Electronic Supplementary Material (ESI) for Soft Matter

Supplementary Information

DNA conformational flexibility study using phosphate backbone neutralization model

Shiyan Xiao,[a] Hong Zhu,[a] Lei Wang,[a] and Haojun Liang[*b,c]

[a] *CAS Key Laboratory of Soft Matter Chemistry, Department of Polymer Science and Engineering, University of Science and Technology of China, Hefei, Anhui 230026, P. R. China.*
[b] *CAS Key Laboratory of Soft Matter Chemistry, Collaborative Innovation Center of Chemistry for Energy Materials, Department of Polymer Science and Engineering, University of Science and Technology of China, Hefei, Anhui 230026, P. R. China.*
[c] *Hefei National Laboratory for Physical Sciences at Microscale, University of Science and Technology of China, Hefei, Anhui, 230026, P. R. China.*

*To whom correspondence should be addressed: E-mail: hjliang@ustc.edu.cn

1. DNA bending variables

In order to characterize the bending degree and direction of the double-stranded oligomer, an algorithm is proposed here based on the screw-axis bending variable proposed by Curuksu, Zakrzewska and Zacharias.[1] The orientation of one nucleotide is defined by a reference axis system located on the base plane with unit vectors ($\overrightarrow{e_1}, \overrightarrow{e_2}, \overrightarrow{e_3}$), where $\overrightarrow{e_1}$ is the vector ($\overrightarrow{N_9 \rightarrow C_8}$), and $\overrightarrow{e_2}$ is vector ($\overrightarrow{N_9 \rightarrow C_1'} \times \overrightarrow{N_9 \rightarrow C_8}$), for the nucleotide at strand 1, but ($\overrightarrow{N_9 \rightarrow C_8} \times \overrightarrow{N_9 \rightarrow C_1'}$) for the nucleotide at strand 2, $\overrightarrow{e_3}$ equals ($\overrightarrow{e_2} \times \overrightarrow{e_3}$) (see Figure S3). $\overrightarrow{e_1}$ and $\overrightarrow{e_3}$ is parallel to the base plane, $\overrightarrow{e_3}$ approximately points along the direction of the long base pair axis from strand 1 to strand 2, and $\overrightarrow{e_2}$ is perpendicular to the base plane. The "handle" vector at each end of the double strand oligomer is the sum of four $\overrightarrow{e_2}$ on two-bp terminal fragment (referred as $\overrightarrow{H_1}$ and $\overrightarrow{H_2}$). The bending angle (θ) of the double-stranded oligomer is defined as the angle between two handle vectors, that is, $\cos(\theta) = \overrightarrow{H_1} \cdot \overrightarrow{H_2}/(|\overrightarrow{H_1}| \cdot |\overrightarrow{H_2}|)$. Then, we can define $\vec{H} = \overrightarrow{H_1} - \overrightarrow{H_2}$, and $\vec{S}$ as the sum of the four $\overrightarrow{e_2}$ vectors of the two base pairs located at the central fragment of the DNA duplex, and $\vec{r}$ as the sum of the vectors $\overrightarrow{C_1'C_1'}$ from strand 1 to strand 2 at the two central base pairs. $\vec{S}$ directs towards the major groove approximately, and $\vec{r}$ directs towards the backbone of strand 2 from that of strand 1. Using $\vec{H}$, $\vec{S}$ and $\vec{r}$, we can derive ϕ by $\cos(\phi) = \vec{H} \cdot \vec{r}/(|\vec{H}| \cdot |\vec{r}|)$, and α by $\cos(\alpha) = \vec{H} \cdot \vec{S}/(|\vec{H}| \cdot |\vec{S}|)$. Then, 1) $\phi \rightarrow 0^\circ$: DNA bending points towards the backbone of strand 1; 2) $\phi \rightarrow 180^\circ$: DNA bending points towards the backbone of strand 2; 3) $\phi \rightarrow 90^\circ$, if $\alpha < 90^\circ$, then DNA bends towards the major groove, and if $\alpha > 90^\circ$, then DNA bends towards the minor groove.

2. PCA and entropy calculations

The PCA method is employed to investigate the conformational flexibility of nucleic acids in physiological condition. The essential dynamics of DNA duplex is derived from the covariance ma trix,[2, 3-4] which is constructed using the Cartesian molecular coordinates of all atoms of nucleic acids after removing the rotational and translational motions by least-square fitting [5] of the DNA to the first frame of the corresponding MD trajectory. Diagonalization of the covariance matrix yields a set of of eigenvectors v_i and eigenvalues λ_i (in nm^2), both of which are associated with the essential movement mode i. The analysis of the eigenvalues provides the

direct information on the flexibility of DNA and the harmonic force constant associated with the essential deformation movement i.[3,6-8]

$$K_i = K_B T/\lambda_i \tag{S1}$$

where K_i is the force constant of the i^{th} essential movement mode, K_B is the Boltzmann's constant (in kcal/mol K), and T is the absolute temperature. Similarly, the eigenvalues derived by diagonalization of the mass-weighted covariance matrix correspond to frequencies (ω_i), which can be used to calculate the entropies in the pseudo-harmonic limit: [9]

$$S = K_B T \sum_i \frac{\alpha_i}{e^{\alpha_i} - 1} - \ln(e^{-\alpha_i}) \tag{S2}$$

With $\alpha_i = \hbar\omega_i/K_B T$, and the sum extends to all the non-trivial vibrations. Since the limited molecular dynamics trajectory, the time dependence is intrinsic to the estimate of entropy. In the present work, an exponential correction formula developed by Harris *et al.*[10] is used to estimate the entropy at infinite simulation time (S_∞). The entropy estimate S_∞ is obtained using $S(t) = S_\infty - \alpha/t\beta$, with α and β as the fitted parameters and t is the simulation time.

In the present work, home developed program using the methodology of 3DNA [11-13] are employed to calculate the helical parameters for DNA helices. Let x_j and x_j be two helicoidal variables $< x_i x_j >$ be their correlation, and elastic force constants associated with helical parameters can be calculated using $< x_i x_j >= K_B T(F_{ij}^{-1})$, where F is the stiffness matrix whose diagonal elements provide the stiffness constants correspond to pure rotational (twist, roll and tilt) and translational (rise, shift and slide) deformations.[14]

3. Anharmonic corrections to the quasi-harmonic approximations.

In the present work, the anharmonic correction to the entropies calculated by principle component analysis (PCA) is achieved using the method developed by Hünenberger, McCammon and their coworkers [15-16].

