JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

Structure of Triplex DNA in the Gas Phase

Annalisa Arcella,[†] Guillem Portella,[†] Maria Luz Ruiz,[‡] Ramon Eritja,[‡] Marta Vilaseca,[§] Valérie Gabelica,[⊥] and Modesto Orozco*[,†,¶]

[†]Joint IRB BSC Research Program in Computational Biology, Institute for Research in Biomedicine, Baldiri Reixach 10, Barcelona 08028, Spain

[‡]Chemistry and Molecular Pharmacology Program, Institute for Research in Biomedicine, IQAC-CSIC, CIBER-BBN, Barcelona 08028, Spain

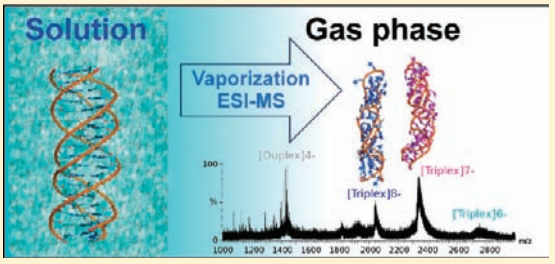
[§]Mass Spectrometry Core Facility, Institute for Research in Biomedicine, Barcelona 08028, Spain

[⊥]Department of Chemistry, University of Liège, Allée de la Chimie, Building B6c, B-4000 Liège, Belgium

[¶]Departament de Bioquímica i Biología Molecular, Facultat de Biología, Universitat de Barcelona, Avgda Diagonal 645, Barcelona 08028, Spain

Supporting Information

ABSTRACT: Extensive (more than 90 microseconds) molecular dynamics simulations complemented with ion-mobility mass spectrometry experiments have been used to characterize the conformational ensemble of DNA triplexes in the gas phase. Our results suggest that the ensemble of DNA triplex structures in the gas phase is well-defined over the experimental time scale, with the three strands tightly bound, and for the most abundant charge states it samples conformations only slightly more compact than the solution structure. The degree of structural alteration is however very significant, mimicking that found in duplex and much larger than that suggested for G-quadruplexes. Our data strongly supports that the gas phase triplex maintains an excellent memory of the solution structure, well-preserved helicity, and a significant number of native contacts. Once again, a linear, flexible, and charged polymer as DNA surprises us for its ability to retain three-dimensional structure in the absence of solvent. Results argue against the generally assumed roles of the different physical interactions (solvent screening of phosphate repulsion, hydrophobic effect, and solvation of accessible polar groups) in modulating the stability of DNA structures.



INTRODUCTION

Most of the available structural information on biomacromolecules (among them nucleic acids) has been obtained by high-resolution experimental techniques such as X-ray crystallography and NMR spectroscopy. In favorable cases, these methods solve the structure of biomacromolecules with uncertainties below 0.1 nm. However, despite all the power of X-ray crystallography and NMR spectroscopy, we cannot ignore that the practical use of these techniques is limited by many technical problems, such as the difficulty to obtain crystals or well-defined nuclear overhauser effect (NOE) maps, often making necessary the use of low-resolution methods as sources of structural data. Mass spectrometry (MS) experiments have become one of the sources of such low-resolution structural information, both through electrospray-MS (ESI-MS) and through its combination with ion mobility spectrometry (IMS-MS).[1−13] MS is fast and requires low sample consumption, but in contrast to other techniques, the information is recorded from samples in the gas phase, which raises the question of to what extent does gas phase structural information reflect the most populated conformation in solution. This question will become even more crucial in the near future when X-ray free electron lasers[14,15] will provide high-resolution structural information of molecules in the gas phase.

A series of studies with proteins[1−13] have demonstrated that if vaporization is done under mild conditions, the structure becomes stable during long periods of time, and gas phase ensembles can be used to accurately model the solution structure.[16−18] In other words, despite the absence of the hydrophobic effect, protein structure remains quite stable in the gas phase. The question is whether these findings also stand for a highly flexible and charged nonglobular molecule as DNA, whose conformation is known to be very dependent on the solvent environment.[19] In this respect, early ESI-MS experiments by different authors[1−3,20−24] provided convincing evidence that the two strands of a DNA duplex remain bound in the gas phase and that even noncovalent drug−DNA complexes remain stable in the complete absence of water. Molecular dynamics (MD) simulations suggested that duplexes are severely distorted when transferred to the gas phase[25,26] but

Received: October 18, 2011
Published: March 15, 2012

   dx.doi.org/10.1021/ja209786t | J. Am. Chem. Soc. 2012, 134, 6596−6606

confirmed that the two strands remain bound by several native contacts, maintaining the general helical shape at least in the sub-microsecond time scale. Very surprisingly, similar calculations with G-quadruplex DNA (four-stranded DNA helices) suggested a full maintenance of native helical structure in the gas phase, at least in the microsecond time scale,[27] a striking finding that is fully consistent with ESI-MS and ESI-IMS-MS experiments.[28−32] The behavior of the other canonical DNA structure, the DNA triplex, in the gas phase remains unknown.

DNA triplexes are formed when a single strand of DNA (named as the triplex forming oligonucleotide; TFO) recognizes a polypurine segment of duplex DNA (the triplex target sequence; TTS) through specific major-groove-mediated hydrogen bonds. The third strand can be either polypurine (antiparallel triplexes) or polypyrimidine (parallel triplexes). Under normal laboratory conditions, the parallel triplexes are much more stable than the antiparallel[33−36] ones, and accordingly, they are better characterized both experimentally[37−45] and theoretically.[46−50] Several NMR and MD studies[37−51] have shown that triplex DNA forms a right-handed helix with general features close to B-form of DNA, including bases perpendicular to the helical axis, a narrow minor groove (mG), and *south* sugar puckerings. The most obvious effect of the third strand is the partition of the duplex major groove (MG) in two new grooves: the major part of the major groove (MMG), similar in width to the major groove of DNA, but with quite different recognition properties, and the minor part of the major groove (mMG), a very narrow groove that can coordinate small polar groups.[46,49]

The presence of the third strands has very important consequences in the functionality of the DNA. The most important one is the dramatic modification of its ability to be recognized by specific major-groove binder proteins.[46,52] For example, DNA repairing systems do not recognize triplex DNAs well, explaining the use of triplexes to generate random mutations in DNA.[53] Furthermore, proteins involved in gene regulation also have large difficulties recognizing DNA sequences when placed on triplexes, and in fact triplexes generated in promoter regions can knock-down or even knockout the expression of the corresponding genes.[54−57] This behavior, combined with the high density of triplex target sequences in promoters[58,59] explains the interest of triplexes for antigene therapy.[53−60]

To our knowledge, first ESI-MS studies of the DNA triplex structures in the gas phase are due to Rosu et al.,[24] who found that the triplex remains stable in the gas phase under mild vaporization conditions. More detailed studies confirmed early results and suggested that triplex maintains noncovalent binding properties in the gas phase,[61−63] which points toward a certain maintenance of its original structure in solution. Unfortunately, MS experiments alone cannot determine the magnitude of the structural degradation induced by vaporization, whether they are large as in duplex DNA or very small as found for G-quadruplex DNA.

In this paper, we present a very extensive atomistic MD study on two stable parallel triplexes (12-mer and 18-mer) based on the (T,C) motif, i.e., triplexes based on the repetition of d(TC$^+$)−d(GA)•d(TC) triads (the dot indicates Watson−Crick base pairing, while the dash stands for Hoogsteen base pairing). The computational results presented here, supported by experimental measurements, demonstrate that DNA triplex is not unfolded in the gas phase, but on the contrary, retains approximate helical structure and a non-negligible amount of native contacts. However, the magnitude of structural distortion found for triplexes is very important, similar to that in duplexes[25−28] and larger than that detected for G-quadruplexes.[27] The present study completes the atlas of gas phase structures of DNA and confirms that basic concepts on the role of electrostatic repulsion on DNA structure need to be revisited.

