2412–2423 *Nucleic Acids Research, 2015, Vol. 43, No. 4* *Published online 6 February 2015*
doi: 10.1093/nar/gkv080

Analyzing ion distributions around DNA: sequence-dependence of potassium ion distributions from microsecond molecular dynamics

Marco Pasi[1], John H. Maddocks[2] and Richard Lavery[1,*]

[1]Bases Moléculaires et Structurales des Systèmes Infectieux, CNRS UM 5086/Université Lyon I, IBCP, 7 passage du Vercors, 69367 Lyon, France and [2]Section de Mathématiques, Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland

Received December 22, 2014; Revised January 20, 2015; Accepted January 21, 2015

ABSTRACT

Microsecond molecular dynamics simulations of B-DNA oligomers carried out in an aqueous environment with a physiological salt concentration enable us to perform a detailed analysis of how potassium ions interact with the double helix. The oligomers studied contain all 136 distinct tetranucleotides and we are thus able to make a comprehensive analysis of base sequence effects. Using a recently developed curvilinear helicoidal coordinate method we are able to analyze the details of ion populations and densities within the major and minor grooves and in the space surrounding DNA. The results show higher ion populations than have typically been observed in earlier studies and sequence effects that go beyond the nature of individual base pairs or base pair steps. We also show that, in some special cases, ion distributions converge very slowly and, on a microsecond timescale, do not reflect the symmetry of the corresponding base sequence.

INTRODUCTION

A distinctive feature of nucleic acids is their charge. DNA and RNA are anionic polyelectrolytes that bear one negative charge per nucleotide and the resulting high charge density generates strong repulsive forces that must be reduced for nucleic acids to perform their biological function. This reduction is achieved through electrostatic screening by water and counterions and it has long been known that the solvent environment plays a significant role in the structure of DNA (1). These effects are particularly important for double-stranded nucleic acids (2,3), and are likely to modulate the binding of proteins and small molecules to double-stranded DNA (4,5) and therefore influence its biological activity.

The structure of the ionic atmosphere around DNA has a long history of physical modeling, marked by the establishment of counterion condensation theory (6) and polyelectrolyte continuum models based on the Poisson–Boltzmann equation (7). More recently, a large number of experimental studies have been performed to examine the structural details of cations binding to DNA (8–15), (see also the recent reviews by Hud and Engelhart (8), Pollack (16) and Lipfert *et al.* (17)).

The behavior of monovalent cations has remained particularly elusive, owing to the difficulties of observing them by either X-ray crystallography or nuclear magnetic resonance spectroscopy. Nevertheless, thanks to clever experimental setups, often involving the use of non-physiological ions (such as Tl^+, Rb^+, Cs^+ and $^{15}NH_4^+$), it has been possible to show that partially dehydrated cations bind deep in both grooves in a highly sequence-dependent manner, often competing with water for the same binding sites and interacting with accessible electronegative atoms of DNA either directly or through shared first-shell water molecules. Experiments have shown that Tl^+ (a mimic of K^+) binds in the major groove of B-DNA at GC steps, with estimated occupancies between 0.20 and 0.35 (9). Several monovalent cations have been observed in the minor groove (K^+, Na^+, Tl^+, Rb^+, Cs^+, NH_4^+), especially in regions where the groove is narrow (e.g. 3′ end of poly-A tracts) (12,15,18), with occupancies as high as 0.5 (19).

Molecular dynamics (MD) simulations have been used extensively to obtain atomic-level insight into DNA–counterion interactions. Since the pioneering work of Young *et al.* (20), a large number of simulation studies have been performed (see the recent reviews by Auffinger and Hashem (21) and, Mocci and Laaksonen (22)). However, the MD picture of the ionic atmosphere remains the subject of some controversy. Along with potential concerns about the reliability of the empirical force-field parameters, and, in most cases, the absence of a polarizability term (23), there is also the issue of lack of convergence given that the slow dif-

*To whom correspondence should be addressed. Tel: +33 4 7272 2637; Fax: +33 4 7272 2604; Email: richard.lavery@ibcp.fr

fusion of ions in water will cause problems for nanosecond-scale simulations (24,25).

Early, short MD simulations (< 2 ns) of alternating GC and AT oligomers in explicit solvent with near-physiological concentrations of KCl were used to identify the preferential binding of K^+ to the major groove of DNA (26,27). Ion populations in the major groove were found to be about 0.1 for A:T and 0.3 for G:C base pairs, while no ions were identified in the minor groove (26). Longer simulations of 15 ns found lower K^+ populations in the major groove and some ion density in the minor groove (28), while 50 ns simulations (24) found maximum populations below 0.13 in both grooves and suggested that the sequence dependence of ion distributions might in fact go beyond the base-pair step.

Turning to sodium cations, 10 ns simulations of the so-called Drew–Dickerson dodecamer (5′-CGCGAATTCGCG-3′, hereafter DDD) showed maximum Na^+ populations of 0.2 in the minor groove of the central AT-rich region (29). Longer simulations of DDD (15 ns), again found lower sodium populations (below 0.1) in the minor groove (30). Extending the simulation time even further to 60 ns (31) increased the maximum occupation fractions to 0.15 and 0.3 for the major and minor groove, respectively, while only 0.04 was found for the optimal site using the first microsecond trajectory for DDD (32). A recent microsecond-scale analysis focusing on CG steps also identified populations around 0.3 in the minor groove for both sodium and potassium (33), and also provided clear evidence for the role of monovalent cations in triggering DNA backbone conformational transitions (34).

The convergence issue has driven the development of alternative approaches to MD, including enhanced sampling (35), continuum Poisson–Boltzmann (36) and intermediate-resolution approaches (37,38). However, thanks to developments in force field parameterization (39,40) and improvements in high performance computing resources (41) convergence has now become reachable using atomic-level MD. Multiple microsecond-scale MD simulations of DNA (42) have however shown that several hundreds of nanoseconds are generally required to obtain converged ion populations (43).

Apart from convergence issues, another problem contributing to discrepancies in observed ion populations is the difficulty in clearly defining physically meaningful spatial regions around DNA, since measuring ion distributions from MD simulations is commonly limited to observing their occurrence within regions surrounding specific DNA atoms. Furthermore, as many of the experimental and simulation studies focused on the DDD oligomer, on other A-tract sequences or on simple dinucleotide repeats, the available data does not provide a comprehensive view of base sequence effects on the ion atmosphere around DNA.

Here, we present a significant step toward understanding the sequence dependence of K^+ binding to B-DNA at room temperature under physiological salt conditions. Thanks to a recently developed curvilinear helicoidal coordinate (CHC) system and associated analysis tools we are able to accurately quantify ion populations and densities within the grooves and in the immediate surroundings of the double helix (43). We apply these tools to a set of microsecond MD trajectories for 39 oligomers containing the 136 distinct tetranucleotide base sequences. These trajectories are converged both in terms of DNA conformation (42) and with one notable exception, in terms of ion atmosphere (43). Our results show ion populations within the grooves that are significantly higher than those typically observed in earlier studies. Furthermore, we find sequence effects on ion populations that go beyond the nature of individual base pairs or base pair steps. Finally, we show that ion distributions associated with long A-tracts converge very slowly and do not reflect the symmetry of the base sequence at least on a microsecond timescale.

MATERIALS AND METHODS

Molecular dynamics simulations

We study the sequence dependence of the potassium cation atmosphere around B-DNA by analyzing MD trajectories obtained for 39 double-stranded B-DNA oligomers, each containing 18 bp. The sequence of each oligomer follows the pattern: 5′-gc-*CD-ABCD-ABCD-ABCD*-gc-3′, where upper case letters indicate sequences that vary between oligomers and lower case letters indicate fixed sequences (dashes have been added for clarity). Every oligomer contains a different 4-bp sequence, *ABCD*, that is repeated three and a half times. This sequence is also used as the name for each oligomer. The full list of the 39 oligomers is given in (42). Since taken together, these oligomers contain all 136 distinct tetranucleotide sequences, the analysis of this dataset of simulations allows us to fully describe the sequence dependence of cation binding up to the tetranucleotide level.