Following the PCA method, fist, a generalized system $\boldsymbol{q}$ is selected, and the averaged configuration $\bar{\boldsymbol{q}}$ and the covariance matrix $\boldsymbol{C_q}$ in this coordinates system are evaluated from a molecular dynamics simulations, as

$$\bar{\boldsymbol{q}} = < \boldsymbol{q} > \quad \text{and} \quad \boldsymbol{C_q} = < (\boldsymbol{q} - \bar{\boldsymbol{q}}) \otimes (\boldsymbol{q} - \bar{\boldsymbol{q}}) >$$

(S3)

Diagonalization of the mass-weighted Cartesian covariance matrix yields a set of $3N - 6$ eigenvectors, with N as the number of atoms used for the construction of covariance matrix. That is,

$$\boldsymbol{V_q^T A_q^{1/2} C_q A_q^{1/2} V_q} = F_q$$

(S4)

Where, $\boldsymbol{V_q}$ is a $M \times M$-dimensional orthogonal matrix, and the columns of which represent the M components of the eigenvectors $\{\boldsymbol{v_{q,m}} | m = 1, M\}$, i.e. the quasi-harmonic modes of the original coordinate system. $F_{\boldsymbol{q,m}}$ is the corresponding eigenvalue, and used to calculate the entropies in the pseudo-harmonic limit using equation (S2).

Based on the above treatment, the simulated trajectory will be projected onto the quasi-harmonic modes, i.e., one considers the transformed coordinates $\boldsymbol{a_q}$ defined as:

$$\boldsymbol{a_q} = \boldsymbol{V_q^T A_q^{1/2}} (\boldsymbol{q} - \bar{\boldsymbol{q}})$$

(S5)

These transformed coordinates will be referred to as quasi-harmonic coordinates.

Ideally, the associated configurational probability distribution in the canonical ensemble can be written as

$$p_o'(\boldsymbol{a_q}) = \frac{e^{-\frac{1}{2}\boldsymbol{a_q^T F_q a_q}}}{\int d\boldsymbol{a_q} e^{-\frac{1}{2}\boldsymbol{a_q^T F_q a_q}}} = \prod_{m=1}^{M} p_{o,m}'(\boldsymbol{a_{q,m}})$$

(S6)

With,

$$p_{o,m}'(\boldsymbol{a_{q,m}}) = (2\pi F_{\boldsymbol{q},m}) e^{-\frac{1}{2}\boldsymbol{a_q^T F_{q,m} a_{q,m}}}$$

(S7)

The probability distribution $p_{o,m}'(\boldsymbol{a_{q,m}})$ is a normalized origin-centered Gaussian along each of the coordinates of the essential motion modes.

Then we can assess the effect of anharmonicities in each essential mode as below.

1) From equation (S5) , we can derive the actual distribution $p'_m(\boldsymbol{a_{q,m}})$ associated with the individual quasi-harmonic coordinate from the molecular dynamics trajectory. And check how close the $p'_m(\boldsymbol{a_{q,m}})$ are to the model Gaussians $p'_{o,m}(\boldsymbol{a_{q,m}})$. More precisely, the function

$$f(\boldsymbol{a_{q,m}}) = \left[-F_{q,m}\ln\left(\frac{p'_m(\boldsymbol{a_{q,m}})}{p'_m(\boldsymbol{0})}\right)\right] \tag{S8}$$

should be a straight line of unit slope going through the origin if the mode is perfectly harmonic.

2) The actual distribution $p'_m(\boldsymbol{a_{q,m}})$ can also be used to calculate the exact entropy. The exact contribution of single eigenmode to the classical entropy is calculated using

$$S^{ah}_{cl,m} = k_B\left[\frac{1}{2}\left(1 - ln\frac{\beta h^2}{2\pi}\right) - \int d\boldsymbol{a_{q,m}} p'_m(\boldsymbol{a_{q,m}})\ln p'_m(\boldsymbol{a_{q,m}})\right] \tag{S9}$$

And the overall entropy correction for the non-harmonic behavior of the investigated system is derived using

$$\Delta S^{ah}_{cl} = \sum_{m=1}^{M} S^{ah}_{cl,m} - S \tag{S10}$$

with S as the entropy in the pseudo-harmonic limit (equation S2)

Reference

[1] Curuksu, J.; Zakrzewska, K.; Zacharias, M. *Nucleic Acids Res.* 2008, *36*, 2268–2283.

[2] Pérez, A.; Blas, J. R.; Rueda, M.; López-Bes, J. M.; de la Cruz, X.; Orozco, M. *J. Chem.Theory Comput.* 2005, *1*, 790–800.

[3] Orozco, M.; Pérez, A.; Noy, A.; Luque, F. J. *Chem. Soc. Rev.* 2003, *32*, 350–364.

[4] Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct., Bioinf.* 1993, *17*, 412–425.

[5] Horn, B. K. P. *J. Opt. Soc. Am. A* 1987, *4*, 629–642.

[6] Noy, A.; Pérez, A.; Márquez, M.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* 2005, *127*, 4910–4920.

[7] Pérez, A.; Noy, A.; Lankas, F.; Luque, F. J.; Orozco, M. *Nucleic Acids Res.* 2004, *32*,

6144–6151.

[8] Noy, A.; Pérez, A.; Lankas, F.; Luque, F. J.; Orozco, M. *J. Mol. Biol.* 2004, *343*, 627 – 638.

[9] Andricioaei, I.; Karplus, M. *J. Chem. Phys.* 2001, *115*, 6289.

[10] Harris, S. A.; Gavathiotis, E.; Searle, M. S.; Orozco, M.; Laughton, C.A. *J. Am. Chem. Soc.* 2001 *123* (50), 12658-12663

[11] Olson, W. K. et al. *J. Mol. Biol.* 2001, *313*, 229 – 237.

[12] Lu, X.-J.; Hassan, M. E.; Hunter, C. *J. Mol. Biol.* 1997, *273*, 668 – 680.

[13] Lu, X.-J.; Olson, W. K. *Nucleic Acids Res.* 2003, *31*, 5108–5121.

