Supporting Information

Flexible 5-Guanidino-4-nitroimidazole DNA Lesions: Structures and Thermodynamics †

† This research is supported by NIH Grant CA-75449 to S. B. and NIH Grant ES-11589 to V. S. and N. E. G.

Lei Jia[‡], *Vladimir Shafirovich*[‡], *Robert Shapiro*[‡], *Nicholas E. Geacintov*[‡], *and Suse Broyde*[‡, §, *]

[‡] Department of Chemistry, New York University, New York, New York 10003

[§] Department of Biology, New York University, New York, New York 10003

E-mail address: broyde@nyu.edu

Running title: Simulation of 5-Guanidino-4-nitroiminidazole damaged DNA

[*]Corresponding author: Suse Broyde: tel. (212)998-8231, fax (212)995-4015, e-mail broyde@nyu.edu

Figure Legends

Figure S1: Representative starting models for the MD simulations: NI(*anti*)•C(*anti*) and NI(*syn*)•C(*anti*). The NI is colored by atom. The partner C is pink. The other duplexes differ only in the partner base. Hydrogens have been removed for clarity.

Figure S2: Stereo views of geometry-optimized NI nucleosides whose energies are between 6.6 and 14.4 kcal/mol.

Figure S3: RMSD vs time plot for each molecular dynamics simulation.

Figure S4: Stereo views of the central 5-mer of the NI trajectory-average duplex structures. Relative energies are given in Table 3. The NI base is colored by atom. The NI base partners are colored as yellow (A), pink (C), purple (G), and orange (T). The NI adjacent bases C5 and C7 are also marked. The view is into the major groove. All stereo figures are prepared for viewing with a stereoviewer.

Figure S5: Hydrogen bonds (black lines) between NI and partner base. Relative energies are given in Table 3. The base pairs are obtained from the trajectory average structures of the selected simulation window. Hydrogen bonds shown have occupancy greater than 50%.

Figure S6: Average bend angles of all NI damaged DNA duplexes. Energetically favored structures (Table 3) are designated by colored asterisks.

Figure S7: Major and minor groove dimensions of all NI damaged DNA duplexes. Energetically favored structures (Table 3) are designated by colored asterisks.

Figure S8: Torsion angles δ, θ, and χ and sugar pucker pesudorotation angle **P** vs time plots for each molecular dynamics simulation. Energetically favored structures (Table 3) are designated in red.

