk water. *Science* **2005**, *310*, 1658–1661.

[70] J. Wang, P. Cieplak, P. A. Kollman. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049–1074.

[71] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

[72] Z. Hall, A. Politis, M. F. Bush, L. J. Smith. C. V Robinson. Charge-state dependent compaction and dissociation of protein complexes: insights from ion mobility and molecular dynamics. *J. Am. Chem. Soc.* **2012**, *134*, 3429–3438.

[73] M. Miteva, P. A. Demirev, A. D. Karshikoff. Multiply-protonated protein ions in the gas phase: calculation of the electrostatic interactions between charged sites. *J. Phys. Chem. B* **1997**, *101*, 9645–9650.

[74] A. Patriksson, E. Marklund, D. van der Spoel. Protein structures under electrospray conditions. *Biochemistry* **2007**, *46*, 933–945.

[75] M. Rueda, S. G. Kalko, F. J. Luque, M. Orozco. The structure and dynamics of DNA in the gas phase. *J. Am. Chem. Soc.* **2003**, *125*, 8007–8014.

[76] M. Rueda, F. J. Luque, M. Orozco. G-quadruplexes can maintain their structure in the gas phase. *J. Am. Chem. Soc.* **2006**, *128*, 3608–3619.

[77] A. Arcella, J. Dreyer, E. Ippoliti, I. Ivani, G. Portella, V. Gabelica, P. Carloni, M. Orozco. Structure and dynamics of oligonucleotides in the gas phase. *Angew. Chemie* **2015**, *127*, 477–481.

[78] D. A. Case, T. A. Darden, T. E. C. Iii, 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. Götz, 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, 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, P. A. Kollman. *Amber 12*, University of California: San Francisco, **2012**.

[79] A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton, M. Orozco. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.* **2007**, *92*, 3817–3829.

[80] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

[81] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.

[82] A. Pérez, F. J. Luque, M. Orozco. Dynamics of B-DNA on the microsecond time scale. *J. Am. Chem. Soc.* **2007**, *22*, 14739–14745.

[83] T. Darden, D. York, L. Pedersen. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089.

[84] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684.

[85] D. R. Roe, T. E. Cheatham. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

[86] D. A. Case, T. A. Darden, T. E. Cheatham III, 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. Götz, 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, 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, P. A. Kollman. *Amber13*, University of California: San Francisco, **2012**.

[87] W. Humphrey, A. Dalke, K. Schulten. VMD: visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

[88] M. F. Mesleh, J. M. Hunter, A. A. Shvartsburg, G. C. Schatz, M. F. Jarrold. Structural information from ion mobility measurements: effects of the long-range potential. *J. Phys. Chem.* **1996**, *100*, 16082–16086.

[89] A. A. Shvartsburg, G. C. Schatz, M. F. Jarrold. Mobilities of carbon cluster ions: critical importance of the molecular attractive potential. *J. Chem. Phys.* **1998**, *108*, 2416.

[90] A. A. Shvartsburg, M. F. Jarrold. An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Lett.* **1996**, *261*, 86–91.

[91] C.-K. Siu, Y. Guo, I. S. Saminathan, A. C. Hopkinson, K. W. M. Siu. Optimization of parameters used in algorithms of ion-mobility calculation for conformational analyses. *J. Phys. Chem. B* **2010**, *114*, 1204–1212.

[92] Optimization of Algorithms for Ion Mobility Calculations A. A. Shvartsburg, S. V. Mashkevich, E. S. Baker, R. D. Smith. *J. Phys. Chem. A* **2007**, *111*, 2002–2010.

[93] H. I. Kim, H. Kim, E. S. Pang, E. K. Ryu, L. W. Beegle, J. A. Loo, W. A. Goddard, I. Kanik. Structural characterization of unsaturated phosphatidylcholines using traveling wave ion mobility spectrometry. *Anal. Chem.* **2009**, *81*, 8289–8297.

[94] H. Kim, H. I. Kim, P. V. Johnson, L. W. Beegle, J. L. Beauchamp, W. A. Goddard, I. Kanik. Experimental and theoretical investigation into the correlation between mass and ion mobility for choline and other ammonium cations in N2. *Anal. Chem.* **2008**, *80*, 1928–1936.

[95] S. E. Anderson, C. Bleiholder, E. R. Brocker, P. J. Stang, M. T. Bowers. A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (III): Application to supramolecular coordination-driven assemblies with complex shapes. *Int. J. Mass Spectrom.* **2012**, *330-332*, 78–84.

[96] C. Bleiholder, S. Contreras, M. T. Bowers. A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (IV). Application to polypeptides. *Int. J. Mass Spectrom* **2013**, *354-355*, 275–280.

[97] C. Bleiholder, S. Contreras, T. D. Do, M. T. Bowers. A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (II). Model parameterization and definition of empirical shape factors for proteins. *Int. J. Mass Spectrom* **2013**, *345-347*, 89–96.

[98] C. Bleiholder, T. Wyttenbach, M. T. Bowers. A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (I). *Method. Int. J. Mass Spectrom.* **2011**, *308*, 1–10.

[99] C. Larriba, C. J. Hogan. Free molecular collision cross section calculation methods for nanoparticles and complex ions with energy accommodation. *J. Comput. Phys.* **2013**, *251*, 344–363.

[100] Y. Alexeev, D. G. Fedorov, A. A. Shvartsburg. Effective ion mobility calculations for macromolecules by scattering on electron clouds. *J. Phys. Chem. A* **2014**, *118*, 6763–6772.

[101] The Martin Jarrold Research Group http://www.indiana.edu/~nano/software.html (accessed Dec 10, 2014).

[102] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, W. M. Skiff. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.

[103] The Bowers Research Group http://bowers.chem.ucsb.edu/theory_analysis/cross-sections/sigma.shtml (accessed Dec 10, 2014).

[104] WebPSA http://luschka.bic.ucsb.edu:8080/WebPSA/index.jsp (accessed Dec 10, 2014).

[105] S. C. Gibson, C. S. Feigerle, K. D. Cook. Fluorometric measurement and modeling of droplet temperature changes in an electrospray plume. *Anal. Chem.* **2014**, *86*, 464–472.

[106] M. Z. Steinberg, K. Breuker, R. Elber, R. B. Gerber. The dynamics of water evaporation from partially solvated cytochrome c in the gas phase. *Phys. Chem. Chem. Phys.* **2007**, *9*, 4690–4697.

[107] M. Z. Steinberg, R. Elber, F. W. McLafferty, R. B. Gerber, K. Breuker. Early structural evolution of native cytochrome c after solvent removal. *Chembiochem* **2008**, *9*, 2417–2423.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.