## BASE SEQUENCE SPECIFICITY OF COUNTERION BINDING TO DNA: WHAT CAN MD SIMULATIONS TELL US?

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Canadian Journal of Chemistry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>cjc-2016-0296.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>31-Aug-2016</td>
</tr>
</tbody>
</table>
| Complete List of Authors: | Atzori, Alessio; Università degli Studi di Cagliari, Dipartimento di Scienze Chimiche e Geologiche  
Liggi, Sonia; Università degli Studi di Cagliari, Dipartimento di Scienze Chimiche e Geologiche  
Laaksonen, Aatto; Universita degli Studi di Cagliari, Dipartimento di Scienze Chimiche e Geologiche; Stockholm University, Division of Physical Chemistry, Arrhenius Laboratory; Science for Life Laboratory  
Porcu, Massimiliano; Università degli Studi di Cagliari, Dipartimento di Scienze Chimiche e Geologiche; Stockholm University, Division of Physical Chemistry, Arrhenius Laboratory  
Lyubartsev, Alexander; Stockholm University, Division of Physical Chemistry, Arrhenius Laboratory  
Saba, Giuseppe ; Università degli Studi di Cagliari, Dipartimento di Scienze Chimiche e Geologiche  
Mocci, Francesca; Università degli Studi di Cagliari, Dipartimento di Scienze Chimiche e Geologiche; Stockholm University, Division of Physical Chemistry, Arrhenius Laboratory |
| Keyword: | DNA, MD simulations, Solvation, Ion Interactions, sequence specificity |
BASE SEQUENCE SPECIFICITY OF COUNTERION BINDING TO DNA: WHAT CAN MD SIMULATIONS TELL US?

Alessio Atzori,1 Sonia Liggi,1 Aatto Laaksonen,1,2,3 Massimiliano Porcu,1,2
Alexander P. Lyubartsev,2 Giuseppe Saba,1 and Francesca Mocci1,2,3

1 Dipartimento di Scienze Chimiche e Geologiche, Università di Cagliari, Cittadella Universitaria di Monserrato, I-09042 Monserrato (CA), Italy
2 Division of Physical Chemistry, Arrhenius Laboratory, Stockholm University, S-10691 Stockholm, Sweden
3 Science for Life Laboratory, 171 21 Solna, Sweden

Corresponding author: Francesca Mocci, fmocci@unica.it, phone: 00393280110799, fax: 00390706754388
Abstract: Nucleic acids are highly charged biopolymers whose secondary structure is strongly dependent on electrostatic interactions. Solvent molecules and ions are also believed to play an important role in mediating and directing both sequence recognition and interactions with other molecules, such as proteins and a variety of ligands. Therefore, to fully understand the biological functions of DNA, it is necessary to understand the interactions with the surrounding counterions. It is well known that monovalent counterions can bind to the minor groove of DNA with consecutive sequences of four, or more, Adenine and Thymine (A-tracts) with relatively long residence times. However, much less is known about their binding to the backbone and to the major groove. In this work, we used molecular dynamics simulations to both investigate the interactions between the backbone and major groove of DNA and one of its physiological counterions (Na\(^{+}\)), and to evaluate the relationship between these interactions and the nucleotide sequence. Three dodecamers, namely CGAAAAATTTTCG, CGCTCTAGACCG, and CGCGAATTCGCG, were simulated using the Toukan-Rahman flexible SPC water model and Smith and Dang parameters for Na\(^{+}\), revealing a significant sequence-dependence on the ion binding to both backbone and major groove. In the absence of experimental data on the atomistic details of the studied interactions, the reliability of the results was evaluated performing the simulations with additional sets of potential parameters for ions and solvent, namely the Åqvist or the Joung and Cheatham ion parameters and the TIP3P water model. This allowed us to evaluate the results by verifying which features are preserved independently from the parameters adopted.