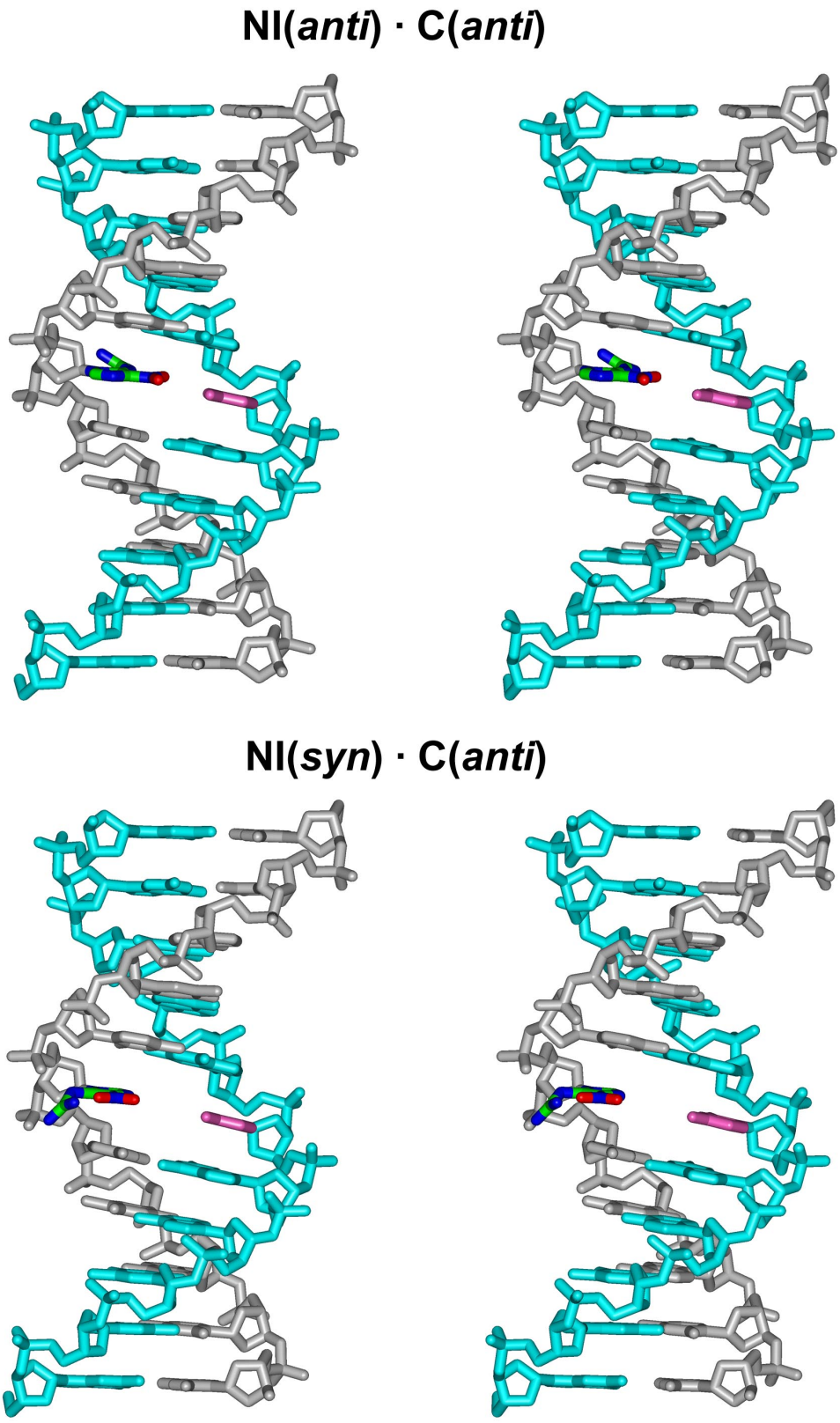

Figure S1

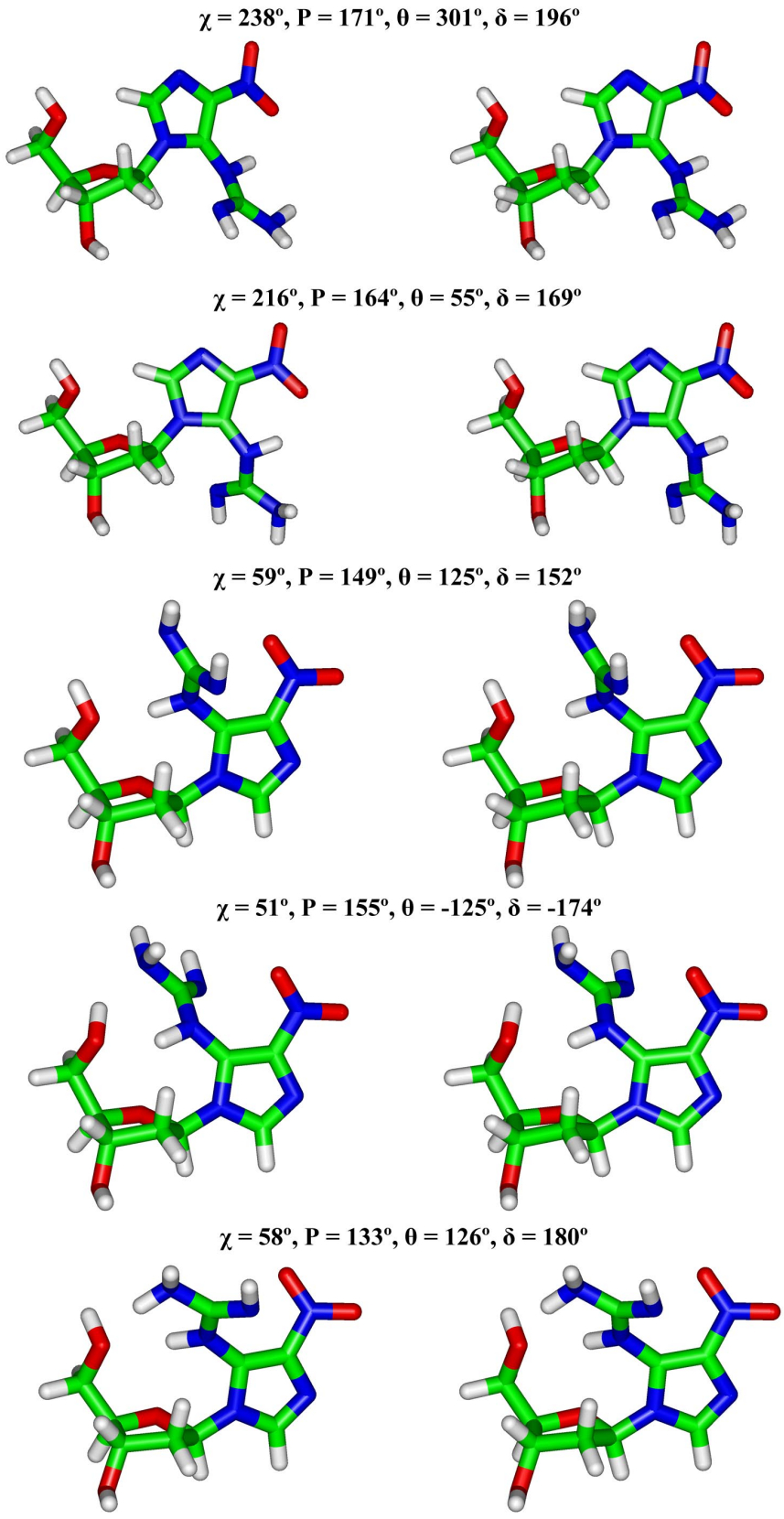
χ = 238°, P = 171°, θ = 301°, δ = 196°
χ = 216°, P = 164°, θ = 55°, δ = 169°
χ = 59°, P = 149°, θ = 125°, δ = 152°
χ = 51°, P = 155°, θ = -125°, δ = -174°