METHODS

Molecular Dynamics Simulations in Solution. The 12-mer and 18-mer systems were simulated in aqueous solution to establish a reference state for the structural study of their behavior in the gas phase. Initial structures were built using a previously optimized template for triplexes[46−48] and were subsequently solvated in a truncated octahedron box and neutralized with 33 sodium ions for triplex 12-mer, and 51 for the 18-mer, distributed around the phosphate groups at optimum positions according to Poisson−Boltzmann calculations. The resulting 12-mer system contains roughly 5300 water molecules, and the 18-mer system encloses approximately 9300 water molecules. The 12-mer system was simulated at 300 K, whereas the 18-mer was simulated both at 300 K and at 372 K. The systems were equilibrated using our multistep protocol[46] and were simulated for 100 ns. Transferable interaction potential 3 point (TIP3P) parameters for water and Dang's parameters for ions were used;[64,65] previous studies[66] demonstrated that no changes in the simulations can be expected related to the addition of 50−200 mM extra salt or to the use of alternative water or ion models.

Gas-Phase Simulations. We have used snapshots collected after 5−10 ns of MD simulations in water of the corresponding triplex as starting guess structures in the gas phase (typical heavy atoms root mean square deviation (RMSd) between snapshots around 0.05 nm). We made the reasonable assumption that all cytosines in the Hoogsteen strand remained protonated after the triplexes have been transferred to the gas phase and that extra protons would migrate to the phosphate groups, without cytosine deprotonation. At this point two crucial decisions had to be taken: (i) the total charge at DNA and (ii) the protonation state. For the present simulations, we selected the expected most populated charge ions that showed complete desolvation and declustering (loss of all ammonium adducts as ammonia): −6 for the 12-mer, −7, −8 for the 18-mer. The determination of the protonation state (i.e., which phosphates are protonated) is more complex, since the total number of different states for a given charge is equal to $m!/[n!(m-n)!]$, where n is the number of protons that are required for a given total charge and m is the number of phosphates (all Hoogsteen cytosines are considered protonated). The simplest way to solve the problem is to assume a distributed neutralization, i.e., to assign equal fractional charge to all phosphates; this leads to the "distributed charged" scheme,[25] which was found to be reasonable for other forms of DNA. However, for triplexes, we found it necessary to consider phosphate−phosphate hydrogen bond across the minor part of the major groove (mMG), which requires individual titration of phosphates (the "localized charge scheme"[25]). This was done using a Monte Carlo procedure (in the order of 500 million Metropolis tests were done), from which we retained the 10 lowest energy configurations, assuming that they are representative of the ensemble of quasi-degenerated charge states detected experimentally for a given total charge.

A second important decision in defining the simulation conditions is the temperature of the system. Because of the loss of kinetic energy during vaporization and the unclear partition of thermal energy among internal and external degrees of freedom, there is no clear connection between simulation and experimental temperatures. Thus, following our previously used strategy,[28,29] we considered two working temperatures: T = 300 K, which simulates ideally mild vaporization conditions, and a higher temperature (T = 372 K), which was used to simulate the heating of ions that could have occurred in the electrospray source region,[67] as well as ion heating inside the traveling wave IMS cell.[68] Contrary to drift tube ion mobility spectrometry, where the ion temperature is equal to the bath gas temperature,

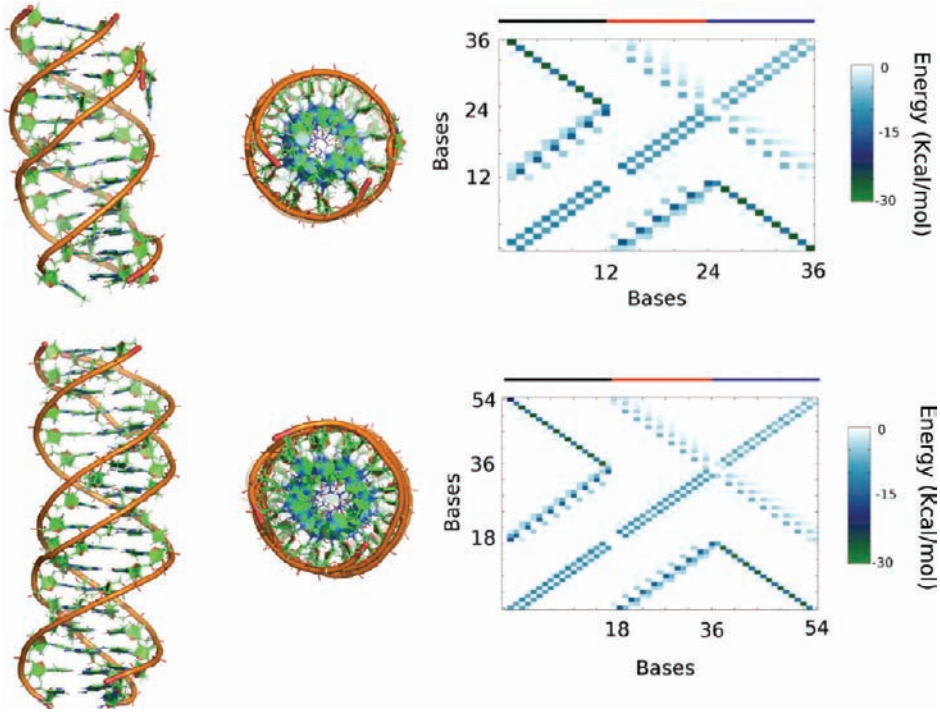


Figure 1. Triplex structure in aqueous solution at room temperature. Left: orthogonal views of the MD averaged (last 5 ns) structures obtained for $d(TC^+)_6$–$d(GA)_6$•$d(TC)_6$ (12-mer at the top) and $d(TC^+)_9$–$d(GA)_9$•$d(TC)_9$ (18-mer at the bottom). Right: energy-based contact maps for 12- and 18-mer triplexes (energy values in kcal/mol). Reference lines are added to label the three strands: black for the Watson–Crick (WC), red for the WC pyrimidine, and blue for the Hoogsteen (H) one.

significant field heating of the ions can occur in traveling wave ion mobility spectrometers as used here, leading to an effective temperature equal to the following eq 1:[69]

$$T_{\text{eff}} = T_{\text{gas}} + \frac{3\pi}{128k^2T}\left(1 + \frac{M}{m}\right)\left(\frac{zeE}{\rho\Omega}\right)^2 \approx T_{\text{gas}} + C\frac{z^2}{\Omega^2} \qquad (1)$$

where T_{gas} is the bath gas temperature (298 K), k is the Boltzmann constant, M is the mass of the gas, z is the ion nominal charge, e is the charge of an electron, E is the effective electric field, ρ is the gas number density, m is the ion mass, and Ω is the ion collision cross section. The effective temperature data of Morsa et al.[68] were used to extrapolate the value of the constant C at the working conditions considered in this work. This extrapolation leads, for the ions of the highest mobility (or z/Ω ratio), to an effective temperature (T_{eff}) around 372 K, which was enforced in our simulations.

After 3 ns of thermalization in gas phase, the systems were simulated for 1 μs. In order to investigate equilibration issues, we extended the simulation time of one of the systems up to 30 μs. Finally, to check for dependence on microscopic initial conditions, five replicas obtained using different starting coordinates and velocities were followed for a selected charge state of the 12-mer. The set of simulations performed in this paper are summarized in Table S1 (Supporting Information (SI)).

Simulation Details. PARMBSC0 revision of amber parm99 force field[70,71] was used to describe DNA, taking parameters for protonated cytosine[48] and neutralized phosphates from previous works.[23−27] As found in other studies,[25] preliminary tests showed that scaling down the charges at neutral nucleosides by a factor of 0.8 (to better simulate gas phase charge distribution) does not significantly change the results. Simulations in water were carried out using periodic boundary conditions and the particle mesh Ewald method[72] for long-range electrostatic treatment (0.12 nm grid size), combined with a cutoff radius of 1 nm for Lennard–Jones interactions. No cutoff was used for either electrostatic or Lennard–Jones interactions in the gas phase. All aqueous simulations were carried out at the isothermal–isobaric ensemble (T = 300 K and P = 1 atm) using Berendsen thermostat and barostat.[73] The SHAKE algorithm[74] was used to constrain all bond distances to their equilibrium value, allowing a 2 fs integration time step for solution conditions. Since forces are stronger in the absence of water, a more conservative 1 fs integration step (using also SHAKE) was used for gas phase calculations. All calculations in solution were carried out using AMBER9.0 computer program,[75] while gas phase calculations were performed using the most efficient GROMACS 4.5 program.[76]

Analysis of Data. A wide variety of tools have been used to process more than 90 μs of data. This includes standard geometrical analysis, stacking and hydrogen bond calculations performed using "in house" programs, clustering analysis by AMBER9/GROMACS 4.5.3 tools, and theoretical determination of collision cross sections (CCS; the experimental observable), which was carried out for the different MD ensembles using the Sigma program, which uses a projection approximation procedure.[77−79]

