Biophysical Journal, Volume *113*

Supplemental Information

Structure and Dynamics of DNA and RNA Double Helices of CAG and GAC Trinucleotide Repeats

Feng Pan, Viet Hoang Man, Christopher Roland, and Celeste Sagui

Structure and dynamics of DNA and RNA double helices obtained from the CAG and GAC trinucleotide repeats [Supporting Material]

Feng Pan, Viet Hoang Man, Christopher Roland, Celeste Sagui*

Department of Physics, North Carolina State University, Raleigh, NC 27695-8202, USA

* E-mail: sagui@ncsu.edu

1 Definitions of twist and handedness

We use the 3DNA software package(1, 2) to calculate the twist angle of the duplexes. Since non-Watson-Crick base pairs are our main object, we choose the "simple" step parameters, which are "intuitive" for non-Watson-Crick base pairs and were introduced into 3DNA as of v2.3-2016jan01. The regular z-axis defined in 3DNA is used here, which is the average of two base normals, taking into consideration the M-N vs M+N base-pair classification. The "simple" inter-base-pair step parameter calculation uses consecutive C1'-C1' vectors. Since the A-A mismatches may have a anti-syn or syn-syn conformation, the z-axis turns out opposite to the normal one. Thus, one must add 180 degrees to the twist angles involving A(anti)-A(syn) and A(syn)-A(syn) mismatches.

Handedness is a natural choice for investigating left- (H $<$ 0) and right-handed (H $>$ 0)) helical structures, based on a former investigation of the B-Z DNA transition(3). For the double helix, the position of the phosphorus (P) atoms of the backbone phosphate groups was found to be a good choice for the definition of handedness. In brief, the definition of handedness for a portion of DNA/RNA between the base pairs n and m makes use of a sequence of P atoms: P^1_n,P^2_n,P^1_{n+1},P^2_{n+1},...,P^1_m,P^2_m, where the upper index indicates the strand number (1 or 2, labeled arbitrarily) and the lower index indicates the base-pair number labeled in the $5 \rightarrow 3$ direction of strand 1. Note that this definition of handedness is independent of the labeling of the strands. Supplementary Fig. S1 (right) shows the P atoms involved in the definition of handedness of a DNA segment between base pairs n and m; the red and purple balls in this figure are the first and last elements in the sequence. The position of these P atoms then defines the handedness via

$$H(p_1p_2p_3...p_n) = \frac{\overrightarrow{AB}}{|\overrightarrow{AB}|} \times \frac{\overrightarrow{CD}}{|\overrightarrow{CD}|} \cdot \frac{\overrightarrow{EF}}{|\overrightarrow{EF}|}, \tag{1}$$

in which each p_i is a point in the sequence discussed above, and

$$H(ABCD) = \sum_{i=1}^{n-3} H(p_ip_{i+1}p_{i+2}p_{i+3}). \tag{2}$$

In this last equation, the points A, B, C, D define the vectors $\overrightarrow{AB}$ and $\overrightarrow{CD}$ and the midpoints of these vectors, called E and F , in turn form the vector $\overrightarrow{EF}$. Supplementary Fig. S1 illustrates this definition for the first term of the sum in the relation (Eq. (1)). The cross product of the unit vectors of $\overrightarrow{AB}$ and

$\overrightarrow{CD}$ defines the (purple) vector whose dot product with the unit vector of $\overrightarrow{EF}$ forms the first term of the sum in the definition of handedness.