χ = 58°, P = 133°, θ = 126°, δ = 180°

Figure S2

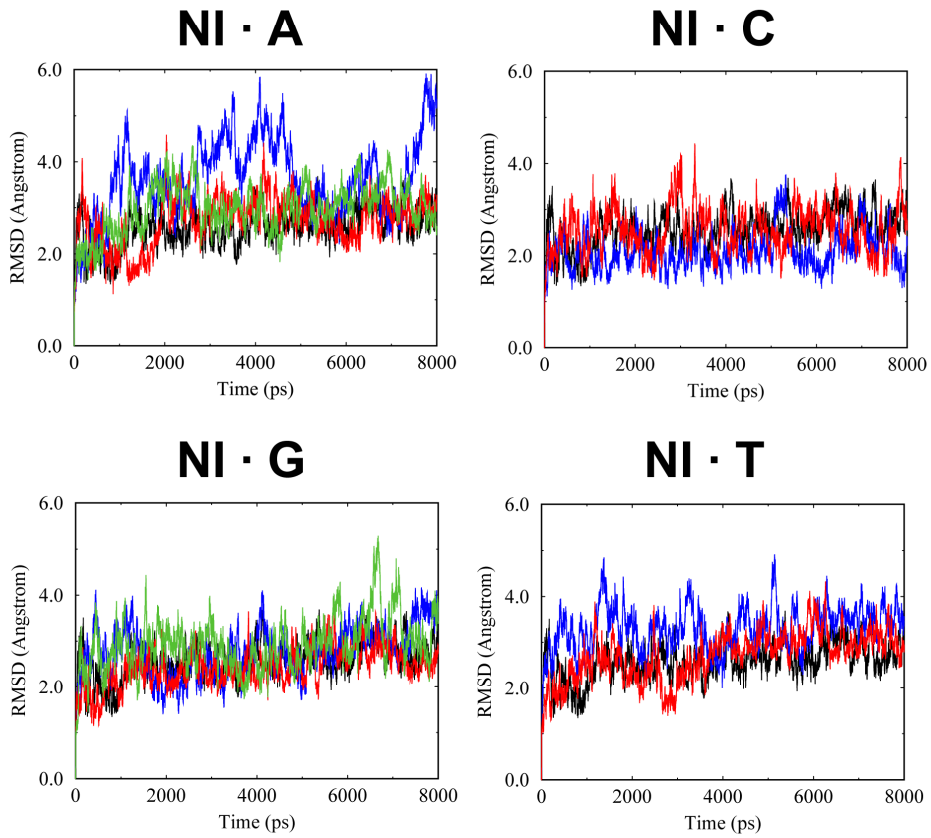
NI · A
NI · C
NI · G
NI · T
RMSD (Angstrom)
Time (ps)
0.0
2.0
4.0
6.0
0
2000
4000
6000
8000
Color Code: Unmodified Control NI(anti) · X(anti)
NI(syn) · X(anti) NI(anti) · X(syn)
X = A, C, G, or T