[14] Lankaš, F.; Šponer, J.; Langowski, J.; Cheatham, T. E. *Biophys. J.* 2003, *85*, 2872–2883.

[15] Baron R.; van Gunsteren W. F.; Hünenberger P. H. *Trends Phys. Chem.* 2006,11, 87.

[16] Baron, R.; Hünenberger P. H.; McCammon, J.A.; *J. Chem. Theory. Comput.* 2009, 5, 3150-3160.

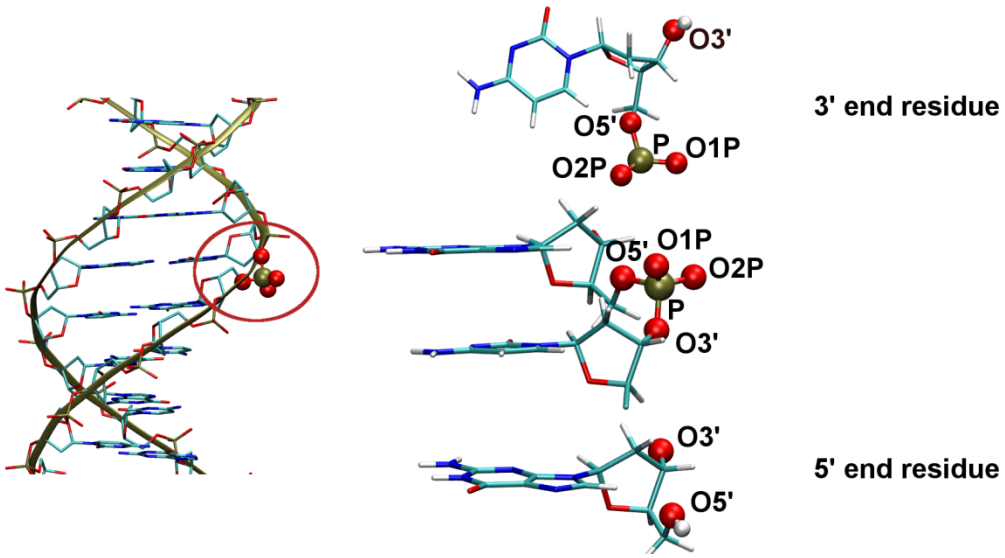

Figure S1: Schematic illustration of the preparation of the neutral atomistic DNA molecule. Posphorus atoms, oxygens, and hydrogens are shown as tan, red, and white balls, respectively. The charges of the atoms on the phosphate group are reduced in order to achieve neutralized DNA residues.

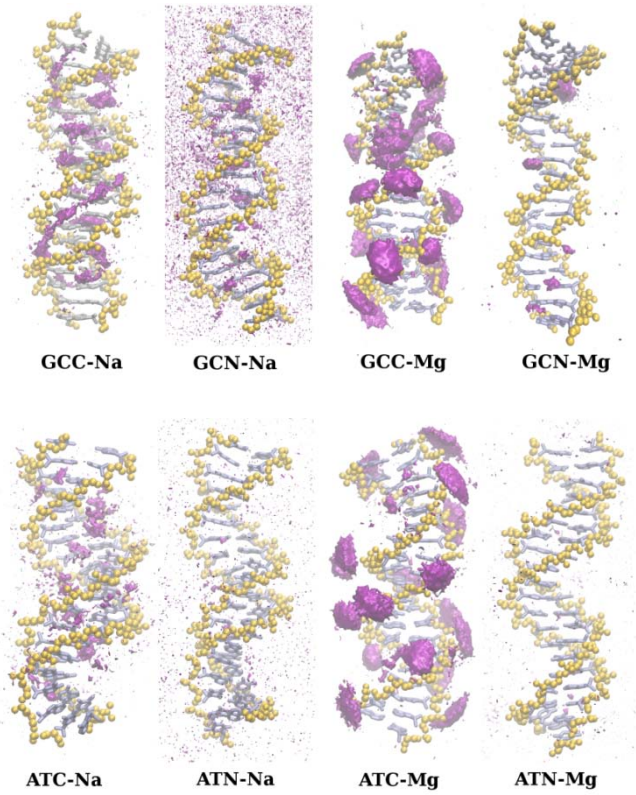

Figure S2: The spatial distribution of metal ions around DNA. The upper picture is corresponded to the charged-full and charge-reduced $d(GC)_{10}{\cdot}d(CG)_{10}$ in Na^{+} and Mg^{2+} electrolytes, respectively. The bottom picture is corresponded to the $d(AT)_{10}{\cdot}d(TA)_{10}$.

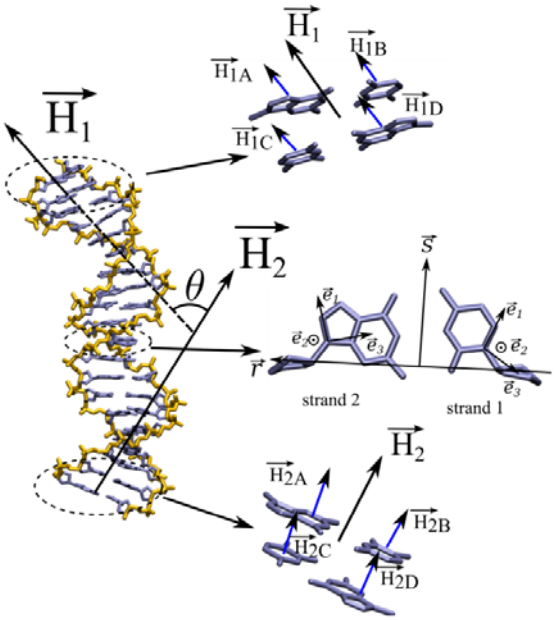

Figure S3 Illustration of the bending angle definition. The global bending angle θ is defined by $\cos(\theta) = \overrightarrow{H_1} \cdot \overrightarrow{H_2} / |\overrightarrow{H_1}| \cdot |\overrightarrow{H_2}|$. If we define $\vec{H} = \overrightarrow{H_1} - \overrightarrow{H_2}$. We can derive ϕ using $\cos(\phi) = \vec{H} \cdot \vec{r} / (|\vec{H}| \cdot |\vec{r}|)$ and α using $\cos(\alpha) = \vec{H} \cdot \vec{S} / (|\vec{H}| \cdot |\vec{S}|)$. If $\alpha \to 0^{\circ}$, DNA bends to the backbone of strand 2; $\alpha \to 180^{\circ}$, DNA bends to the backbone of strand 1; $\phi \to 90^{\circ}$, then DNA bends to the major ($\alpha < 90^{\circ}$) or minor ($\alpha > 90^{\circ}$) groove.