Sample Preparation for Experimental Validation. MS experiments cannot determine alone the structure of the triplex in the gas phase. However, low-resolution structural data derived from IMS-MS experiments can help us validate our theoretical models. The triplexes were prepared in neutral or acidic conditions. Single stranded $d(GA)_6$, $d(GA)_9$, $d(TC)_6$, and $d(TC)_9$ were bought from either Sigma-Aldrich or Eurogentec. For the neutral conditions, the strands were mixed in suitable stoichiometry and resuspended in 150 mM ammonium acetate (pH = 5.5) to a final 2 μM concentration. Samples were heated at 90 degrees and annealed overnight to form 12- and 18-mer d(TC)–d(GA)•d(TC) triplexes, which were easily detected in 20% polyacrylamide gels. Melting curves revealed clear triplex signatures with melting temperatures around 50 (12-mer) and 65 (18-mer) degrees in UV-melting experiments (data not shown). Triplex samples were lyophilized before use and resuspended for MS analysis in water and mixed 1:1 (v/v) with solution A to a final concentration of 150 μM (solution A: 100 mM NH_4OAc neutralized with triethylamine (pH 7.11)/isopropanol (9:1)) immediately prior to the IMS-MS analysis. Neutral pH and 5% isopropanol favored ionization in the negative mode and significantly improved signal-to-noise ratio with chip-based nanoelectrospray in the negative mode.[80] For acidic

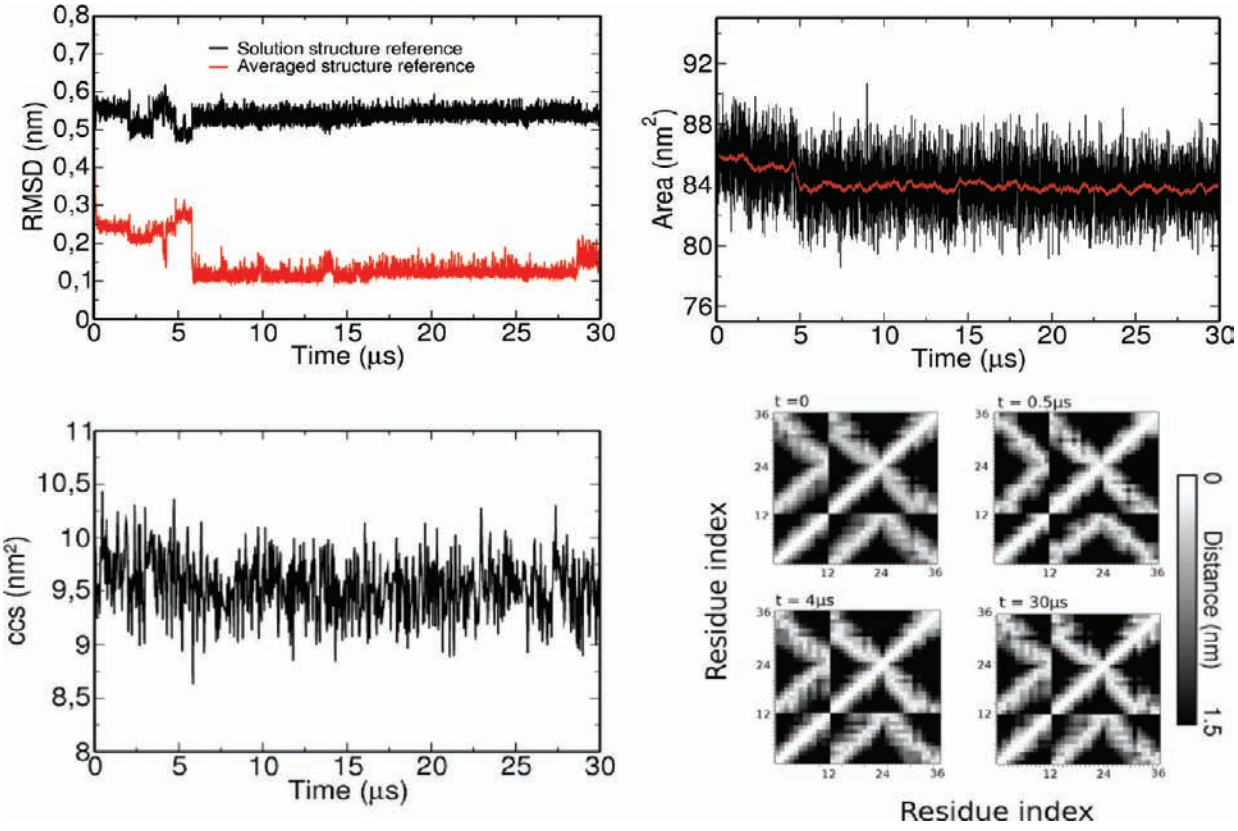


Figure 2. Key simulation results obtained in the 30 μs trajectory of [12-mer triplex]$^{6-}$ in the gas phase. Left-top: evolution with time of the RMSd (in nm) from the solution structure (black) and the MD-averaged one (red). Right-top: variation of the solvent accessible surface (in nm^2) along time. Left-bottom: evolution with time of the collision cross section. Right-bottom: distance-based contact maps (distances are color coded) obtained from different 5 μs windows along simulation compared to that found in the original solution structure.

conditions, strands were dissolved in water, and triplexes were prepared by annealing in solution B to a final concentration of 100 μM (solution B: 150 mM NH_4OAc supplemented with acetic acid to pH = 3). Samples under neutral conditions were injected at a concentration of 150 μM, while those at acidic pH (where triplex is more stable) were diluted before injection to a concentration of 10 μM (12-mer) or 15 μM (18-mer).

Ion-Mobility Mass Spectrometry (IMS-MS). Two sets of experiments were performed in Barcelona and Liege, with the triplexes in neutral (Barcelona) and acidic (pH = 3; Liege) conditions. Both sets of experiments were performed in the negative ion mode on Synapt G1 HDMS traveling wave ion mobility mass spectrometers (Waters, Manchester, U. K.). The main difference is the ionization source used. Experiments on neutral triplexes were performed using a 384-well plate refrigerated at 15 °C and introduced by automated chip-based nanoelectrospray using a Triversa NanoMate (Advion BioSciences), whereas experiments at pH = 3 were performed using the standard electrospray source (flow rate of 4 μL/min) at room temperature. Each instrument was tuned so as to obtain a proper ion signal in the softest possible conditions. The details of voltages, pressures, and temperatures used for the ion transfer to the IMS cell are given in Table S2 (SI). On both instruments, the traveling wave IMS cell was operated at a wave velocity s = 300 m/s and a wave height WH = 8 V. The instruments were mass-calibrated using sodium iodide or cesium iodide and mobility-calibrated using oligonucleotides of known collision cross sections. Different calibrants lists and calibration procedures were used for each instrument (details in Figure S1, Tables S2 and S3 (SI)), and therefore, the two sets of results are those of totally independent traveling wave IMS experiments.

RESULTS AND DISCUSSION

Control Simulations in Aqueous Solvent. MD simulations performed using a fully hydrated environment yield stable trajectories in the 0.1 μs time scale for both the 12- and 18-mer triplexes. Helicity is fully preserved (see Figure 1), defining a structure with a general B-shape form (see Figures 1 and S2, Table S4 (SI)), as previously suggested by shorter MD simulations and NMR experiments (see the Introduction) and in clear disagreement with early models derived from X-ray fiber diffraction data (see ref 46 for discussion). Sugar puckerings are in general in the south to southeast region, but pyrimidine nucleotides reach around 9−11% of the north region, indicating that sugar puckering is quite flexible in triplexes. The largest structural distortions found are located at the ends of the helices, where fraying of a protonated cytosine is detected in both 12- and 18-mer simulations (see Figures 1 and S2 (SI)). At 372 K, the triplex shows indications of unfolding (data not shown), but the slow kinetics of the process preclude a direct visualization of these events in the accessible simulated time scale. Very interestingly, end-fraying effects involving Watson−Crick interactions are largely reduced compared to the situation in duplexes, affecting only marginally d(T−A•T) triads. In summary, extended MD simulations suggest that the aqueous triplex is very stable at room temperature, is quite rigid, and displays a general B-like conformation that fits well with the experimental and previous MD data.[37−45]