Figure S3

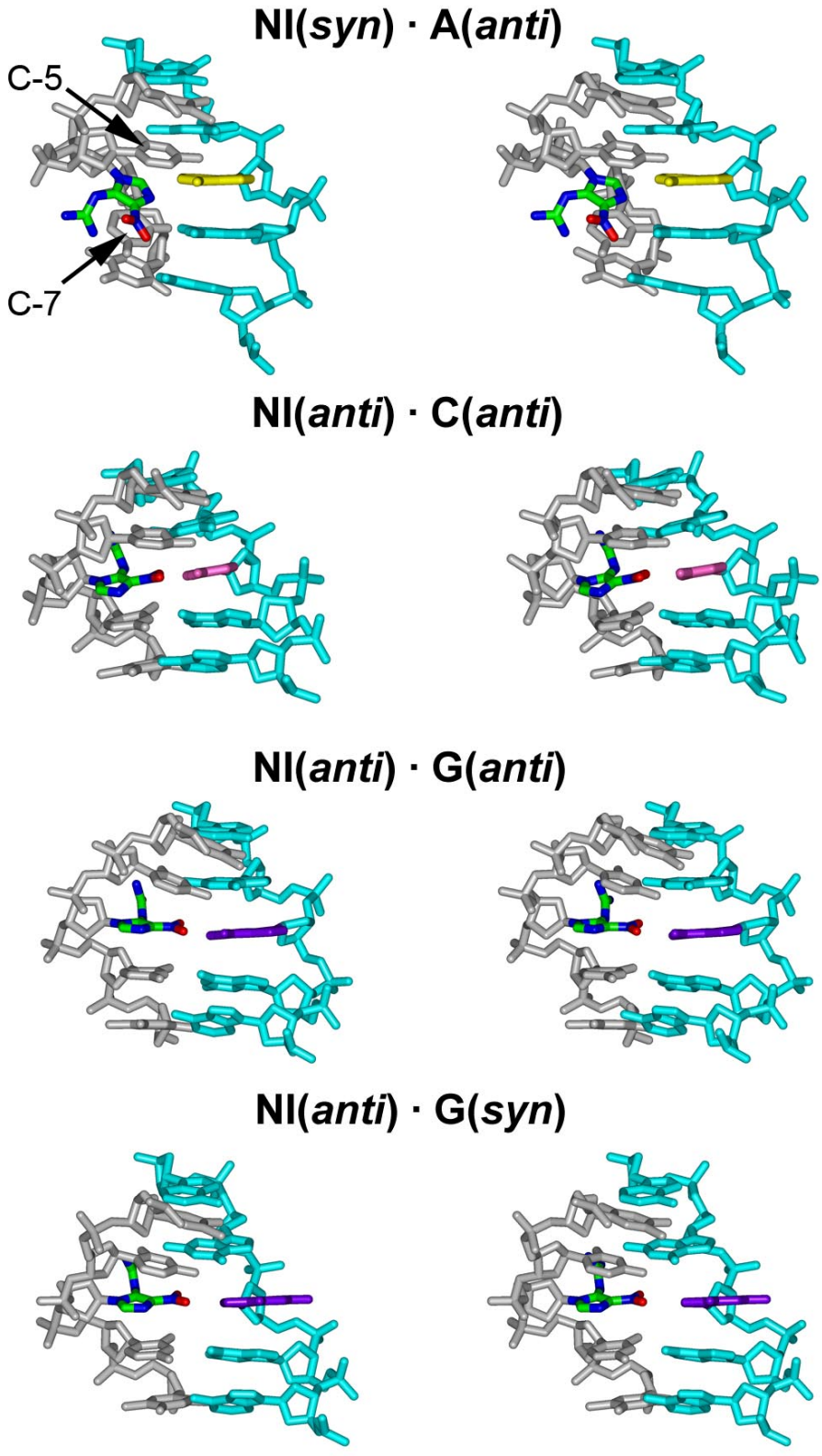

Figure S4

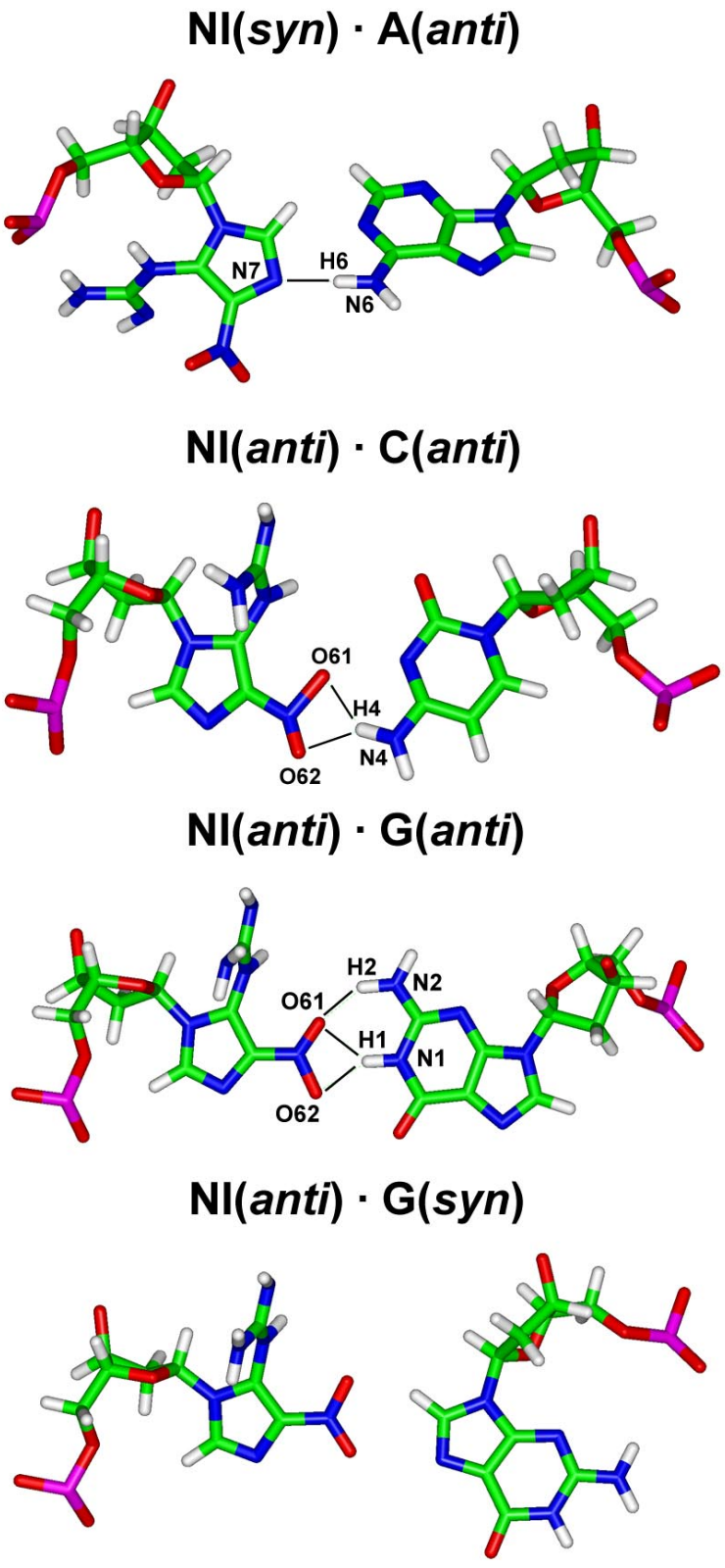
NI(*syn*) · A(*anti*)
N7
H6
N6
NI(*anti*) · C(*anti*)
O61
H4
N4
O62
NI(*anti*) · G(*anti*)
H2
N2
O61
H1
N1
O62
NI(*anti*) · G(*syn*)

Figure S5

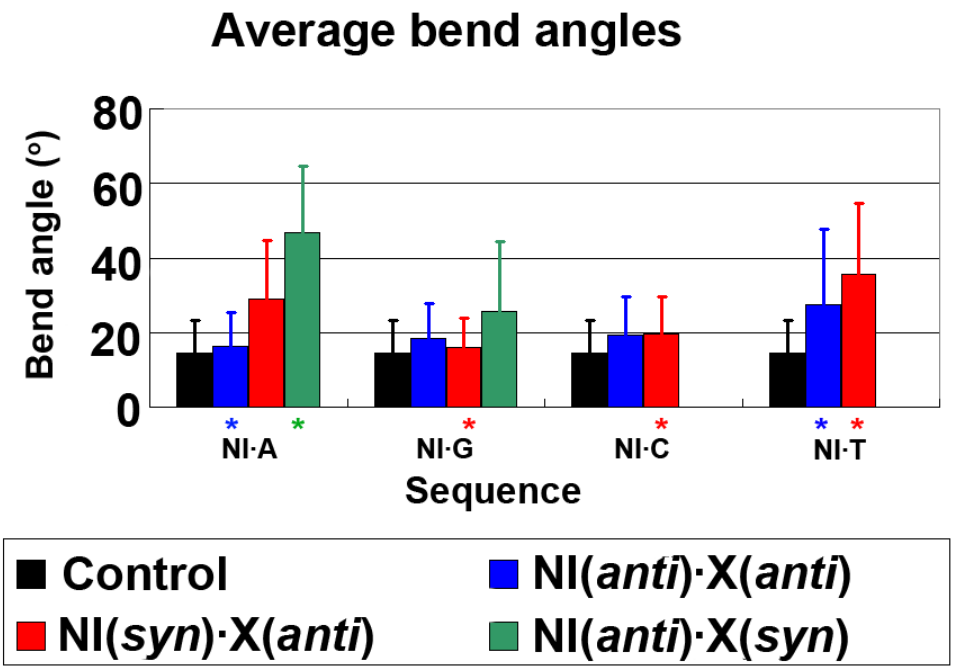
Average bend angles
Bend angle (°)
80
60
40
20
0
NI·A
NI·G
NI·C
NI·T
Sequence
Control
NI(anti)·X(anti)
NI(syn)·X(anti)
NI(anti)·X(syn)