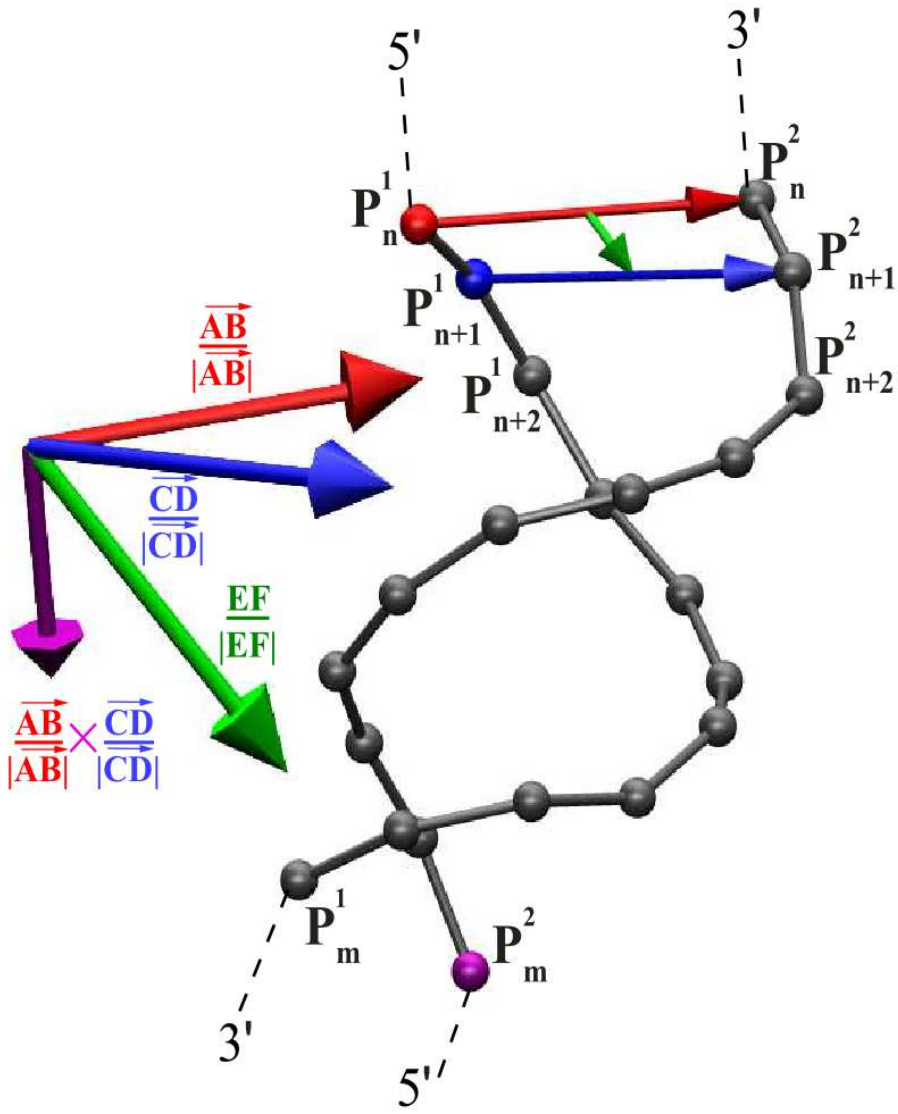

Figure S1: **Definition of handedness**

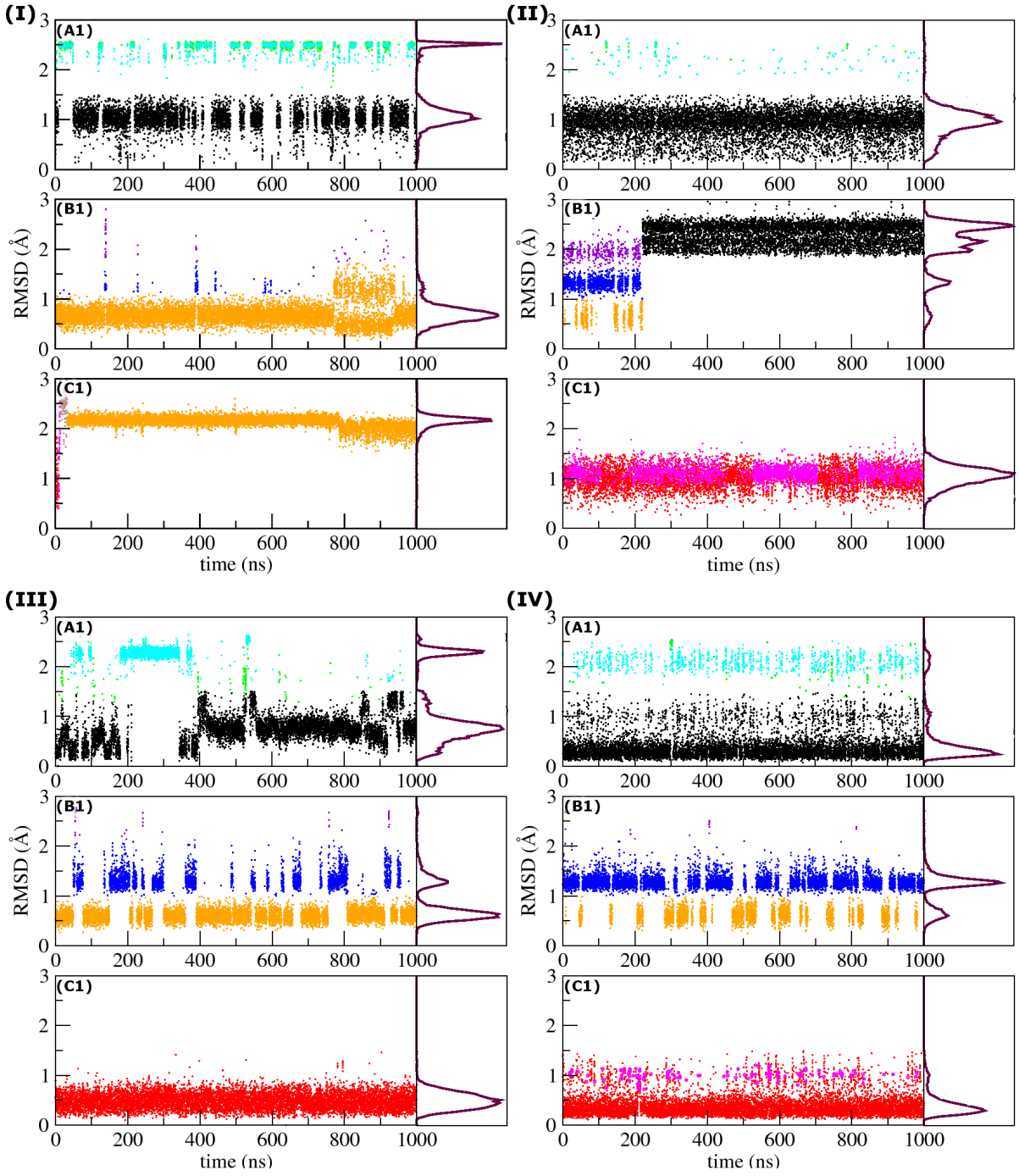

Figure S2: **RMSD for the single internal mismatch A_5-A_{14} during 1 μs simulations.** (I) RNA-CAG (II) RNA-GAC (III) DNA-CAG (IV) DNA-GAC Conformations are color coded to agree with the mismatch conformations in Fig. 2. Initial conformations for