Gas Phase Simulations. As discussed above, we planned to explore the diversity of the gas phase conformational space by

dx.doi.org/10.1021/ja209786t | J. Am. Chem. Soc. 2012, 134, 6596−6606

Replica	Averaged rms (nm)	Averaged ccs (nm²)	Hbond%	
			Watson-Crick	Hoogsteen
1	0.54 ± 0.01	9.6 ± 0.3	42	15
2	0.74 ± 0.02	9.5 ± 0.3	18	6
3	0.70 ± 0.01	9.4 ± 0.2	17	16
4	0.82 ± 0.01	9.7 ± 0.2	22	6
5	0.78 ± 0.01	9.8 ± 0.3	18	16
Av.	**0.6 ± 0.1**	**9.6 ± 0.3**	**23**	**12**

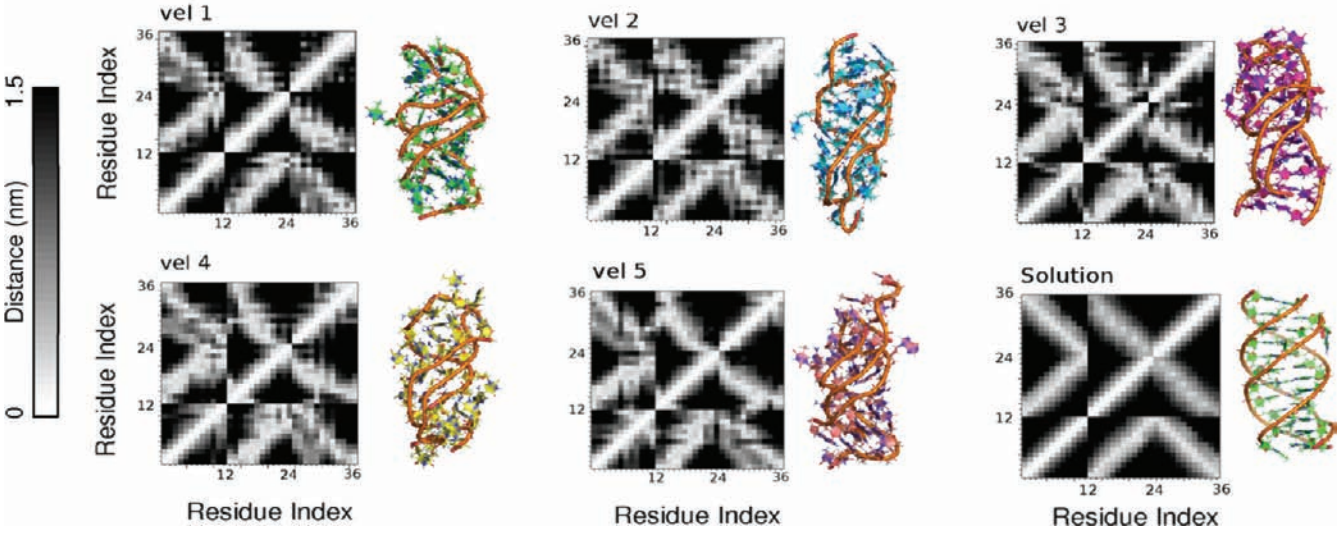


Figure 3. Key simulation results obtained from 5 replicas of a single charge state of the 12-mer triplex in gas phase. Top: MD-averaged values of RMSd (from solution structure), collision cross-section, and percentage of Watson−Crick and Hoogsteen hydrogen bonds. Bottom: averaged structures obtained in the last 50 ns of the replicas and distance contact maps. As a reference, we have included the water structure and its associated contact map.

selecting a series of different reasonable simulation conditions, which can define the boundaries of "real experimental conditions" (372 K) and of "ideally mild gas phase conditions" (300 K); see the Methods. Additionally, we explored the impact of simulation details and of the length of the triplex, with the ultimate goal to characterize the gas phase conformational space of triplexes as accurately and universally as possible.

Time Dependence. The time dependence was explored first, since current computers do not allow us to reach experimental time scales, forcing us to perform multiple shorter simulations (in this paper, 1.0 μs long), which might ignore slow conformational rearrangements. To check the expected impact of these very slow conformational arrangements in the gas phase, we perform a 30 μs simulation for one charge state (−6) of the 12-mer triplex (T = 300 K). Results (Figures 2 and S3−S5 (SI)) clearly demonstrate that no strand dissociation happens even after 10^{10} integration of Newton's equations of motion. The helix remains quite well-defined for the entire period, and structural rearrangements from the beginning to the end of the simulation are minuscule. This is noted (see Figures 2 and S4−S5 (SI)) in structural representations, RMSd plots, macroscopic descriptors, or even in the hydrogen bond pattern: 47% of Watson−Crick (WC) and 24% of Hoogsteen (H) hydrogen bonds are maintained after the first μs of MD, values that remain unaltered in the last microsecond of the trajectory (40% of WC and 30% of H hydrogen bonds are maintained). Very interestingly, we do not observe a loss of structure or an increase in internal mobility as time advances (Figures 2 and S4 (SI)), and in fact the compactness (as measured by radii of gyration, solvent accessible surface, or CCS) even increases slightly from 1 to 30 μs, suggesting that no unfolding is expected to happen in the millisecond time scale periods either, i.e., in the typical experimental time region. In summary, vaporization leads to a very fast transition from the canonical triplex to compact structures, which remain quite stable in the sub-millisecond time scales (looking at the extreme stiffness of the structure, probably even at the millisecond scale). Furthermore, the lack of significant differences between 30 and 1 μs structures strongly suggest that useful information can be derived from shorter simulations as those performed in the rest of this work.

Trajectory Heterogeneity. Trajectory heterogeneity was another source of concern in our simulations. MD is stochastic in nature, and it is a priori unclear whether different trajectories starting from slightly different starting conditions converge to the same ensemble. To investigate this point, we performed 5 × 1.0 μs trajectories for the 12-mer duplex using different starting configurations (taken always from the solution ensemble) and velocities (see the Methods) and a common charge state (for completeness, we selected one different to that studied in the 30 μs simulations). Results summarized in Figure 3 clearly show that trajectories separate in all cases quickly from the solution structure (typical RMSd around 0.7 nm), adopting different stable folds, which are very distant in terms of RMSd but share many common structural characteristics (Figure 3). As found in the very long trajectory, the helical structure is well