MD simulations were carried out using the AMBER suite of programs (44,45), with periodic boundary conditions inside a truncated octahedral cell with the parmbsc0 modifications (39) to the parm99 force field (45,46), Dang parameters (40) for the ions and SPC/E water (47). Each system was neutralized with 34 potassium ions and then an appropriate number of K^+Cl^- ion pairs were added to reach a salt concentration of 150 mM. Note that this implies that the number of K^+ exceeds the number of Cl^- ions in the simulation cell, as would be expected close to a highly negatively charged solute molecule and also implies an effective bulk K^+ molarity of 336 mM. The ions were initially placed at random within the simulation cell, but at least 5 Å from DNA and at least 3.5 Å from one another. The complex was then solvated with a layer of water at least 10 Å thick. A typical simulation involved ∼11 500 water molecules and 37 000 atoms in total. Electrostatic interactions were treated using the particle mesh Ewald method (48) with a real-space cutoff of 9 Å and cubic B-spline interpolation onto the charge grid with a spacing of 1 Å. Lennard-Jones interactions were truncated at 9 Å and the pairlist was built with a buffer region and a triggered list update whenever a particle moved more than 0.5 Å from the previous update.

Each oligomer was constructed in a canonical B-DNA conformation, solvated, then equilibrated by energy minimization of the solvent and then of the solute and solvent together, followed by a slow thermalization, using the protocol described earlier (49–51). Simulations were carried out

in an NPT ensemble, using the Berendsen algorithm (52) to control temperature and pressure, with a coupling constant of 5 ps for both parameters. All chemical bonds involving hydrogen atoms were restrained using SHAKE (53), allowing for stable simulations with a 2 fs time step. Center of mass motion was removed every 500 steps to avoid kinetic energy building up in translational motion (54) and to keep the solute centered in the simulation cell. Each of the 39 simulations was run for 1 μs, saving snapshots every 1 ps; the analysis described here is performed using the snapshots corresponding to the final 900 ns of each simulation.

Ion distribution analysis

The distribution of potassium cations around B-DNA was determined from the MD trajectories, and analyzed using a recently developed CHC system (43). The latest version of Curves+ (55,56) has the ability to calculate the position of ions, water molecules and any chosen solute atoms along a trajectory using a CHC system with respect to the instantaneous helical axis of each snapshot. This data can then be analyzed with the Canion utility program, which accumulates the ion positions to generate a 3D CHC histogram. In the CHC system, the position of an ion is defined in terms of its longitudinal (D), radial (R) and angular (A) coordinates; since these are measured based on the local helical axis of DNA, the CHC system allows the generation of distributions that naturally account for the coupling between ion motions and the global deformations of the DNA molecule. The D coordinate describes the position of the ion along the DNA molecule; it is measured in units of base pair steps, and therefore varies continuously between 1 and N within a DNA segment of N base pairs The R and A coordinates are the distance and angle, respectively, in a polar coordinate system in the plane perpendicular to the helical axis. The R coordinate is the distance from the helical axis, measured in Å, while A is measured in degrees and in canonical B-DNA, $A = 0$ is the line joining the two strands, oriented in the Crick → Watson direction.

In the analysis that follows we use two ways of describing ion distributions around DNA, namely ion populations, in units of (usually fractional) numbers of ions, and ion densities, in units of molarity. Ion populations are computed by taking the time average of the numbers of ions at each simulation snapshot within the given region as defined by chosen ranges of the three CHC coordinates. In the particular case of a sufficiently small region, there will effectively be at most one ion present at each time snapshot, so that the ion population reduces to an occupancy, i.e. the fraction of time in which an ion is actually present in the given region. To obtain molarities we must divide by a volume, and we consistently do this by computing the 3D Cartesian volume of the given CHC region with respect to the time-averaged structure of the DNA (and, in particular, its corresponding helical axis system). Furthermore, when we compute ion densities close to the DNA we reduce the Cartesian volume of the given CHC region by excluding volume inside the van der Waals envelope of the DNA (computed using standard Pauling atomic radii). When the time-averaged DNA structure is itself curved, the physical volume of bins corresponding to uniformly discretized CHC coordinates will vary with each of the three coordinates, but in a known way. Isosurfaces of the resulting molarity distribution can then be visualized using molecular graphics software.

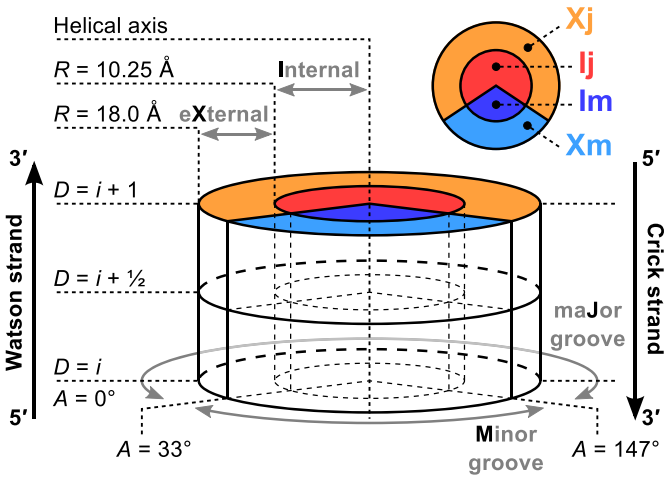

Figure 1. Schematic view of a base-pair step. Untwisted view of a base-pair step detailing the CHC space-partitioning scheme used to measure ion populations around DNA. In order to distinguish ion binding at the level of a base pair i, or between base pairs i → i + 1, we split the base-pair steps in two halves along the D coordinate. The angular space is also split in two regions: the minor groove, $33^\circ < A < 147^\circ$ (abbreviated 'm', blue) and major groove, $147^\circ \leq A \leq 33^\circ$ (abbreviated 'j', red). Finally, extending out from the helical axis, we consider two regions: 'internal' (I), for distances from the helical axis up to the phosphorous radius ($R \leq 10.25$ Å), and 'external' (X), stretching out to the counterion condensation theory radius ($10.25 < R \leq 18$ Å, see also Figures 2 and 3).

CHC distributions can also be analyzed in terms of their 1D and 2D marginals, which are obtained by integrating over ranges of respectively two or one of the CHC coordinates. By selecting these ranges appropriately, we can describe the ion populations within well-defined, physically relevant regions. This strategy becomes particularly useful when studying the sequence-dependent features of the ionic atmosphere. Figure 1 presents a schematic view of the CHC space discretization used for sequence-dependence analyses. According to our previous study (43), we define the limits of the grooves using the average positions of the C1' atoms, with the minor groove spanning $33^\circ < A < 147^\circ$. We define the 'internal' (or 'I') region of the grooves to include everything that is closer to the helical axis than the average distance of the phosphorous atoms, $R = 10.25$ Å. Following counterion condensation theory (6,57), we also define an 'external' (or 'X') region in the interval $10.25 < R \leq 18$ Å, given that 18 Å corresponds to the 76% neutralization of the phosphate charges (see results section), which is predicted by Manning's theory to correspond to the limit of monovalent ion condensation. When necessary we use the short notation Im and Ij to distinguish the internal regions of the minor and major grooves and Xm and Xj to indicate the corresponding external regions. Finally, to distinguish ions binding at the level of the base pair, or at base-pair steps (i.e. between base pairs), the D coordinate is discretized in steps of 0.5. Binding events at levels are analyzed by integrating over the two half steps flanking the level (i.e. for level $D =$ i, the half step $\mathrm{i}-\frac{1}{2} < D \leq \mathrm{i}$ and $\mathrm{i} < D \leq \mathrm{i}+\frac{1}{2}$), and grouped

in terms of their trinucleotide sequence dependence (i − 1, i, i + 1). On the other hand, binding events at steps are analyzed by summing the two half step values between the two base pairs (i.e. for base pairs i and i + 1, the half steps $i < D \leq i + \frac{1}{2}$ and $i + \frac{1}{2} < D \leq i + 1$) and grouped according to their tetranucleotide sequence (i − 1, i, i + 1, i + 2).

Population analyses for specific regions were performed using the relevant flags of the Canion program, which is available, together with the present versions of Curves+ and other related software, at http://bisi.ibcp.fr/tools/curves_plus/. Post-processing and plotting were performed using numpy (58) and matplotlib (59). Molecular graphics were created using Chimera (60,61).