each of the three panels in a column are as follows. Top: anti-anti (conformation A1 in Fig. 2); Middle: anti-syn (B1); Bottom: syn-syn (C1).

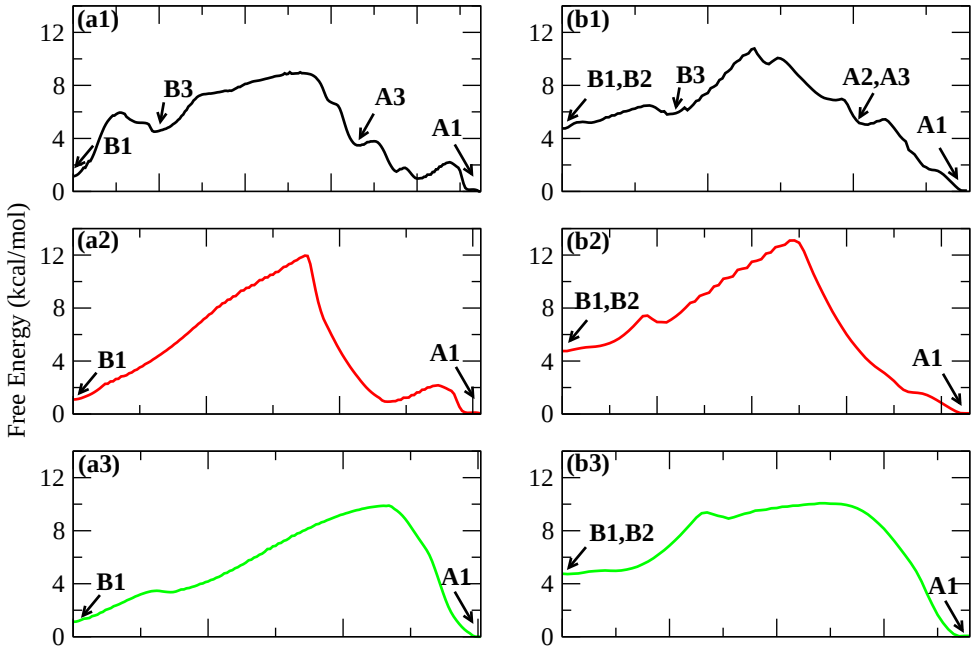

Figure S3: **Possible paths for the B1 → A1 transition on the (Ω,χ_{14}) free energy maps.** (a) RNA-CAG; (b) RNA-GAC. Different colors indicate different paths on the (Ω,χ_{14}) maps.