dx.doi.org/10.1021/ja209786t | J. Am. Chem. Soc. 2012, 134, 6596−6606

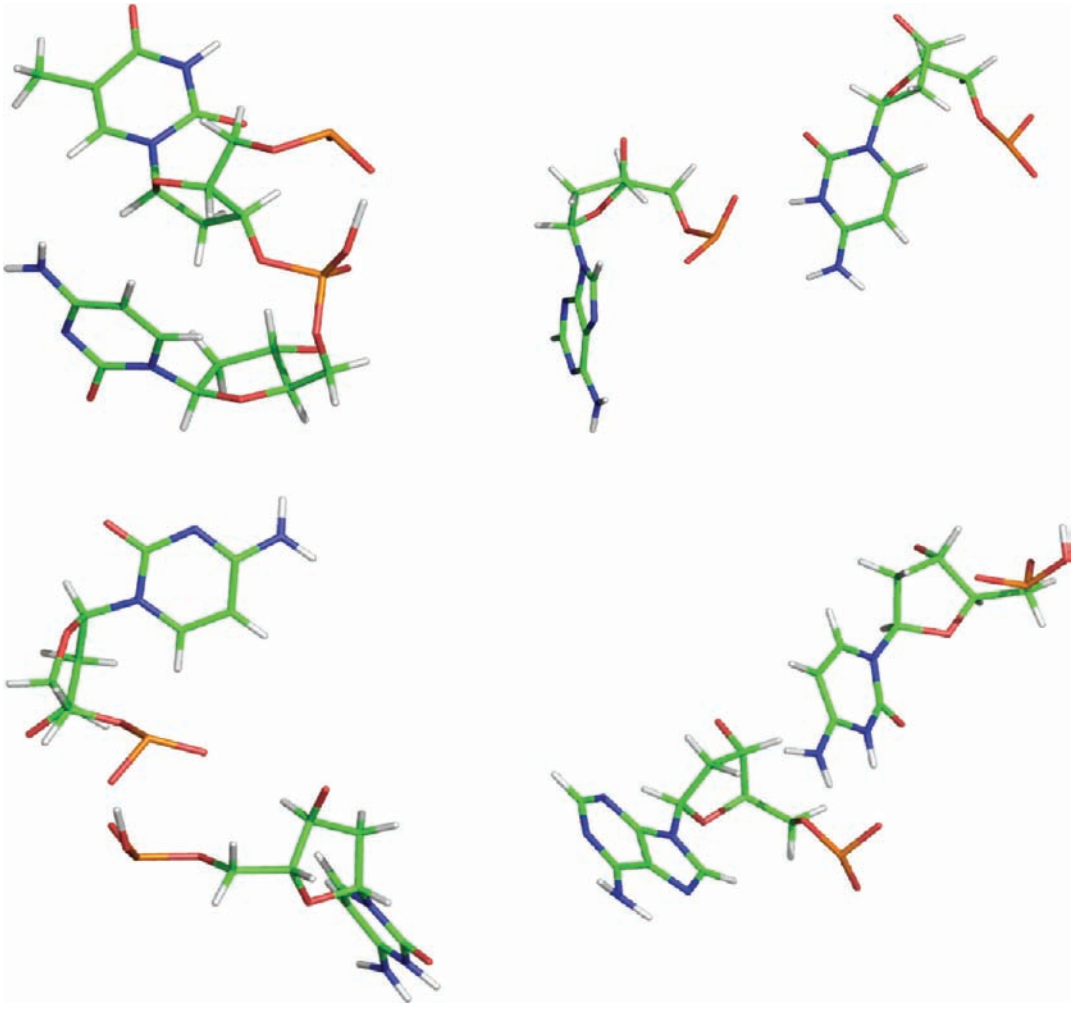


Figure 4. Schematic representation of typical molecular environments around charged phosphates and protonated cytosines. On the left side the images illustrate hydrogen bond mediated interaction between phosphate groups, whereas on the right side they depict the interaction between protonated cytosines and phosphate groups.

maintained in all replicas, even though the minor grooves are strongly reduced, mostly as a consequence of the formation of hydrogen bonds between charged and neutral phosphates (Figure 4). Additional helical distortions are due to the formation of saline bridges between Hoogsteen cytosines and anionic phosphates. These bridges induce changes in the placement of the nucleobases, which in some cases are extruded outside the pseudohelix, while in others they remain stacked inside. This arrangement allows a non-negligible amount of canonical hydrogen bonding to be kept, especially the Watson–Crick ones (Figure 3), while establishing many other noncanonical interactions. Overall, replica simulations illustrate that conformational transitions induced by vaporization have a quite stochastic nature, leading then to quite diverse conformers (in terms of RMSd). However, most of the RMSds are due to terminal movements (for example base flipping) or different packing of nucleobases around the helix axis. Very good convergence is achieved in the gas phase sampled structures in terms of global descriptors, like overall shape and type of contacts, including those properties experimentally measurable.

Effect of Charge Partitioning at Constant Total Charge. Vaporization is a fast process, and there are many phosphate groups in DNA triplex than can capture protons from the solvent or from the ammonium counterions during the transition to gas phase. Accordingly, many quasi-degenerated charge distributions can be populated for each charge state, which might lead to different structures. That is, we cannot expect structural diversity due only to the different total charge of the different triplex ions (see experimental results below), but also to the existence of different ways to distribute a given total charge over the structure. In order to cover this source of variability, we selected 10 alternative charge partitionings among the most stable found in our Monte Carlo titration procedure (see the Methods) and performed 1 μs MD simulations with all of them.

We detected fast and major structural rearrangements upon vaporization of the 12-mer triplex at room temperature (see RMSd data and Figures 4 and 5) but not strand detachment or unfolding for any of the studied charge partitioning possibilities. Gas phase structures are more rigid than the ones obtained from solution (collected at the same temperature), as noted in the mean RMSd with respect to their respective averages (Figure 5). The global helicity is maintained, and all the structures collected for the 10 charge partitioning possibilities display a clear memory of the solution conformation (see Figures 6 and S5 (SI)). As described above, the tendency of negatively charged phosphates to interact with neutral

dx.doi.org/10.1021/ja209786t | *J. Am. Chem. Soc.* 2012, 134, 6596–6606

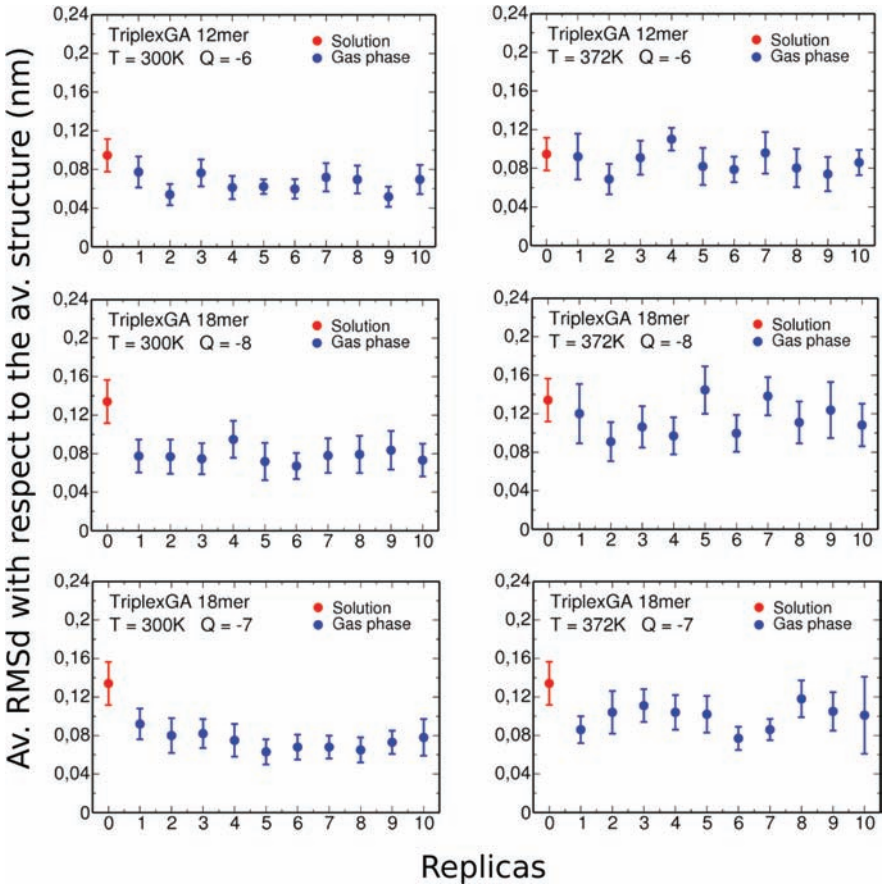


Figure 5. RMSd (bars correspond to associated standard derivations) with respect to the respective MD-averaged structures obtained for the 10 different charge states of the 12-mer and 18-mer triplexes in the gas phase at two different temperatures and charged states. The values obtained for the 100 ns MD simulations in water are indicated in all cases in red.

phosphates and protonated cytosines led to a dramatic narrowing of the minor and major groove. This narrowing causes the disruption of several nucleobase arrangements, some of which become extra-helical, justifying the slight increase in radii of gyration upon vaporization. Around 32% of canonical Watson−Crick and 14% of Hoogsteen hydrogen bonds are preserved in these simulations, values in close agreement with those extracted from multiple replicas and from the very long trajectory. Furthermore, around 50% native contacts are preserved, and many non-native contacts emerge. This leads to a small but consistent compression (around 5% in the CCS) of the triplex with respect to the solution situation, matching again the results found in the replicas of a single charge state and in the very extended MD simulation. Despite all the heterogeneity and the uncertainties in defining simulation setup, the reproducibility of the overall MD results is impressive.

Validation with a Larger Triplex and Different Total Charges. The overall results obtained for the 18-mer triplex agree with those obtained for the 12-mer (Figures 3−6 and S5 (SI)), and they are qualitatively equivalent for the charge states simulated that correspond to mild vaporization conditions ($Q = -7$ and $Q = -8$). Vaporization leads to major changes in the structure (RMSd around 0.7 nm from solution conformation), but the general helical conformation is well maintained. As in the 12-mer simulations, the largest source of distortion is the narrowing of the minor grooves as a consequence of phosphate−phosphate and cytosine$^+$−phosphate contacts (examples in Figure 4), which leads again to the disruption of many triads. A very significant number of canonical hydrogen bonds (around 31 (WC) and 24% (H)) and native contacts (around 43 and 21%, respectively) are maintained, which combined with a myriad of non-native interactions justify the compactness of the structure (average CCS in vacuum and solution are almost the same), as found for the 12-mer triplex. In summary, the results seem well converged with respect to the triplex length, and we cannot expect dramatic changes in the behavior of triplexes in the gas phase related to their different sizes, at least in a range of triplex lengths below the persistence length of DNA, which are typically used in mass spectroscopy.