Keywords: DNA-counterion interactions, DNA sequence-specificity, Ion parameters, molecular dynamics
Introduction

The highly negatively charged nature of DNA affects its structure, flexibility and functions, causing strong intra- and interstrand repulsions, which are offset by the presence of solvent and ions surrounding DNA in physiological conditions.\(^1\) Consequently, the properties of nucleic acids cannot be fully understood without considering their interactions with the solvent molecules and counterions.\(^{2-5}\) Ever since the first analysis of a small DNA fragment using atomic resolution X-ray performed in 1973,\(^6\) the structural and functional roles of counterions were greatly stressed. In the following decade, it has been proposed that the selectivity of alkaline ions for specific portions of DNA, rich in adenine and thymine, could explain the stabilization of the B-DNA conformation.\(^7\) Nonetheless, until mid-1990s, the important role of the interactions with monovalent counterions on the sequence-dependent properties have been neglected in most investigations focusing on the DNA structure. Indeed, these interactions were commonly believed to be "delocalized" and independent from the base sequence.\(^8-9\) The lack of studies on the sequence-specific effects of these interactions can be partially explained by difficulties connected with their observations. More specifically, X-ray diffraction studies of nucleic acids cannot distinguish between the isoelectronic Na\(^+\) ions and water molecules, as well as due to their irregular coordination structure around DNA.\(^{10}\) Similarly, high resolution solution NMR techniques are not able to provide atomistic details of the interactions of DNA with alkaline ions. This is due to both fast ion exchange in NMR time scale and to the lack of modification of the detected DNA signals by such ions.\(^{10}\)

Given these experimental limitations, the development of both computational resources and molecular modeling techniques have boosted the theoretical studies of nucleic acids and their binding with physiological ions.\(^{11}\) A vital interest to explain the role of monovalent counterions on DNA sequence specific features\(^{11-13}\) originated from the molecular dynamics (MD) investigations performed by Young et al.\(^{14-15}\) In particular, their investigation showed
how counterions interact preferentially with the minor groove of consecutive sequences of AT base pairs without any 5'-TpA-3' steps (defined as A-tracts). Following these results, several experimental studies were performed\textsuperscript{15-18} provoking a significant debate especially among crystallographers.\textsuperscript{12-13, 19-21} Most of the recent MD studies on sequence specificity\textsuperscript{12, 22-27} have been focusing on the interactions of counterions with the minor groove, while those with the backbone and major groove regions have received less consideration. In order to bridge this gap, the relationships between the DNA sequence and the interactions of ions with the major groove and the backbone have been evaluated in this study. To this purpose, three different sequences were simulated (See Figure 1):

- CGAAAATTTCG (from now on called A4T4), containing an 8 bases long A-tract;
- CGCGAATTCGCG (known as Drew Dickerson Dodecamer and hence called DDD\textsuperscript{28}), with a 4 bases long A-tract;
- CGCTCTAGAGCG (from now on called TA), having the same nucleotide composition of DDD but different order of the bases, and hence lacking an A-tract.

To partly compensate for the lack of experimental data on the atomistic details of the studied interactions, the results of the simulations were cross-checked by performing the simulations using three combinations of water model (TIP3P\textsuperscript{29} and flexible SPC\textsuperscript{30}) and ion parameters (Smith and Dang,\textsuperscript{31} Joung and Cheatham,\textsuperscript{32} and Äqvist\textsuperscript{33}), as detailed in the computational methods section. The used combinations of potential parameters were chosen due to their wide use in the previous investigations aimed to study the interactions of B-DNA with counterions (see the recent review by Mocci and Laaksonen, Table 1\textsuperscript{11}), and to compare them with a recently developed potential model\textsuperscript{32} to verify its performance on this type of investigation.
Therefore, while obtaining a representative and comprehensive picture of the ion/DNA interactions using MD, we also tried to establish a well-defined protocol using different parameters to study interactions between counterions and specific DNA sequences, as we did recently in a study on DNA quadruplexes.\textsuperscript{34}

**Computational methods**

The initial structures of the three duplexes were built using the Nucgen module of the AMBER package.\textsuperscript{35} The experimental data for TA and DDD suggests that they are in the B form.\textsuperscript{28} Even if no crystallographic structures are available for A4T4, it was generated in a canonical B-DNA form, similarly to the other two oligomers. Each duplex was immersed in a truncated octahedral box containing 4500 water molecules and 22 Na\(^+\) counterions. Three different simulations were performed for each sequence using the *parmbsc0* force field\textsuperscript{36} for the DNA and three different combinations of solvent models and ions parameters:

**Set 1**: a refined model of water, the Toukan-Rahman flexible SPC model,\textsuperscript{30} with the Smith and Dang parameters for Na\(^+\);\textsuperscript{31}

**Set 2**: Na\(^+\) parameters recently developed by Joung and Cheatham\textsuperscript{32} along with TIP3P water model;\textsuperscript{29}

**Set 3**: the most widely used combination used in studies of B-DNA interaction with counterions: Åqvist parameters\textsuperscript{33} for Na\(^+\) together with the TIP3P water model.\textsuperscript{29}

Numerical values of the used parameters are reported in Table 1.