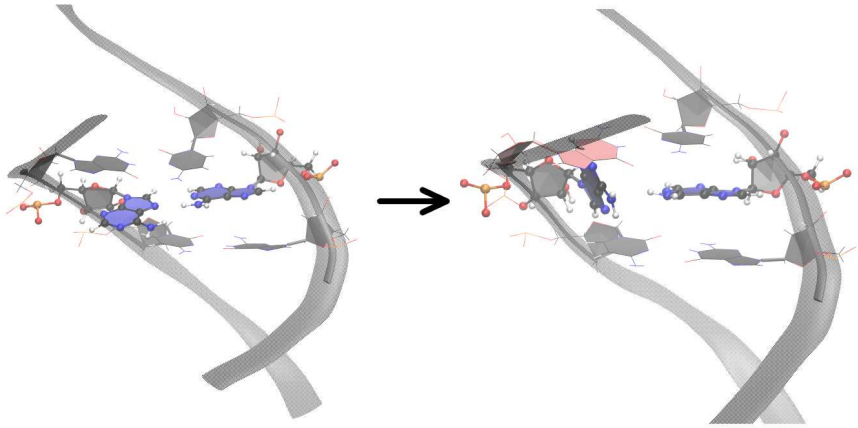

Figure S4: **syn→ anti rotation in a clockwise direction.** This rotation results in a clash betwen A14 and the pink G13.

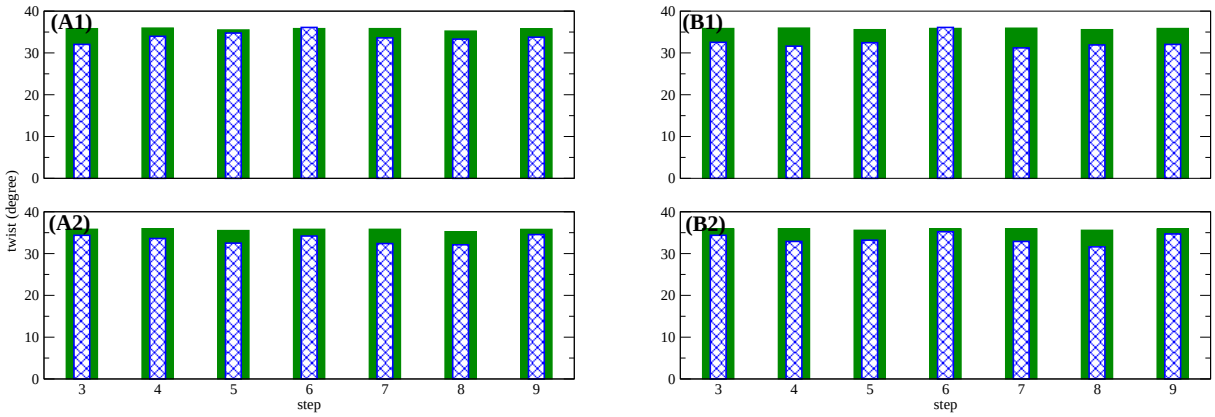

Figure S5: **Simple twist of the middle eight basepairs in DNA with initial anti-syn and syn-syn mismatch conformations.** (A1) anti-syn $(CAG)_4$; (A2) syn-syn $(CAG)_4$; (B1) anti-syn $(GAC)_4$; (B2) syn-syn $(GAC)_4$. Green bars show the initial values. Blue bars show the average value taken from the final 200ns.