Figure S6

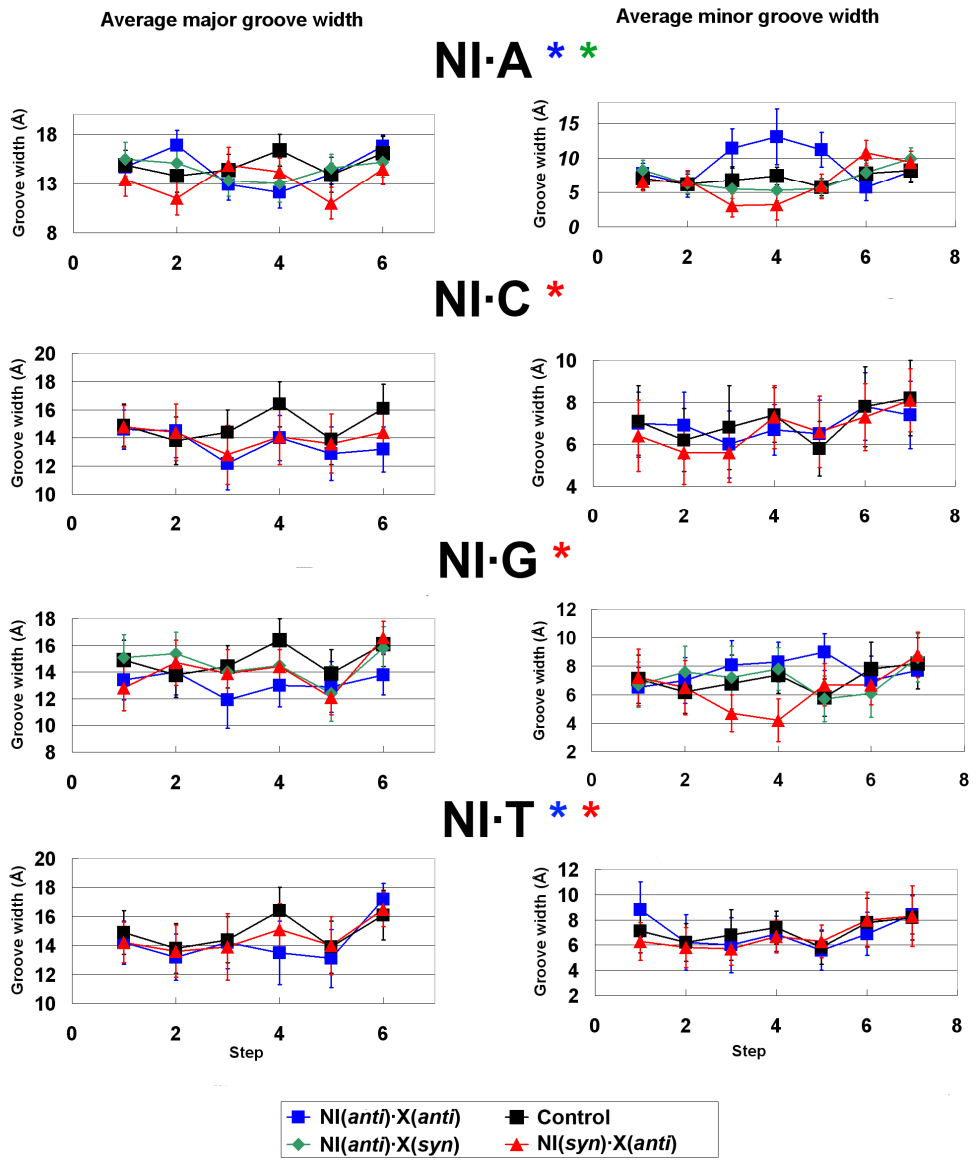
Average major groove width
Average minor groove width
NI·A * *
NI·C *
NI·G *
NI·T * *
Groove width (Å)
Step
NI(anti)·X(anti)
NI(anti)·X(syn)
Control
NI(syn)·X(anti)

Figure S7

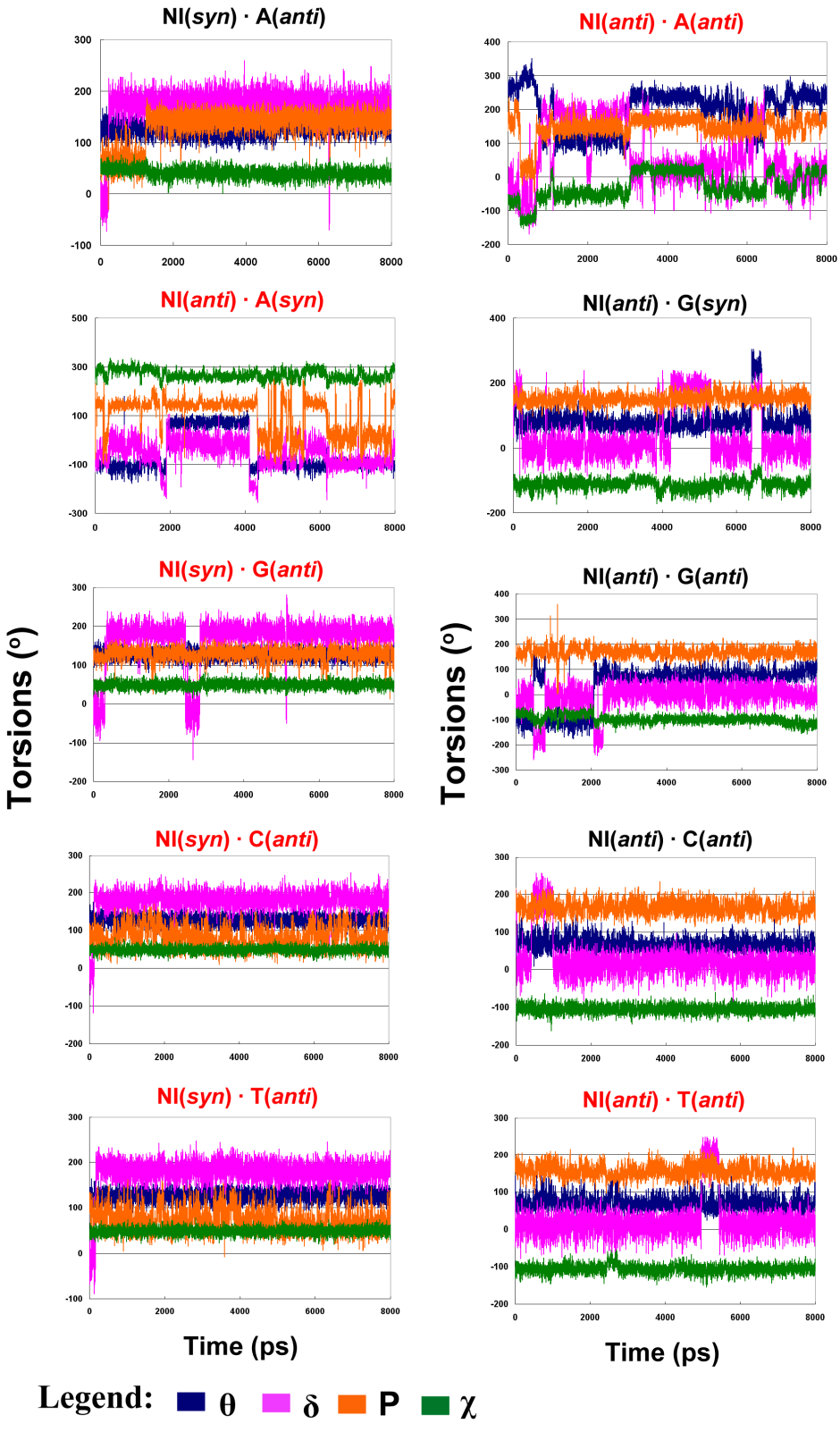
NI(syn) · A(anti)
NI(anti) · A(anti)
NI(anti) · A(syn)
NI(anti) · G(syn)
NI(syn) · G(anti)
NI(anti) · G(anti)
NI(syn) · C(anti)
NI(anti) · C(anti)
NI(syn) · T(anti)
NI(anti) · T(anti)
Torsions (°)
Torsions (°)
Time (ps)
Time (ps)
Legend: θ δ P χ