Sequence fragments of double stranded DNA are always written in the 5′-3′ direction as a simple string of letters (e.g. the tetranucleotide AGAG or the dinucleotide step GA), which corresponds to the sequence along the Watson strand. Since we only consider canonically base-paired DNA, the sequence on the opposite (Crick) strand is implied. When referring to a dinucleotide step within a longer sequence, we underline the step in question (e.g. ACGT). Purines and pyrimidines are indicated using the letters R and Y respectively, while X indicates any of the four possible bases. When we refer to the base pairs flanking a given sequence, we separate them by two dots (e.g. A..G). Base pairs are distinguished from base pair steps by using a colon (A:T or G:C).

RESULTS AND DISCUSSION

Before discussing the detailed results obtained in this study we should note that the ion populations we present appear to converged (with one notable exception, see below). As we have shown in an earlier publication (43), convergence is typically achieved in roughly 300–400 ns. The results presented here are derived from the final 900 ns of 1 μs trajectories carried out for each of the 39 oligomers studied by the groups participating in the ABC project (42,49–51). Ion distributions, which in the present study are limited to an analysis of potassium ions (K^+), are described in terms of the CHC D, R and A presented in the preceding section.

We begin by considering the K^+ population around B-DNA as a function of the distance R from the helical axis. Figure 2 shows the cumulative population for six oligomers belonging to the ABC set that contain the 10 distinct dinucleotide steps. These populations are averaged over all angles A and over the full length D of each of the oligomers. From the upper panel, we can see that there are roughly between five and nine K^+ ions within the grooves ($R = 10.25$, the average phosphorus radius) of these 18-bp oligomers. In the following text, we refer to the region within the grooves ($R \leq 10.25$) as the internal or 'I' region. In order to fully neutralize the net charge of the oligomers (−34, corresponding to two strands, each containing 17 anionic phosphate groups) it is necessary to integrate the ion distribution up to $R = 24$ Å. However, if we consider the 76% neutralization predicted by counterion condensation theory for monovalent ions (6,57), this value is achieved at 18 Å. This distance should contain all nominally 'bound' counterions and this indeed seems to be the case given the radial distribution plots of K^+ molarity shown in the left panel of

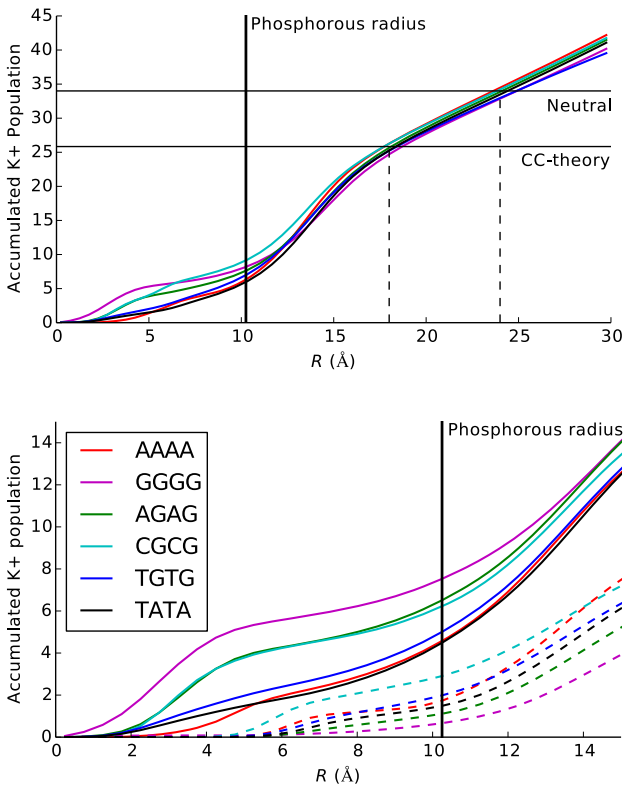

Figure 2. Accumulated K^+ populations. Cumulative K^+ ion population as a function of distance from the helical axis (R, in Å). Populations are integrated along the full length of six 18-mers (see legend of bottom panel) and either over the full angular range (top panel), or discriminating the major and minor grooves (bottom panel, solid and dashed lines, respectively). The solid vertical line indicates the phosphorous radius at $R = 10.25$ Å. Horizontal lines in the top panel show the accumulated ion population required to fully neutralize the DNA (one negative charge on each of the 34 phosphates of an 18-mer), or to achieve 76% neutralization (CC-theory). Vertical dashed lines indicate the average distances (24 and 18 Å, respectively) at which these conditions are achieved.

Figure 3, since by $R = 18$ Å we have virtually reached the bulk K^+ molarity (0.336 M, taking into account the extra potassium ions added to achieve electroneutrality). In the following text, we refer to the zone corresponding to bound ions outside the grooves ($10.25 < R \leq 18$ Å) as the external or 'X' region.

Returning to Figure 2, we see little sequence dependence beyond the phosphorus radius, however the differences become clearer below this radius, and, in particular, when we separate the minor (dotted lines) and major groove (solid lines) populations, as in the lower panel of Figure 2. The major groove populations are generally significantly larger than those of the minor groove since the former covers an angular range of 246°, while the latter covers only 114° (and consequently the accessible volume is significantly larger in the major groove). However, the base sequence modulates the populations and, for example, the major groove of the GGGG oligomer contains three more ions than that of AAAA, whereas its minor groove contains one ion less.

We now consider the distribution of K^+ density around DNA averaged over the full set of 136 distinct tetranucleotides. The left hand panels of Figure 3 show this dis-

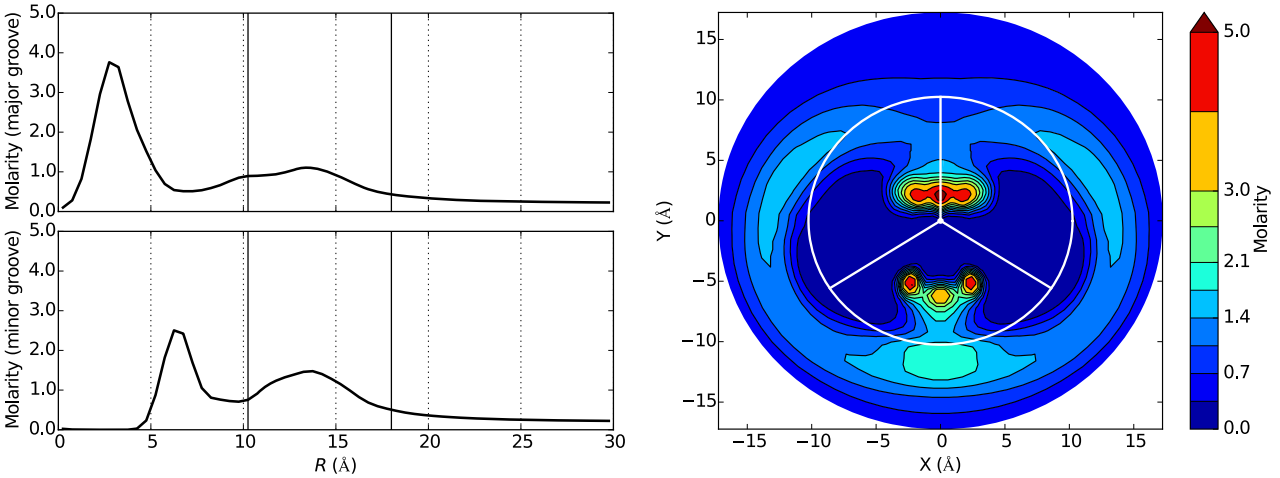

Figure 3. Average K^+ atmosphere. Structure of the K^+ atmosphere around B-DNA, obtained averaging over the 136 distinct tetranucleotides. Left: K^+ molarity distribution as a function of the distance from the helical axis (R, in Å) for the major (upper plot) and the minor (lower plot) grooves. The vertical bars show distances of 10.25 and 18 Å from the helical axis. Right: 2D K^+ molarity distribution in the *RA* plane. The molarity increases going from blue to red, in discrete, non-uniform steps chosen to highlight the structure of the distribution. The white circle indicates the phosphorous radius ($R = 10.25$), while white radial vectors indicate the groove limits and the center of the major groove (vertical vector); see also Figure 1.