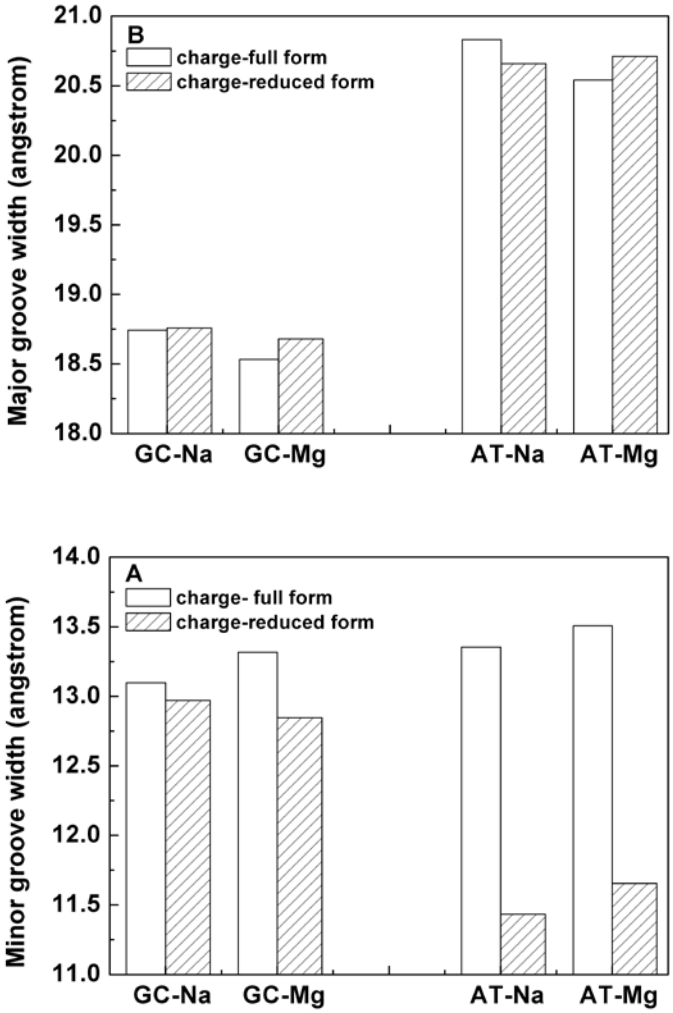

Figure S4 Minor groove width (A) and major groove width (B) of isolated DNA models. The two base pairs at each end of isolated DNA duplex have been excluded in the study, and groove width values presented here are average values calculated for all investigated base pairs. The deviations between groove widths of designated base pairs for each DNA duplex are presented in Table 2 in the main text.

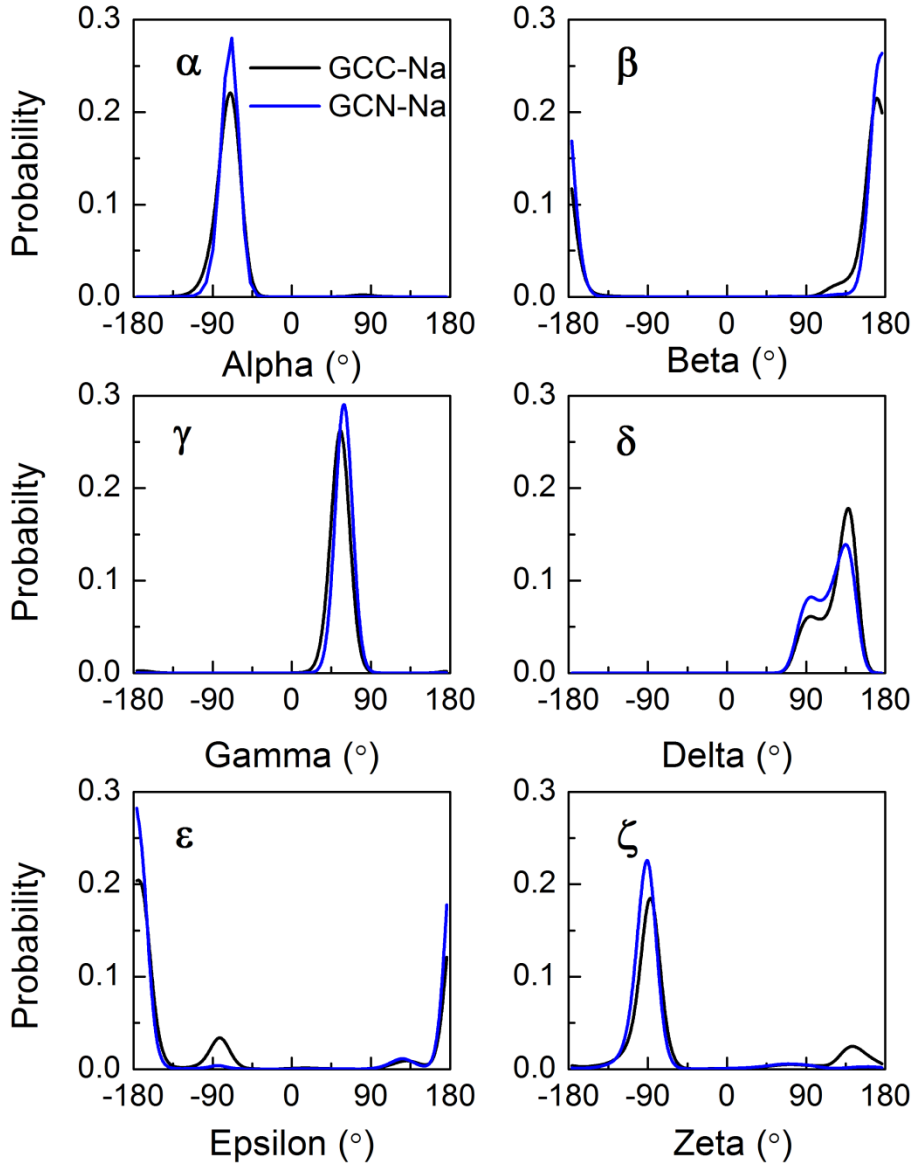

Figure S5 Distributions of the backbone torsion angles for $d(GC)_{10}$·$d(CG)_{10}$ (GCC and GCN) in Na^+.