Figure S8

Table S1: Values of torsions for NI employed in duplex initial models

	χ(°)	θ(°)	δ(°)
NI(*anti*) · A(*anti*)	**154**	**15**	**195**
NI(*anti*) · A(*syn*)	**154**	**15**	**195**
NI(*anti*) · C(*anti*)	**154**	**15**	**195**
NI(*anti*) · G(a*nti*)	**154**	**15**	**195**
NI(*anti*) · G(*syn*)	**154**	**15**	**195**
NI(*anti*) · T(a*nti*)	**157**	**15**	**195**
NI(*syn*) · A(*anti*)	**84**	**135**	**0**
NI(*syn*) · C(*anti*)	**93**	**135**	**68**
NI(*syn*) · G(*anti*)	**84**	**135**	**0**
NI(*syn*) · T(*anti*)	**84**	**135**	**0**

Table S2: Added force field parameters for the NI lesion

Atom Parameters		
Atom Type	Atomic Mass	Polarizability(Å^3)
NO	14.01	0.53

Bond Parameters		
Bond	Force Constant K_r (kcal mol^{-1} Å^{-2})	Equilibrium Bond Length r_{eq} (Å)
CD-N*	506.4	1.33
CC-CD	522.6	1.40
CC-NO	327.6	1.42
NO-O2	761.2	1.23
CD-N2	449,0	1.37

Angle Parameters		
Angle	Force Constant K_θ (kcal mol^{-1}rad^{-2})	Equilibrium Bond Angle θ_{eq} (º)
CT-N*-CD	67.6	120.3
CK-N*-CD	68.6	106.9
CK-NB-CC	70.5	104.6
NB-CC-CD	71.0	111.2
NB-CC-NO	68.6	124.2
CD-CC-NO	67.6	124.6
CC-NO-O2	69.3	120.0
O2-NO-O2	76.4	120.0
CC-CD-N2	68.4	129.9
CC-CD-N*	71.0	104.8
CD-N2-H	49.0	120.0
CD-N2-CD	65.5	120.0
N2-CD-N*	72.6	120.0
N2-CD-N2	73.3	120.0
CD-N*-H	52.6	111.7

Torsional Parameters				
Torsion	Number of paths	Torsion Barrier $V_n/2$ (kcal mol^{-1})	Phase Angle γ (º)	Periodicity *n*
X -N*-CD-X	2	9.5	180	2
X -CC-CD-X	4	16.0	180	2
X -NB-CC-X	2	9.5	180	2
X -CC-NO-X	4	3.0	180	2
X -N2-CD-X	4	2.7	180	2

Improper Torsions			
Torsion	Torsion Barrier $V_n/2$ (kcal mol^{-1})	Phase Angle γ (º)	Periodicity *n*
CC-O2-NO-O2	10.5	180	2

Van der Waals Parameters		
Atom Type	Van der Waals R (Å)	Van der Waals Well Depth ε (kcal mol^{-1})
NO	1.824	0.17

Table S3: AMBER atom type, connection type, and partial charge assignments for the NI lesion[a].

Atom Name	Atom Type	Connection Type	Partial Charge
P	P	M	1.221333
O1P	O2	E	-0.79232
O2P	O2	E	-0.79232
O5'	OS	M	-0.5057
C5'	CT	M	-0.02002
H5'1	H1	E	0.092955
H5'2	H1	E	0.092955
C4'	CT	M	0.055017
H4'	H1	E	0.14164
O4'	OS	E	-0.24723
C3'	CT	M	0.07792
H3'	H1	E	0.116204
C2'	CT	3	-0.11809
H2'1	HC	E	0.062163
H2'2	HC	E	0.062163
C1'	CT	B	0.00508
H1'	H2	E	0.192481
N9	N* (nd)	S	0.064975
C8	CK (cd)	B	0.100283
H8	H5 (h5)	E	0.194863
N7	NB (nc)	S	-0.54865
C5	CC (cc)	B	0.238038
N6	NO (no)	B	0.617587
O61	O2 (o)	E	-0.50899
O62	O2 (o)	E	-0.38634
C4	CD (cd)	S	0.038302
N3	N2 (nh)	B	-0.38224
H3	H (hn)	E	0.351142
C2	CD (c2)	B	0.606928
N1	N* (n2)	S	-0.6783
H1	H (hn)	E	0.321915
N2	N2 (na)	B	-0.90549
H21	H (hn)	E	0.383818
H22	H (hn)	E	0.383818
O3'	OS	M	-0.5359

[a]: Analogous atom types from the GAFF (59) force field are shown in parentheses. Torsional parameters in Table S2 were taken from GAFF.