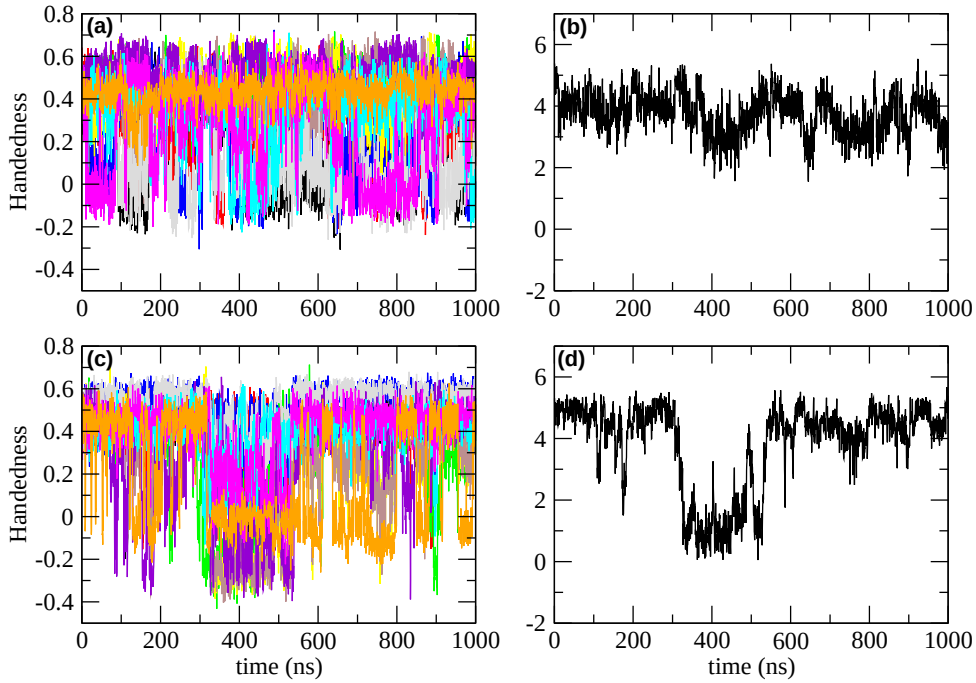

Figure S6: **Handedness of the middle six basepairs in DNA with initial anti-anti mismatch conformations.** Top: $(CAG)_4$; Bottom: $(GAC)_4$. The left column shows local handedness, with different colors representing different turns. The right column shows the total handedness.

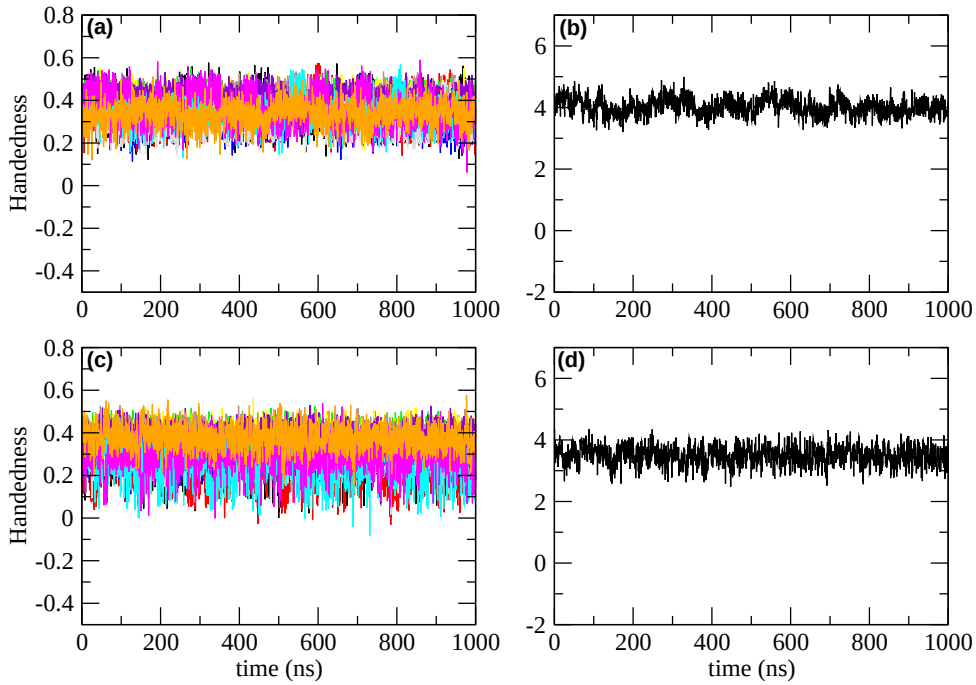

Figure S7: **Handedness of the middle six basepairs in RNA with initial anti-anti mismatch conformations.** Top: $(CAG)_4$; Bottom: $(GAC)_4$. The left column shows local handedness, with different colors representing different turns. The right column shows the total handedness.

