SUPPORTING INFORMATION

Analyzing ion distributions around DNA

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

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

1 The curvilinear helicoidal coordinate system

By the helical axis of a nucleic acid fragment we mean a continuous curve $\boldsymbol{\gamma}(D)$ that is piecewise helical, and which has an associated orthonormal, right-handed frame $(\mathbf{d}_1(D), \mathbf{d}_2(D), \mathbf{d}_3(D))$. Here, $D \in [1, n]$ is a continuous parameterisation of a DNA fragment with n base pairs that is expressed in units of base pair steps, and which is measured from the first base pair to the last. The helical axis is formed from segments of (circular) helices within each junction, that are matched continuously at each base pair level. For any DNA fragment atomistic configuration, and for integer values $D = i, \quad i = 1, \ldots, n$ both the helical axis points and associated frames take prescribed values $\boldsymbol{\gamma}(i) = \boldsymbol{\gamma}^i$ and $\mathbf{d}_j(i) = \mathbf{d}_j^i$, with the data $\boldsymbol{\gamma}^i$ and $\mathbf{d}_j^i$ computed by Curves+ using the axis frames at each base pair level. At all intermediate values $i < D < i + 1$, $\boldsymbol{\gamma}(D)$ and $\mathbf{d}_j(D)$ are obtained by continuous helicoidal interpolation using Euler's Screw Axis $F(i)$, which is a line associated with the two base pair levels i and $i + 1$. Specifically the orientations $\mathbf{d}_j(D)$ are obtained by uniformly rotating the axis frames around the Screw Axis $F(i)$, while the helical segment $\boldsymbol{\gamma}(D)$ is obtained by simultaneously and uniformly translating along, and rotating about, $F(i)$ (see Figure S1). In mathematical terms this means that in each junction $i = 1, \ldots, n - 1$ there are three constants u_j such that the vector $\mathbf{u} = \sum_{j=1}^{3} u_j \mathbf{d}_j(D)$ is the (constant) tangent to the line $F(i)$, and the interpolating frames satisfy

$$\frac{d\mathbf{d}_j}{dD} = \mathbf{u} \times \mathbf{d}_j, \quad i < D < i + 1, \quad \mathbf{d}_j(i) = \mathbf{d}_j^i, \quad \mathbf{d}_j(i + 1) = \mathbf{d}_j^{i+1}. \tag{1}$$

Similarly in each junction $i = 1, \ldots, n - 1$ there exist three other constants v_j such that the vector $\mathbf{v} = \sum_{j=1}^{3} v_j \mathbf{d}_j(D)$ satisfies

$$\frac{d\boldsymbol{\gamma}}{dD} = \boldsymbol{\gamma}' = \mathbf{v}, \quad i < D < i + 1, \quad \boldsymbol{\gamma}(i) = \boldsymbol{\gamma}^i, \quad \boldsymbol{\gamma}(i + 1) = \boldsymbol{\gamma}^{i+1}, \tag{2}$$

where the $\mathbf{d}_j(D)$ are computed in (1). The above conditions uniquely define $\boldsymbol{\gamma}(D)$ and $\mathbf{d}_j(D)$ as smoothly varying quantities within each junction, that are continuous, but no better, across each base pair. The numbers u_i and v_i are constant within each junction, but discontinuous across base pairs. As a result $\boldsymbol{\gamma}'$ is discontinuous at base pair levels and $\boldsymbol{\gamma}$ can have a corner (see Figure S1). For this reason, $\mathbf{d}_3$ cannot be taken to be both continuously varying and parallel to $\boldsymbol{\gamma}'$; specifically, the plane $(\mathbf{d}_1, \mathbf{d}_2)$ is not in general

orthogonal to the helical segment $\boldsymbol{\gamma}$. Finally, we note that the vector $\mathbf{d}_3(i)$, which is called the local helical axis in [1], is not simply related to the local junction Screw Axis $F(i)$.