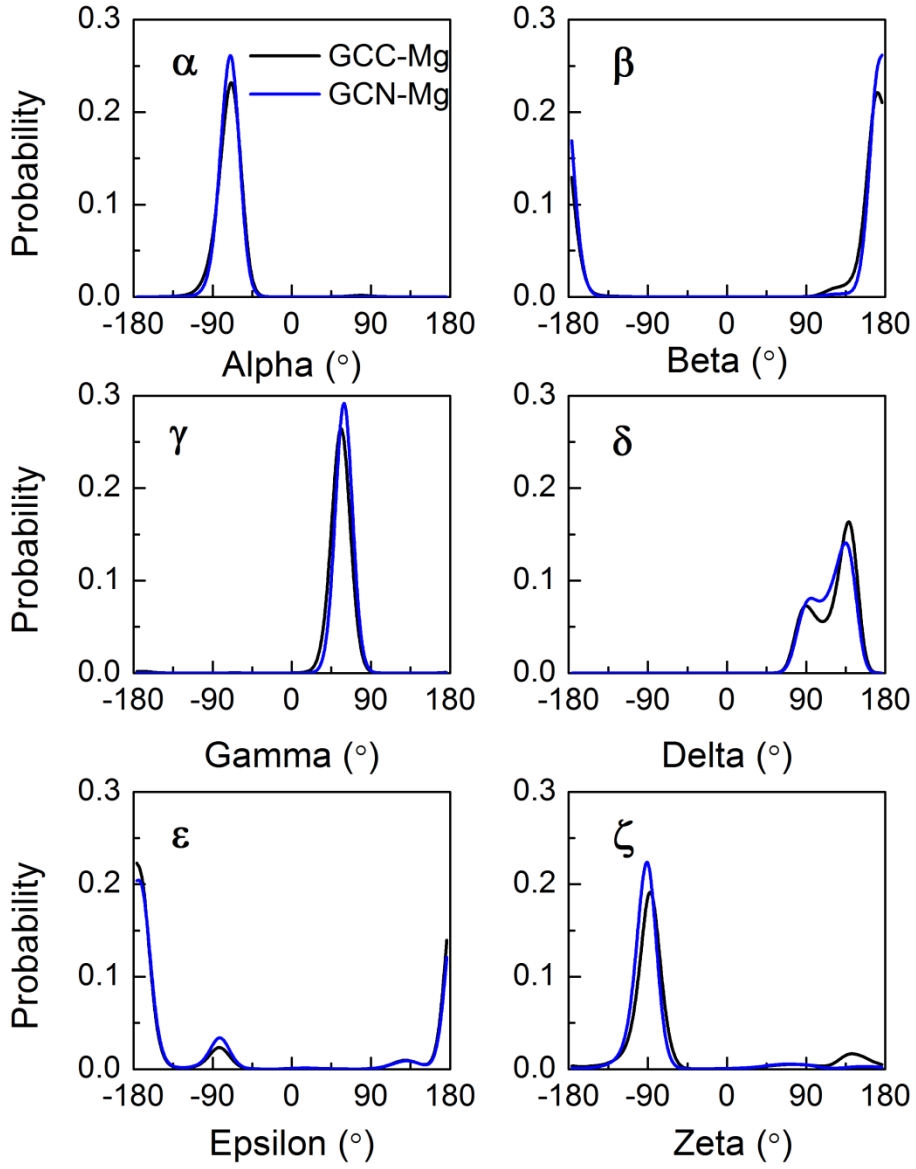

Figure S6 Distributions of the backbone torsion angles for $d(GC)_{10}$·$d(CG)_{10}$ (GCC and GCN) in Mg^{2+}.

