Novel eGZ-motif formed by regularly extruded guanine bases in a left-handed Z-DNA helix as a major motif behind CGG trinucleotide repeats: Supporting Information

Ashkan Fakharzadeh, Jiahui Zhang, Christopher Roland, and Celeste Sagui*

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

E-mail: sagui@ncsu.edu

Biased molecular dynamics simulation details

Here, we further outline our simulation protocols for obtaining free energy landscales via the Adaptively Biased Molecular Dynamics (ABMD) method[1,2] and the work from Steered Molecular Dynamics (SMD)[3] simulations. The initial coordinates for the 5'-GC-(GGC)-GC-3' mismatched DNA helices obtained via BIOVIA Discovery Studio were minimized and equilibrated according to the following protocol. First, we minimized the energy of the initial conformations at zero temperature. Subsequently, the temperature was gradually raised from 0 to 300 K over 50 ps run with a 1 fs time step, followed by another 50 ps run where the temperature was still at 300 K. Then a 300 ps run was used to gradually reduce the restraining harmonic constants on the DNA helix. In addition to the hydrogen bonds in the Watson-Crick base pairs, the χ (and ϕ dihedrals in relevent simulations) were restrained to preserve the initial conformations. In order to speed up the biased simulation, we first used the generalized Born[4,5] implicit solvent model (PBrandii mbondi) with the OL15[6] and

BSC1[7] AMBER force fields with a collision frequency of 1 ps^{-1} along with a 999 Å cutoff for the nonbonded interactions. After equilibration, SMD[8] runs were used to flip the GG mismatches outside of the helical core. These runs made use of a steering force constant of 200 $kcal/mol(rad^2)$ over 200 ns runs using the (ϕ_4, ϕ_{11}) pair as collective variables. For the free energy calculations, we randomly selected 8 configurations, equilibrated and used them as the initial configurations for the ABMD runs. The ABMD runs made use of the Multiple Walker algorithm[9,10] with 8 walkers and the Well-Tempered (WT) algorithm[11] with a 0.2 radian resolution for the reaction coordinates, a pseudo temperature of 10,000 K and a flooding timescale of 2 ps, all at 300 K. Each run lasted at least for 1 μs. During the above SMD and ABMD steps, in addition to the χ restrains, constraints of 1 kcal/mol on hydrogen bonds of W-C base pair, as well as root-mean-squared deviation of phosphate atoms of CG W-C base pair were used to avoid large-scale twisting of the whole structure during the long simulations. These simulations provided a rough estimate of the biasing potential. We then followed this up with at least 300 ns WT-simulation (parameters $\tau_f = 1$ ps, $4\Delta\xi = 0.2$ rad ($\sim 10^\circ$), pseudotemperature 10,000 K) at constant volume explicit solvent and 300 K. Convergance of free energy landscapes were confirmed by noting that both the position and differences in the free energy values of the minima remain approximately constant as further ABMD cycles are performed. The final biasing potential was processed by the nfe-umbrella-slice tool of AmberTools[12] to obtain the two-dimensional free energy landscapes.

The free energy for flipping a single base in $GGC8_{ZZ,in}$ was calculated again using well tempered ABMD in low (0.15 M), and high (5 M) salt concentration. A 1-dimensional center-of-mass pseudo-dihedral angle, CPD,[13] was chosen as a collective variable defined by the center of mass of two flanking base pairs, the flanking sugar groups, and the five-member ring of the flipping G. The choice of this CV was crucial in calculating free barriers and their convergence.[13] The initial structure came from the final structure of $GGC8_{ZZ,in}$ equilibrium molecular dynamics using OL15 FF, which was employed throughout this ABMD as well. The ABMD runs made use of parameters $4\Delta\xi = 0.2$ radian, $T' = 10,000$ K and $\tau = 2$ ps for

a total simulation length of at least 700 ns per system at constant volume explicit solvent and 300 K.The 1-dimensional free energy profile was reconstructed by inverting biasing potential energy via nfe-umbrella-slice tool. The estimated free energies are relatively accurate but the absolute values may be associated with inaccuracies given the typical force field and sampling imperfections.

Definition of handedness

Handedness (H) represents a useful way to investigate left- and right-handed helical structures. The definition of handedness is based on a former investigation of the B-Z DNA transition.[14] Handedness in right-handed helical turns is positive, and in left-handed helical turns it is negative. The overall handednesss of a helix is obtained by summing over all turns and quantitatively quite close to the overall helical twist. 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: $P_n^1, P_n^2, P_{n+1}^1, P_{n+1}^2, ..., P_m^1, P_m^2$, 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. S10 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

$$H(p_1, p_2, p_3, ..., p_n) = \sum_{i=1}^{n-3} H(p_i p_{i+1} p_{i+2} p_{i+3}) \tag{1}$$

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

$$H(ABCD) = \frac{\overrightarrow{AB}}{|\overrightarrow{AB}|} \times \frac{\overrightarrow{CD}}{|\overrightarrow{CD}|} \cdot \frac{\overrightarrow{EF}}{|\overrightarrow{EF}|} \tag{2}$$

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 Figure S10 left illustrates this definition for the first term of the sum in the above relation. The cross product of the unit vectors of $\overrightarrow{AB}$ and $\overrightarrow{CD}$ defines the (purple) vector (direction of handedness) whose dot product with the unit vector of $\overrightarrow{EF}$ forms the first term of the sum in the definition of handedness.

Hydrogen Bond Analysis

Hydrogen bond (Hbond) analysis was performed using the CPPTRAJ program of the AMBER simulation package. A 3Å distance and a 135° angle cutoff were considered for the calculations. The Tables S3 to S6 list the observed Hbonds between the mismatches; bonds with fractions less than 10% are not listed. The last column for the *out* configurations indicates the type of Hbond.

References

(1) Babin, V.; Roland, C.; Sagui, C. Adaptively Biased Molecular Dynamics for Free Energy Calculations. *Journal of Chemical Physics* **2008**, *128*, 134101.

(2) Babin, V.; Karpusenka, V.; Moradi, M.; Roland, C.; Sagui, C. Adaptively biased molecular dynamics: an umbrella sampling method with a time dependent potential. *Inter. J. Quantum Chem.* **2009**, *109*, 3666–3678.

(3) Izrailev, S.; Stepaniants, S.; Isralewitz, B.; Kosztin, D.; Lu, H.; Molnar, F.; Wrig-

gers, W.; Schulten, K. *Steered molecular dynamics. Computational Molecular Dynamics: Challenges, Methods, Ideas*; Springer-Verlag: Berlin, Germany, 1998; pp 39–65.

(4) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004**, *55*, 383–394.

(5) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. *The Journal of Physical Chemistry B* **2000**, *104*, 3712–3720.

(6) Zgarbová, M.; Sponer, J.; Otyepka, M.; Cheatham III, T. E.; Galindo-Murillo, R.; Jurecka, P. Refinement of the sugar–phosphate backbone torsion beta for amber force fields improves the description of z-and b-dna. *Journal of chemical theory and computation* **2015**, *11*, 5723–5736.

(7) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portella, G. e. Parmbsc1: a refined force field for DNA simulations. *Nature methods* **2016**, *13*, 55.

(8) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophysical Journal* **1997**, *72*, 1568 – 1581.

(9) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. *The Journal of Physical Chemistry B* **2006**, *110*, 3533–3539.

(10) Minoukadeh, K.; Chipot, C.; Lelièvre, T. Potential of Mean Force Calculations: A Multiple-Walker Adaptive Biasing Force Approach. *Journal of Chemical Theory and Computation* **2010**, *6*, 1008–1017.

(11) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys Rev Lett* **2008**, *100*, 020603.

(12) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Ceritti,; S., D.; Coetz, A.; Greene, D. e. *AMBER 20*; University of California: San Francisco, 2020.

(13) Song, K.; Campbell, A. J.; Bergonzo, C.; de los Santos, C.; Grollman, A. P.; Simmerling, C. An Improved Reaction Coordinate for Nucleic Acid Base Flipping Studies. *Journal of Chemical Theory and Computation* **2009**, *5*, 3105–3113.

(14) Moradi, M.; Babin, V.; Roland, C.; Sagui, C. Reaction path ensemble of the B-Z-DNA transition: a comprehensive atomistic study. *Nucleic Acids Res.* **2013**, *41*, 33–43.

Table S1: Summary of the radius of gyration data for the different helical duplexes

Structure	Initial	Final	Average (SD)
GGC8$_{BZ,out}$ OL15	20.34	19.64	20.29(0.64)
GGC8$_{BZ,out}$ BSC1	19.93	20.10	19.70(0.52)
GGC8$_{ZZ,out}$ OL15	25.66	19.18	19.64(0.63)
GGC8$_{ZZ,out}$ BSC1	22.73	25.94	25.15(1.0)
GGC8$_{ZZ,in}$ OL15	31.11	30.59	30.41(0.45)
GGC8$_{ZZ,in}$ BSC1	29.46	30.27	29.91(0.70)
CGG4$_{ZZ,in}$ OL15	15.08	18.68	17.87(0.65)
CGG4$_{ZZ,in}$ BSC1	15.84	15.02	15.29(0.61)
GGC4$_{ZZ,alt}$ BSC1	11.56	11.89	11.88(0.15)
GGC4$_{ZZ,alt}$ OL15	11.71	12.03	11.90(0.22)

All values given are in Å. The radius of gyration has been measured using P atoms. The initial, final, and average values correspond to the last 1 μs of simulation time, and SD the standard deviation.

Table S2: π-π stacking parameters between the stacked mismatches over the last 200ns

Structure	Force Field	Residues	Ave. Distance (Å)	Ave. Angle(°)	Ave. VDW Energy (kcal/mol)
$GGC8_{BZ,out}$	OL15	G14::G41	5.41(0.90)	38.59(46.36)	-3.14(1.03)
$GGC8_{ZZ,out}$	BSC1	G20::G32	4.35(0.71)	21.22(49.10)	-4.45(1.07)
$GGC8_{ZZ,out}$	BSC1	G17::G26	4.15(0.62)	28.08(38.18)	-3.65(1.24)
$GGC4_{ZZ,alt}$	BSC1	G8::G24	3.56(0.16)	7.35(6.08)	-5.34(0.56)
$GGC4_{ZZ,alt}$	OL15	G8::G24	4.00(0.48)	14.37(59.12)	-5.37(0.69)
$GGC4_{ZZ,alt}$	OL15	G15::G18	3.88(0.35)	15.29(13.97)	-3.75(0.78)

The numbers represent the average distance, angle, and Van der Waals energy between $\pi - \pi$ stacked bases of the mismatches for the different GGC_{out} and GGC_{alt} helices. All the calculations were obtained from the last 200 ns of the simulations.