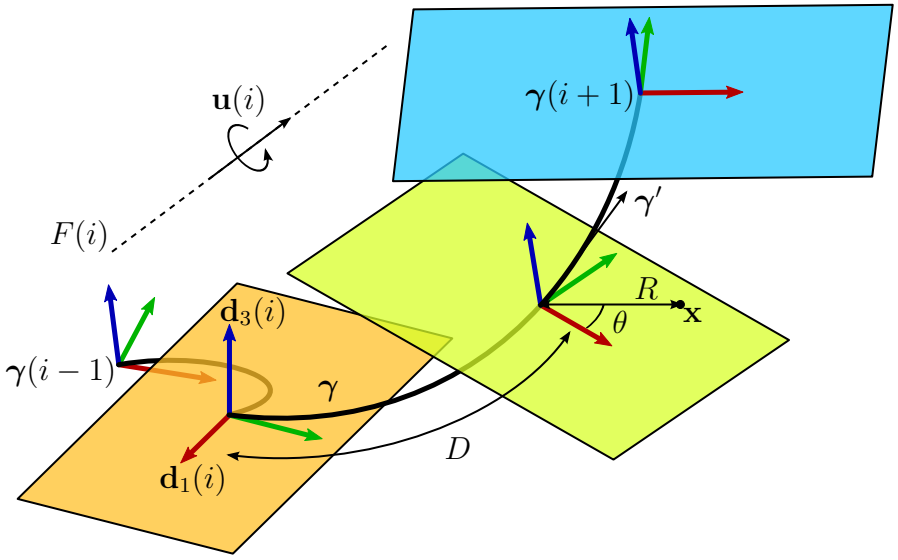


Figure S1: Illustration of the helicoidal interpolation between given axis frames at base pair levels i and $i+1$ (the axis frame at base pair level $i-1$ is also shown). The non-orthogonal curvilinear coordinate system for Euclidean space comprising D along with polar coordinates (R, θ) in each $(\mathbf{d}_1, \mathbf{d}_2)$ plane is free from singularity within each junction, provided that R is sufficiently small.

2 Computing curvilinear helicoidal coordinates

The explicit formula for the change of coordinates between Cartesian coordinates $\mathbf{x} = (x, y, z)$ and curvilinear helicoidal coordinates (D, R, θ) is:

$$\mathbf{x} = \boldsymbol{\alpha}(D, R, \theta) = \boldsymbol{\gamma}(D) + R\cos\theta\,\mathbf{d}_1(D) + R\sin\theta\,\mathbf{d}_2(D). \tag{3}$$

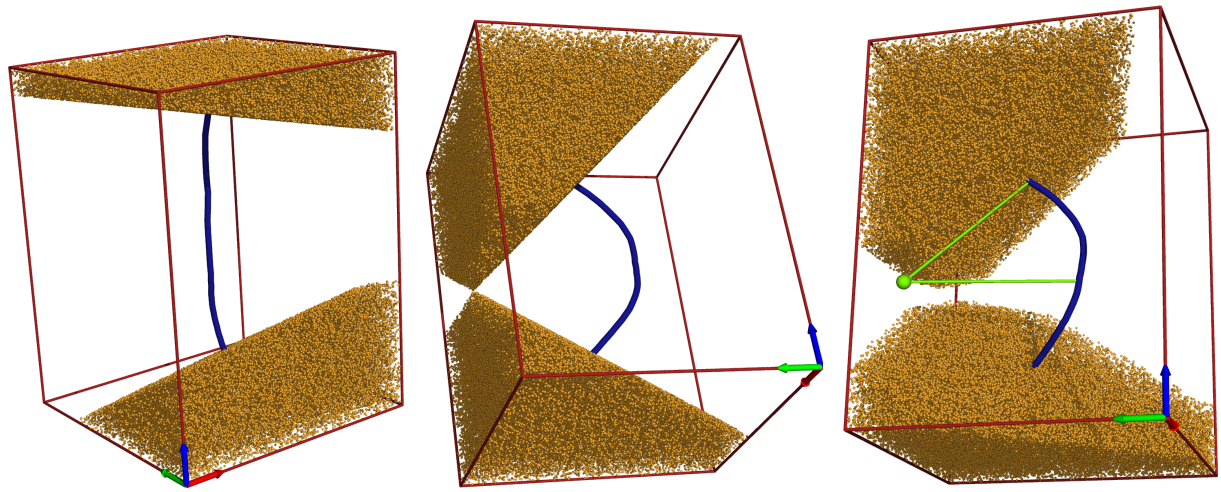
where $\boldsymbol{\gamma}$, $\mathbf{d}_1$ and $\mathbf{d}_2$ are defined in (2) and (1). As any two planes $(\mathbf{d}_1(D_1), \mathbf{d}_2(D_1))$ and $(\mathbf{d}_1(D_2), \mathbf{d}_2(D_2))$ intersect if they are not parallel, this change of coordinates will have singularities, which can be computed explicitly using vector calculus and linear algebra. In particular, if $v_3 > 0$ (as it can be taken to be for non-planar helices), then locally the transformation (3) will be invertible provided that R is sufficiently small, so that $(D, R, \theta) = \boldsymbol{\alpha}^{-1}(\mathbf{x})$. To explicitly compute $\boldsymbol{\alpha}^{-1}(\mathbf{x})$, Curves+ searches for the value D^* such that:

$$(\mathbf{x} - \boldsymbol{\gamma}(D^*)) \cdot \mathbf{d}_3(D^*) = 0. \tag{4}$$

The helicoidal coordinates of the point $\mathbf{x}$ are then (D^*, R^*, θ^*), where R^* is the length of the vector $\mathbf{x} - \boldsymbol{\gamma}(D^*)$, and the angle θ^* is measured anti-clockwise around $\mathbf{d}_3(D^*)$ from $\mathbf{d}_1(D^*)$.

The existence of curvilinear helicoidal coordinates for a given point is subject to there being (at least) one value of D such that the point lies in the $(\mathbf{d}_1(D), \mathbf{d}_2(D))$ plane. The

extent of the region of space for which curvilinear helicoidal coordinates exist and are free from singularity depends therefore on the helical axis. Figure S2 (a) shows the average helical axis of an octadecamer (the ATGC oligomer in the ABC library of simulations), in which the $(\mathbf{d}_1(D), \mathbf{d}_2(D))$ planes at all values of D are close to being parallel one to another. The region of space for which curvilinear helicoidal coordinates exist and are unique then extends beyond the boundary of the box, and is limited only by the two $(\mathbf{d}_1, \mathbf{d}_2)$ planes at the first and last base pair levels. In contrast panels (b) and (c) of Figure S2 show the example of the highly deformed helical axis of the protein-bound TATA box (PDB code 1CDW). Here, the $(\mathbf{d}_1, \mathbf{d}_2)$ planes at the first and last base pair levels intersect in a line within the boundary of the box. The green sphere in panel (c) is a point lying on the intersection of two distinct $(\mathbf{d}_1, \mathbf{d}_2)$ planes so that there are two distinct values D^* for which (4) is satisfied. The green lines are the two vectors $(\mathbf{x} - \boldsymbol{\gamma}(D^*))$ for the two corresponding values of D^*. The helicoidal coordinates are only valid sufficiently close to the DNA axis such that no two distinct $(\mathbf{d}_1, \mathbf{d}_2)$ planes can intersect.



(a) (b) (c)

Figure S2: The orange spheres fill the entire region within a prescribed box (of size 60x60x80Å centred around the helical axis shown in blue) for which curvilinear helicoidal coordinates do not exist, for (a) a comparatively undeformed helical axis of an octadecamer, and (b),(c) the helical axis of the highly deformed TBP-bound TATA box. The green sphere in panel (c) has two sets of helicoidal coordinates (see text).

3 Calculating ion concentrations

Given an ensemble of ion locations, the corresponding ensemble of ion helicoidal coordinates (D, R, θ) is found according to the procedure described in section 2, and is stored in a 3D histogram $H(D, R, \theta)$. In order to obtain ion concentrations expressed as a molarity, it is necessary to divide the ion count in each bin of the histogram H by its corresponding Euclidean volume V with respect to the helical axis of the reference structure. For the bin with the ranges $D_1 < D < D_2$, $R_1 < R < R_2$, $\theta_1 < \theta < \theta_2$ (see Figure S3)

$$V = \int_{D_1}^{D_2} \int_{R_1}^{R_2} \int_{\theta_1}^{\theta_2} \det\left(\frac{\partial \boldsymbol{\alpha}}{\partial(D, R, \theta)}\right) dD dR d\theta, \tag{5}$$

where $\partial\boldsymbol{\alpha}/\partial(D, R, \theta)$ is the Jacobian matrix of the transformation (3). A simple computation using equations (1) and (2) provides the expression

$$\det\left(\frac{\partial\boldsymbol{\alpha}}{\partial(D, R, \theta)}\right) = R(v_3 - u_2 R\cos\theta + u_1 R\sin\theta) > 0, \tag{6}$$

where the last condition is always satisfied for R sufficiently small, i.e. for volume elements located sufficiently close to the helical axis, because v_3 is positive by construction.