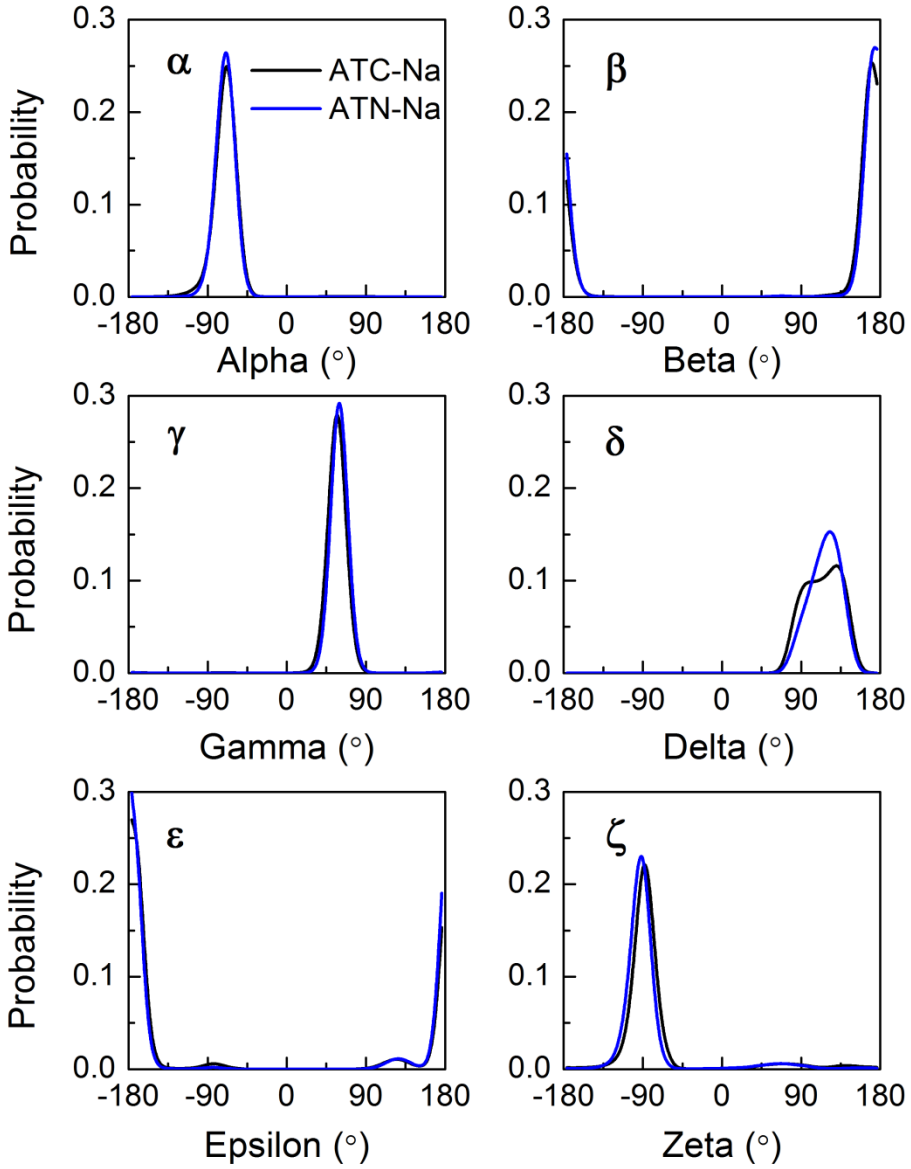

Figure S7 Distributions of the backbone torsion angles for $d(AT)_{10}$·$d(TA)_{10}$ (ATC and ATN) in Na^+.

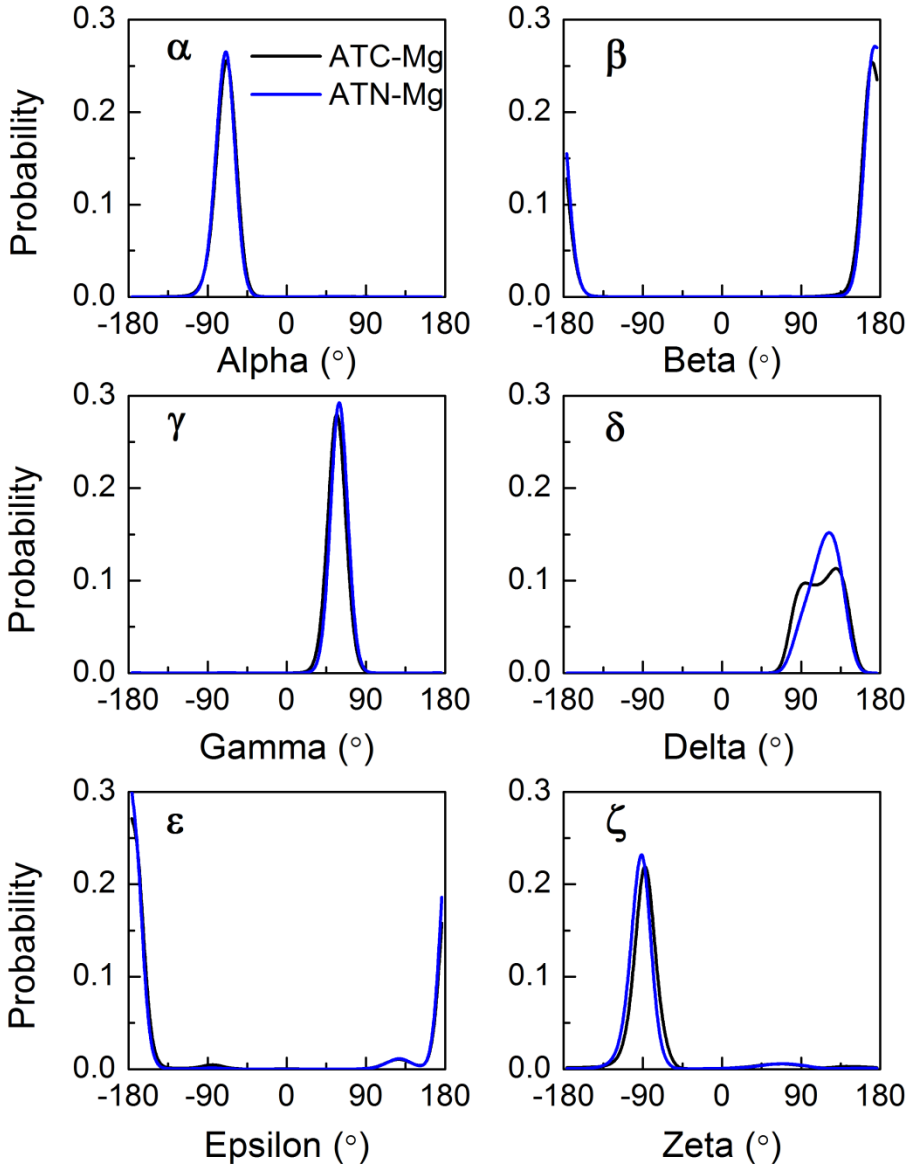

Figure S8 Distributions of the backbone torsion angles for $d(AT)_{10}$·$d(TA)_{10}$ (ATC and ATN) in Mg^{2+}.

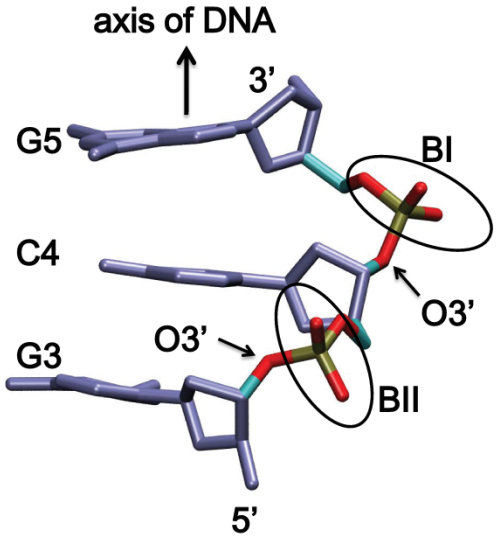

Figure S9 Illustration of BI and BII phosphate group. For BI conformer, the orientation of O1P-O2P vector is roughly perpendicular to the axis of DNA, and the O3' atom points toward outside from the center of the double helix. In BII conformer, O1-O2P vector directs roughly along the axis of DNA, and O3' atom points toward the center of the double helix.