Table S3: H-bonds formed by the mismatches in the inner-GGC8 *out* configurations

Structure	Acceptor	Donor	Fraction	Type
$GGC8_{BZ,out}$ BSC1	C_15:OP1	G_17:N2	0.92	Backbone with the closest cytosine
	C_15:OP1	G_17:N1	0.15	
	C_30:OP1	G_32:N2	0.52	
	C_30:OP2	G_32:N1	0.40	
	C_30:OP1	G_32:N1	0.18	
	C_30:OP2	G_32:N2	0.18	
	C_36:OP1	G_38:N2	0.46	
	C_36:OP1	G_38:N1	0.32	
	G_35:N7	G_8:N1	0.24	Mismatches from two strands
	G_14:O3'	G_17:N1	0.20	Backbone with another mismatch
	G_38:N7	G_35:N2	0.12	Mismatches from the same strand (weak, disappears)
$GGC8_{BZ,out}$ OL15	C_6:OP1	G_8:N1	0.55	Backbone with the closest cytosine
	C_6:OP2	G_8:N2	0.45	
	C_6:OP2	G_8:N1	0.26	

	C_15:OP2	G_17:N1	0.92	
	G_14:O3'	G_17:N2	0.55	
	C_30:OP1	G_32:N1	0.44	
	C_30:OP2	G_32:N1	0.41	
	C_30:OP2	G_32:N2	0.35	
	C_39:OP2	G_41:N2	0.35	
	C_39:OP2	G_41:N1	0.27	
	G_11:O6	G_35:N2	0.22	Mismatches from two strands
	G_14:O3'	G_17:N2	0.55	Backbone with another mismatch
	G_17:OP1	G_38:N2	0.12	Backbone with a mismatch from the other strand (last 200ns)
	G_14:O5'	G_41:N1	0.12	Base stacking (backbone, last 200ns)
	G_38:O3'	G_14:N1	0.18	Base stacking (backbone, 200ns-800ns)
$GGC8_{ZZ,out}$ BSC1	C_6:OP2	G_8:N1	0.50	Backbone with the closest cytosine
	C_6:OP2	G_8:N2	0.41	
	C_6:OP1	G_8:N2	0.19	
	C_6:OP1	G_8:N1	0.12	
	C_12:OP2	G_14:N1	0.35	
	C_12:OP2	G_14:N2	0.18	
	C_33:OP2	G_35:N1	0.38	
	C_33:O5'	G_35:N2	0.36	
	C_33:OP1	G_35:N1	0.25	
	C_33:OP1	G_35:N2	0.10	
	C_39:OP1	G_41:N1	0.33	
	C_39:OP1	G_41:N2	0.31	
	C_39:OP2	G_41:N1	0.30	
	C_39:O5'	G_41:N2	0.26	
	C_39:OP2	G_41:N2	0.11	

	G_11:O6	G_38:N1	0.47	Flipped in
	G_11:O6	G_38:N2	0.21	
	G_35:OP1	G_21:N2	0.24	Base stacking (backbone)
	G_35:OP1	G_21:N1	0.20	
	G_21:OP1	G_35:N2	0.23	
$GGC8_{ZZ,out}$ OL15	C_9:O2	G_11:N1	0.37	Backbone with the closest cytosine
	C_9:O2	G_11:N2	0.24	
	C_39:OP2	G_14:N1	0.20	Backbone with a cytosine from the other strand
	C_39:OP1	G_14:N2	0.12	
	G_41:OP1	G_14:N2	0.11	Backbone with a mismatch from the other strand
	G_17:OP1	G_14:N2	0.11	Backbone with another mismatch

Table S4: H-bonds formed by the mismatches for the inner-GGC8 *in* configurations

Structure	Acceptor	Donor	Fraction
$GGC8_{ZZ,in}$ BSC1	G_41:O6	G_8:N1	0.38
	G_41:O6	G_8:N2	0.23
	G_8:O6	G_41:N1	0.36
	G_8:O6	G_41:N2	0.22
	G_38:O6	G_11:N1	0.46
	G_38:O6	G_11:N2	0.28
	G_11:O6	G_38:N1	0.28
	G_11:O6	G_38:N2	0.17
	G_35:O6	G_14:N1	0.28
	G_35:O6	G_14:N2	0.17
	G_14:O6	G_35:N1	0.45
	G_14:O6	G_35:N2	0.27
	G_32:O6	G_17:N1	0.28

	G_32:O6	G_17:N2	0.17
	G_17:O6	G_32:N1	0.46
	G_17:O6	G_32:N2	0.28
$GGC8_{ZZ,in}$ OL15	G_41:O6	G_8:N1	0.46
	G_41:O6	G_8:N2	0.25
	G_8:O6	G_41:N1	0.30
	G_8:O6	G_41:N2	0.16
	G_38:O6	G_11:N1	0.28
	G_38:O6	G_11:N2	0.15
	G_11:O6	G_38:N1	0.47
	G_11:O6	G_38:N2	0.26
	G_35:O6	G_14:N1	0.40
	G_35:O6	G_14:N2	0.21
	G_14:O6	G_35:N1	0.35
	G_14:O6	G_35:N2	0.19
	G_32:O6	G_17:N1	0.37
	G_32:O6	G_17:N2	0.20
	G_17:O6	G_32:N1	0.38
	G_17:O6	G_32:N2	0.20

Table S5: H-bonds formed by the mismatches in the $CGG4_{ZZ,in}$ configurations

Structure	Acceptor	Donor	Fraction
$CGG4_{ZZ,in}$ BSC1	G_3:O6	G_26:N1	0.62
	G_3:O6	G_26:N2	0.48
	G_6:O6	G_23:N1	0.66
	G_6:O6	G_23:N2	0.39
	G_12:O6	G_17:N1	0.62

	G_12:O6	G_17:N2	0.47
CGG4$_{ZZ,in}$ OL15	G_3:O6	G_26:N1	0.53
	G_3:O6	G_26:N2	0.32
	G_23:O6	G_6:N2	0.22
	G_6:O6	G_23:N2	0.11
	G_20:O6	G_9:N2	0.21
	G_9:O6	G_20:N2	0.18
	G_9:O6	G_20:N2	0.18
	G_17:O6	G_12:N2	0.15
	G_12:O6	G_17:N2	0.20

Table S6: H-bonds formed by single guanines in the GGC4$_{ZZ,alt}$ configurations

Structure	Acceptor	Donor	Fraction	Type
GGC4$_{ZZ,alt}$ BSC1	G_24:O4'	G_8:N2	0.13	Base stacking (backbone)
GGC4$_{ZZ,alt}$ OL15	G_24:O5'	G_8:N1	0.18	
	G_24:OP2	G_8:N2	0.17	
	G_24:OP1	G_8:N2	0.14	
	G_18:OP2	G_15:N1	0.77	
	G_18:O5'	G_15:N2	0.75	
	G_13:OP1	G_21:N2	0.78	Backbone with a guanine from the other strand
	C_12:OP2	G_21:N2	0.31	Backbone with a cytosine from the other strand
	C_12:OP2	G_21:N1	0.17	
	C_12:OP1	G_21:N1	0.15	

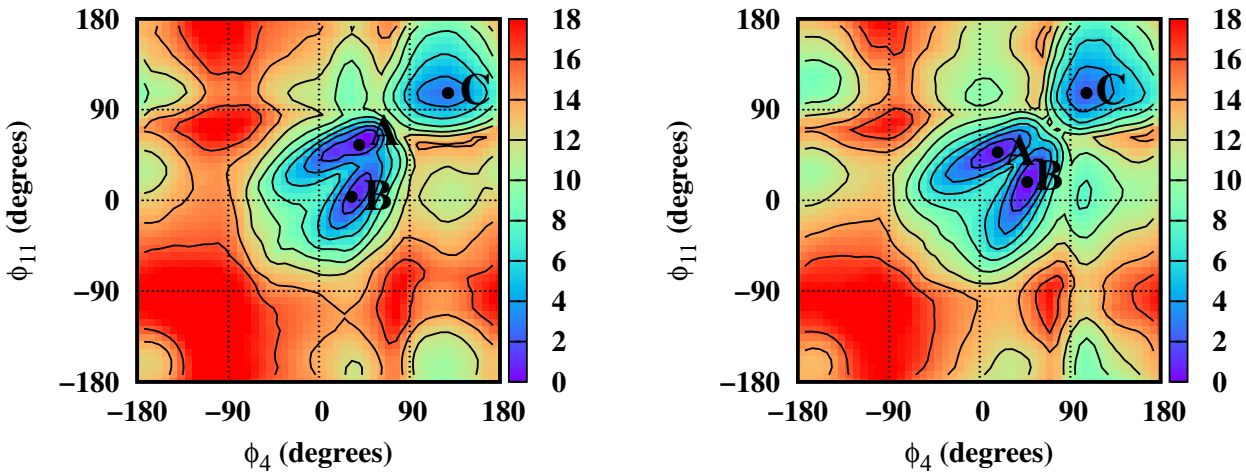


Figure S1: Two-dimensional free energy maps for a single GG mismatch with the center-of-mass pseudo-dihedral angle Φ as collective variables for both residues 4 and 11. The GG mismatches are in B-form (thus, it is a BZ-junction) and restrained to anti-syn (left) and syn-syn (right) conformations. The simulations used the BSC1 force field in water using the GB/SA solvation model. The letters (A, B, C) represent the local minima.

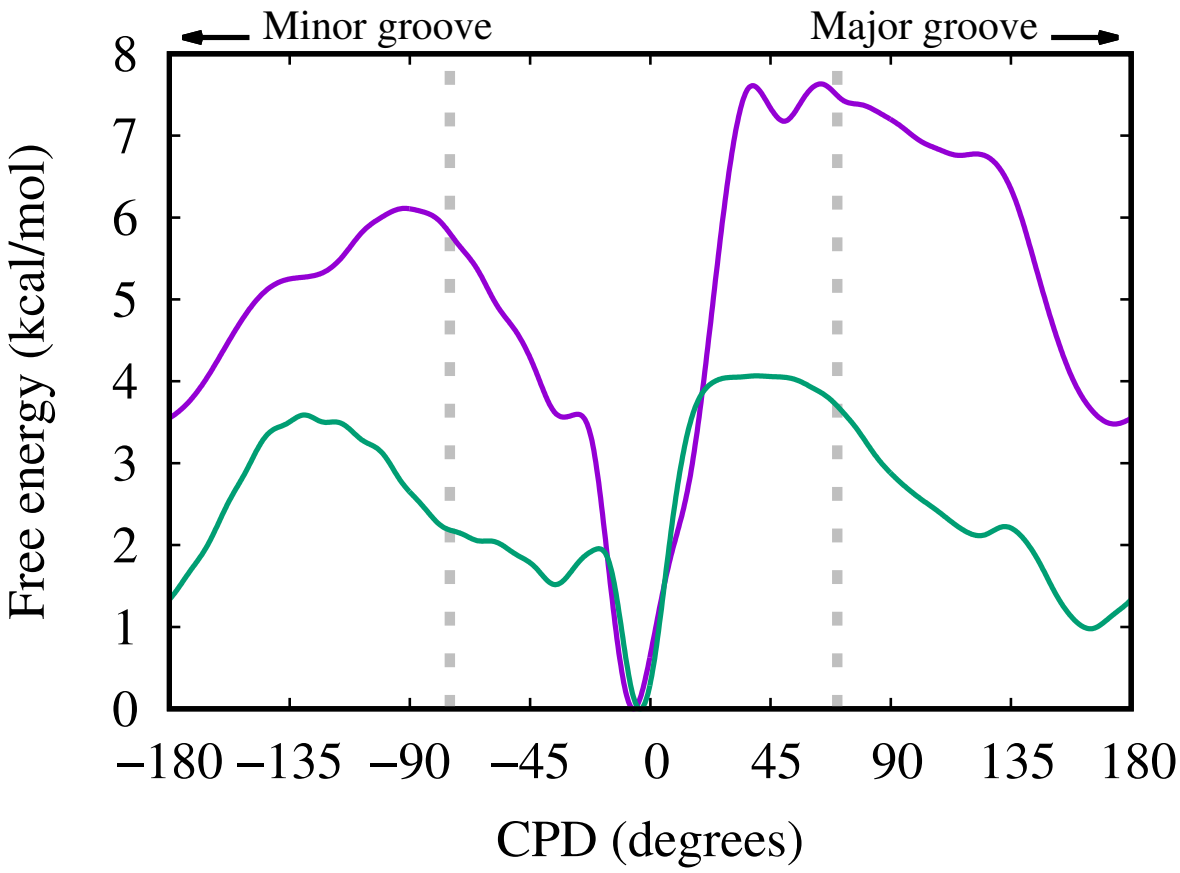


Figure S2: Free energy for the base flipping of single mismatched G in GGC8$_{ZZ,in}$ in low (0.15 M, purple) and high (5 M, green) NaCl salt concentration. The collective variable is the modified center-of-mass pseudo-dihedral angle, CPD,[13] where positive and negative values reflect flipping into the major and minor grooves, respectively. The area between gray dashed lines, -75° to +75°, corresponds to the inner helical core.