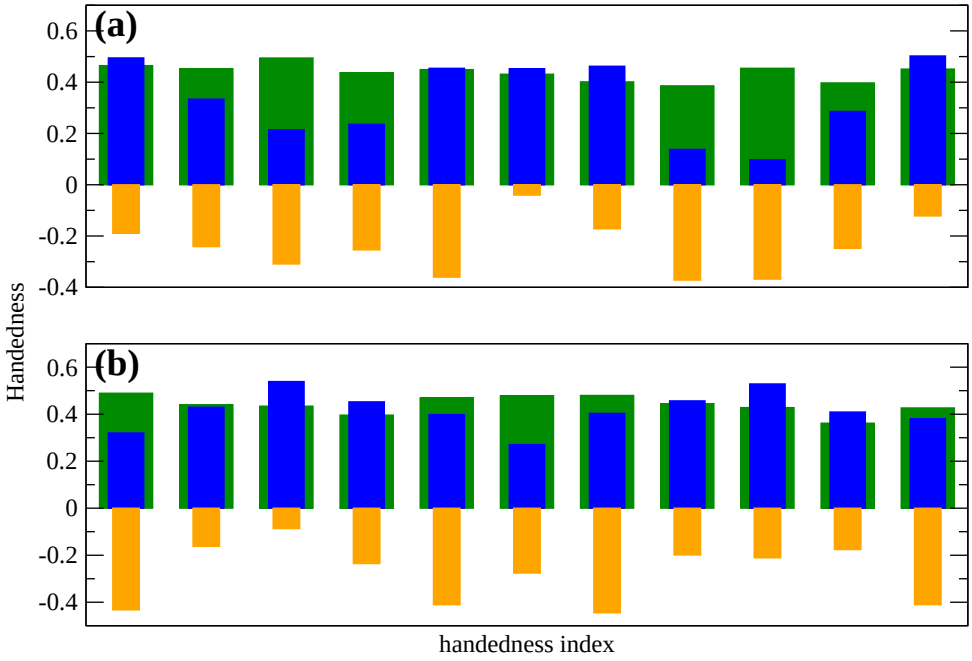

Figure S8: **Average local handedness of the middle six basepairs in DNA with initial anti-anti mismatch conformations.** (a) $(CAG)_4$; (b) $(GAC)_4$. Green bars show initial values taken from the first 10ns. Blue bars show average values taken from the final 200ns. Orange bars show the minimum value throughout the run. The index is specified by the P atoms of different residues, and acts as a sliding window through successive four residues. Thus, the first index is defined by $P_3P_{16}P_4P_{15}$ using Eq. (1) where the lower number represents residue index, the second is defined by $P_{16}P_4P_{15}P_5$, etc.