Effect of Temperature. As commented above, it is difficult to map the spectrometer temperature into a simulation temperature, as the first one is a macroscopic concept, whereas the second is a microscopic magnitude derived exclusively from the internal velocities of the atoms in the macromolecule. However, we can define a range of temperatures (spanning from $T = 300$ to 372 K, see the Methods) that will cover from the "ideally" mild (low field drift tube ion mobility) to more disruptive (traveling wave ion mobility) conditions. Thus, to make our conclusions extensible to a reasonable range of temperatures (see the Methods and refs 25−27), we repeated the simulation of the 10 charge distributions for 12- and 18-mer triplexes using a higher effective temperature ($T = 372$ K).

Temperature has a dramatic effect in the conformation of DNA in solution, visible already in the multinanosecond

dx.doi.org/10.1021/ja209786t | J. Am. Chem. Soc. 2012, 134, 6596−6606

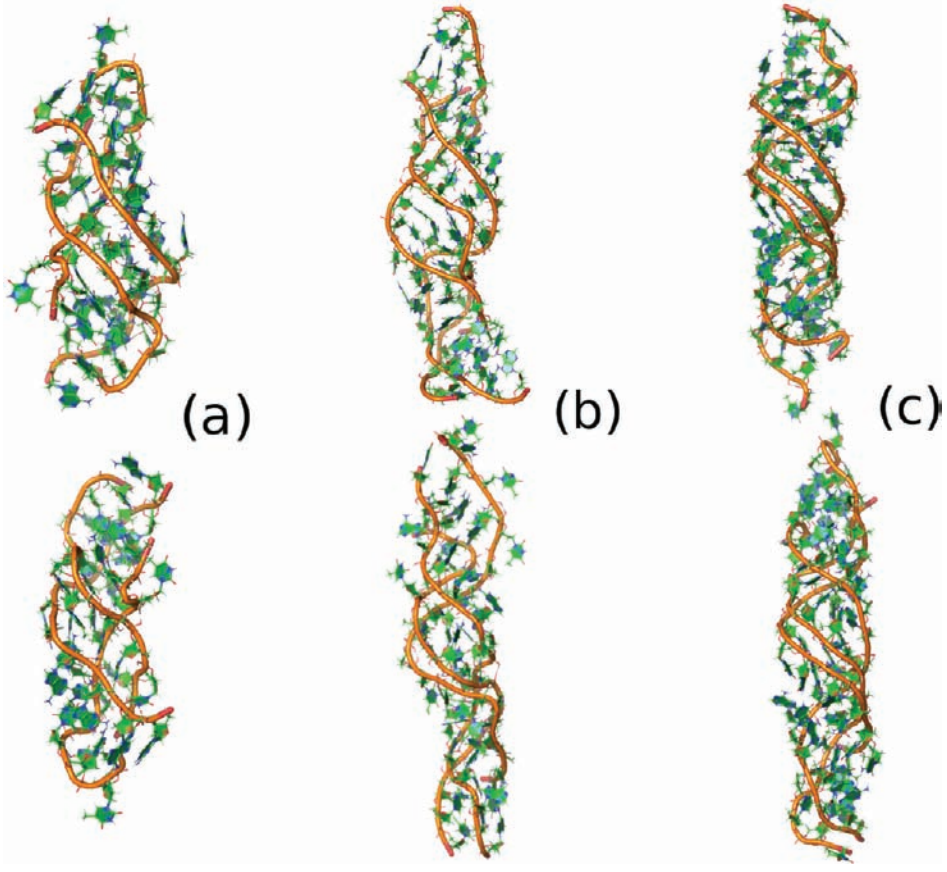


Figure 6. Representative averaged structures obtained from the last 50 ns of the trajectories of the 10 different charge distributions of the 12-mer and 18-mer triplexes in the gas phase at two different temperatures (above at 300 K, below at 372 K): (a) 12-mer, (b) 18-mer and $Q = -7$, (c) 18-mer and $Q = -8$. The RMSd between the shown structures and the rest are below 0.8 nm for each state and temperature.

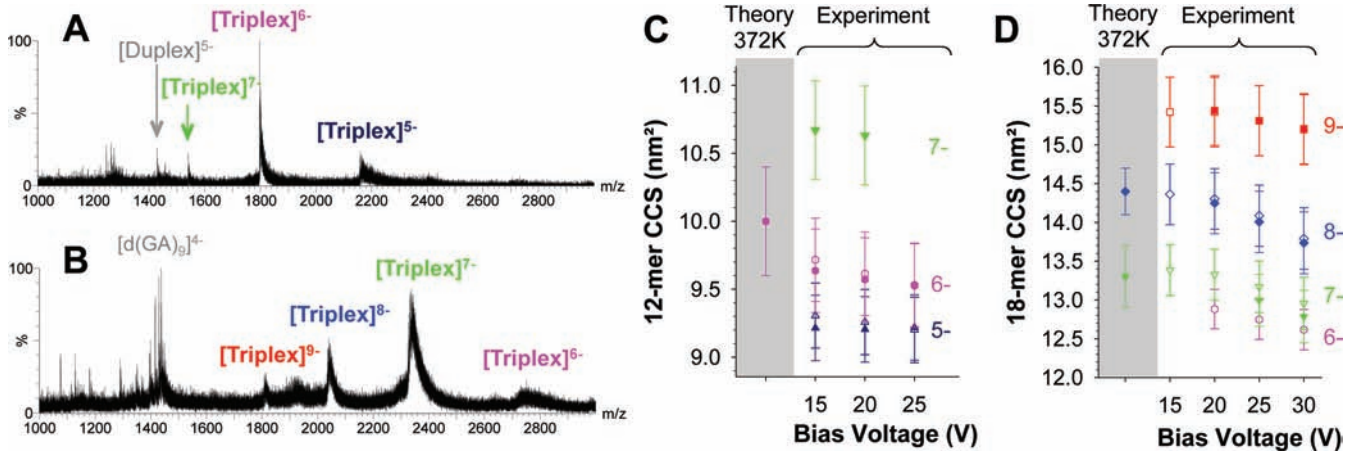


Figure 7. Comparison between simulations and experiments. (A,B) ESI-MS spectra of the 12-mer triplex (A) and 18-mer triplex (B) obtained from acidic conditions at a bias voltage of 20 V. (C,D) Experimental and theoretical collision cross sections obtained for the different charge states of the 12-mer triplex (C) and 18-mer triplex (D), as a function of the bias voltage. Dark blue, 5−; pink, 6−; green, 7−; navy blue, 8−; red, 9−; filled symbols, triplex with no ammonium adducts; open symbols, triplexes including the whole adduct distribution visible on the respective mass spectra.

scale,[19,81] but quite interestingly, it has little impact in the gas phase. All macroscopic descriptors (Rg, or CCS) are quite robust to changes in temperature, and in fact it is difficult to find any clear and systematic difference in the global shapes of the triplexes at low and high temperature (see Figure 6). Contact plots confirm that the global helical structure is not altered (Figure S5 (SI)) when gas phase simulations are run at high temperature and the only difference is related to the reduction of directional hydrogen bond contacts due to thermal vibrations. In summary, gas phase structures seem to be quite resistant to temperature variations within reasonable margins, and mass spectrometric experiments are probably not far from following "ideally mild" vaporization conditions, supporting the reliability of simulation conditions followed in the paper.