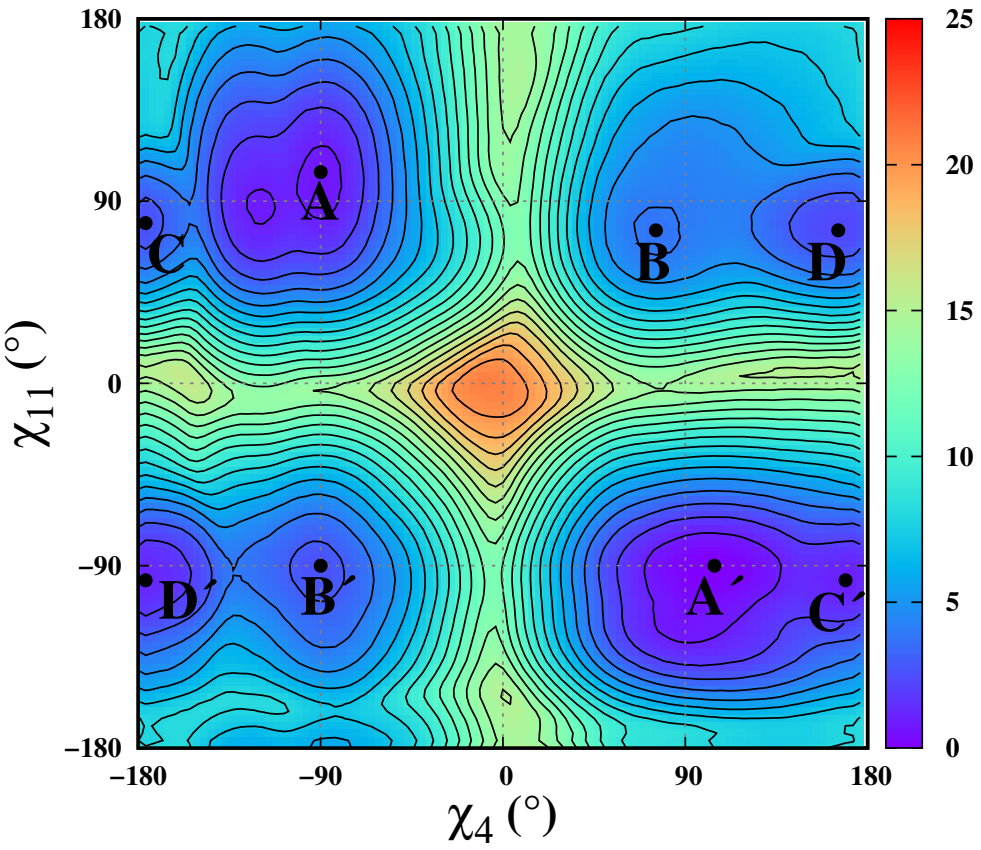


Figure S3: Two-dimensional BSC1 free energy map for a single GG mismatch with the glycosidic torsion angle χ as collective variable for both residues 4 and 11. The GG mismatches were restrained to remain inside the helical core. The letters (A, B, C, D) represent the local minima.

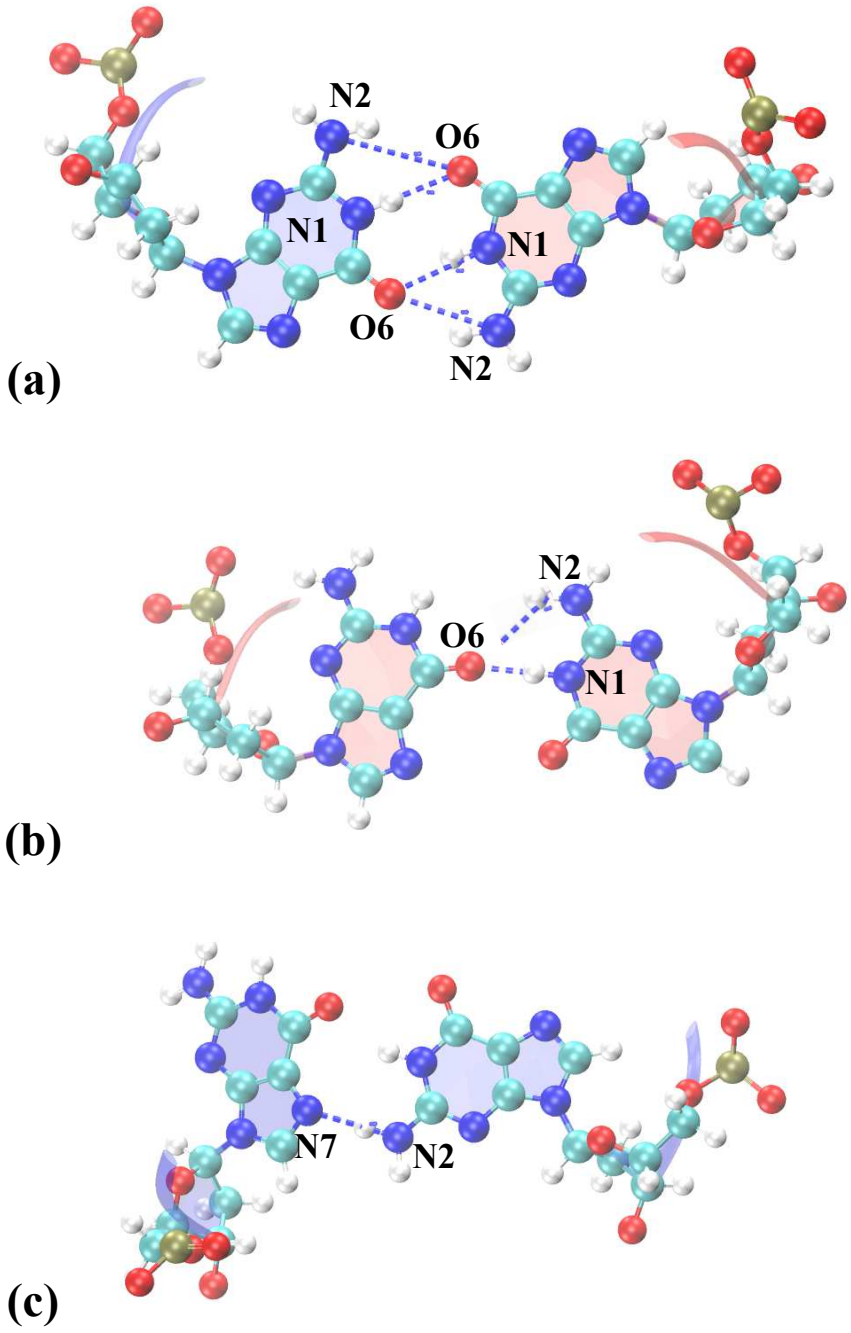
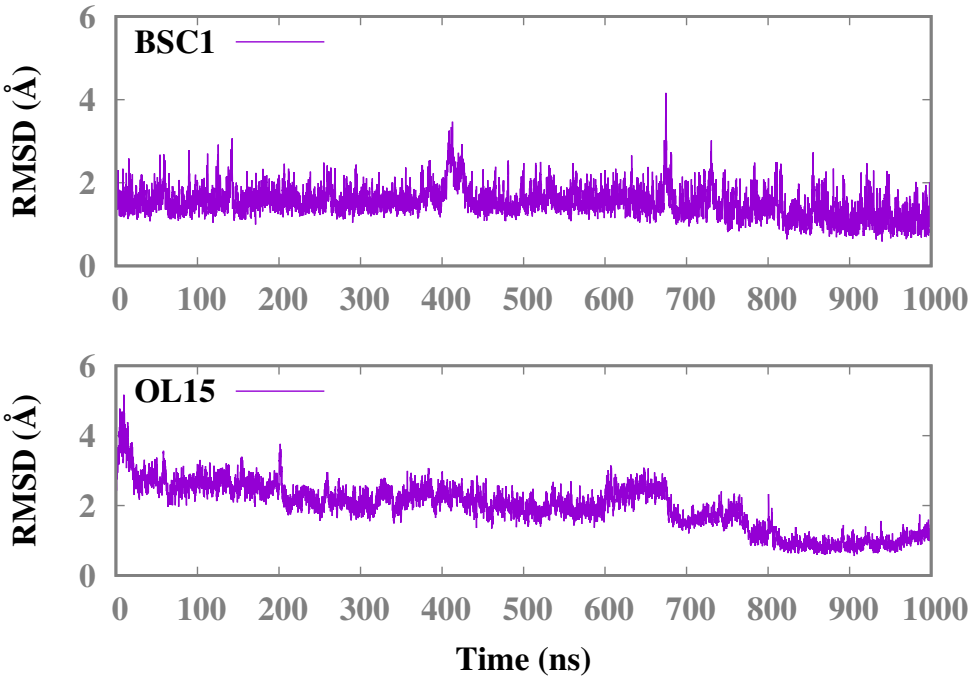
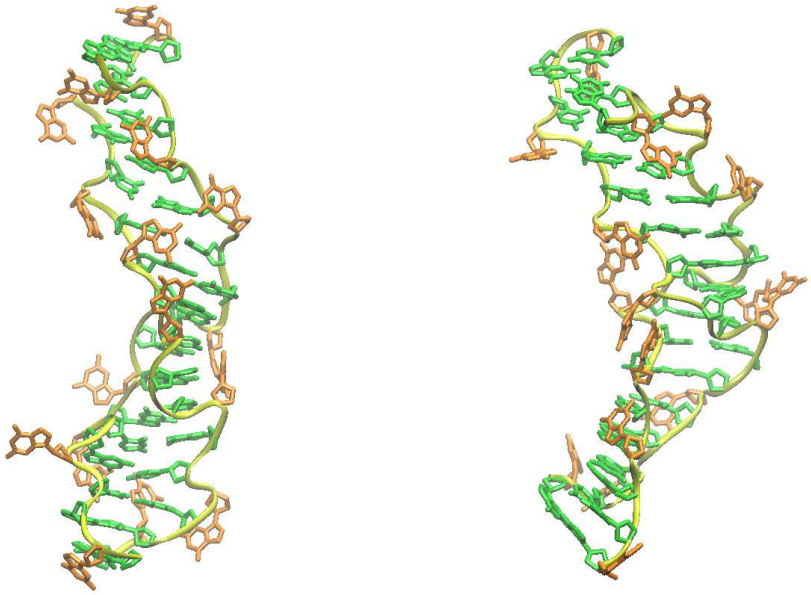


Figure S4: Hydrogen bonds for the single mismatched GG in the homoduplex d(5'-GC-G**G**C-GC-3'). (a) Two bonds for the anti-syn conformation, O6...N1 or O6...N2; (b) Single bond for the syn-syn conformation, O6...N1 or O6...N2; (c) Single bond for the rare anti-anti conformation (not a minimum), N2...N7. Anti bases are shaded blue, and syn bases, red.