tribution plotted in units of molarity, separately for the major and minor grooves. In each case, we see two molarity peaks, one within the groove and one between the phosphorus radius and the counterion condensation limit of 18 Å. Note that a small shoulder occurs in the major groove distribution due to a restricted volume at the entrance to the groove. Note also that the K^+ ions can approach the helical axis more closely on the side of the major groove. This can be clearly seen in the 2D *RA* plot on the right of Figure 3, where the major groove is at the top of the figure. This difference is less marked for molarity peaks beyond the phosporus radius and, in consequence, the two ion density peaks are much closer to one another on the minor groove side (lower left panel) than they are for the major groove (upper left panel). We will return to this point later. Looking at the *RA* plot in more detail we can also see that the highest molarity in the major groove lies in a continuous zone spanning the center of the groove, while the minor groove density shows three distinct peaks. The reason for this will become clear when we consider base sequence effects. Looking at the K^+ density in the external region of the grooves ($10.25 < R \leq 18$ Å) we again see that the highest molarity corresponds to the grooves and not to the phosphate groups (in particular to the anionic oxygens, which lie at $R \approx 12$ Å and in this representation are slightly toward the major groove side of the white lines delimiting the grooves; see Figure 3 of our earlier work (43)). In this outer zone the molarity peaks occur at the center of the narrow minor groove, but offset to the sides of the wider major groove, presumably as a result of the greater separation between the anionic phosphates. This absence of molarity peaks associated with the phosphate groups can also be seen in Supplementary Figure S1 where we have plotted the 0.6 M isomolarity surface for the AGAG oligomer (similar results are found with all the other oligomers studied). These isosurfaces are almost cylindrical and only deformed by a helical indentation on the side of the major groove. This result is contrary to the notion that cations should accumulate around the anionic phosphate groups and is the consequence of the superposition of the electrostatic potentials from both phosphodiester strands of a B-DNA duplex (62,63). The presence of negative potentials in the helical grooves is supported by high-resolution crystallographic data showing monovalent ions bound in the grooves (8,10,64) and by the preferred groove location of small cationic ligands (65). The existence of these localized binding sites, associated with high ion populations also explains why, as we will see below, we can observe high local molarities, despite the fact that we are simulating a salt solution in the millimolar range.

We can now look in more detail at some specific examples of ion binding that will also serve to introduce a simplified representation adapted to comparing all the 39 ABC oligomers we study. Figure 4 shows 3D isomolarity plots for four different oligomers: AGAG, CGCG, AAAA and GGGG. For each oligomer, two molarity isosurfaces are shown, 5 M as a green wire mesh and 15 M as a solid red surface. Further examples of this data are provided in Supplementary Figure S2 that shows the 3 M isosurfaces for the six oligomers containing the ten distinct dinucleotide steps, but also uses two different representations of DNA to make it easier to see the groove location of the molarity peaks.

Returning to Figure 4, we can begin to observe the sequence-dependence of K^+ ion distributions. Two of the oligomers shown, AGAG and CGCG, exhibit strongly localized molarity peaks. For AGAG, the peaks are located in the major groove at the level of the G:C base pairs and close to the electronegative N7 and O6 atoms. In CGCG, the peaks occur in both grooves, but at base pair steps, GC (i.e. 5′-GpC-3′) in the major groove and CG in the minor groove. In both these oligomers the molarity peaks reflect the tetranucleotide repeating sequence with the exception of weaker binding due to the end-effects involving two to three terminal base pairs. The other two oligomers presented are rather different. In the case of GGGG, the molarity peaks

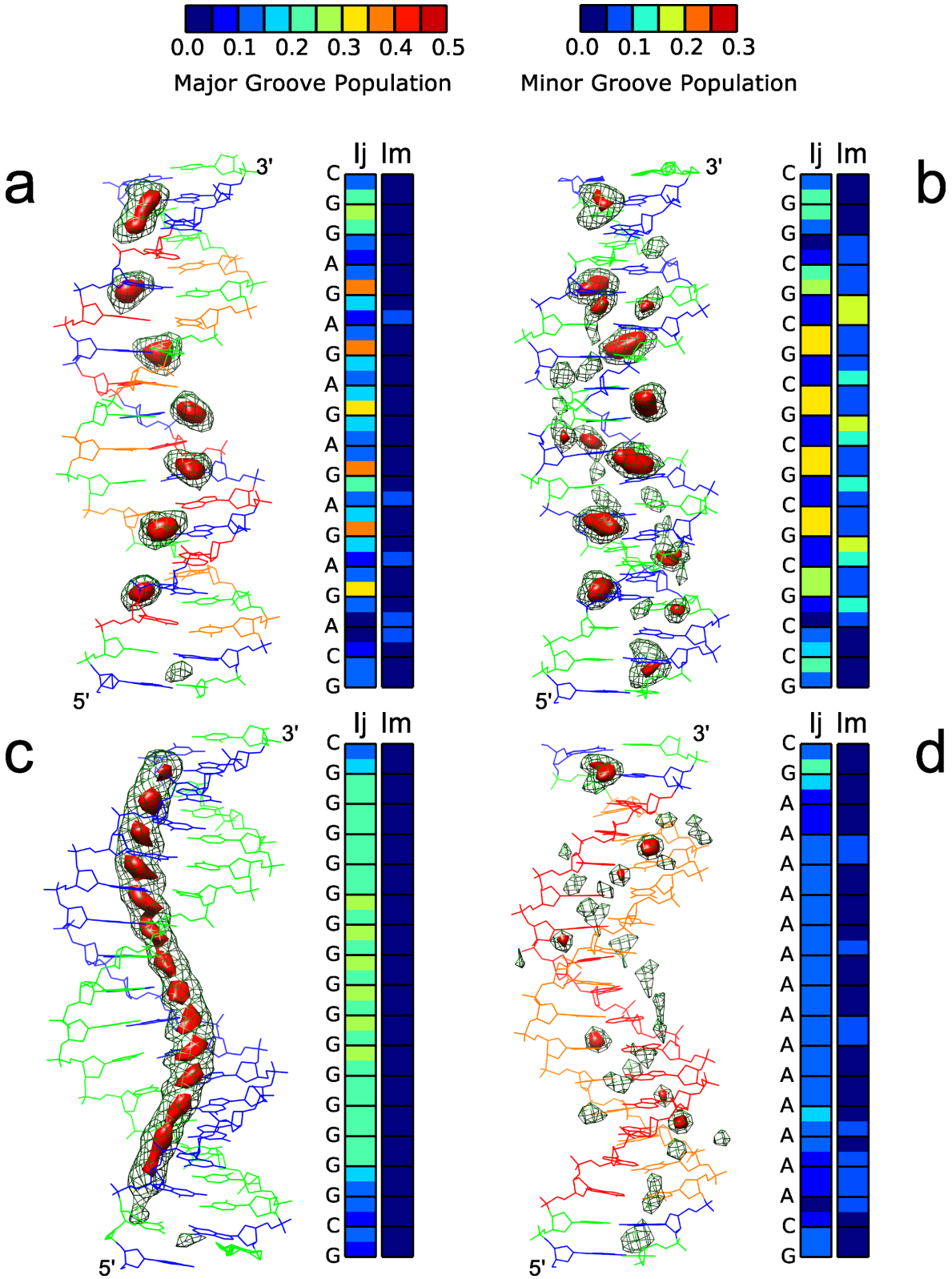

Figure 4. K^+ distributions along the helix. Two representations of the K^+ atmosphere around the AGAG (**a**), CGCG (**b**), GGGG (**c**) and AAAA (**d**) oligomers. (a–d), left: Cartesian K^+ isomolarity surfaces at 15 M (red) and 5 M (green mesh) reconstructed from CHC histograms with respect to the average structure, shown in stick representation and colored according to sequence (A = red, C = green, G = blue, T = orange); the 5′ and 3′ ends of the Watson strand are marked. (a–d), right: K^+ population along the oligomer (5′-3′ direction upward) within the major (left column) and minor (right column) grooves, using the color scheme defined at the top: population increases from blue to red in steps of 0.05. Each base pair step is split into two half steps; see Figure 1 for a precise definition of the grooves.

are concentrated in the major groove, as for AGAG, but now they almost coalesce into a continuous distribution along the groove, even if the peaks still lie at the level of the base pairs and close to the guanine strand. For AAAA, the highest molarity peaks lie at base pair steps in the minor groove, but strikingly these peaks do not occur at every step and thus do not reflect the mononucleotide repeating symmetry of the base sequence. We will return to this point below.