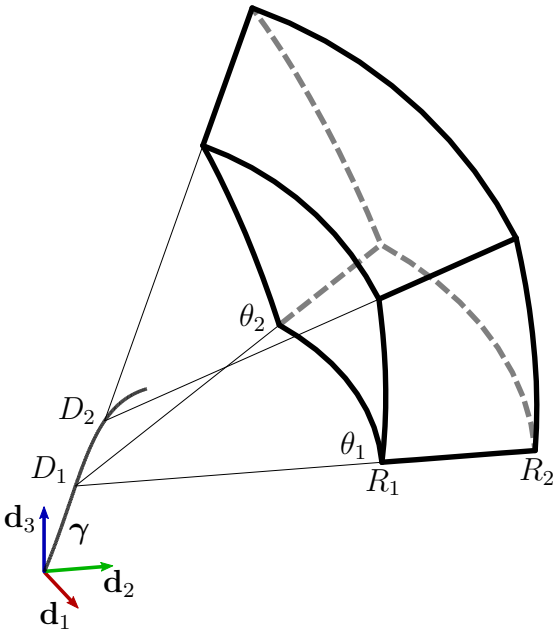


Figure S3: Shape of an (artificially large) histogram bin with curvilinear helicoidal coordinate ranges $D_1 < D < D_2$, $R_1 < R < R_2$, $\theta_1 < \theta < \theta_2$ with respect to the helical axis $\boldsymbol{\gamma}$.

Evaluation of the definite integral (5) using (6) then yields the formula

$$V = (D_2 - D_1)\left(\frac{(\theta_2 - \theta_1)}{2}(R_2^2 - R_1^2)v_3 - \frac{(R_2^3 - R_1^3)}{3}\big(u_1(\cos\theta_2 - \cos\theta_1) + u_2(\sin\theta_2 - \sin\theta_1)\big)\right), \tag{7}$$

which reduces to the more familiar expression

$$V = (D_2 - D_1)\frac{\theta_2 - \theta_1}{2}(R_2^2 - R_1^2)v_3. \tag{8}$$

when $u_1 = u_2 = 0$, corresponding to the cases where the helical segment $\boldsymbol{\gamma}$ is straight, with the corresponding coordinates then being cylindrical, and the bin accordingly being a torsionally-sheared truncated wedge, as opposed to the distorted case shown in Figure S3.

4 Supplementary figures

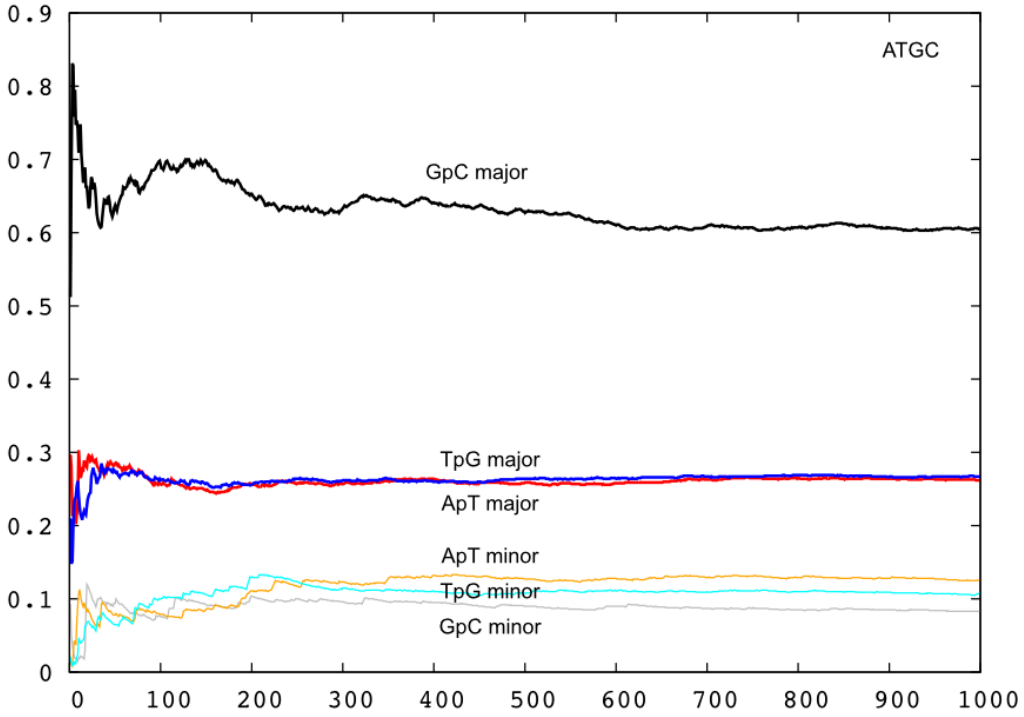


Figure S4: Time-averaged K+ populations within the DNA grooves for the unique base pair steps (A9pT10, T10pG11, G11pC12) belonging to the central tetranucleotide of the ATCG oligomer for increasing durations (ns) of the molecular dynamics trajectory.