Table S4: Box sizes and numbers of waters in MD simulation initial models[a].

	NI(*syn*)•X(*anti*)	NI(*anti*)•X(*anti*)	NI(*anti*)•X(*syn*)
NI•A	50Å x 55Å x 67Å 4206 H_2O	50Å x 50Å x 67Å 3791 H_2O	50Å x 51Å x 67Å 3874 H_2O
NI•C	50Å x 51Å x 67Å 3874 H_2O	50Å x 50Å x 67Å 3790 H_2O	
NI•G	49Å x 55Å x 67Å 4097 H_2O	49Å x 51Å x 67Å 3758 H_2O	49Å x 50Å x 67Å 3680 H_2O
NI•T	49Å x 51Å x 67Å 3756 H_2O	49Å X 51Å x 67Å 3757 H_2O	
G•C (Unmodified Control)		50Å x 50Å x 67Å 3789 H_2O	

[a]: X is A, C, G, or T.

Table S5: Trajectories (ps) employed for thermodynamic and structural analyses [a].

	NI(*syn*)•X(*anti*)	NI(*anti*)•X(*anti*)	NI(*anti*)•X(*syn*)
NI•A	4000-7000	3200-6200	5000-8000
NI•C	4000-7000	2000-5000	
NI•G	3000-6000	3000-6000	2100-3100, 5400-6400, and 7000-8000
NI•T	4000-7000	2900-4900 and 6000-7000	
G•C (Unmodified Control)		3000-6000	

[a]: X is A, C, G, or T.

Table S6: Hydrogen bonds and occupancies involving NI[a].

Sequence	Hydrogen bond and occupancy involving NI (%)		
Control	(G6)N2-H21…O2(C17)	(G6)N1-H1…N3(C17)	(C17)N4-H42…O6(G6)
	99.3	99.7	96.1
NI(*syn*) • A(*anti*)	(NI6)N7…H61-N6(A17)	(NI6)O2P…H22-N2(NI6)	(NI6)O4'…H3-N3(NI6)
	55.8	54.8	54.8
	(NI6)O61…H1-N1(G16)	(C5)O3'…H3-N3(NI6)	
	76.1	89.4	
NI(*syn*) • C(*anti*)	(NI6)N7…H41-N4(C17)	(NI6)O2P…H22-N2(NI6)	(NI6)O5'…H3-N3(NI6)
	74.7	59.2	67.4
	(NI6)O4'…H3-N3(NI6)		
	70.4		
NI(*syn*) • G(*anti*)	(NI6)N7…H21-N2(G17)	(NI6)O62…H1-N1(G17)	(NI6)O2P…H22-N2(NI6)
	94.7	87.0	59.2
	(NI6)O4'…H3-N3(NI6)	(C5)O3'…H3-N3(NI6)	
	59.4	64.5	
NI(*syn*) • T(*anti*)	(NI6)O5'…H3-N3(NI6)	(NI6)O4'…H3-N3(NI6)	
	84.8	67.9	
NI(*anti*) • A(*anti*)	(NI6)O62…H61-N6(A17)		
	58.9		
NI(*anti*) • A(*syn*)	(A17)N7…H1-N1(NI6)	(NI6)O61…H21-N2(NI6)	
	63.8	62.1	
NI(*anti*) • C(*anti*)	(NI6)O61…H41-N4(C17)	(NI6)O62…H41-N4(C17)	
	91.6	57.3	
NI(*anti*) • G(*anti*)	(NI6)O61…H21-N2(G17)	(NI6)O61…H1-N1(G17)	(NI6)O62…H1-N1(G17)
	97.2	71.7	82.4
NI(*anti*) • G(*syn*)			
NI(*anti*) • T(*anti*)	(NI6)O61…H3-N3(T17)	(NI6)O62…H3-N3(T17)	
	80.3	51.6	

[a]: Summed values are totals for the indicated hydrogen bonds. Only hydrogen bonds with occupancy greater than 50% are listed.

Table S7: MM-PBSA free energy analysis components[a] for each DNA duplex (kcal/mol).

	NI(*syn*)•A(*anti*)	NI(*anti*)•A(*anti*)	NI(*anti*)•A(*syn*)	NI(*syn*)•C(*anti*)	NI(*anti*)•C(*anti*)
E_{elec}	453.0(50.7)	367.4(55.3)	299.9(33.6)	262.2(51.9)	273.3(44.0)
E_{vdw}	-207.1(10.0)	-200.3(9.49)	-199.2(8.6)	-196.8(10.2)	-197.1(9.6)
E_{int}	1030.0(20.7)	1032.9(22.9)	1031.7(19.7)	1023.1(18.8)	1033.4(20.6)
E_{MM}	1275.9(50.3)	1200.0(57.2)	1132.4(37.7)	1088.5(53.3)	1109.6(41.0)
$G_{nonpolar}$	19.2(0.3)	19.4(0.2)	19.5(0.2)	19.3(0.2)	19.36(0.2)
G_{PB}	-5812.8(47.3)	-5739.4(52.7)	-5673.8(30.6)	-5716.3(47.1)	-5731.7(38.1)
G_{sol}	-5793.7(47.4)	-5720.0(52.8)	-5654.3(30.6)	-5696.9(47.2)	-5712.3(38.2)
$<G_{PB}+E_{elec}>$	-5359.9(13.4)	-5372.0(11.3)	-5373.9(10.8)	-5454.1(11.9)	-5458.3(12.9)
$<E_{MM}+G_{sol}>$	-4517.8(17.3)	-4520.1(20.4)	-4521.8(20.0)	-4608.4(18.8)	-4602.7(16.4)
-TS	570.1(2.6)	581.1(0.9)	580.8(2.0)	575.4(1.75)	574.2(1.0)
G_{tot}	-5087.8(19.9)	-5101.2(21.3)	-5102.6(22.0)	-5183.8(20.6)	-5176.9(17.4)