Figure 4 also introduces a simplified representation of the K^+ ion population that is useful for rapidly comparing different base sequences. This representation, shown on the right of each 3D molarity plot, shows the ion population in the major (Ij) and minor (Im) grooves ($R \leq 10.25$ Å) as a color code for each half-step along the oligomer. Note that a separate color scale is used for each groove (with the blue to red showing increasing populations), in order to allow for the accessible volume difference. This compact representation shows several things. First, it shows that K^+ molarity peaks and ion populations are closely related to one another. Second, it shows whether populations (and molarity peaks) are localized in the major or minor grooves and whether they occur mainly at the level of base pairs (indicated by the sequence letters for the 'Watson' strand, with the 5′-3′ direction pointing upward), or at base pair steps (i.e. between two base pairs). Third, it shows whether a regularly repeating base sequence leads to regularly repeating ion distributions (clearly not the case for AAAA) and, fourth, it helps in visualizing end-effects. Using this representation it is possible to compare the ion distributions for all 39 ABC oligomers, using a common color scale for each of the two grooves as shown in Supplementary Figure S3.

We can now begin to discuss sequence-dependence in more detail. Until now, most discussions of ion binding to DNA have been based on the simple chemical notions, namely cations within the grooves of the double helix will accumulate preferentially around electronegative base atoms. If this was a sufficient explanation, sequence-dependence would be limited to the nature of a given base pair or eventually to the nature of a given base pair step. In fact, sequence-dependence is significantly more complex and more distant base pairs can also influence ion distributions. This can be seen if we now concentrate our attention on the central tetranucleotides of each oligomer (thus avoiding end-effects and setting aside for the moment unusual cases such as AAAA where the observed ion distribution does not reflect the sequence symmetry). The results are presented graphically in Figure 5. Once again, K^+ populations are color-coded with the blue to red scale for increasing values. The upper panels of Figure 5 represent the ion populations within the major (Ij) and minor (Im) grooves of the 10 distinct dinucleotide steps and the two distinct base pairs (along the horizontal axis) as a function of the flanking base pairs (along the vertical axis). By 'steps', we imply that the sampling is done between two base pair levels ($i < D \leq i+1$), while by 'levels', we imply that the sampling involves the two half-steps surrounding a given base pair ($i - \frac{1}{2} < D \leq i + \frac{1}{2}$). The lower panels show the same data for the external region of each groove (Xj and Xm), within the space delimited by the outer molarity peaks shown in Figure 2 ($10.25 < R \leq 18$ Å). This data has been extracted from the central tetranucleotides of the 39 ABC oligomers. The numerical data corresponding to Figure 5 is given in Supplementary Figure S4.

Starting with the upper panels of Figure 5 we can firstly see what has already been remarked, namely that the highest K^+ populations occur for GC steps in the major groove (fifth column of the top left-hand panel of Figure 5), with a maximum population of 0.84 within the AGCA tetranucleotide. (Note that for base pair steps with inversion symmetry, there will be two strictly identical tetranucleotides, in this case AGCA and TGCT). The effect of the flanking bases is significant since making a single change from AGCA to TGCA decreases the ion population by more than 30%, down to 0.57. Other GX steps also have relatively high populations, and these are again sensitive to the flanking bases and, as an example, changing GGGA to TGGG decreases the population by almost 50%, from 0.55 to 0.29. Lastly, we note that several steps have very low K^+ populations in the major groove whatever their environment, notably CG (maximum population 0.17), AA (0.26), TG (0.27) and AT (0.29).

We see a similar situation in the minor groove (Im, top right-hand panel of Figure 5). The highest K^+ populations occur for the TG and CG steps, with 0.49 for CTGA and 0.48 for CCGG. Both these populations can again be dramatically reduced, by almost 90%, by changing the flanking bases, to 0.05 for ATGT and to 0.06 for ACGT. Apart from these two steps, only the remaining YR step (i.e. TA) has a significant maximum population (0.34 for CTAG), no other step exceeding 0.17 whatever the sequence environment.

Moving to the external region of the grooves (Xj and Xm, see the lower two panels of Figure 5), the ion populations increase because of the approximately doubled volume, but, as might be expected, the sequence dependence is much weaker. On the major groove side the overall range for the base pair steps is only 0.66 (GGCC) to 0.86 (ATAT and ATGT). On the minor groove side there is slightly more variation from 0.30 (CCGG) to 0.56 (AAAA), but even these ranges remain small with respect to those within the grooves (0.72 for the major groove and 0.48 for the minor groove).

If we finally compare the upper and low panels of Figure 5, we can see that there is a tendency for the internal and external populations on the minor groove side to be anticorrelated, while this is not apparent on the major groove side. This can be explained by the fact that the inner and outer density peaks are much closer together on the minor groove side (see left hand panels of Figure 3) and are thus more directly coupled.

It is also interesting to remark that the sequence-dependent ion populations in the grooves do not necessarily compensate one another for a given base pair or base pair step. If we integrate the ion populations over the full range of A for $R \leq 10.25$ Å, we see total populations ranging from 0.22 (ACGT) to 0.92 (AGCA), while moving out to $R \leq 18$ Å results in a similar range of 1.43 (GAGG) to 2.15 (AGCA). Thus, if we consider a charged wire model of DNA, these results suggest that the total charge per base pair step can vary by as much as 0.7 as a function of base sequence.