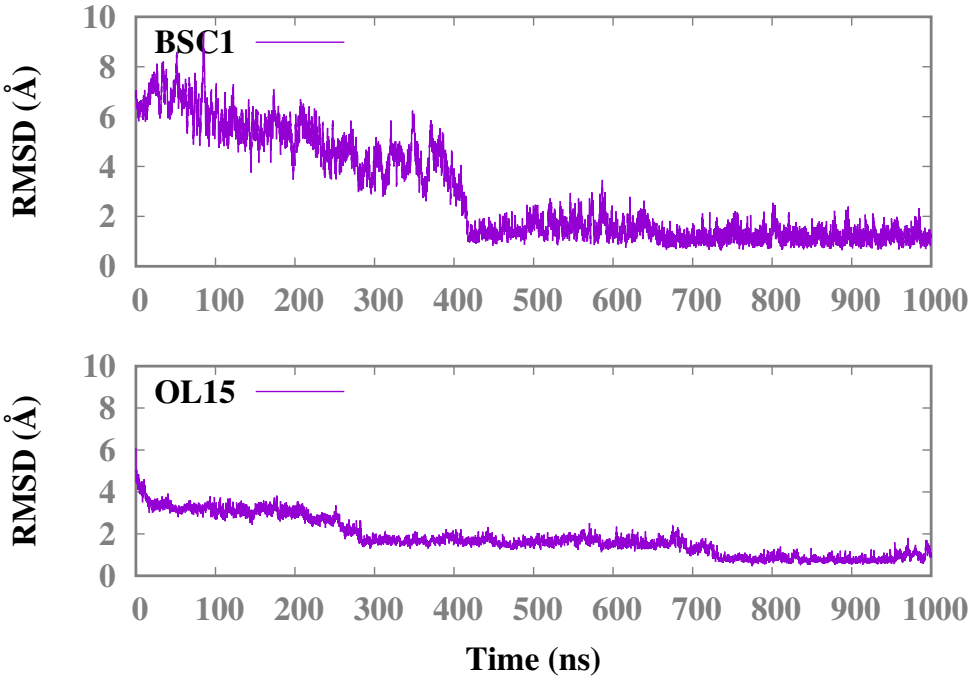


(a)

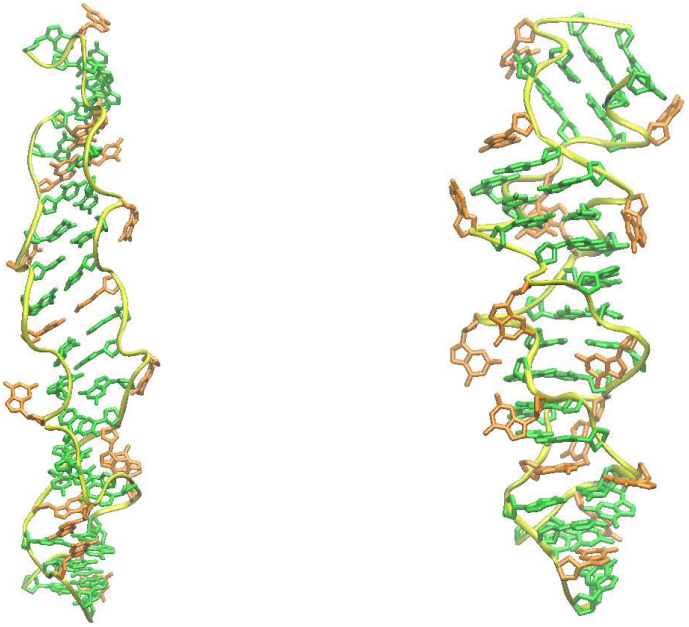


(b)

Figure S5: RMSD as a function of time, and a snapshot of GGC8$_{BZ,out}$. (a) RMSD as function of time (ns) with respect to the average frame over the last 200 ns; only data for the last 1 μs is shown. (b, left) A snapshot of GGC8$_{BZ,out}$ BSC1 after 2 μs, and (b, right) a snapshot of GGC8$_{BZ,out}$ OL15 after 2 μs, The mismatches, WC bases, and backbone are shown by colors brown, green, and yellow respectively.

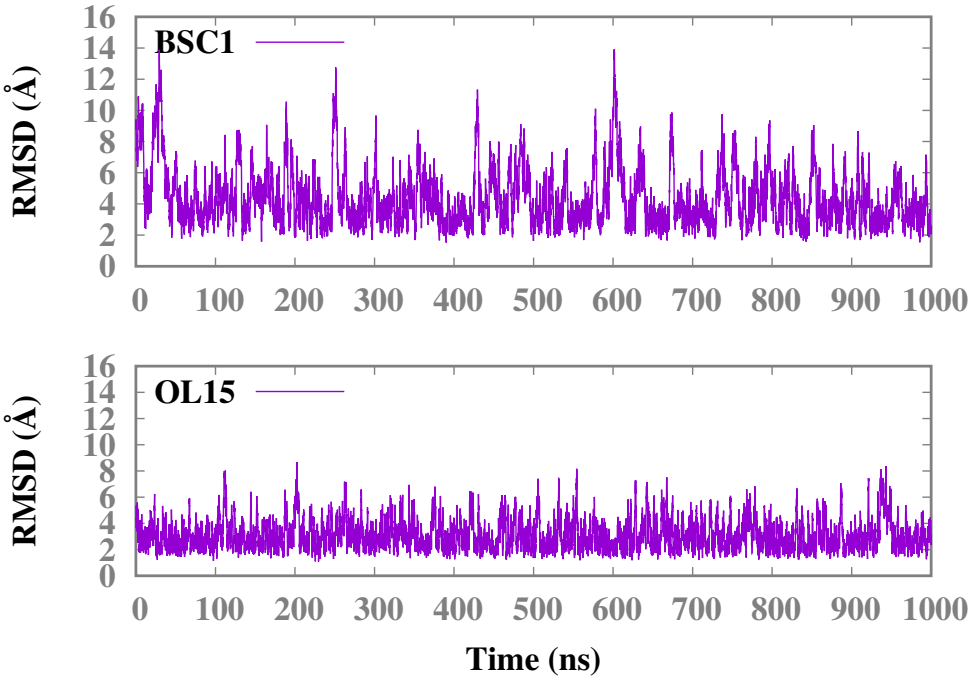


(a)

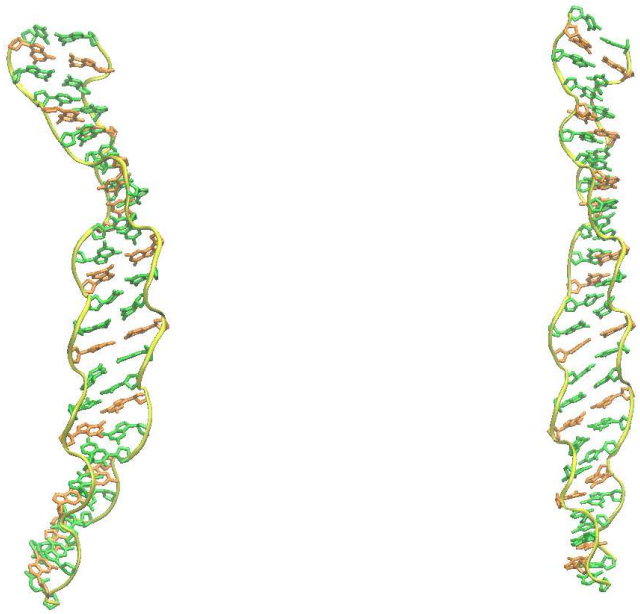


(b)

Figure S6: RMSD as a function of time, and a snapshot of GGC8$_{ZZ,out}$. (a) RMSD as function of time (ns) with respect to the average frame over the last 200 ns; only data for the last 1 μs is shown. (b, left) A snapshot of GGC8$_{ZZ,out}$ BSC1 after 1.2 μs, and (b, right) a snapshot of GGC8$_{ZZ,out}$ OL15 after 1 μs, The mismatches, WC bases, and backbone are shown by colors brown, green, and yellow respectively.

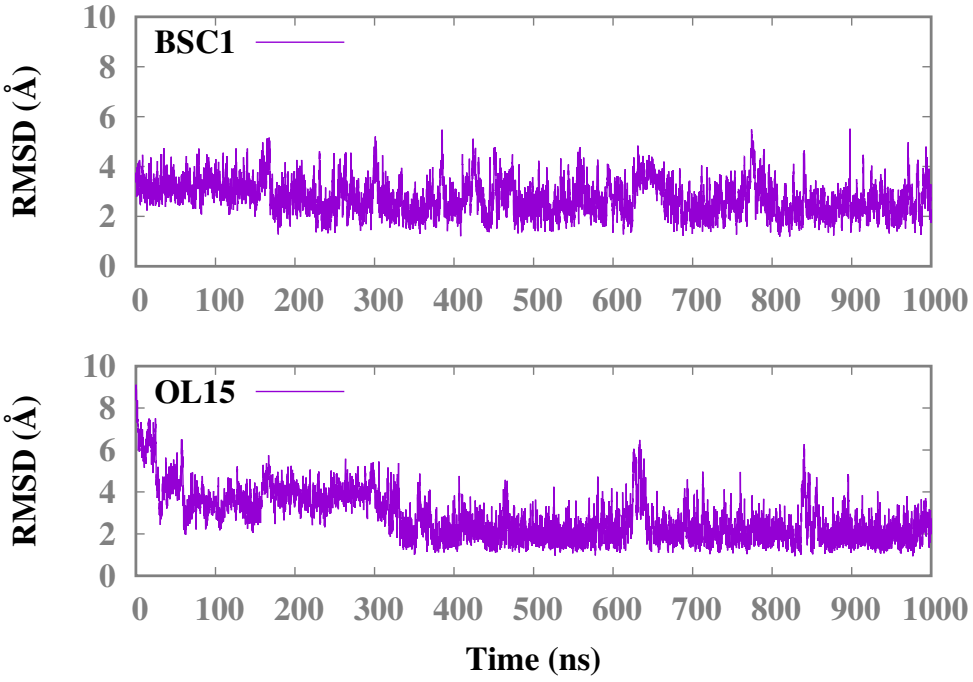


(a)

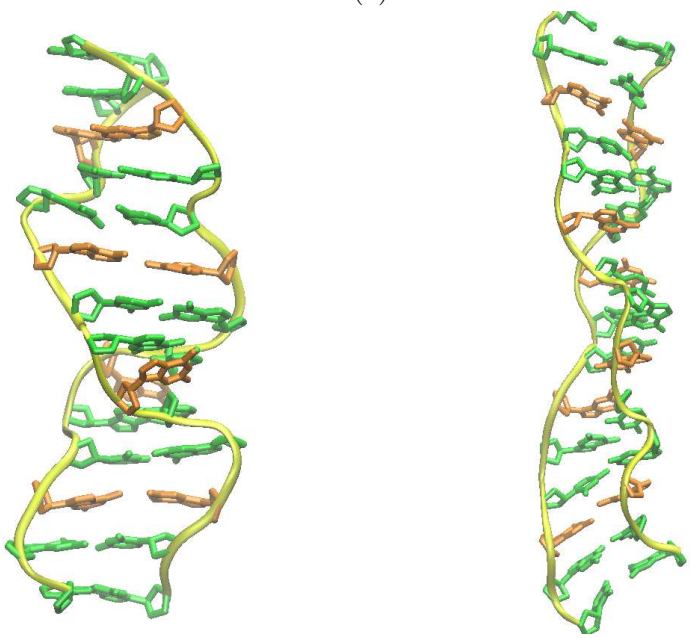


(b)

Figure S7: RMSD as a function of time, and a snapshot of GGC8$_{ZZ,in}$. (a) RMSD as function of time (ns) with respect to the average frame over the last 200 ns; only data for the last 1 μs is shown. (b, left) A snapshot of GGC8$_{ZZ,in}$ BSC1 after 1 μs, and (b, right) a snapshot of GGC8$_{ZZ,in}$ OL15 after 1 μs, The mismatches, WC bases, and backbone are shown by colors brown, green, and yellow respectively.