	NI(*syn*)•G(*anti*)	NI(*anti*)•G(*anti*)	NI(*anti*)•G(*syn*)	NI(*syn*)•T(*anti*)	NI(*anti*)•T(*anti*)
E_{elec}	294.6(50.0)	315.8(41.0)	327.4(39.8)	381.2(54.5)	395.5(59.3)
E_{vdw}	-207.1(9.0)	-204.9(8.6)	-206.5(8.9)	-202.9(11.3)	-208.6(11.1)
E_{int}	1028.1(20.7)	1038.9(18.2)	1031.2(19.9)	1028.4(21.3)	1028.2(21.6)
E_{MM}	1115.6(53.9)	1149.8(41.3)	1152.1(44.7)	1206.7(53.5)	1215.1(60.0)
$G_{nonpolar}$	19.4(0.2)	19.5(0.2)	19.4(0.2)	19.3(0.4)	19.1(0.2)
G_{PB}	--5716.5(48.3)	-5738.6(38.7)	-5746.6(37.4)	-5708.7(50.4)	-5721.7(57.3)
G_{sol}	-5697.1(48.4)	-5719.1(38.7)	-5727.2(37.4)	-5689.4(50.3)	-5702.6(57.4)
$<G_{PB}+E_{elec}>$	-5421.9(9.5)	-5422.8(10.3)	-5419.3(11.3)	-5327.5(12.9)	-5326.3(13.1)
$<E_{MM}+G_{sol}>$	-4581.5(19.6)	-4569.4(18.9)	-4575.1(17.3)	-4482.7(19.5)	-4487.5(17.8)
-TS	577.9(3.3)	582.6(3.0)	574.0(3.2)	577.3(3.1)	577.3(3.1)
G_{tot}	-5159.4(22.8)	-5152.0(22.0)	-5149.1(20.5)	-5060.0(22.6)	-5060.8(20.3)

[a]: E_{elec} : electrostatic energy, E_{vdW} : van der Waals energy, E_{int} : internal energy, E_{MM} : molecular mechanical energy, $G_{nonpolar}$: nonpolar solvation free energy, G_{PB} : electrostatic solvation free energy, G_{sol} : total solvation free energy, -TS: the product of temperature (K) and entropy, G_{tot} : total free energy. The standard deviations are given in parentheses.

Molecular Dynamics Protocol:

MD simulations were carried out with the SANDER module from the AMBER 8.0 suite (*53*). The cutoff value of the nonbonded interactions was set at 9.0 Å. The SHAKE algorithm (*65*) was employed to constrain bonds involving hydrogen with a tolerance of 10^{-6} Å. A 2 fs time step was used in the dynamics simulations, and the translational motion of the center of mass was removed every 1 ps (*63*). Visual inspection of the trajectories revealed no abnormal overall rotation of the DNA duplexes, indicating that energy leakage from the internal motion to global rotation through the "flying ice cube effect" (*63*) was not contributing in this case. The particle mesh Ewald (PME) method (*61, 62*) was applied to treat long range electrostatic interactions. To neutralize the solute, 20 Na^+ counterions were added to the system with the LEAP module. The system was solvated by adding a rectangular box of TIP3P water (*64*) 10 Å from the boundary of the DNA in each direction. The sizes of the box and the numbers of water molecules added to the system are shown in Table S4.

To relax the solvent, the solute DNA was held fixed by 25 kcal/mol restraints, and the system was minimized with 1000 steps of steepest descent minimization. This was followed by a 25ps MD simulation with the same constraints on the DNA under constant volume. At this stage, the temperature was raised from 10 K to 300 K during the first 20 ps, using the Berendsen coupling algorithm (*60*) with a coupling parameter of 1.0 ps. Then, 5 rounds of minimization with 600 steps of steepest descent minimization in each round were performed to uniformly decrease the restraints on the DNA from 25 kcal/mol to 0 kcal/mol (*66*). Next, a 60 ps MD simulation was applied to equilibrate the system under constant volume, with the temperature raised uniformly from 10 K to

300 K in the first 40 ps. Finally, an MD simulation of 8 ns was carried out under constant pressure for data production.

MM-PBSA Protocol:

We employed the following standard protocol: First, solvent and counterions were removed from the system. Then, 100 snapshots were extracted at 15 ps intervals from the trajectory window of 1.5 – 3.0 ns. These were employed to compute individual molecular mechanical energies (E_{MM}) and solvation free energies ($G_{solvation}$), and then to obtain their average values. The same force field and parameters (*53-55*) used in the MD simulations was applied to calculate the molecular mechanical energies(E_{MM}), but without cutoff for nonbonded interactions. To calculate the G_{PB}, we used a grid spacing of 0.5 Å with a cubic lattice in which the largest linear dimension of the molecule occupied 80% of the lattice. The dielectric constants were set to 1 and 80 for the gas and solution phases, respectively. To consider the effect of salt on the solvation free energy, we added a physiological 0.15 M salt concentration. Partial charges employed were the same set as for the MD simulation. For the G_{PB} calculation, we carried out 300 iterations using the linear PB equation followed by 1000 iterations with non-linear PB equation (*85*). To calculate the $G_{nonpolar}$, we applied the PARSE parameter set for atomic radii, and used a probe radius of 1.4 Å to define the dielectric boundary (*82*). To calculate the solute entropy, we employed the following strategy: first, we energy-minimized each of the 100 snapshots from the MD trajectory using 1000 steps of steepest descent followed by conjugate gradient until the RMSD of the Cartesian elements of the gradient converged to less than 10^{-4} kcal/(mol^{-1} · Å). The dielectric constant was set to be 4r (where r is the interatomic distance in Å) to mimic the solution environment (*85*). Then, we selected five

minimized structures to compute the entropy with the NMODE module of the AMBER 8.0 suite and averaged the resultant values.