**Preparation of the system for set 1**

Twenty-two Na\(^+\) counterions were placed using the utility Xleap of the AMBER14 package\textsuperscript{35} at the points of lowest Coulombic potential, which were calculated using a distance-
dependent dielectric constant on a grid of 1 Å resolution. 4500 water molecules were added to fill a truncated octahedral box of side length ~65 Å. The system was equilibrated with an initial steepest descent minimization, followed by a 3 ps dynamics with positional constraints on the Na\(^+\) and DNA atoms; the constraints were then gradually removed by several minimization runs, those on Na\(^+\) being relaxed faster. The system was then heated to 300 K over 2 ps and followed by 400 ns of dynamics using Mdynamix\(^37\) in NPT ensemble at 300 K and 1 bar (using Nose-Hoover algorithm with time constants of 30.0 fs for thermal fluctuations and 700.0 fs for volume fluctuations). The double time step algorithm of Tuckerman was used, with a time step of 0.2 fs for short (cut off=5 Å) intra and inter-molecular interactions, and a time step of 2 fs for the others (cut off=14 Å). Electrostatic interactions were calculated using the particle mesh Ewald method.\(^38\) The cutoff radius in the real space (Rcut) was set to 14 Å, while the cutoff radius in the reciprocal space, k\(_{\text{max}}\), was determined from the condition \(\exp[-\alpha^2 k_{\text{max}}^2/ R_{\text{cut}}^2] = \exp[-8.1]\) \(\alpha\) was taken as \(\alpha R_{\text{cut}}=2.7\).

**System preparation for set 2 and 3**

System setup and equilibration are the same as the ones described for set 1. Four-hundred ns of dynamics were carried out using the PMEMD module of AMBER 14\(^35\) with a time step of 2 fs and SHAKE constraints on hydrogens (tolerance = 0.00005). The Berendsen algorithm\(^39\) with a coupling constant of 3 ps was utilized to maintain constant temperature (300 K) and pressure (1 bar). In order to reduce the computational cost of the performed calculations, we evaluated the effect of the reduction of the cutoff radius for non-bonded interactions from 14 Å (as in set 1) to 9 Å. This reduction did not lead to any significant variation in the interactions trend of DNA with Na\(^+\) counterions (Supplementary Information Figure 1). Electrostatic interactions were calculated by the particle mesh Ewald method as implemented in AMBER 14,\(^35\) with a cubic B-spline interpolation order and 0.00001 tolerance for the...
direct space sum cutoff. Periodical removal of the net center-of-mass momentum was applied in order to eliminate the “flying ice cube” phenomenon. Volume, pressure, density, and temperature all reached their stationary values within 15 ps.

Analysis of the results

Normalized radial distribution functions (RDFs) between the Na\(^+\) ions and selected reference atoms of DNA in the backbone and in the major groove (highlighted in Figure 2) and residence times of Na\(^+\) to the oxygens of the phosphate groups of each sequence were calculated using an in-house version of the Tranal module of Mdynamix. In the calculation of the RDFs the two terminal base pairs of each duplex were neglected, considering only the eight central base pairs. The residence time of direct binding events of counterions with DNA were calculated for each filament in the 5’→ 3’ direction. The binding events were estimated as a weighted average of the contact residence times being equal or longer than 3 ps. Contacts are considered to be interrupted if the ion moves away from the reference electronegative atom for more than 4 ps and 3.2 Å. VMD was used to visualize trajectories and to generate graphical representations of each system.

Results and discussion

During the simulations different types of interactions between counterions and DNA can occur, such as direct or solvent mediated contacts. In order to quantify and classify these interactions, we performed a set of analyses providing complementary information. In the following paragraphs, we first discuss the results obtained employing the set 1 parameters for the three oligomers analyzing the sequence dependence of the interactions in the backbone and in the major groove using RDFs. Thereafter, we compare the results obtained with set 1 with those from set 2 and 3 parameters using both RDFs and weighted average residence times.
Are Na\(^+\) interactions with the backbone and major groove of DNA sequence-dependent?