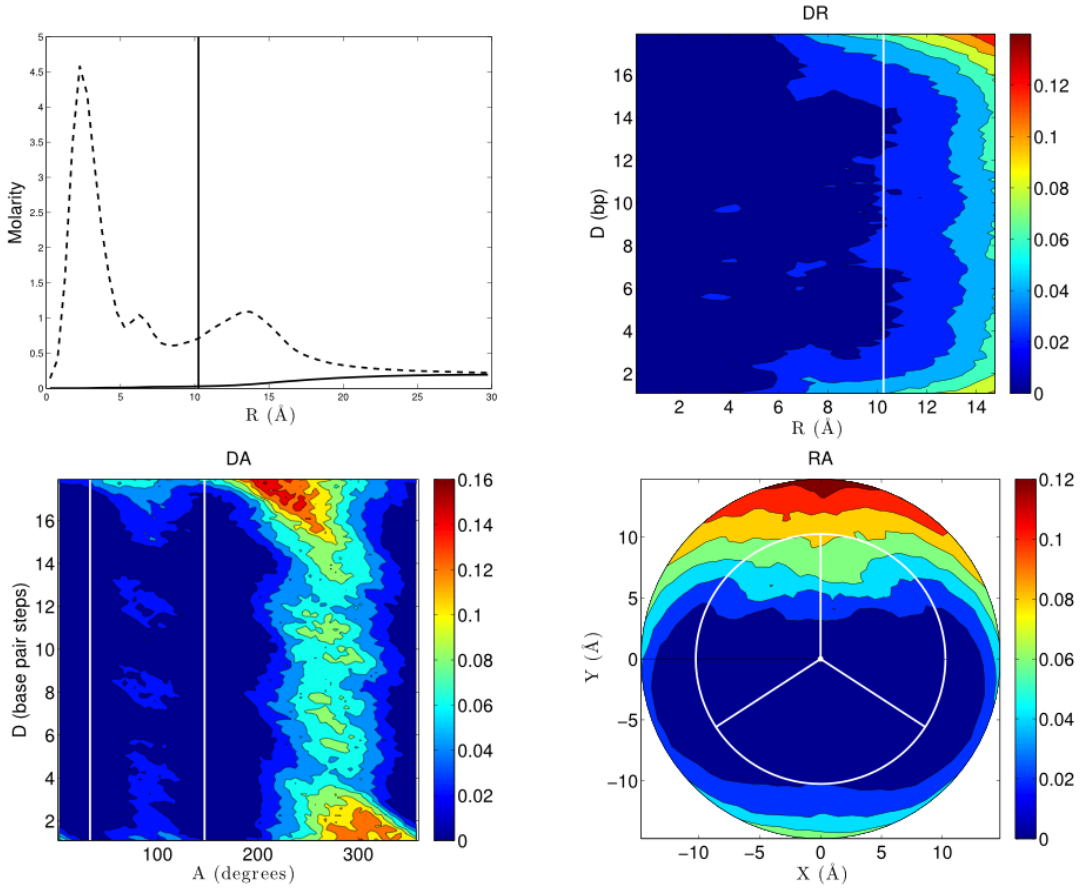


Figure S5: Chlorine ion distributions for the 1 μs AGCT trajectory: 1D R distribution (top left, solid line, contrasted with the K+ distribution, dotted line line), 2D DR distribution (top right), 2D DA distribution (bottom left), 2D RA distribution (bottom right). All values are molarities, and, for the 2D plots, the blue to red color scale indicates increasing values.