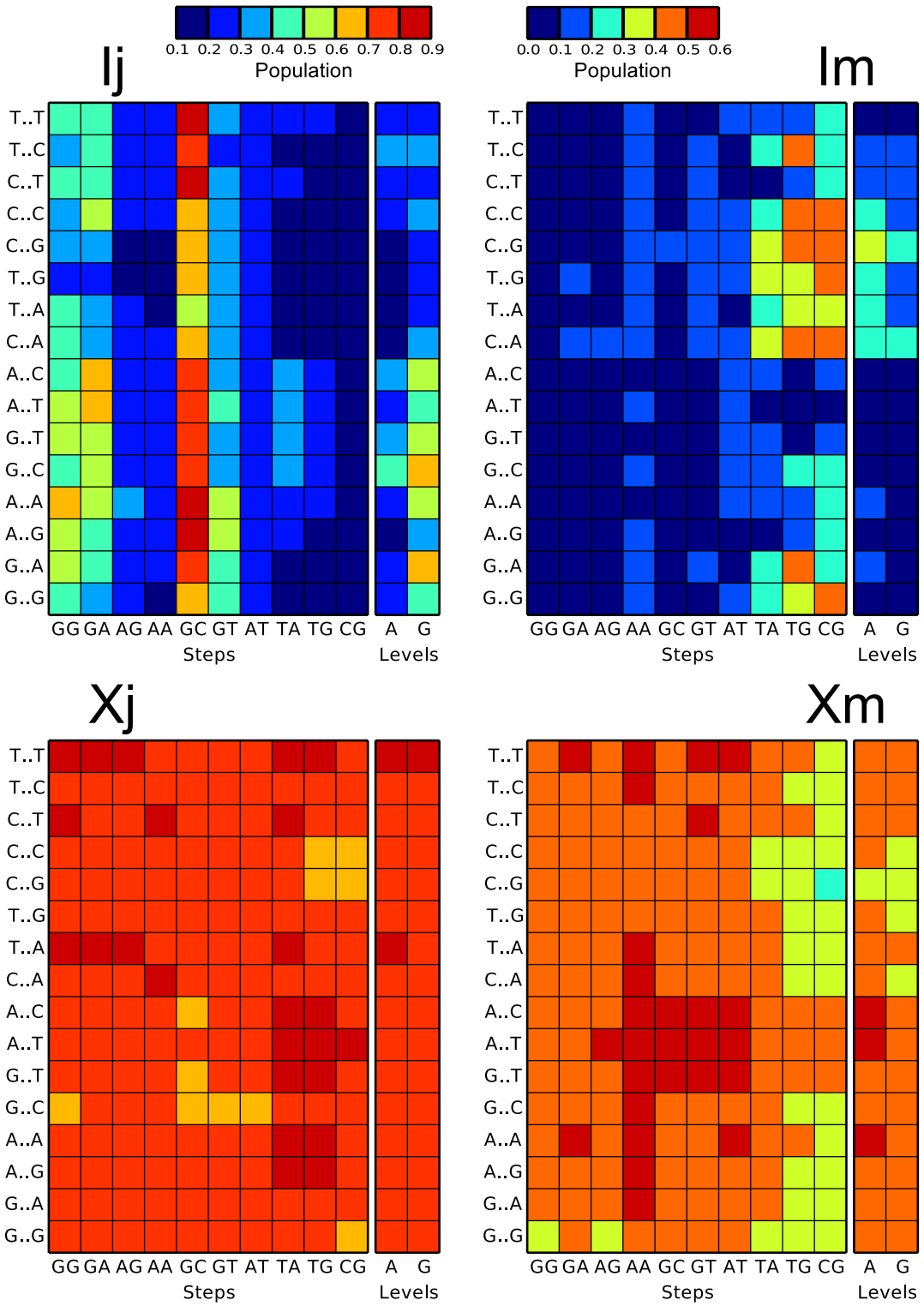

Figure 5. Average K^+ populations in the grooves. Average K^+ population in the internal ('I', top) and external ('X', bottom) region of the major ('j', left) and minor ('m', right) groove. The populations are measured at the central steps of all distinct tetranucleotide sequences (first 10 columns of each panel) and at the central levels of all distinct trinucleotide sequences (last two columns), and colored according to the scheme defined at the top: population increases from blue to red in steps of 0.1. See also Figure 1 for a precise definition of step and level, of the grooves, and of the internal and external regions. Tetranucleotide sequences are arranged so that each column represents one of 10 distinct dinucleotide steps, and each row corresponds to one of the 16 possible flanking sequences. Correspondingly, trinucleotide sequences are arranged so that each column represents one of the two distinct base pairs and each row corresponds to one of the 16 possible flanking sequences. Note that for base pair steps GC, AT, TA and CG, six population values appear twice due to symmetry (e.g. CTAC is identical to GTAG).

We finally return to the interesting case of K^+ distributions in the minor groove (Im) of the AAAA oligomer. The data in Figure 4 showed that the molarity peaks along the minor groove did not repeat with the mononucleotide symmetry of the oligomer base sequence (excluding the GC ends). This becomes even clearer in the upper left-hand panel of Figure 6 where we plot the average minor groove molarity along the oligomer (*D* from 3 to 16). Instead of reflecting the sequence, the molarity peaks show a pattern that repeats over 4-bp steps (between base pairs $5 \rightarrow 9$ and $11 \rightarrow 15$), separated by two steps with very low ion density (between base pairs 9 and 11). When we look at the time evolution of the cumulative ion populations at these positions (upper right-hand panel of Figure 6) we see a very different behavior to that of other ions (see Supplementary Figure S5). Instead of rapidly damped oscillations, AAAA shows sawtooth-like curves that only begin to smooth out after 600 ns of simulation. In order to test for full convergence, we extended the simulations on the AAAA oligomer for a further 500 ns. The results in the lower panels of Figure 6 show that the sawtooth oscillations of cumulative ion population continue leading to a new pattern, which again exhibits distinctly different populations at adjacent steps. The explanation for this behavior is not clear for the moment, but may be related to an incompatibility between the location of the ion binding sites and the optimal separation of K^+ ions in an aqueous environment. This will require further analysis. For the present, it is however interesting to compare the AAAA oligomer with the other mononucleotide repeat sequence GGGG. For GGGG, the principal K^+ binding site is in the major groove so the comparison is not direct, however we can note significant differences. Although the ion population is significantly higher for Ij GGGG (0.49 versus < 0.2 for Im AAAA), the average residence time in each binding site is much shorter (4.6 versus 16.4 ps for Im AAAA, see Figure S6). This may be related to the fact that ions in the GGGG major groove move easily along the oligomer without leaving the groove (note that neighboring GGGG sites are separated on average by 3.5 Å, whereas sites along the AAAA minor groove are roughly 2 Å further apart). In contrast, ions in the AAAA minor groove move principally between the internal and external regions of this groove (results of a Markov state analysis, data not shown). Moving in and out of the narrow AAAA minor groove could imply the need for a more substantial change in hydration and explain the longer residence times and the slower convergence. Confirming this possibility requires a coupled ion/water analysis that is now underway.

CONCLUSIONS

Our earlier work on a limited set of B-DNA oligomers and the current more comprehensive summary show that microsecond-scale MD trajectories are necessary to obtain converged monovalent ion distributions and even this timescale may be insufficient in some particular cases. When the simulation data is analyzed using CHC it is possible to overcome problems associated with the overall dynamics of the double helix (bending, twisting and stretching) and to obtain precise data on ion populations and ion molarities for well-defined volumes within the grooves, or surrounding the double helix, without the necessity of referring to the positions of specific DNA atoms. We have also been able to define internal ('I') and external ('X') regions associated with the major ('Ij, Xj') and minor grooves ('Im, Xm') on the basis of the sequence-averaged radial distributions of potassium ions that are also compatible with counterion condensation theory.

The results of the analysis carried out here lead to a number of interesting conclusions.

(i) The highest K^+ molarity peaks occur within the grooves, close to electronegative base sites, rather than close to the anionic phosphate groups. Depending on the base sequence, the molarity peaks within the grooves may be centered at base pair levels or at the steps between successive base pairs.
(ii) K^+ populations can be surprisingly high compared to previous work, particularly within the major groove of GC steps where they can exceed 0.8. The maximum minor groove populations approach 0.5 and are associated with YG steps.
(iii) Within both grooves the ion populations at specific base pairs or base pair steps can be strongly influenced by the flanking base sequence. As an example, the GC step major groove population ranges from 0.84 when contained within a TGCT segment, to 0.57 within TGCA. Similarly, moving from CCGG to ACGT changes the CG minor groove population by almost 90%, from 0.48 to 0.06.
(iv) K^+ populations in the external region (Xm and Xj) of the helical grooves are higher than in the internal regions (Im and Ij), due to the increased volume sampled (the volume is almost exactly doubled), however the populations in the external regions are much less sequence dependent.
(v) The internal and external radial distribution peaks are much closer on the minor groove side of B-DNA than on the major groove side, which leads to more significant coupling of the ion populations on the minor groove side. This is particularly striking in the case of YR steps.
(vi) The oligo-A tracts are associated with minor groove ion distributions that converge very slowly, do not reflect the sequence symmetry on the microsecond timescale and are associated with long residence times.

The results we have obtained in this study constitute a comprehensive state of the art view of sequence-dependent potassium ion distributions around B-DNA. Work is already underway to extend this analysis to other monovalent ions, notably sodium, to water distributions and to the interplay between water and ions. We are also looking at how far molecular electrostatic potentials can be used to explain ion distributions and what role ion distributions may play in the interaction of DNA with ligands or proteins. Lastly, the present results also constitute a reference for testing the impact of changes in force field parameters or formulation and notably the role of polarization.