The two negatively charged oxygen atoms of the phosphate groups of DNA, namely O1P and O2P, possess an opposite spatial orientation: O1P faces the minor groove, while O2P is directed toward the major groove (Figure 3a). Consequently, their interactions with the counterions are expected to be different.

The calculated RDFs show that, independently from the sequence, the direct contact between Na\(^+\) and the phosphate oxygen atoms (peak at \(~2.5\) Å) occurs preferentially with O1P (Figure 3b, black line), confirming what has been observed in earlier investigations.\(^{26, 42}\) A sequence specific feature can be observed by comparing the intensity of the first RDF peak among the three sequences: the probability of binding to O1P increases with the length of the A-tract. Indeed, the first peak intensity is highest for the A4T4 oligomer, followed by that for DDD and the lowest being that for the TA (See Table S1 in the supplementary information for the numeric values of the maxima for each RDF peak). On the contrary, this trend is not observed for the binding of Na\(^+\) to O2P.

Concerning the major groove, the RDFs of Na\(^+\) with respect to the selected electronegative atoms (Figure 4a) are shown in Figure 4b; in Table S2 of the supplementary information are listed the numeric values of the maxima for each RDF peak. The comparison of the RDFs of Na\(^+\) indicates that the probability of Na\(^+\) to bind electronegative atoms of DNA is dependent on the base pair sequences. For instance, the direct binding to T\(\text{O}4\) is a rare event in the major groove of A-tracks (absence of a green peak at 2.5 Å for A4T4). On the contrary, when a thymine is not embedded in an A-tract sequence, the probability of a direct binding is moderately high (green peaks at 2.5 Å for TA and DDD).

By comparing the intensity of the peaks corresponding to the direct binding to any of the atoms in the major groove of TA and DDD, it can be observed that the intensities are always...
higher for the direct contact with the former. On the overall, in contrast with what has been observed for the interactions of \(\text{Na}^+\) with backbone atoms, the presence of an A-tract is associated with a lower probability of finding an ion directly bound to the major groove.

Noteworthy, the peaks with the highest intensity for each oligomer belong to \(A_{N7}\) independently from the sequence considered. When the guanine base is present in the sequence, the binding to \(G_{N7}\) is less probable than to \(G_{O6}\). Direct contacts between counterions and \(C_{N4}\) and \(A_{N6}\) are negligible: in fact, the presence of partial positive charges on the nitrogen of these atoms causes an electrostatic repulsion of the counterions.

In order to interpret the peaks at distances greater than 2.5 Å, further analyses are required. Generally, the second peak can indicate a water mediated contact (~5 Å, Figure 5a), or an interaction with an electronegative atom located in a complementary or nearby base (from ~4 to ~6 Å Figure 5b). An example of the latter is the direct contact of \(\text{Na}^+\) with \(A_{N7}\), which is reflected also on the peaks of \(A_{N6}\) (yellow peaks around ~4 Å) and \(T_{O4}\) (green peaks around ~6 Å) in Figure 4b. The third peak may originate from a direct contact of the ion in the opposite groove. In particular, the peaks at ~7 Å for \(T_{O4}\) and ~8 Å for \(A_{N7}\) in DDD are caused by a \(\text{Na}^+\) interacting with \(T_{O2}\) located in the minor groove (Figure 5c). Indeed, the RDFs calculated for \(T_{O2}\) (as shown in SI Figure 2) indicate that the intensity of its first peak is correlated to those observed in the third peaks of the major groove.

**Comparison of the results obtained with different ion parameters and solvent models**

In order to evaluate the impact of different parameters for sodium and water on the interactions between DNA and counterions, we compared the results obtained with the three different combinations of parameters described in the computational methods section.
The RDFs represented in Figure 6 indicate how the affinity of Na\textsuperscript{+} for the oxygen atoms of the backbone is highly dependent on the set of parameters employed, with set 3 leading to a remarkable lower affinity than the other two sets. This result is in agreement with what was observed by Noy et al.\textsuperscript{43} when comparing Åqvist with Joung & Cheatham parameters for Na\textsuperscript{+}.

Even if the binding affinity varies with the set, the main sequence specific features of the ion binding are observed with all sets. Namely, all sets show a preferential binding of the counterions to O1P compared to O2P. Further, this preference increases when A-tracts are present, and with the length of the A-tract. To better quantify this trend, in Table S1 of the supplementary information we report the heights of the peaks, and the ratio between the peaks maxima of the O1P and O2P RDFs. It can be seen that the mentioned ratio decreases with the length of the A-tract, with the following trend within all the combination of parameters: A4T4 > DDD > TA.