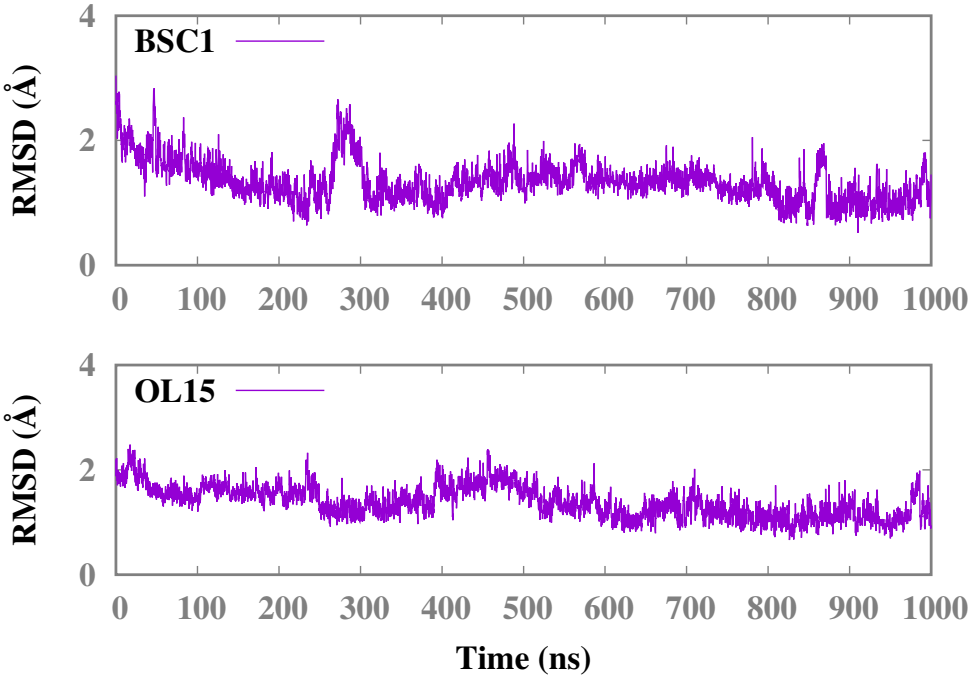


(a)

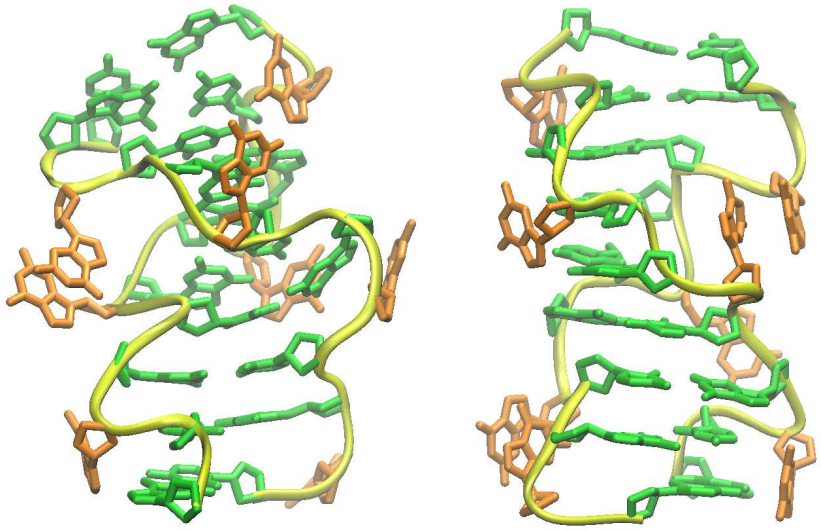


(b)

Figure S8: RMSD as a function of time, and a snapshot of CGG4$_{ZZ,in}$. (a) RMSD as function of time (ns) with respect to the average frame over the last 200 ns; only data for the last 1 μs is shown. ,(b, left) A snapshot of CGG4$_{ZZ,in}$ BSC1 after 1 μs, and (b, right) a snapshot of CGG4$_{ZZ,in}$ OL15 after 1 μs, The mismatches, WC bases, and backbone are shown by colors brown, green, and yellow respectively.



(a)



(b)

Figure S9: RMSD as a function of time, and a snapshot of GGC4$_{ZZ,alt}$. (a) RMSD as function of time (ns) with respect to the average frame over the last 200 ns; only data for the last 1 μs is shown. (b, left) A snapshot of GGC4$_{ZZ,alt}$ BSC1 after 1.2 μs, and (b, right) a snapshot of GGC4$_{ZZ,alt}$ OL15 after 1.2 μs, The mismatches, WC bases, and backbone are shown by colors brown, green, and yellow respectively.

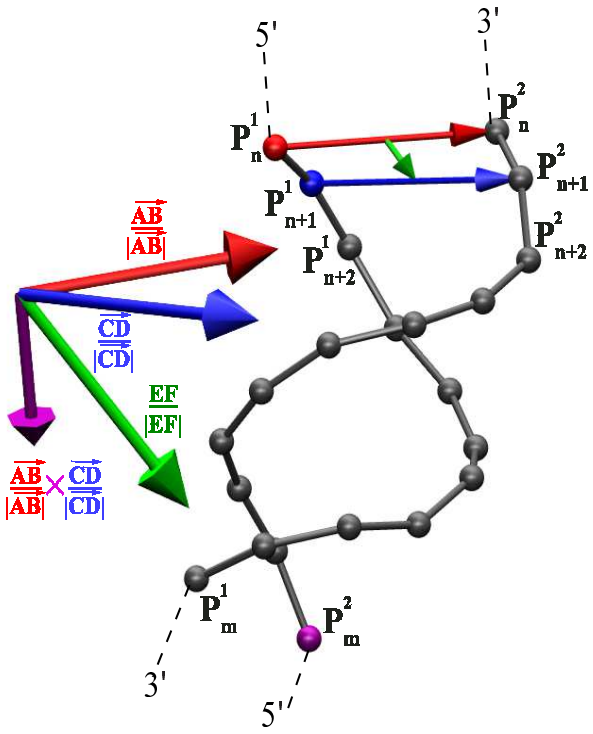


Figure S10: Phosphorus atoms (shown as grey spheres) used to define the collective variable of handedness (H) for a segment between two DNA base pairs n and m. Here, P_j^i represents the phosporus atom of the backbone phosphate group of the j^{th} base pair of strand i. Labeling of the base pairs is in the $5' \to 3'$ direction for strand 1, and the opposite direction for strand 2. This labeling of the helices is of course arbitrary, and may be interchanged, leaving H unchanged. Red, blue, and green arrows on the right represent the vectors involved in the definition of the first term of the sum in Eq. (1). The arrows on the left represent the unit vectors corresponding to those on the right. On the left, the cross product of the red and blue vectors gives the purple vector whose dot product with the green vector gives the first term of the sum (which is positive in this example corresponding to a right-handed turn).

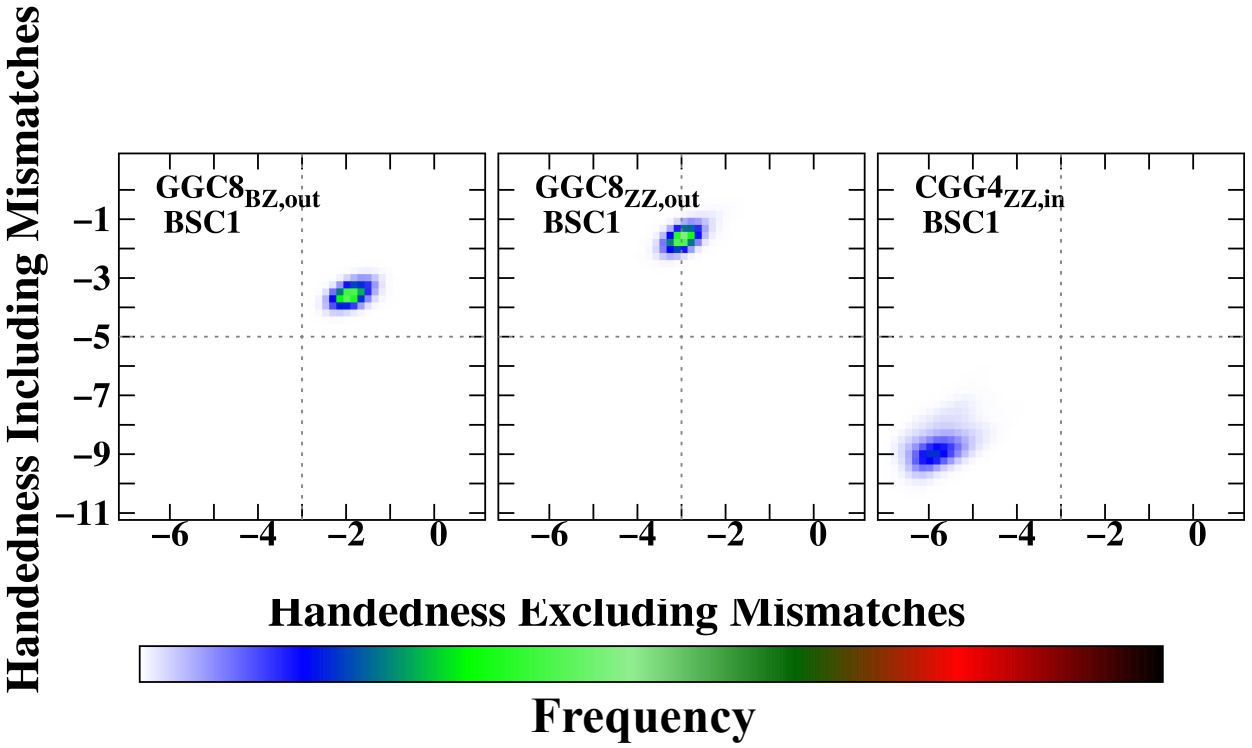


Figure S11: Two-dimensional histogram of handedness including mismatches versus handedness excluding mismatches. The data are based on calculations obtained from the middle residues during the last 1 μ simulation time for each helix. The total number of considered residues is constant for all calculations, which means that handedness with mismatches includes a larger number of terms (19) than handedness without mismatches (11).

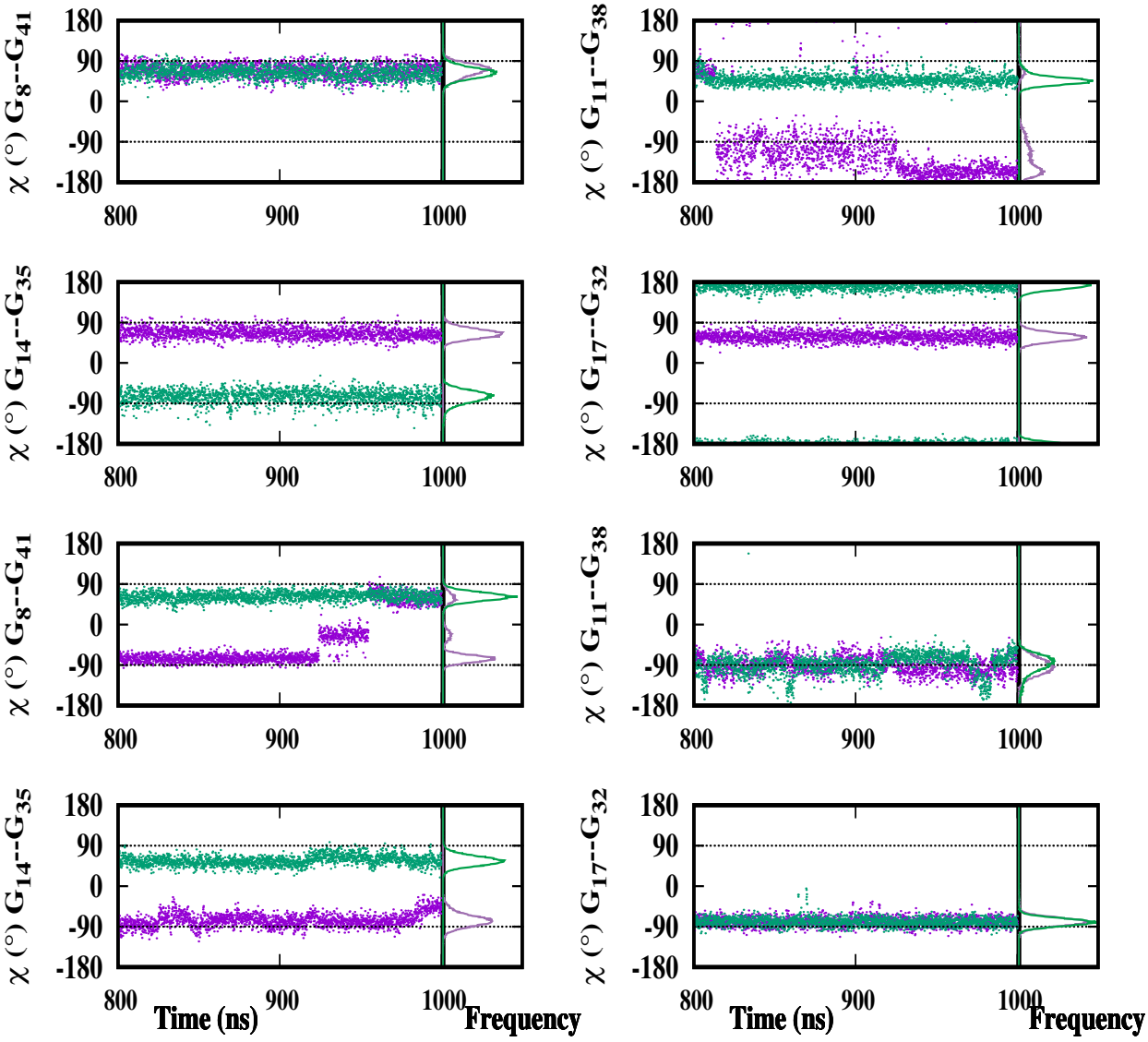


Figure S12: Time series of χ dihedral angles of middle GG mismatches in GGC8$_{BZ,out}$. Top four panels represent BSC1, and bottom OL15. The first strand is shown by purple, and the second one by green color.