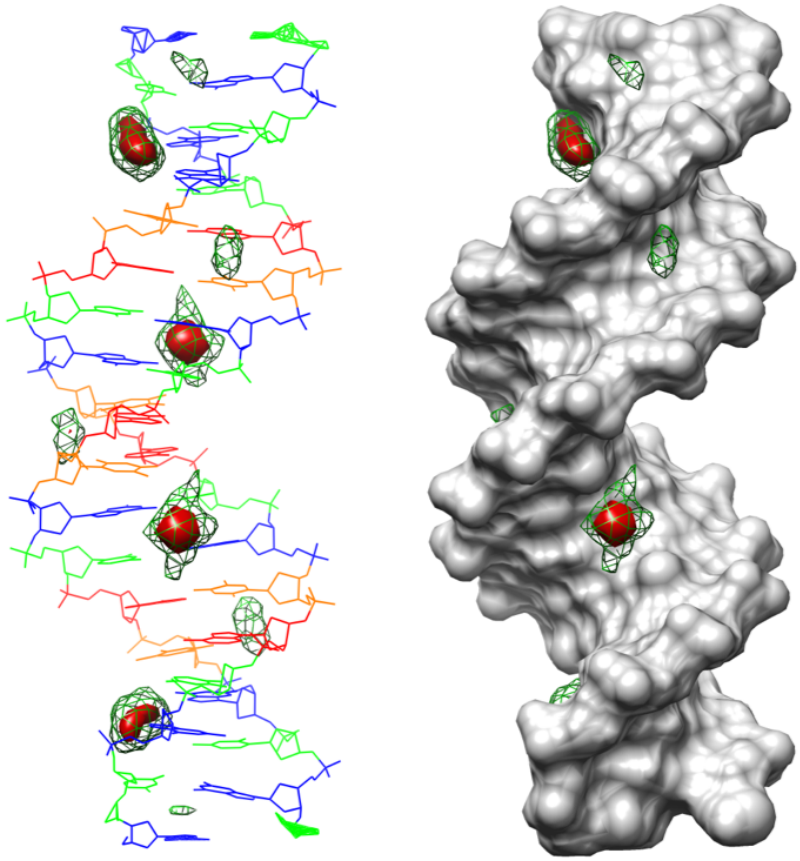


Figure S6: 3D K+ distributions derived from the curvilinear helicoidal analysis of the 1 μs ATGC trajectory. The average DNA structure is shown as a line drawing on the left (G: blue, C: green, A: red, T: orange) and as a grey solvent accessible surface on the right. Molarity isodensity surfaces are plotted at 15 M (solid red) and 5 M (green mesh).

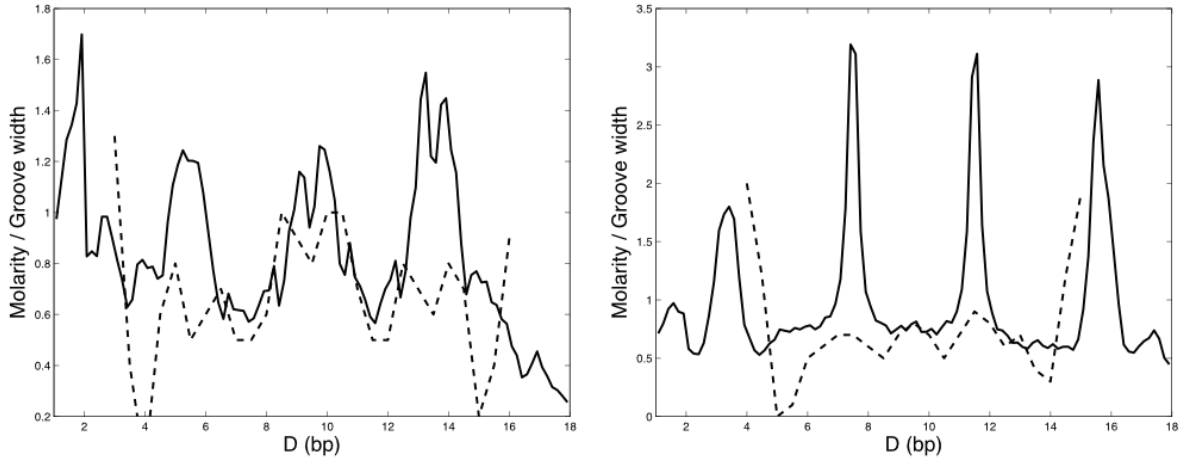


Figure S7: Variations in K+ molarity along the DNA grooves (solid lines) of the ATGC oligomer. Values are averaged over the 1 μs ATGC trajectory and compared with variations in groove width (dotted lines). Groove width variations are plotted in Å with respect to the respective minimal values (major: 9.4 Å, minor: 2.5 Å) on the same scale as the molarities. Left: major groove. Right: minor groove.

5 References

1. 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.