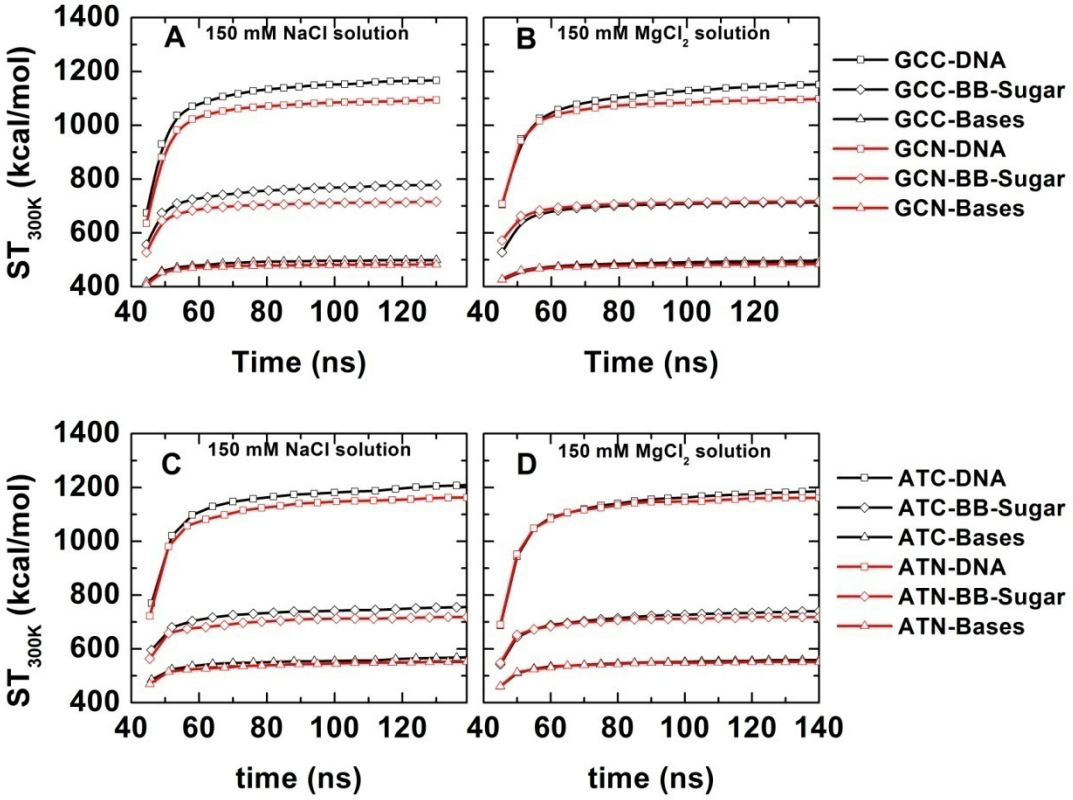

Figure S10 Time evolution of conformational entropy of the whole DNA, the backbone and sugar part (labeled as BB-Sugar), and bases ford$(GC)_{10}$·d$(CG)_{10}$ (upper panel) and d$(AT)_{10}$·d$(TA)_{10}$ (lower panel).

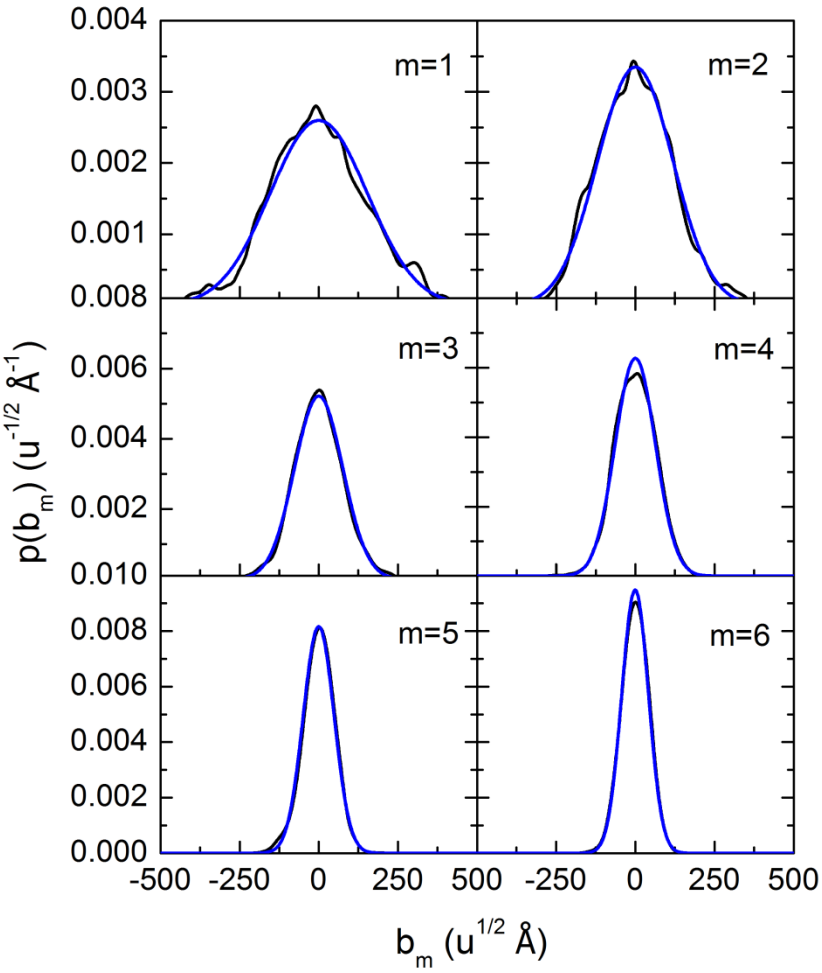

Figure S11 Probability distributions along selected components of the quasi-harmonic coordinates b_m for the charge-full state of $d(GC)_{10}{\cdot}d(CG)_{10}$ in Na^+. The actual distributions from the whole simulation (black line) are displayed together with the corresponding approximate Gaussians for increasing component indices m. All probability distributions are normalized. The letter "u" stands for the atomic mass unit.