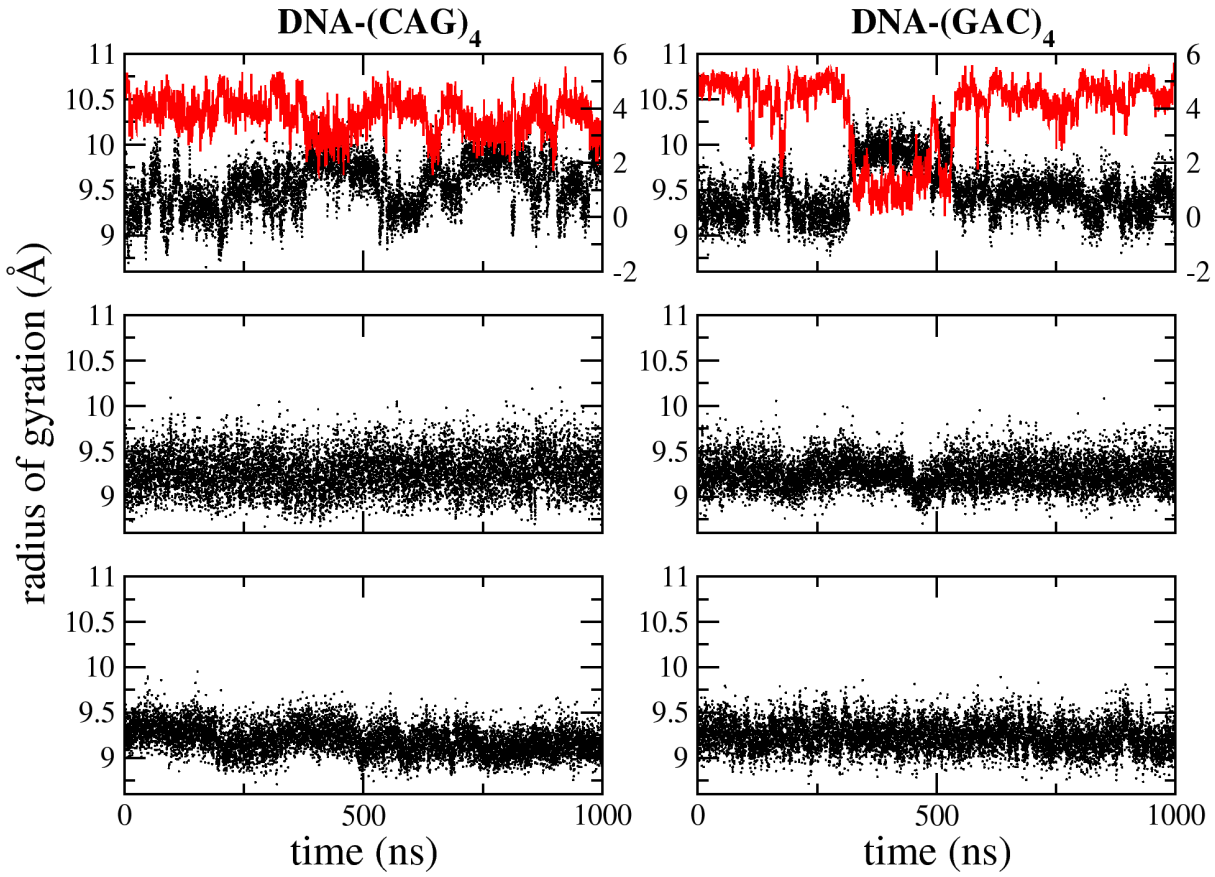

Figure S9: **Radius of gyration during 1 μs simulations** Considered here are the residues 4-9 on one strand and the complementary residues 16-21 on the other. Left column: DNA-$(CAG)_4$; Right column: DNA-$(GAC)_4$. Top: anti-anti; Middle: anti-syn; Bottom: syn-syn. The red lines show the total handedness as comparison, with its scale shown on the right side.

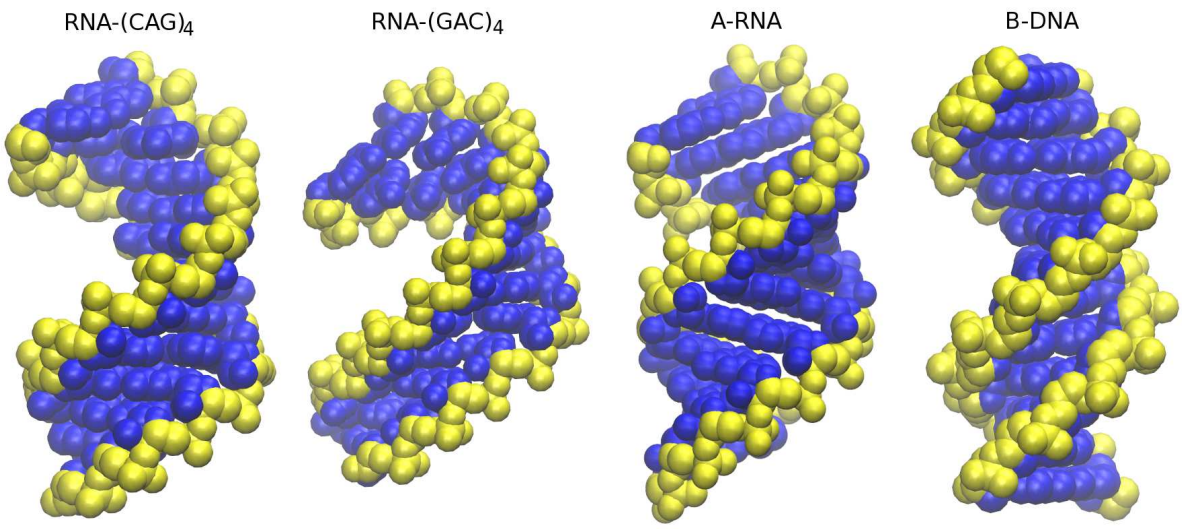

Figure S10: **Comparison between RNA-$(CAG)_4$, RNA-$(GAC)_4$ and standard B-DNA, A-RNA helices in ball model.** The RNA-$(CAG)_4$ and RNA-$(GAC)_4$ structures are determined by choosing the lowest combined RMSD value of the middle two AA mismatches, with respect to the anti-anti minimum A1.