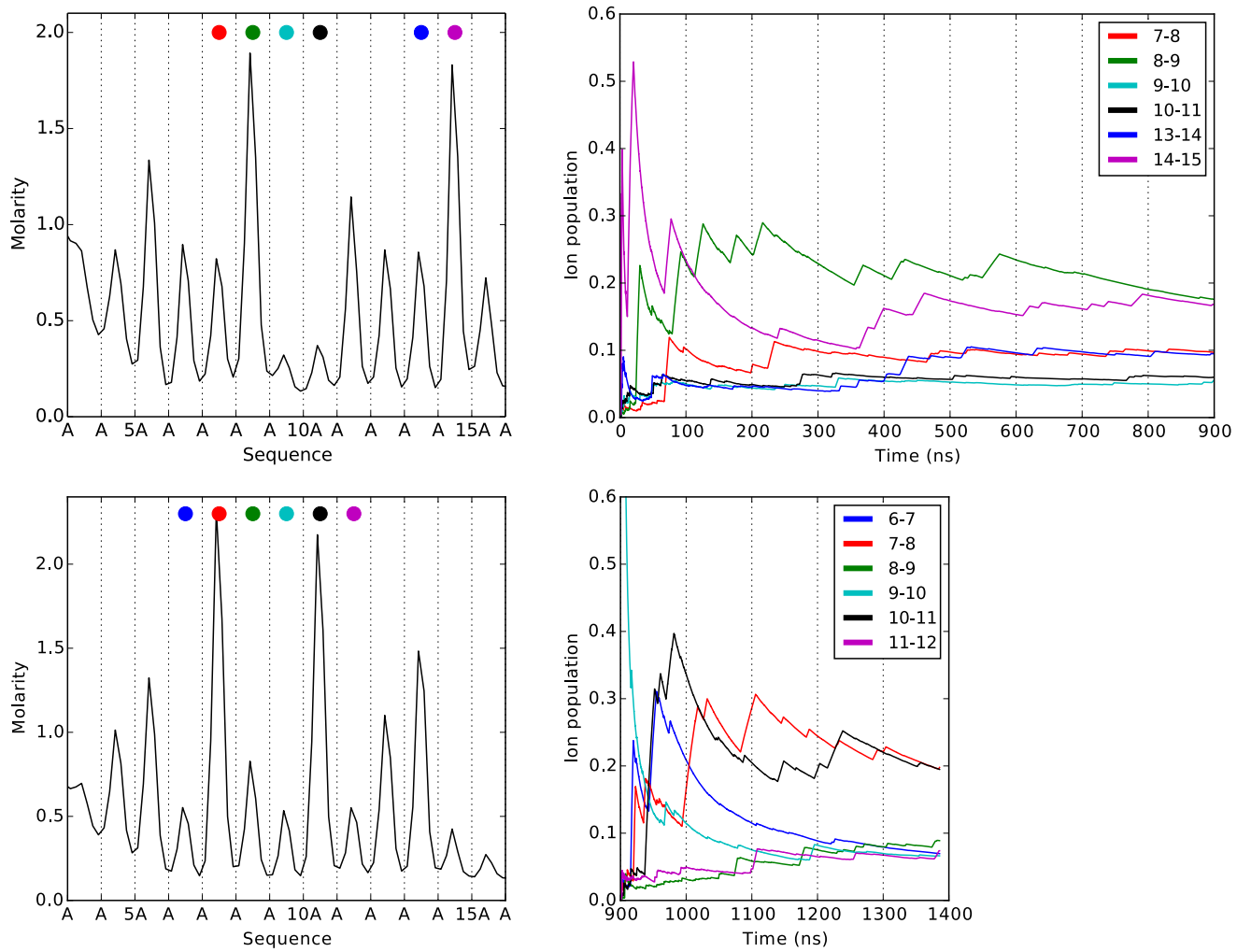

Figure 6. Ion distributions in the AAAA minor groove. (Left panels) Average K^+ molarity in the internal minor groove region (Im) along the AAAA oligomer, excluding two terminal base pairs at either end ($3 \leq D \leq 16$). (Right panels) Time evolution of the average ion population at selected steps along the AAAA oligomer. The selected steps are highlighted in the panels on the left with correspondingly colored dots. The top two panels refer to the 900 ns simulation discussed throughout the article, while the bottom two panels represent a further 500 ns extension of the same simulation, analyzed separately.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

The authors thank Prof Gerald Manning for helpful discussions.

FUNDING

CNRS (to R.L. and M.P.); ANR Project CHROME [ANR-12-BSV5-0017-01 to R.L. and M.P.]; Swiss National Science Foundation [200020-143613 to J.H.M. and M.P.]. Funding for open access charge: ANR Research Project CHROME [ANR-12-BSV5-0017-01].
Conflict of interest statement. None declared.

REFERENCES

1. Saenger,W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, NY.
2. McFail-Isom,L., Sines,C.C. and Williams,L.D. (1999) DNA structure: cations in charge? *Curr. Opin. Struct. Biol.*, **9**, 298–304.
3. Westhof,E. (1988) Water: an integral part of nucleic acid structure. *Annu. Rev. Biophys. Biophys. Chem.*, **17**, 125–144.
4. Dahirel,V., Paillusson,F., Jardat,M., Barbi,M. and Victor,J.M. (2009) Nonspecific DNA-protein interaction: why proteins can diffuse along DNA. *Phys. Rev. Lett.*, **102**, 228101.
5. Heddi,B., Foloppe,N., Hantz,E. and Hartmann,B. (2007) The DNA structure responds differently to physiological concentrations of K(+) or Na(+). *J. Mol. Biol.*, **368**, 1403–1411.
6. Manning,G.S. (1978) The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.*, **11**, 179–246.
7. Anderson,C.F. and Record,M.T. Jr (1990) Ion distributions around DNA and other cylindrical polyions: theoretical descriptions and physical implications. *Annu. Rev. Biophys. Biophys. Chem.*, **19**, 423–463.
8. Hud,N.V. and Engelhart,A.E. (2009) Sequence-specific DNA-metal ion interactions. In: *Nucleic Acid-Metal Ion Interactions*. Royal Society of Chemistry, London, Vol. **14**, pp. 75–117.