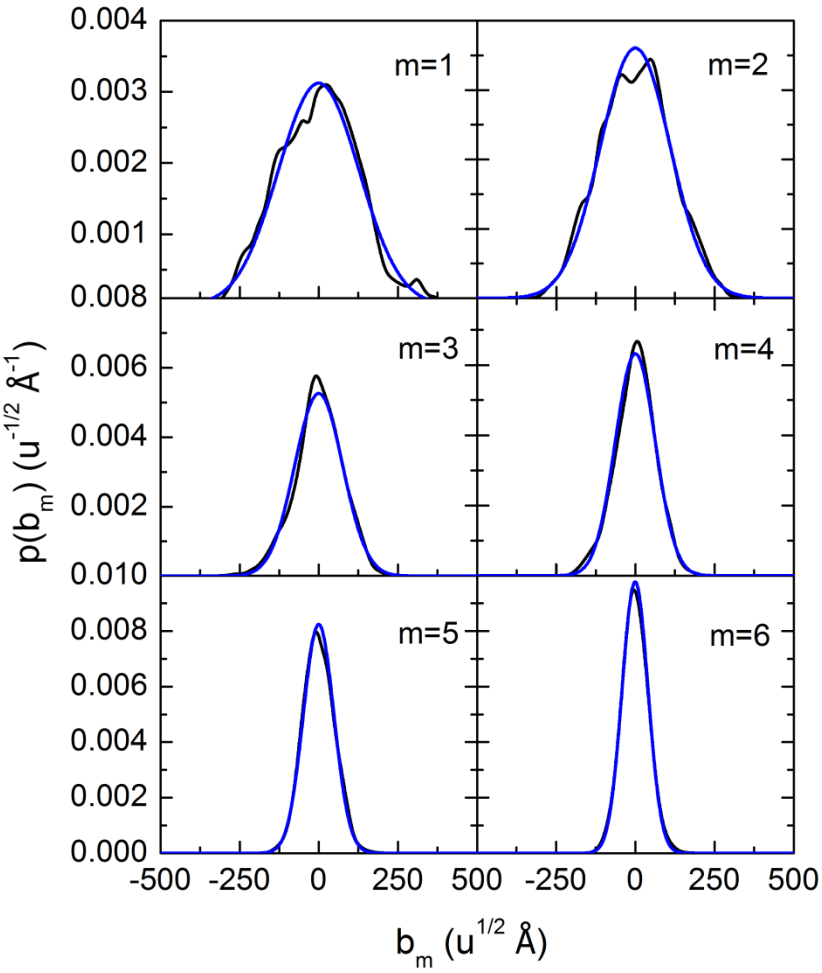

Figure S12 Probability distributions along selected components of the quasi-harmonic coordinates b_m for the charge-reduced state of $d(GC)_{10}{\cdot}d(CG)_{10}$ in Na^+. The actual distributions from the whole simulation (black line) are displayed together with the corresponding approximate Gaussians for increasing component indices m. All probability distributions are normalized. The letter "u" stands for the atomic mass unit.

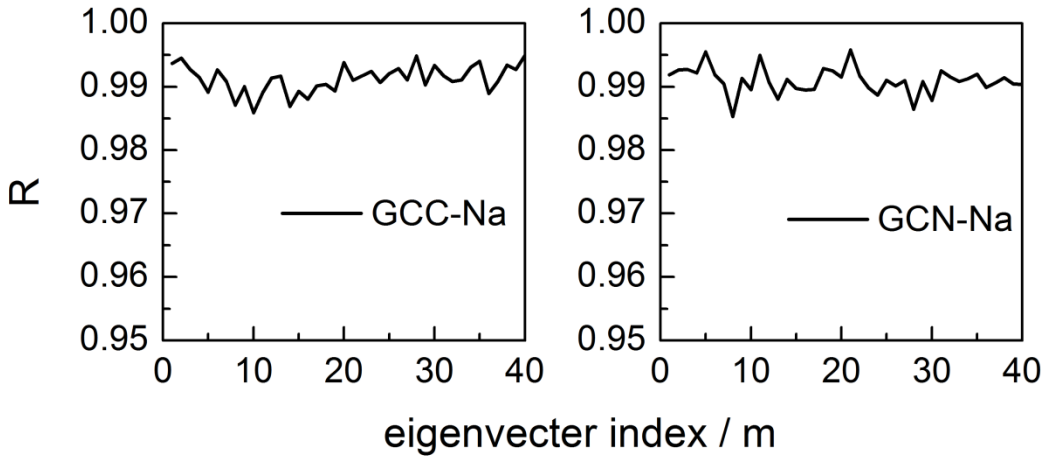

Figure S13 Correlation coefficient R for the fitting of the probability distribution associated with a quasi-harmonic coordinate to a Gaussian displayed as a function of the eigenvector index m from 1 to 40. The coefficient R is the linear regression coefficient corresponding to the fit of equation S8 to a straight line.

Table S1: The charges of the 5’ end residue, 3’ end residue and the central residues are listed.

Residue	Atom type	Amber 99pbsc0	After modification
5’ end	05’	-0.6318	-0.6318
	O3’	-0.5232	-0.1508
central	05’	-0.4954	-0.1428
	P	1.1659	0.3360
	O1P	-0.7761	-0.22365
	O2P	-0.7761	-0.22365
	O3’	-0.5232	-0.1508
3’ end	05’	-0.4954	-0.1428
	P	1.1659	0.3360
	O1P	-0.7761	-0.22365
	O2P	-0.7761	-0.22365
	O3’	-0.6459	-0.6459

Table S2: Averages of helical parameters for all simulated systems.

	Shift	Slide	Rise	Tilt	Roll	Twist
GCC-Na	-0.007	-0.309	3.336	-0.057	3.968	33.094
GCN-Na	-0.003	-0.431	3.326	-0.001	4.404	33.388
GCC-Mg	-0.017	-0.346	3.342	-0.047	4.120	33.106
GCN-Mg	0.001	-0.406	3.327	-0.005	4.176	33.521
ATC-Na	-0.012	-0.797	3.293	0.006	5.682	31.101
ATN-Na	-0.005	-0.777	3.278	-0.033	2.488	33.018
ATC-Mg	0.002	-0.802	3.288	0.070	6.129	31.155
ATN-Mg	0.001	-0.800	3.278	0.001	3.108	32.794