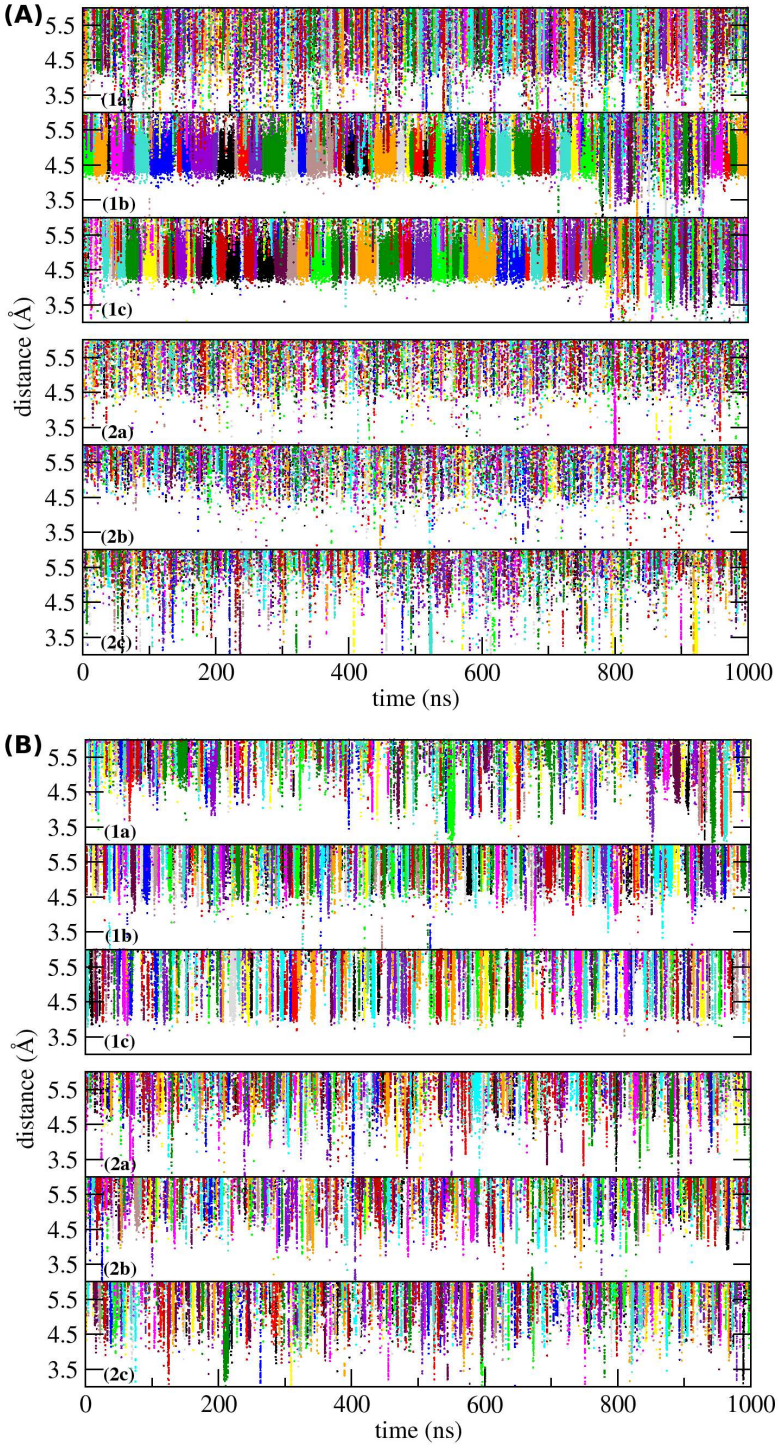

Figure S11: **Distance between Na^+ ions and the center of mass of the A-A mismatches.** The single mismatch duplexes are: (A) RNA-CAG (top) and RNA-GAC (bottom); (B) DNA-CAG (top) and DNA-GAC (bottom). For each duplex, the initial conformations for the top, middle and bottom panels are anti-anti, anti-syn, and syn-syn respectively. Different colors represent different ions to show the binding time for individual ions.

	RNA-$(CAG)_4$(anti-anti)	RNA-$(CAG)_4$(anti-syn)	RNA-$(GAC)_4$(anti-anti)	RNA-$(GAC)_4$(anti-syn)	A-RNA	B-DNA
major groove width (Å) (direct PP distance)	23.1±1.8	16.9±1.2	24.8±1.9	18.8±3.1	15.2	17.3
minor groove width (Å) (direct PP distance)	16.6±0.6	18.1±0.5	16.3±0.6	19.0±1.0	18.8	11.5
inclination (degree)	5.6±4.5	14.4±3.7	4.6±5.0	16.3±4.6	19.0	-5.5

Table S1: **Major/minor groove width and basepair inclination for RNA-$(CAG)_4$, RNA-$(GAC)_4$ and standard B-DNA, A-RNA.** The results of RNA-$(CAG)_4$ and RNA-$(GAC)_4$ are taken from the middle five base pairs and averaged through 50ns.

Supporting References

[1] Lu, X. J., and W. K. Olson, 2003. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* 31:5108 – 5121.

[2] Lu, X., and W. K. Olson, 2008. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat Protoc* 3:1213–1227.

[3] Moradi, M., V. Babin, C. Roland, and C. Sagui, 2012. Reaction path ensemble of the B-Z-DNA transition: a comprehensive atomistic study. *Nucleic Acids Res.* 41:33–43.