Experimental Validation. ESI-MS data clearly confirm previous results[34,61,62] suggesting that vaporization does not

dx.doi.org/10.1021/ja209786t | J. Am. Chem. Soc. 2012, 134, 6596−6606

lead to a disruption of the two triple helices considered here and that despite phosphate−phosphate repulsion, the three strands remain tightly bound. Figure 7 shows that the triplexes sprayed from acidic (pH = 3) conditions (see the Methods) were ionized at charge states −5, −6, and −7 for the 12-mer (Figure 7A) and from −6 to −9 for the 18-mer (Figure 7B). When samples at neutral pH were sprayed (not shown), the main charge states were −6, −7, and −8 for the 12-mer triplex and −7, −8, and −9 for the 18-mer triplex. The collision cross sections of the triplexes were measured as a function of the IMS cell bias voltage (Figure 7C,D). Increasing this voltage causes the ion population to heat up transiently, because of collisions with the nitrogen gas leaking out of the IMS cell, before being thermalized at around 372 K in the IMS cell itself. We see that the CCS decreases slightly when the bias voltage is increased. The CCS were also measured both for the naked triplex (no ammonium adduct, filled symbols) and for the whole ammonium adduct distribution when observable (open symbols). The CCS was found to increase very slightly and steadily with ammonium adducts, suggesting no significant conformational changes. The values of the CCS obtained at bias = 20 V are summarized in Table S5 (SI). Results obtained with the different experimental protocols (different sample preparation, instrument tuning, and calibration curve) agree.

Comparison between experimental CCS and values derived from the MD-derived structural samples can help to detect significant errors in our simulations. Nevertheless, also in case of agreement these results should be taken with caution. The sources of error in the experimental determination of the CCS with traveling wave IMS (errors in the published CCS and errors in traveling wave IMS measurements) are taken into account by using the 95% prediction interval, as described in Figure S1 (SI), which are reported in Figure 7C,D and Table S5 (SI). Also, there are expected errors in atomistic MD simulations, and probably more importantly there are intrinsic errors in the algorithms used to transform coordinates into CCS, which would probably require recalibration for nucleic acids as was already done for peptides.[82]

The analysis of the 55 μs of MD ensembles of the 12-mer triplex gives a theoretical CCS of 10.0 ± 0.4 nm^2 (total charge Q = −6), and the analysis of the cumulated 60 μs of MD trajectories for the 18-mer triplex leads to a theoretical CCS of 13.3 ± 0.4 nm^2 for Q = −7 and 14.4 ± 0.3 for Q = −8. These values are all close to the experimental estimates at bias voltages ≤ 20 V: 9.6 ± 0.3 nm^2 for the 12-mer, 13.3 ± 0.4 nm^2 (Q = −7) and 14.4 ± 0.5 nm^2 (Q = −8) for the 18-mer (see Table S5 (SI)). For the most abundant charge states, the results reproduce the collision cross section increase when the total charge of the ion increases. Keeping in mind all the sources of noise in theoretical or experimental CCS determination, the agreement between MD (using the projection approximation) and IMS-MS data is probably partially fortuitous but clearly supports the validity of our MD-derived conclusions and in general the use of last generation MD methods to obtain structural information under conditions where no high resolution structural data is available. Caution is however necessary to avoid overestimation of the capability of simulating different ions. Thus, for very charged ions, we cannot expect that classical MD simulations with simple force-fields and fixed topology can provide accurate results. In fact, test calculations of Q = −9 (for the 18-mer oligo) yield too compact structures in the microsecond range compared to experimental observable.

CONCLUSIONS

Simulations presented here provide an extended description of the nature of DNA triplexes in the gas phase and complete the atlas of canonical structures of DNA in the gas phase. Our simulations do not support the hypothesis that the triplex adopts a well-defined structure in the gas phase, as found for the DNA G-quadruplex. On the contrary, we suggest that it adopts a large number of stable but distorted structures, more similar to the situation found for duplex DNA. It is likely that different triplexes bearing the same total charge might display different distributions of charges in the molecule, which can evolve in the gas phase to different structures, and that even different molecules with the same distribution of charges might evolve to adopt different conformations. However, all these structures, which can be quite distinct in terms of RMSd, are close in terms of general conformational characteristics and macroscopic structural descriptors such as the collision cross section (CCS). The different metastable structures defining the gas phase ensemble are very rigid and do not interconvert even in the sub-millisecond time scale explored here. Interestingly, the results from the unprecedented amount of MD simulations performed for this work demonstrate that despite major changes in conformation, the triplexes maintain a clear memory of the helical structure in solution when transferred to the gas phase. Irrespective of the temperature and length of the oligonucleotide, a significant amount of native interactions are maintained upon vaporization, including many canonical hydrogen bonds, which combined with a myriad of non-native nucleobase and backbone interactions (especially important are those involving negatively charged phosphates) avoid the elongation of the triplex and leads to a small reduction of CCS with respect to solution state.

The good agreement between experimental and calculated CCS suggests that these structural models adequately describe the triplexes produced by ESI-MS at their most abundant charge states. Very interestingly, our work shows evidence that the key physical interactions that stabilize the secondary structure of nucleic acids, namely, screened Coulombic repulsion and the nucleobase-nucleobase hydrophobic terms, does not lead to the complete unfolding of a highly charged and nonglobular polymer such as triplex DNA when these interactions are tuned by the lack of solvent. On the contrary, when transferred to the gas phase, the DNA triplex maintains very clear signatures of the conformation in solution. This finding opens interesting possibilities for using gas phase structural data to infer solution structure of nucleic acids.

ASSOCIATED CONTENT

Supporting Information

Summary of simulations performed and experimental conditions (Tables S1 and S2). Calibrants for TWIMS (Table S3). Collision cross section calibration and comparison of the theoretical and experimental CCS for the different triplexes in all different conditions (Figure S1 and Table S4). Key structural descriptors of 12-mer and 18-mer triplexes both in solution and in gas phase for each temperature and charge state (Tables S5 and S6). Time evolution of structural descriptors of the triplexes in aqueous solution (Figure S2). RMSd map and time evolution of structural descriptors in 30 μs trajectory for 12-mer triplex (Figures S3 and S4). Examples of contact maps found in gas phase simulations (Figure S5). This information is available free of charge via the Internet at http://pubs.acs.org/.

AUTHOR INFORMATION

Corresponding Author

modesto.orozco@irbbarcelona.org

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We are indebted to Dr. Brandon T. Ruotolo (U. Cambridge) for their strong help and support in setting up IMS-MS experiments at the mass spectrometry core facility at IRB and Dr. Alberto Pérez (Stony Brooks) and Dr. Manuel Rueda (UCSD) for support in setting up MD simulations. This work was supported by the Spanish Ministry of Science and Innovation (BIO2009-10964 and Consolider E-Science), Instituto de Salud Carlos III (INB, COMBIOMED RETICS, Sara Borrell program (GP)), and Fundación Marcelino Botín. V.G. acknowledges the support of the Fonds de la Recherche Scientifique-FNRS (research associate position and FRFC Grant 2.4528.11).