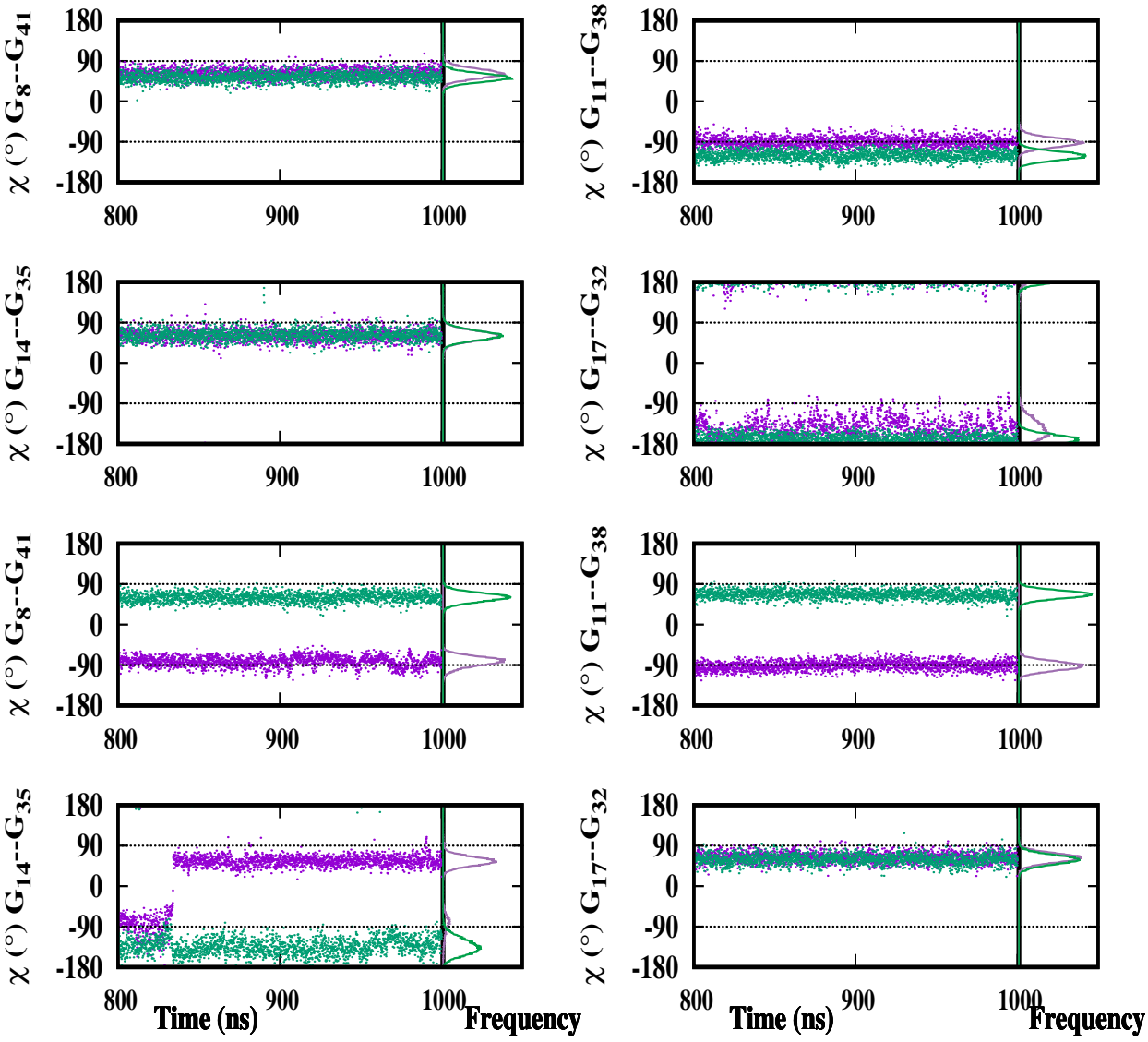


Figure S13: Time series of χ dihedral angles of middle GG mismatches in GGC8$_{ZZ,out}$. Top four panels represent BSC1, and bottom OL15. The first strand is shown by purple, and the second one by green color.

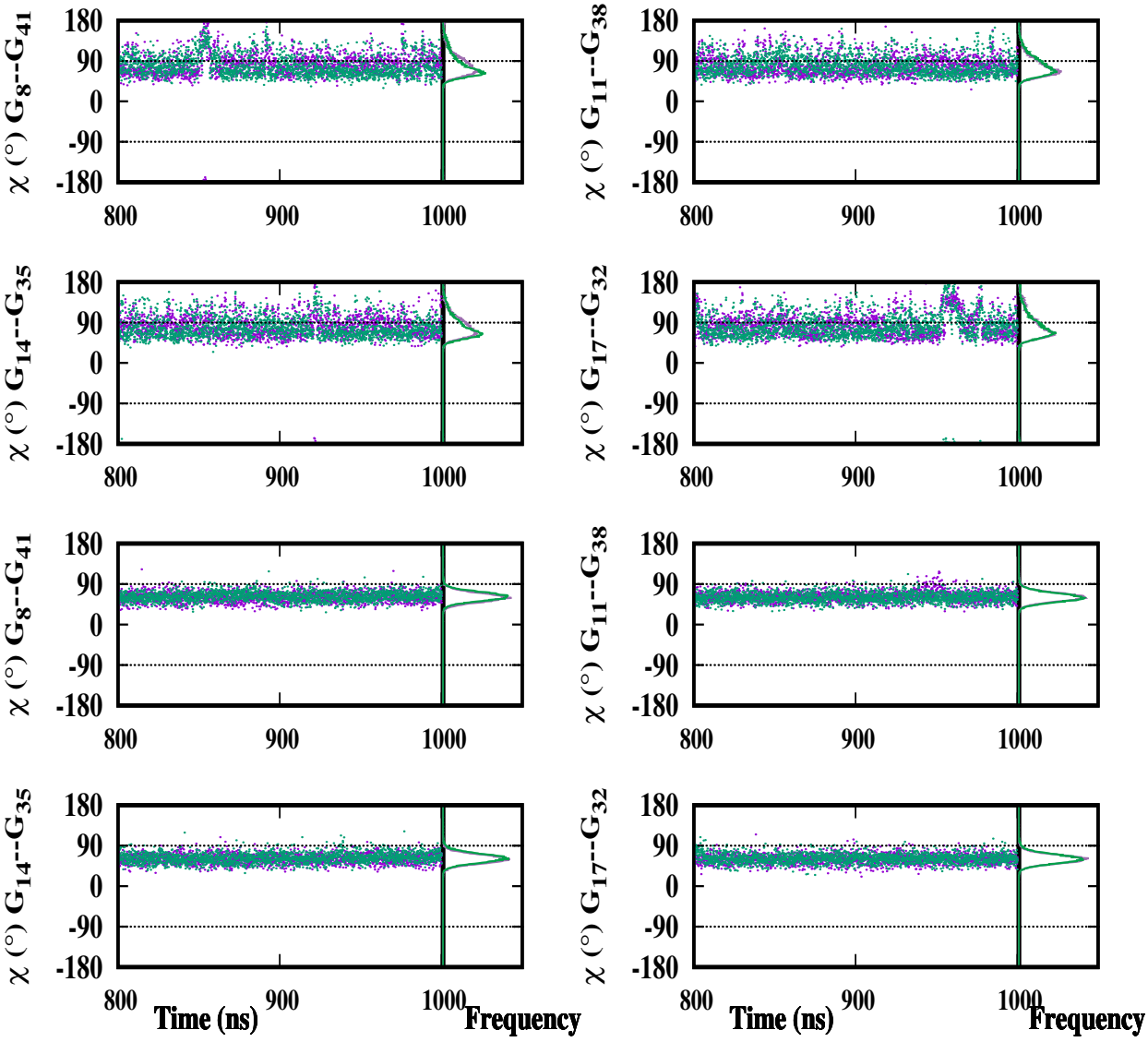


Figure S14: Time series of χ dihedral angles of middle GG mismatches in GGC8$_{ZZ,in}$. Top four panels represent BSC1, and bottom OL15. The first strand is shown by purple, and the second one by green color.

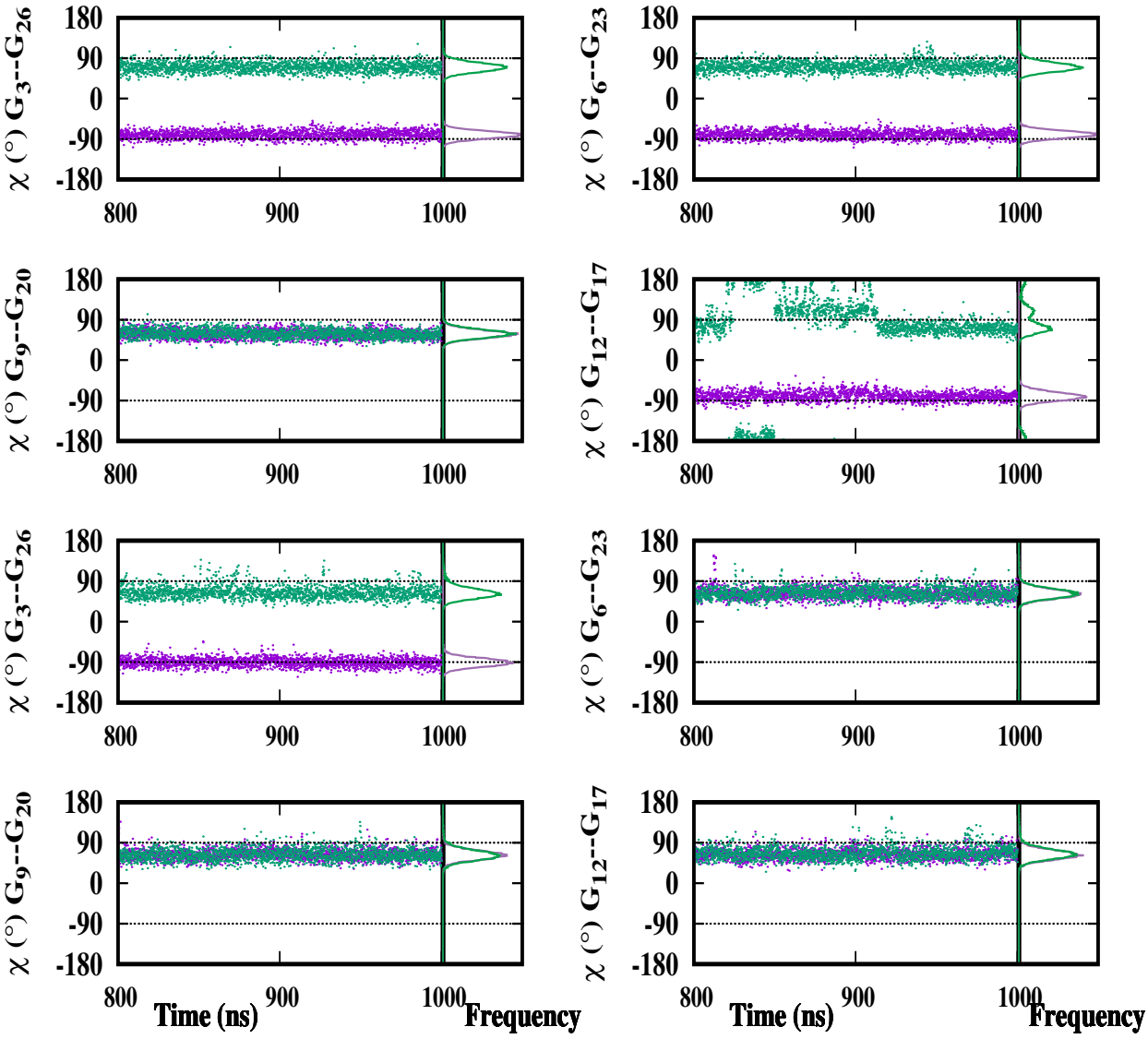


Figure S15: Time series of χ dihedral angles of middle GG mismatches in CGG4$_{ZZ,in}$. Top four panels represent BSC1, and bottom OL15. The first strand is shown by purple, and the second one by green color.