We verified whether the higher probability of binding to the phosphate of A-tract containing sequences is correlated to a longer residence time in the binding sites. In Figure 7 are plotted the weighted average residence times of the direct binding events occurring along the backbone. For more details on this analysis, please refer to the computational methods section. It is useful to recall that the two filaments of each oligomer in this study have the same base sequence; therefore, the same trend in the residence time can be expected. More precisely, we expect the residence time along the two chains to overlap when full convergence is reached. This is observed for A4T4 and DDD with all sets of parameters. For these sequences, residence times of Na\textsuperscript{+} are higher when the ion binds to O1P atoms rather than O2P atoms, especially in A-tracts. In addition, the residence time increases along the A-tract in the 5'→3' direction. However, this trend is observed only with two of the three sets analyzed: the widely used Åqvist parameters (set 3) does not reveal this sequence specific
feature; in addition, the residence times of binding events are generally shorter (~ 100 ps) than those observed with the other two sets. On the other hand, no clear difference in residence times between O1P and O2P is observed for the TA sequence with any of the sets of parameters.

On the basis of these analyses, we can conclude that according to the simulations performed with set 1 and set 2, both the probability of the direct binding event and its residence time are highly affected by the sequence. This does not apply to the simulations performed with set 3, where the sequence-specificity in residence times is not clearly observed.

Concerning the major groove, the comparison of the RDFs calculated from the simulations with the three different sets of parameters indicates that all the combinations of potential models provide the same qualitative information concerning the sequence specificity of the ion binding (Figure 8). In fact, all of the observations made in the previous section concerning the sequence specificity resulting from the simulations with set 1, are still valid with sets 2 and 3 and will not be further commented here. Independently from the sequence and the set of parameters used, A_{N7} is the most favorable binding site of sodium ion, followed by G_{O6} for DDD and TA.

However, there are some differences that deserve attention: with set 3, the binding to A_{N7} in both TA and DDD is highly favored compared to that to other electronegative atoms in the groove, while, this preference is less pronounced with the other two sets of parameters. Another difference concerns the peak at ~7 Å for T_{O4}, which is clearly observed in set 1, especially for DDD, and is less evident in the simulations with set 2 and 3. This is due to the diverse probability of Na^{+} in binding to T_{O2} in the minor groove. As shown in the supporting information (Figure S2), the intensity of the peak for the direct contact between Na^{+} and T_{O2} of DDD is noticeably higher for set 1 than for the other two sets.
Residence time of direct binding events of Na\(^+\) to the electronegative atoms of the major groove for the two filaments of each duplex are shown in Figure 9. As mentioned previously, we expect an overlap between the two filaments when a full convergence is achieved. This is observed with most of the simulations, with only few irregularities observed. This is somehow unexpected considered that the degree of convergence for the interactions in the major groove is higher than the one observed for the backbone atoms.

As observed previously for the binding to the backbone, set 3 parameters leads to short residence times of binding events (~ 100 ps), while set 1 produces the highest residence times (between 300 and 1200 ps).

The residence times exhibit slight sequence specificity, which is possible to observe within all sets of parameters. In detail, the residence time of the binding to the thymine bases in the center of A-tracts is not only a rare event, as indicated by the RDFs, but is also characterized by a short residence time. Longer residence time for the binding to thymine is observed when the base is located at the end of the A-tract, or when is not part of an A-tract. On the contrary, the residence time of Na\(^+\) binding to adenine is slightly higher in the center of the A-tract.

Differently from what has been observed for the interactions with the oxygen atoms of the phosphate, an end effect is observed in the residence time of the binding of the ion to guanine: in fact, the residence time of terminal guanines are shorter than those observed for the same base in the center of the oligomer. However, it cannot be excluded that this result is also due to the lack of adjacent purine type bases.