REFERENCES

(1) Gale, D. C.; Smith, R. D. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 1154.
(2) Hofstadler, S. A.; Griffey, R. H. *Chem. Rev.* **2001**, *101*, 377.
(3) Gabelica, V.; De Pauw, E.; Rosu, F. *J. Mass Spectrom.* **1999**, *34*, 1328.
(4) Sharon, M.; Robinson, V. C. *Annu. Rev. Biochem.* **2007**, *76*, 167.
(5) Banks, J. F.; Whitehouse, C. M. *Methods Enzymol.* **1996**, *270*, 486.
(6) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64.
(7) Benesch, J. L.; Robinson, C. V. *Curr. Opin. Struct. Biol.* **2006**, *16*, 245.
(8) Benjamin, D. R.; Robinson, C. V.; Hendrick, J. P.; Harti, F. U.; Dobson, C. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 7391.
(9) Heck, A. J. R.; van den Heuvel, C. V. *Mass Spectrom. Rev.* **2004**, *23*, 368.
(10) Wyttenbach, T.; Bower, M. T. *Top. Curr. Chem.* **2003**, *255*, 207.
(11) Jurnzecko, E.; Barran, P. E. *Analyst* **2011**, *136*, 20.
(12) Hopper, J. T. S.; Oldham, N. J. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1851.
(13) Koeniger, S. L.; Merembloom, S. I.; Clemmer, D. E. *J. Phys. Chem. B* **2006**, *110*, 7017.
(14) Neutze, R.; Wouts, R.; van der Spoel, D.; Weckert, E.; Hajdu, J. *Nature* **2000**, *406*, 752.
(15) Neutze, R.; Hudt, G.; van der Spoel, D. *Radiat. Phys. Chem.* **2004**, *71*, 905.
(16) Patriksson, A.; Marklund, E; van der Spoel, D. *Biochemistry* **2007**, *46*, 933.
(17) Meyer, T.; de la Cruz, X.; Orozco, M. *Structure* **2009**, *17*, 88.
(18) D'Abramo, M.; Meyer, T.; Bernadó, P.; Fernandez-Recio, J.; Orozco, M. *J. Chem. Theor. Comput.* **2009**, *5*, 3129.
(19) Pérez, A.; Orozco, M. *Angew. Chem., Int. Ed. Engl.* **2010**, *49*, 4805.
(20) Reyzer, M. L.; Brobdbelt, J. S.; Kerwin, S. M.; Kuman, D *Nucleic Acids Res.* **2001**, *14*, 464.
(21) Gabelica, C.; De Pauw, E. *Int. J. Mass Spectrom.* **2002**, *219*, 151.
(22) Wan, K. X.; Shibue, T.; Grom, M. L. *J. Am. Chem. Soc.* **2000**, *122*, 300.
(23) Gidden, J.; Ferzoco, A.; Baker, E. S.; Bower, M. T. *J. Am. Chem. Soc.* **2004**, *126*, 15132.
(24) Rosu, F.; Gabelica, V.; Houssier, C.; Colson, P.; De Pauw, E. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1729.
(25) Rueda, M.; Kalko, S. G.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2003**, *125*, 8007−8014.
(26) Rueda, M.; Luque, J.; Orozco, M. *J. Am. Chem. Soc.* **2005**, *127*, 11690−11698.
(27) Rueda, M.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2006**, *128*, 3608−3619.
(28) Vairamani, M.; Gross, M. L. *J. Am. Chem. Soc.* **2003**, *125*, 42.
(29) Gabelica, C.; Rosu, F.; Witt, M.; Baykut, G.; De Pauw, E. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 201.
(30) Gidden, J.; Baker, E. S.; Ferzoco, A.; Bowers, M. T. *Int. J. Mass Spectrom.* **2005**, *240*, 183.
(31) Rosu, F.; Gabelica, V.; Poncelet, H.; De Pauw, E *Nucleic Acids Res.* **2010**, *38*, 5217.
(32) Gabelica, V.; Baker, E. S.; Teulade-Fichou, M. P.; De Pauw, E.; Bowers, M. T. *J. Am. Chem. Soc.* **2007**, *129*, 895−904.
(33) Scaria, P. V.; Shafer, R. H. *Biochemistry* **1996**, *35*, 10985.
(34) Chandler, S. P.; Fox, K. R. *Biochemistry* **1996**, *35*, 15038.
(35) Washbrook, E.; Fox, K. R. *Biochem. J.* **1994**, *301*, 569.
(36) Cheng, Y. K.; Pettitt, B. M. *J. Am. Chem. Soc.* **1992**, *114*, 4465.
(37) Schultze, P.; Koshlap, K. M.; Feigon, J. *Biochemistry* **1997**, *36*, 2659.
(38) Koshlap, K. M.; Schultze, P.; Brunar, H.; Dervan, P. D.; Feigon, J. *Biochemistry* **1997**, *26*, 2659.
(39) Bartley, J. P.; Brown, T.; Lane, A. N. *Biochemistry* **1997**, *36*, 14502.
(40) Gotfredsen, C. H.; Schultze, P.; Feigon, J. *J. Am. Chem. Soc.* **1998**, *120*, 4281.
(41) Asensio, J. L.; Brown, T.; Lane, A. N *Nucleic Acids Res.* **1998**, *26*, 3677.
(42) Tarkoy, M.; Phipps, A. K.; Schultze, P.; Feigon, J. *Biochemistry* **1998**, *37*, 5810.
(43) Phipps, A. K.; Tarkoy, M.; Schultze, P.; Feigon, J. *Biochemistry* **1998**, *37*, 5820.
(44) Asensio, J. L.; Dosanjh, H. S.; Jenkins, T. C.; Lane, A. N. *Biochemistry* **1999**, *37*, 1588.
(45) Lane, A. N; Jenkins, T. C. *Curr. Org. Chem.* **2001**, *5*, 845.
(46) Shields, G. C.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1997**, *119*, 7463.
(47) Soliva, R.; Laughton, C. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1998**, *120*, 1126.
(48) Soliva, R.; Luque, F. J.; Orozco, M. *Nucleic Acids Res.* **1999**, *27*, 2248.
(49) Cubero, E; Güimil-Garcia, R.; Luque, F. J.; Eritja, R.; Orozco. *Nucleic Acids Res.* **2001**, *29*, 2522.
(50) Spackova, N.; Cubero, E.; Sponer, J.; Orozco, M. *J. Am. Chem. Soc.* **2004**, *126*, 14642.
(51) Orozco, M.; Pérez, A.; Noy, A.; Luque, F. *J. Chem. Soc. Rev.* **2003**, *32*, 350.
(52) Jiménez, E.; Vaquero, A.; Espinás, M. L.; Soliva, R.; Orozco, M.; Bernués, J.; Azorín, F. *J. Biol. Chem.* **1998**, *273*, 24640.
(53) Havre, P. A.; Gunther, E. J.; Gasparro, F. P.; Glazer, P. M. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 7879.
(54) Curiel, D. T.; Wagner, E.; Cotten, M.; Birnstiel, M. L.; Agarwal, S.; Li, C. M.; Loechel, S.; Hu, P. C. *Hum. Gene Ther.* **1992**, 147.
(55) Shen, C.; Buck, A.; Polat, B.; Schmid-Kotsas, A.; Matuschek, C.; Gross, H.; Bachem, M.; Reske, S. N. *Cancer Gene Ther.* **2003**, *10*, 403.
(56) Soyfer, V. N.; Potaman, V. N. *Triple Helical Structures*; Springer: New York, 1995.
(57) Majumdar, A.; Khorlin, A.; Dyatkina, N.; Lin, F. L.; Powell, J.; Liu, J.; Fei, Z.; Khirpine, Y.; Watanable, K. A.; George, J.; Glazer, P. M.; Seidman, M. M. *Nat. Genet.* **1998**, *20*, 212.
(58) Goñi, J. R.; Vaquerizas, J. M.; Dopazo, J.; Orozco, M *BMC Genomics* **2006**, *7*, 63.
(59) Goñi, R.; de la Cruz, X.; Orozco, M *Nucleic Acids Res* **2004**, *32*, 354−360.
(60) Chan, P. P.; Glazer, P. M. *J. Mol. Med.* **1997**, *75*, 267.
(61) Wan, C.; Guo, X.; Liu, Z.; Liu, S. *J. Mass Spectrom.* **2008**, *43*, 164.
(62) Rosu, F.; Nguyen, C.-H.; De Pauw, E.; Gabelica, V. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1052.
(63) Rosu, F.; De Pauw, E.; Gabelica, V. *Biochimie* **2008**, *90*, 1074.
(64) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(65) Smith, D. E.; Dang, L. X. *J. Chem. Phys.* **1994**, *100*, 3757−66.
(66) Noy, A.; Soteras, I.; Luque, F. J.; Orozco, M. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10596.
(67) Gabelica, V.; De Pauw, E. *Mass Spectrom. Rev.* **2005**, *24*, 566.
(68) Morsa, D.; Gabelica, V.; De Pauw, E. *Anal. Chem.* **2011**, *83*, 5775−5782.
(69) Shvartsburg, A. A.; Smith, R. D. *Anal. Chem.* **2008**, *80*, 9689−9699.
(70) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laugthon, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817.
(71) Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845.
(72) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1983**, *98*, 10089.
(73) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Hook, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(74) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(75) Case, D. A.; Darden, T.; Gohlke, H.; Luo, R.; Merz, J. R.; K., M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. *J. Comput. Chem.* **2005**, *26*, 1668−1688.
(76) Hess, B.; Kutzer, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435.
(77) Shvartsburg, A. A.; Jarrold, M. F.; Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. *Chem. Phys. Lett.* **1996**, *261*, 86.
(78) Wyttenbach, T.; Witt, M.; Bowers, M. T. *J. Am. Chem. Soc.* **2000**, *122*, 3458.
(79) Wyttenbach, T.; von Helden, G.; Batka, J.; Carlat, D.; Bowers, M. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 275.
(80) Turner, K. B.; Nagan, N.; Fabris, D. *Nucleic Acids Res.* **2006**, *34*, 1305.
(81) Portella, G.; Orozco, M. *Angew. Chem., Int. Ed. Engl.* **2010**, *49*, 7673.
(82) Siu, C.-K.; Guo, Y.; Saminathan, I. S.; Hopkinson, A. C.; Siu, K. W. M. *J. Phys. Chem. B* **2009**, *114*, 1204−1212.