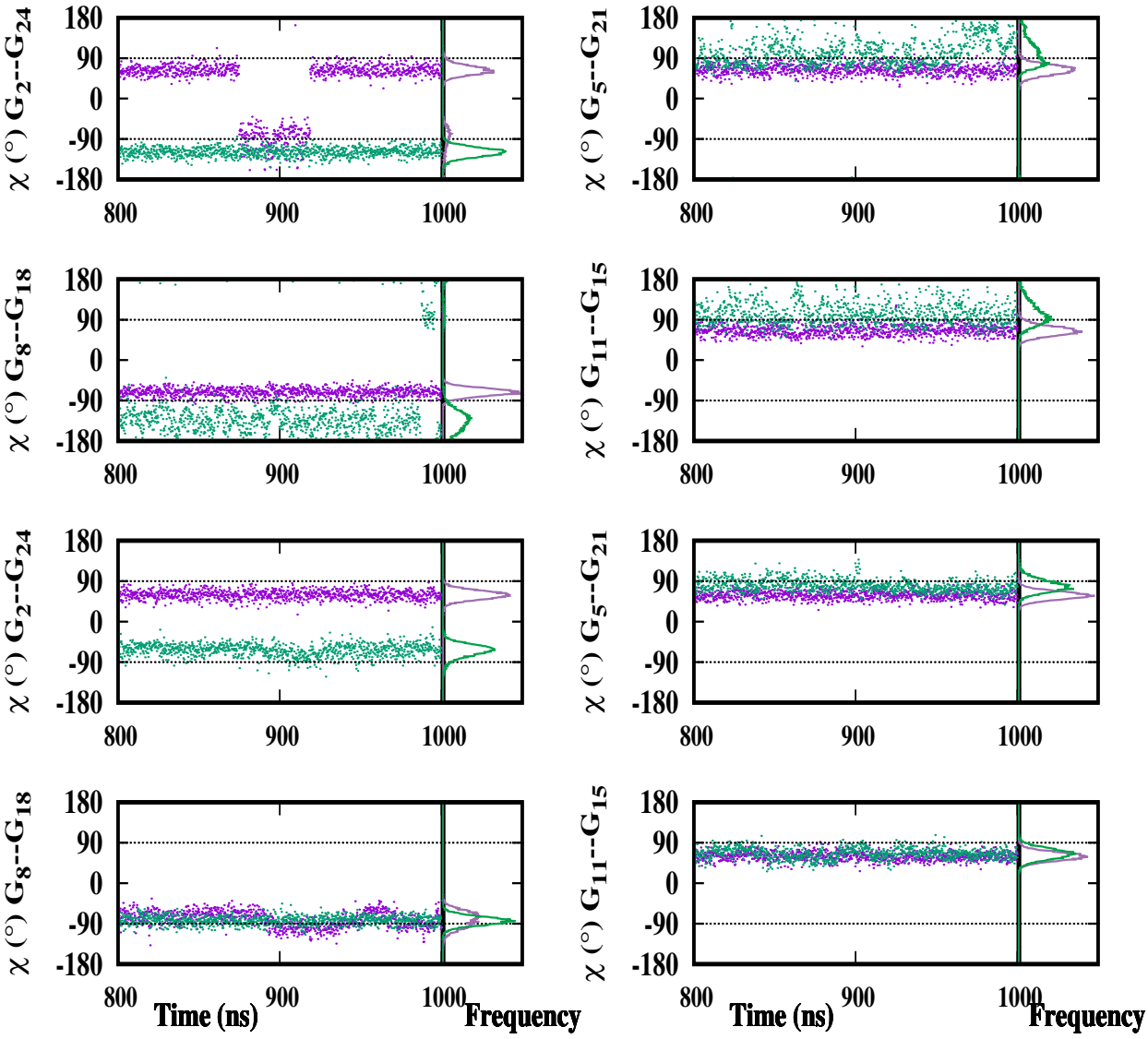


Figure S16: Time series of χ dihedral angles of middle GG mismatches in $GGC4_{ZZ,alt}$. Top four panels represent BSC1, and bottom OL15. The first strand is shown by purple, and the second one by green color.

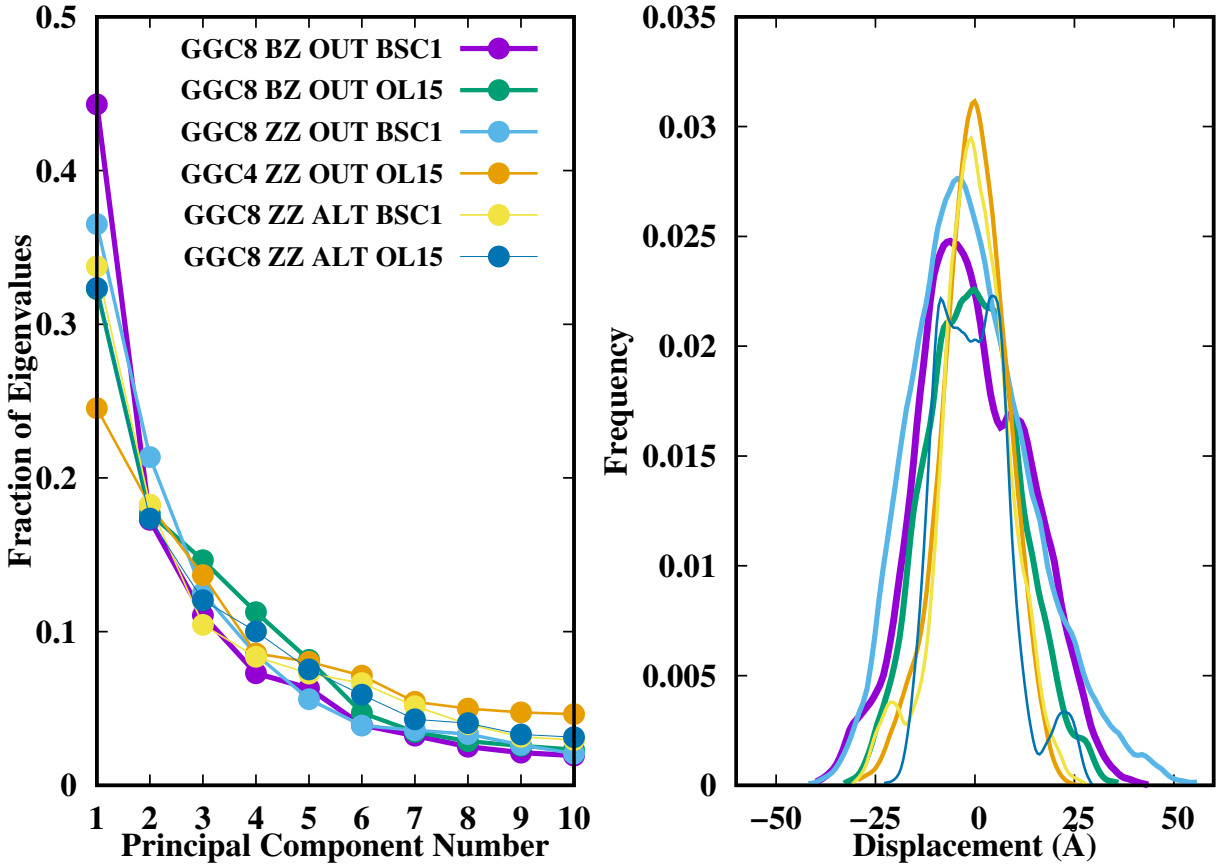


Figure S17: Principal component analysis (PCA) of inner-GGC8$_{BZ,out}$, inner-GGC8$_{ZZ,out}$, and GGC4$_{ZZ,alt}$. Left panel: Plot of first ten principal components (PCs); Right Panel: Displacement along the direction of the first eigenvector over last 200 ns.

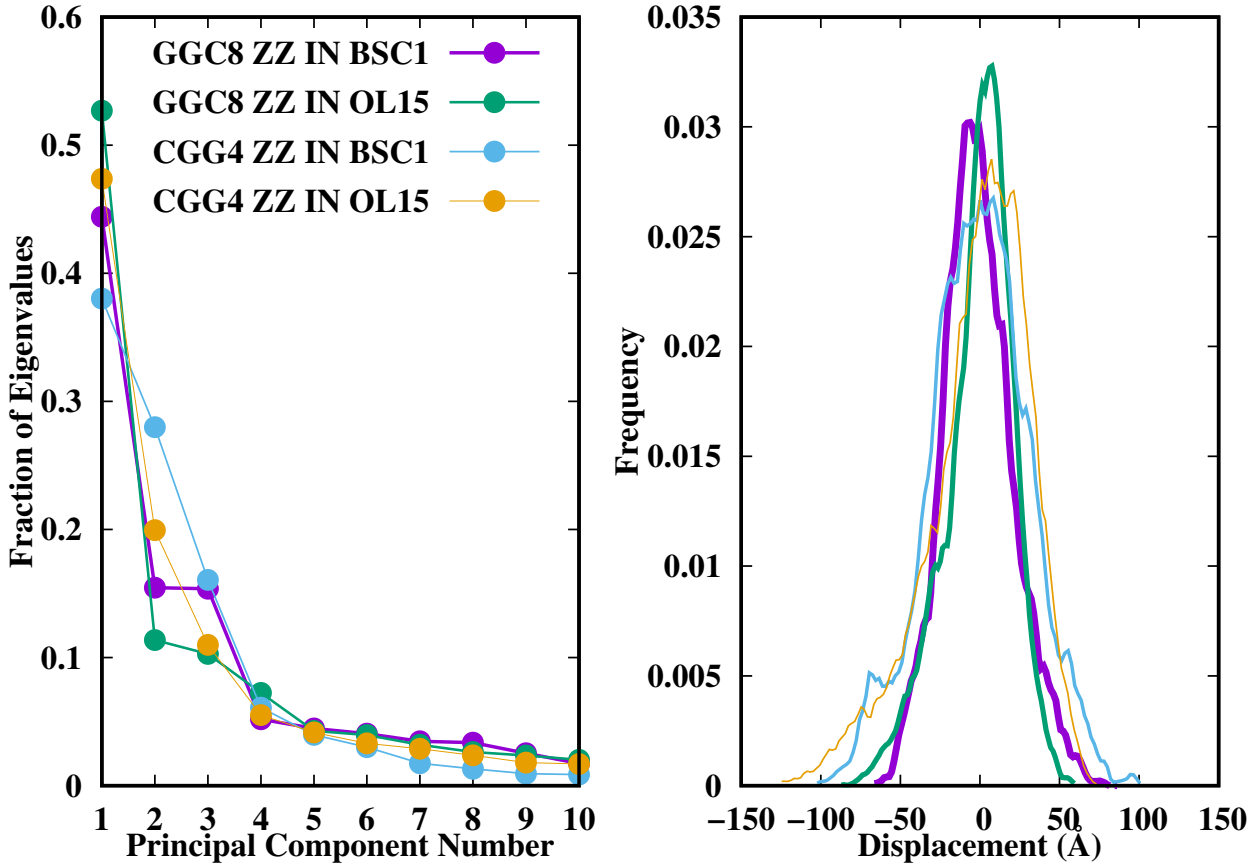


Figure S18: PCA of inner-GGC8$_{ZZ,out}$, and CGG4$_{ZZ,in}$. Left Panel: Plot of first ten PCs; Right Panel: Displacement along the direction of the first eigenvector over last 200 ns.