9. Howerton,S.B., Sines,C.C., VanDerveer,D. and Williams,L.D. (2001) Locating monovalent cations in the grooves of B-DNA. *Biochemistry*, **40**, 10023–10031.
10. Tereshko,V., Wilds,C.J., Minasov,G., Prakash,T.P., Maier,M.A., Howard,A., Wawrzak,Z., Manoharan,M. and Egli,M. (2001) Detection of alkali metal ions in DNA crystals using state-of-the-art X-ray diffraction experiments. *Nucleic Acids Res.*, **29**, 1208–1215.
11. Hamelberg,D., McFail-Isom,L., Williams,L.D. and Wilson,W.D. (2000) Flexible structure of DNA: ion dependence of minor-groove structure and dynamics. *J. Am. Chem. Soc.*, **122**, 10513–10520.
12. Woods,K.K., McFail-Isom,L., Sines,C.C., Howerton,S.B., Stephens,R.K. and Williams,L.D. (2000) Monovalent cations sequester within the A-tract minor groove of $[d(CGCGAATTCGCG)]_2$[†,§]. *J. Am. Chem. Soc.*, **122**, 1546–1547.
13. Tereshko,V., Minasov,G. and Egli,M. (1999) The Dickerson-Drew B-DNA dodecamer revisited at atomic resolution. *J. Am. Chem. Soc.*, **121**, 470–471.
14. Chiu,T.K., Kaczor-Grzeskowiak,M. and Dickerson,R.E. (1999) Absence of minor groove monovalent cations in the crosslinked dodecamer CGCGAATTCGCG. *J. Mol. Biol.*, **292**, 589–608.
15. Hud,N.V., Sklenár,V. and Feigon,J. (1999) Localization of ammonium ions in the minor groove of DNA duplexes in solution and the origin of DNA A-tract bending. *J. Mol. Biol.*, **286**, 651–660.
16. Pollack,L. (2011) SAXS studies of ion-nucleic acid interactions. *Annu. Rev. Biophys.*, **40**, 225–242.
17. Lipfert,J., Doniach,S., Das,R. and Herschlag,D. (2014) Understanding nucleic acid-ion interactions. *Annu. Rev. Biochem.*, **83**, 813–841.
18. Denisov,V.P. and Halle,B. (2000) Sequence-specific binding of counterions to B-DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 629–633.
19. Cesare Marincola,F., Denisov,V.P. and Halle,B. (2004) Competitive Na(+) and Rb(+) binding in the minor groove of DNA. *J. Am. Chem. Soc.*, **126**, 6739–6750.
20. Young,M.A., Jayaram,B. and Beveridge,D.L. (1997) Intrusion of counterions into the spine of hydration in the minor groove of B-DNA: fractional occupancy of electronegative pockets. *J. Am. Chem. Soc.*, **119**, 59–69.
21. Auffinger,P. and Hashem,Y. (2007) Nucleic acid solvation: from outside to insight. *Curr. Opin. Struct. Biol.*, **17**, 325–333.
22. Mocci,F. and Laaksonen,A. (2012) Insight into nucleic acid counterion interactions from inside molecular dynamics simulations is "worth its salt". *Soft Matter*, **8**, 9268–9284.
23. Patel,D.S., He,X. and MacKerell,A.D. (2014) Polarizable empirical force field for hexopyranose monosaccharides based on the classical drude oscillator. *J. Phys. Chem. B.*, **119**, 637–652.
24. Várnai,P. and Zakrzewska,K. (2004) DNA and its counterions: a molecular dynamics study. *Nucleic Acids Re.*, **32**, 4269–4280.
25. Rueda,M., Cubero,E., Laughton,C.A. and Orozco,M. (2004) Exploring the counterion atmosphere around DNA: what can be learned from molecular dynamics simulations? *Biophys. J.*, **87**, 800–811.
26. Auffinger,P. and Westhof,E. (2001) Water and ion binding around r(UpA)12 and d(TpA)12 oligomers–comparison with RNA and DNA (CpG)12 duplexes. *J. Mol. Biol.*, **305**, 1057–1072.
27. Auffinger,P. and Westhof,E. (2000) Water and ion binding around RNA and DNA (C,G) oligomers. *J. Mol. Biol.*, **300**, 1113–1131.
28. Dixit,S.B., Mezei,M. and Beveridge,D.L. (2012) Studies of base pair sequence effects on DNA solvation based on all-atom molecular dynamics simulations. *J. Biosci.*, **37**, 399–421.
29. Feig,M. and Pettitt,B.M. (1999) Sodium and chlorine ions as part of the DNA solvation shell. *Biophys. J.*, **77**, 1769–1781.
30. McConnell,K.J. and Beveridge,D.L. (2000) DNA structure: what's in charge? *J. Mol. Biol.*, **304**, 803–820.
31. Ponomarev,S.Y., Thayer,K.M. and Beveridge,D.L. (2004) Ion motions in molecular dynamics simulations on DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 14771–14775.
32. Pérez,A., Luque,F.J. and Orozco,M. (2007) Dynamics of B-DNA on the microsecond time scale. *J. Am. Chem. Soc.*, **129**, 14739–14745.
33. Dans,P.D., Faustino,I., Battistini,F., Zakrzewska,K., Lavery,R. and Orozco,M. (2014) Unraveling the sequence-dependent polymorphic behavior of d(CpG) steps in B-DNA. *Nucleic Acids Res.*, **42**, 11304–11320.
34. Dans,P.D., Perez,A., Faustino,I., Lavery,R. and Orozco,M. (2012) Exploring polymorphisms in B-DNA helical conformations. *Nucleic Acids Res.*, **40**, 10668–10678.
35. Min,D., Li,H., Li,G., Berg,B.A., Fenley,M.O. and Yang,W. (2008) Efficient sampling of ion motions in molecular dynamics simulations on DNA: variant Hamiltonian replica exchange method. *Chem. Phys. Lett.*, **454**, 391–395.
36. Kirmizialtin,S., Silalahi,A.R., Elber,R. and Fenley,M.O. (2012) The ionic atmosphere around A-RNA: Poisson-Boltzmann and molecular dynamics simulations. *Biophys. J.*, **102**, 829–838.
37. Giambaşu,G.M., Luchko,T., Herschlag,D., York,D.M. and Case,D.A. (2014) Ion counting from explicit-solvent simulations and 3D-RISM. *Biophys. J.*, **106**, 883–894.
38. Howard,J.J., Lynch,G.C. and Pettitt,B.M. (2011) Ion and solvent density distributions around canonical B-DNA from integral equations. *J. Phys. Chem. B*, **115**, 547–556.
39. Pérez,A., Marchán,I., Svozil,D., Sponer,J., Cheatham,T.E., Laughton,C.A. and Orozco,M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys. J.*, **92**, 3817–3829.
40. Dang,L.X. (1995) Mechanism and thermodynamics of ion selectivity in aqueous-solutions of 18-crown-6 ether—a molecular dynamics study. *J. Am. Chem. Soc.*, **117**, 6954–6960.
41. Pérez,A., Luque,F.J. and Orozco,M. (2012) Frontiers in molecular dynamics simulations of DNA. *Acc. Chem. Res.*, **45**, 196–205.
42. Pasi,M., Maddocks,J.H., Beveridge,D., Bishop,T.C., Case,D.A., Cheatham,T., Dans,P.D., Jayaram,B., Lankas,F. *et al.* (2014) μABC: a systematic microsecond molecular dynamics study of tetranucleotide sequence effects in B-DNA. *Nucleic Acids Res.*, **42**, 12272–12283.
43. Lavery,R., Maddocks,J.H., Pasi,M. and Zakrzewska,K. (2014) Analyzing ion distributions around DNA. *Nucleic Acids Res.*, **42**, 8138–8149.
44. Pearlman,D.A., Case,D.A., Caldwell,J.W., Ross,W.S., Cheatham,T.E., DeBolt,S., Ferguson,D., Seibel,G. and Kollman,P. (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.*, **91**, 1–41.
45. Case,D.A., Cheatham,T.E., Darden,T., Gohlke,H., Luo,R., Merz,K.M., Onufriev,A., Simmerling,C., Wang,B. and Woods,R.J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.*, **26**, 1668–1688.
46. Cheatham,T.E. 3rd, Cieplak,P. and Kollman,P.A. (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.*, **16**, 845–862.
47. Berendsen,H.J.C., Grigera,J.R. and Straatsma,T.P. (1987) The missing term in effective pair potentials. *J. Phys. Chem.*, **91**, 6269–6271.
48. Essmann,U., Perera,L., Berkowitz,M.L., Darden,T., Lee,H. and Pedersen,L.G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.*, **103**, 8577–8593.
49. Lavery,R., Zakrzewska,K., Beveridge,D., Bishop,T.C., Case,D.A., Cheatham,T., Dixit,S., Jayaram,B., Lankas,F., Laughton,C. *et al.* (2010) A systematic molecular dynamics study of nearest-neighbor effects on base pair and base pair step conformations and fluctuations in B-DNA. *Nucleic Acids Res.*, **38**, 299–313.
50. Dixit,S.B., Beveridge,D.L., Case,D.A., Cheatham,T.E. 3rd, Giudice,E., Lankas,F., Lavery,R., Maddocks,J.H. and Osman,R. (2005) Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II: sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophys. J.*, **89**, 3721–3740.
51. Beveridge,D.L., Barreiro,G., Byun,K.S., Case,D.A., Cheatham,T.E. 3rd, Dixit,S.B., Giudice,E., Lankas,F. and Lavery,R. (2004) Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys. J.*, **87**, 3799–3813.
52. Berendsen,H.J.C., Postma,J.P.M., van Gunsteren,W.F., DiNola,A. and Haak,J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.
53. Ryckaert,J.P., Ciccotti,G. and Berendsen,H.J.C. (1977) Numerical-integration of cartesian equations of motion of a system with constraints—molecular-dynamics of N-alkanes. *J. Comput. Phys.*, **23**, 327–341.

54. Harvey,S.C., Tan,R.K.Z. and Cheatham,T.E. III (1998) The flying ice cube: velocity rescaling in molecular dynamics leads to violation of energy equipartition. *J. Comput. Chem.*, **19**, 726–740.
55. Blanchet,C., Pasi,M., Zakrzewska,K. and Lavery,R. (2011) CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res.*, **39**, W68–W73.
56. Lavery,R., Moakher,M., Maddocks,J.H., Petkeviciute,D. and Zakrzewska,K. (2009) Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.*, **37**, 5917–5929.
57. Manning,G.S. (1969) Limiting laws and counterion condensation in polyelectrolyte solutions I. Colligative properties. *J. Chem. Phys.*, **51**, 924–933.
58. Oliphant,T.E. (2007) Python for scientific computing. *Comput. Sci. Eng.*, **9**, 10–20.
59. Hunter,J.D. (2007) Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.*, **9**, 90–95.
60. Goddard,T.D., Huang,C.C. and Ferrin,T.E. (2007) Visualizing density maps with UCSF Chimera. *J. Struct. Biol.*, **157**, 281–287.
61. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
62. Lavery,R. and Pullman,B. (1981) Molecular electrostatic potential on the surface envelopes of macromolecules: B-DNA. *Int. J. Quant. Chem.*, **20**, 259–272.
63. Jayaram,B., Sharp,K.A. and Honig,B. (1989) The electrostatic potential of B-DNA. *Biopolymers*, **28**, 975–993.
64. Maehigashi,T., Hsiao,C., Woods,K.K., Moulaei,T., Hud,N.V. and Williams,L.D. (2012) B-DNA structure is intrinsically polymorphic: even at the level of base pair positions. *Nucleic Acids Res.*, **40**, 3714–3722.
65. Kopka,M.L., Yoon,C., Goodsell,D., Pjura,P. and Dickerson,R.E. (1985) The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 1376–1380.