As expected from the RDFs, where the first peak for the binding to cytosine is absent or very low indicating an extremely rare event (Figure 8), residence times are noticeably short, or even equal to zero, since no binding event is observed.
Conclusions

MD simulations of three different DNA sequences allowed us to verify that the interactions of DNA with Na⁺ counterions are highly sequence dependent not only in the minor groove, as shown in previous studies, but also in the backbone and in the major groove. In fact, the probability of the ions to interact with a given base and its close surrounding varies noticeably between the A-tract and non A-tract sequences. The three combinations of solvent and counterions parameters agreed mutually on several results:

(i) Na⁺/DNA direct contact is less probable in the major groove of oligomers containing A-tract;

(ii) Independently from the sequence, AN7 is the most favorite site in the major groove for the interactions of Na⁺;

(iii) The binding of the counterions with the outermost atoms of the duplex DNA, i.e. the oxygen atoms of the phosphates, are also affected by the sequence, with the A-tracts having higher probability of direct binding with Na⁺.

However, the details of the sequence-specific features of the interactions should be judged carefully, since they might vary with the models employed for the solvent and counterions. Indeed, the different sets adopted in this study do not give a unified picture of the sequence dependence of the average residence time of the binding events occurring in the backbone. More specifically, the most widely used set of parameters used in the previous MD simulations of B-DNA interaction with Na⁺, consisting of the combination of the amber adapted Åqvist parameters for the ion with TIP3P solvent model, differs noticeably from the other sets. The differences observed are both quantitative and qualitative: set 3 leads to
considerably lower residence times, and does not display the sequence dependence of the
residence times of the ion binding to the backbone.

**Acknowledgements**

F.M. gratefully acknowledges financial support from the Regione Autonoma della Sardegna,
through the Legge Regionale 07/09/2007 (code CRP-59740) and Prof. Giuseppe Saba for
technical support. A.L. wishes to acknowledge the Visiting Professor Programme financed
by the Regione Autonoma Sardegna and the Swedish Research Council (VR) for financial
support. All authors acknowledge the Swedish National Infrastructure for Computing (SNIC).
References


380 doi: 10.1002/bip.10334.


382 http://dx.doi.org/10.1016/0263-7855(96)00018-5.


384 **2003**, *31* (20), 5971.


Tables

Table 1- Ion potential parameters used in the MD Simulations presented in this study.

Lennard-Jones parameters

<table>
<thead>
<tr>
<th>Set name</th>
<th>Clp&lt;sup&gt;a&lt;/sup&gt;</th>
<th>W&lt;sup&gt;b&lt;/sup&gt;</th>
<th>σ (nm)</th>
<th>ε (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>set 1</td>
<td>SD</td>
<td>SPC</td>
<td>2.350 x 10&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>5.440 x 10&lt;sup&gt;-1&lt;/sup&gt;</td>
</tr>
<tr>
<td>set 2</td>
<td>JC</td>
<td>TIP3P</td>
<td>2.439 x 10&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>3.658 x 10&lt;sup&gt;-1&lt;/sup&gt;</td>
</tr>
<tr>
<td>set 3</td>
<td>AA</td>
<td>TIP3P</td>
<td>3.328 x 10&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>1.159 x 10&lt;sup&gt;-2&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>Counterion parameters with the following abbreviations: SD: Smith and Dang; JC: Joung and Cheatham; AA: Amber-adapted Åqvist parameters.  
<sup>b</sup>Water model: flexible SPC model; TIP3P model.
Figures Caption

**Fig. 1** – The three DNA sequences considered in this study. Backbone atoms are in ribbon and the atoms of the bases in licorice representations, respectively. The bases are color-coded as shown in the legend. Sodium ions are represented as spheres in cyan.

**Fig. 2** - A•T and C•G in the canonical Watson-Crick DNA base pairing; electronegative reference atoms used in the analysis of ion binding in the major groove and in the backbone are labeled.

**Fig. 3** - (a) The DDD sequence with highlighted oxygen atoms of the phosphates O1P (black spheres) and O2P (red spheres). (b) RDFs of Na$^+$ with respect to O1P (black lines) and O2P (red lines) for the three oligomers A4T4, TA and DDD obtained with set 1 parameters.

**Fig. 4** - (a) The DDD sequence with the highlighted reference atoms of the major groove. (b) RDFs of Na$^+$ with respect to the reference atoms of the major groove for the three oligomers A4T4, TA and DDD obtained with set 1 parameters. Color codes for each atom are shown in the legend.