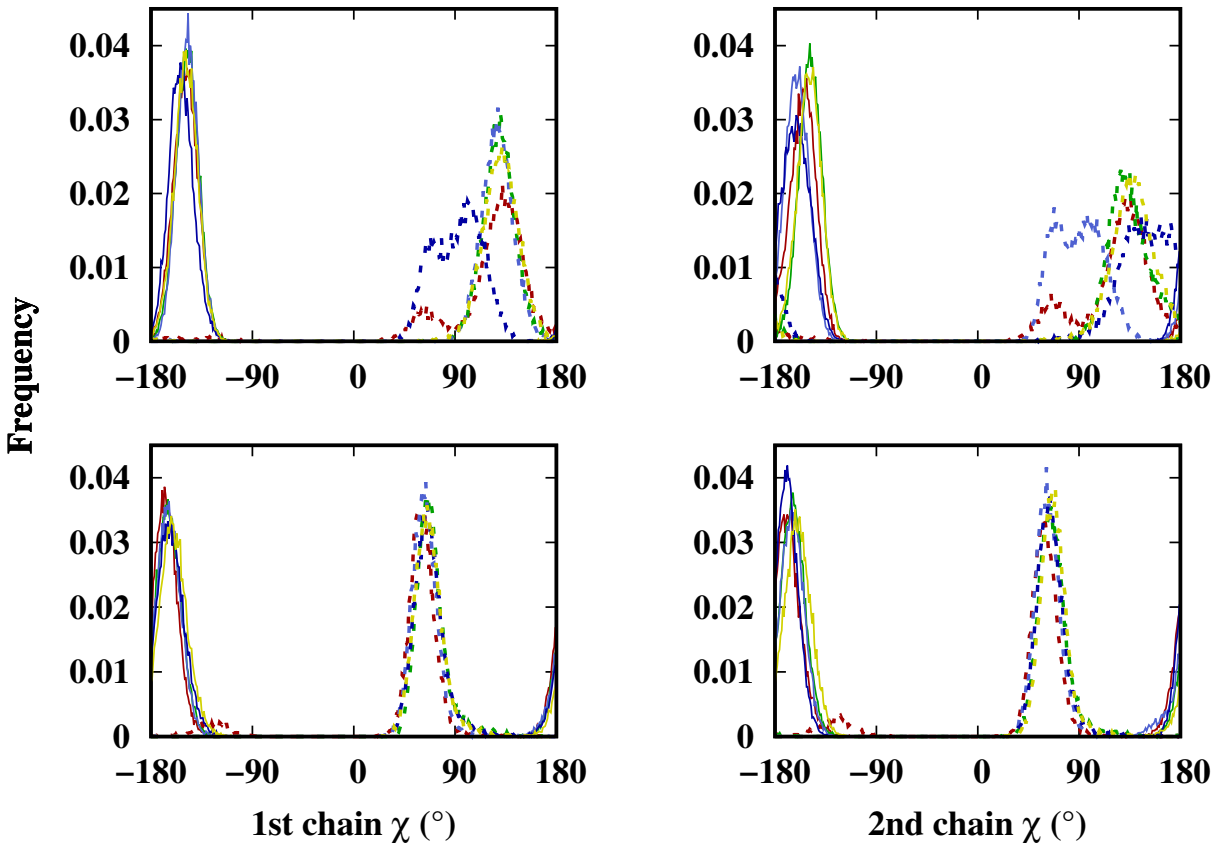


Figure S19: Histogram of χ dihedral angles W-C bases in $CGG4_{ZZ,in}$ (top) BSC1, and (bottom) OL15. The guanine are shown by dot lines, and cytosine are shown by solid lines. Different colors represent different base pairs. The data obtained from the last 200 ns of simulations.

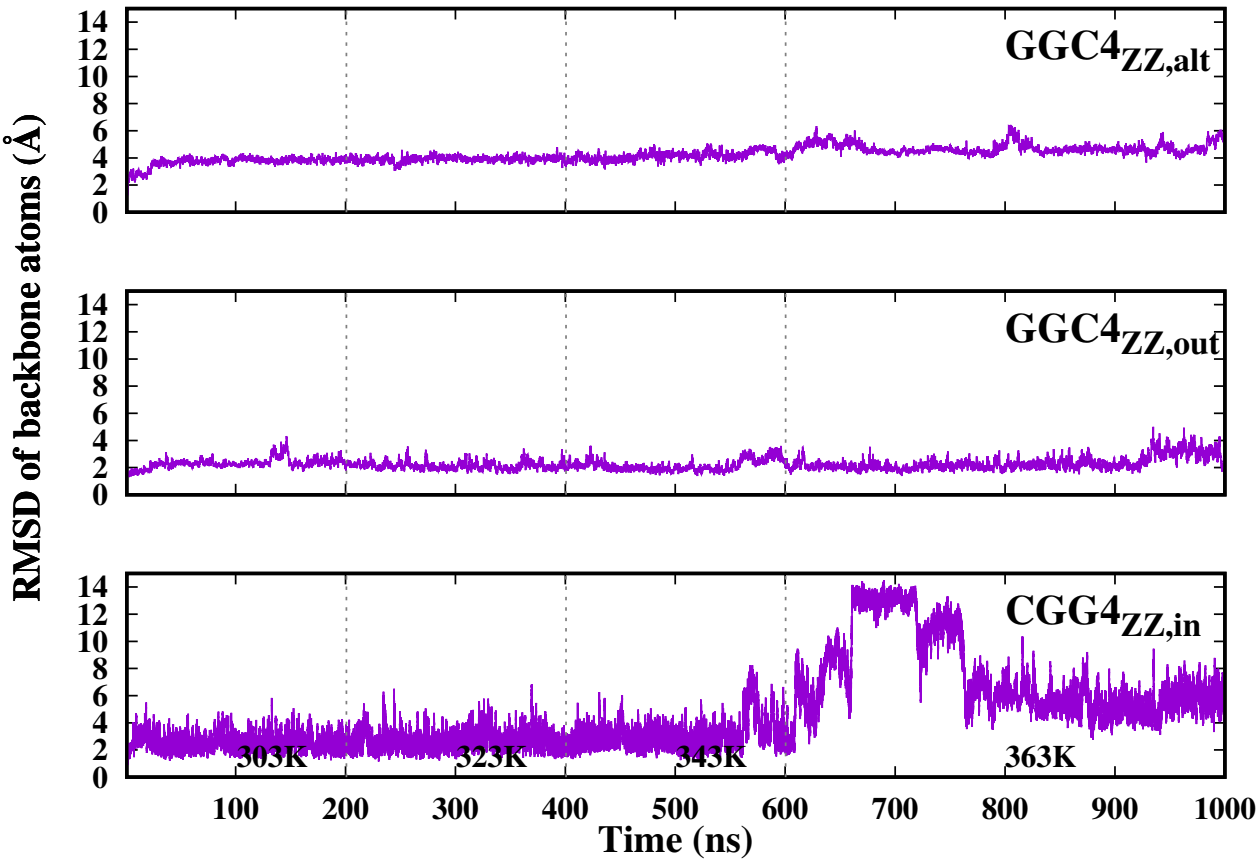


Figure S20: The time evolution of the RMSD of the most stable duplexes with respect to the first frame versus time. The temperature gradually increased up to 363K where the duplexes started to melt under low salt (200 mMol) conditions. The data obtain from a 1μs simulation.

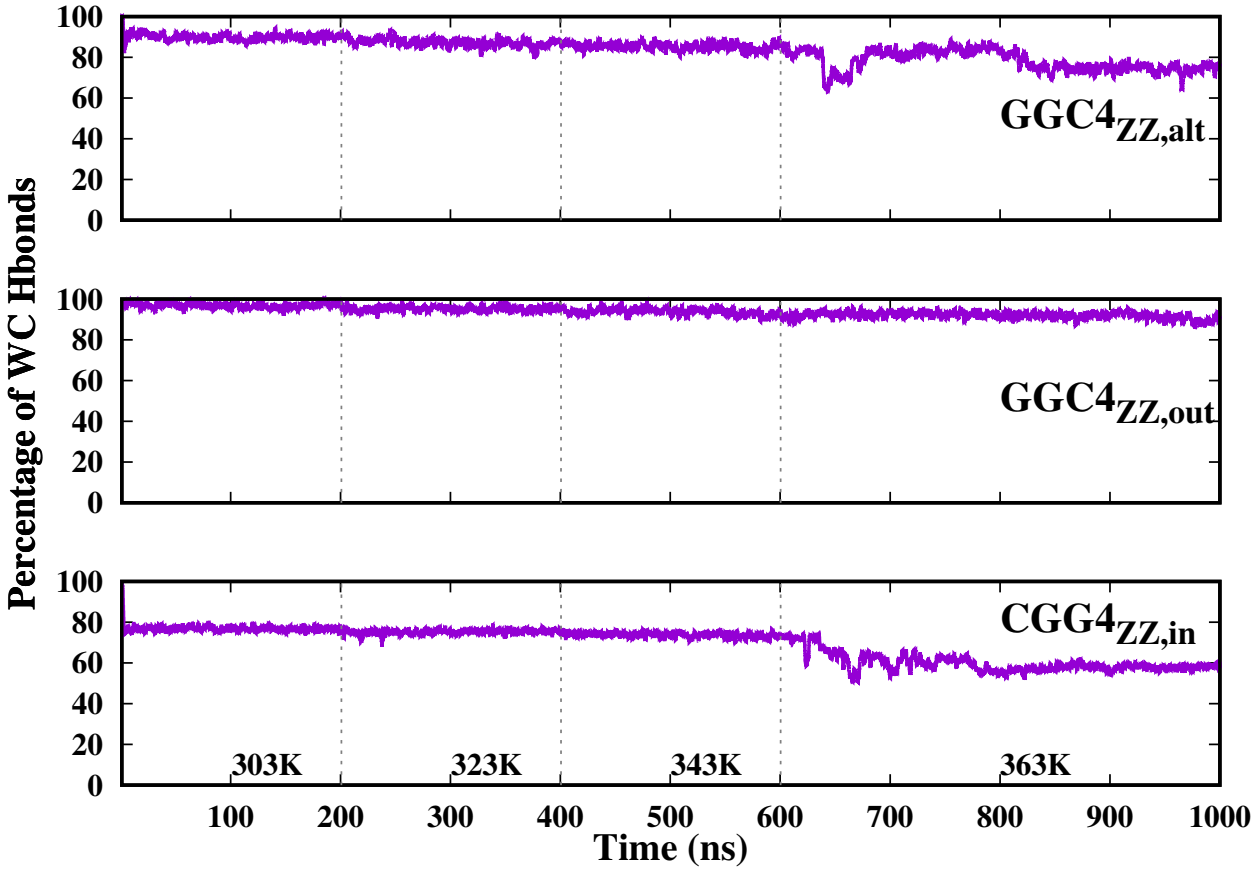


Figure S21: Percentage of hydrogen bonds of WC bases of the most stable duplexes as the simulation temperature is increased (at low 200mM salt concentration). At 363K the duplexes start to melt. The data was obtained from 1μs simulations, where the number of hydrogen bonds were averaged over a 100 step window.