**Fig. 5** – a) Representative conformation of a water-mediated contact of a sodium ion with A$_{N7}$ to explain the peak of the RDF at ~4.5 Å. b) Representative conformation of a direct contact of a sodium ion with A$_{N7}$ to explain the peaks at ~4.5 Å and ~6 Å observed in the peaks of RDF respectively for A$_{N6}$ and T$_{O4}$. c) Representative conformation of a direct contact of a sodium ion with T$_{O2}$ to explain the peaks at ~6 Å observed in RDF for T$_{O4}$.

**Fig. 6** - RDFs of Na$^+$ with respect to O1P (black) and O2P (red) for the three oligomers A4T4, TA and DDD obtained respectively with set 1, 2 and 3 parameters.

**Fig. 7** - Weighted average residence time of Na$^+$ in direct contact with the oxygen atoms of the phosphate groups for the three oligomers A4T4, TA and DDD obtained respectively with
set 1, 2 and 3 parameters. Both filaments are represented in the 5’→3’ direction. Color code:
black, continuous line = O1P filament 1; black dashed line = O1P filament 2; red, continuous
line = O2P filament 1; red dashed line = O2P filament 2.

**Fig. 8** - RDFs of Na$^+$ with respect to reference atoms of the major groove for the three
oligomers A4T4, TA and DDD obtained respectively with set 1, 2 and 3 parameters. Color
codes for each electronegative atom are shown in the legend.

**Fig. 9** - Weighted average residence time of Na$^+$ in direct contact with electronegative atoms
of the major groove for the three oligomers A4T4, TA and DDD obtained respectively with
set 1, 2 and 3 parameters. Colour code: black line = filament 1; red line = filament 2.
Fig. 1 – The three DNA sequences considered in this study. Backbone atoms are in ribbon and the atoms of the bases in licorice representations, respectively. The bases are color-coded as shown in the legend. Sodium ions are represented as spheres in cyan.

Figure 1
171x106mm (300 x 300 DPI)
Fig. 2 - A•T and C•G in the canonical Watson-Crick DNA base pairing; electronegative reference atoms used in the analysis of ion binding in the major groove and in the backbone are labeled.

Figure 2
Fig. 3 - (a) The DDD sequence with highlighted oxygen atoms of the phosphates O1P (black spheres) and O2P (red spheres). (b) RDFs of Na$^+$ with respect to O1P (black lines) and O2P (red lines) for the three oligomers A4T4, TA and DDD obtained with set 1 parameters.

Figure 3
404x137mm (300 x 300 DPI)
Fig. 4 - (a) The DDD sequence with the highlighted reference atoms of the major groove. (b) RDFs of Na\(^+\) with respect to the reference atoms of the major groove for the three oligomers A4T4, TA and DDD obtained with set 1 parameters. Color codes for each atom are shown in the legend.

Figure 4
390x131mm (300 x 300 DPI)
Fig. 5 – a) Representative conformation of a water-mediated contact of a sodium ion with AN7 to explain the peak of the RDF at ~4.5 Å. b) Representative conformation of a direct contact of a sodium ion with AN7 to explain the peaks at ~4.5 Å and ~6 Å observed in the peaks of RDF respectively for AN6 and TO4. c) Representative conformation of a direct contact of a sodium ion with TO2 to explain the peaks at ~6 Å observed in RDF for TO4.

Figure 5
Fig. 6 - RDFs of Na⁺ with respect to O1P (black) and O2P (red) for the three oligomers A4T4, TA and DDD obtained respectively with set 1, 2 and 3 parameters.

Figure 6
326x244mm (300 x 300 DPI)
Fig. 7 - Weighted average residence time of Na+ in direct contact with the oxygen atoms of the phosphate groups for the three oligomers A4T4, TA and DDD obtained respectively with set 1, 2 and 3 parameters. Both filaments are represented in the 5'→3' direction. Color code: black, continuous line = O1P filament 1; black dashed line = O1P filament 2; red, continuous line = O2P filament 1; red dashed line = O2P filament 2.

Figure 7
Fig. 8 - RDFs of Na+ with respect to reference atoms of the major groove for the three oligomers A4T4, TA and DDD obtained respectively with set 1, 2 and 3 parameters. Color codes for each electronegative atom are shown in the legend.

Figure 8
322x262mm (300 x 300 DPI)
Fig. 9 - Weighted average residence time of Na⁺ in direct contact with electronegative atoms of the major groove for the three oligomers A4T4, TA and DDD obtained respectively with set 1, 2 and 3 parameters. Colour code: black line = filament 1; red line = filament 2.

Figure 9