Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Electronic Supporting Information

for

Small molecule binders recognize DNA microstructural variations via an induced fit mechanism

E. Kathleen Carter[a], Sarah Laughlin-Toth[a,b], Thomas Dodd[a], W. David Wilson[a], and Ivaylo Ivanov[a]

[a]Department of Chemistry and Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, GA, USA

*Corresponding author:

Ivaylo Ivanov

Email: iivanov@gsu.edu

Tel: +1 404 413 5529

Fax: +1 404 413 5505

*Corresponding author:

W. David Wilson

Email: wdw@gsu.edu

Tel: +1 404 413 5503

Fax: +1 404 413 5505

Present Address: [b]Aflac Cancer and Blood Disorders Center, Department of Pediatrics, Emory University School of Medicine, Atlanta, GA, USA

Supplementary Methods Information

Ab initio determination of pKa

In order to confirm the protonation state of the azabenzimidazole ring of DB2277, we used the Born-Haber thermodynamic cycle depicted in Figure S3A [1]. Accordingly, the definition of ΔG^0 is given by

$$\Delta G^0 = -\Delta G_{solv}(HA) + \Delta G_{vap}(H_2O) + \Delta G^0_{vac} + \Delta G_{solv}(A^-) + \Delta G_{solv}(H_3O^+)$$

where both $-\Delta G_{solv}(HA)$ and $\Delta G_{solv}(A^-)$ are taken into account to determine the energy difference between optimized geometries in the solution and gas phases (ΔE_{relax}) in accordance with .

$$-\Delta G_{solv}(HA) = -[G_{solv}(HA) - G_{gas}(HA) + \Delta E_{relax}(HA)]$$
$$\Delta G_{solv}(A^-) = G_{solv}(A^-) - G_{gas}(A^-) + \Delta E_{relax}(A^-)$$

The pKa is related to ΔG^0 through

$$pKa = \frac{\Delta G^0}{2.303RT}$$

The values for $\Delta G_{vap}(H_2O)$ and $\Delta G_{solv}(H_3O^+)$ are -6.28 kcal/mol and -264.61 kcal/mol, respectively. Additionally, we accounted for our experimental conditions of 1 atm at 310 K to make the final expression for the determination of pKa to be

$$pKa\ (1\ atm, 310\ K) = \frac{-\Delta G_{solv}(HA) + \Delta G^0_{vac} + \Delta G_{solv}(A^-) - 268.921}{1.418}$$

All calculations were performed using the 6-31G+* basis set at the B3LYP DFT level of theory in Gaussian09. We first built DB2277 in four different protonation states on the azabenzimidazole ring: (1) doubly-protonated at N3 and N2, (2) singly-protonated at N3 (3) singly-protonated at N2 and (4) completely de-protonated. Each molecule separately underwent geometry optimization using the polarizable continuum model (PCM) [2,3] as implemented in Gaussian09 to mimic solution phase conditions. Then the molecule was transferred to the gas phase and the internal energy was calculated while it maintained the PCM optimized configuration. Geometry optimization of the molecule was then performed in the gas phase starting with the optimized PCM structure. The difference in energy between geometry optimized gas phase structure and PCM structure is the relaxation energy, ΔE_{relax}. Calculation of the optimized geometries allowed us to compute theoretical pKa values using the Born-Haber thermodynamic cycle. Results can be found in Table S3 and Figure S3B.

Supplementary Figures and Tables

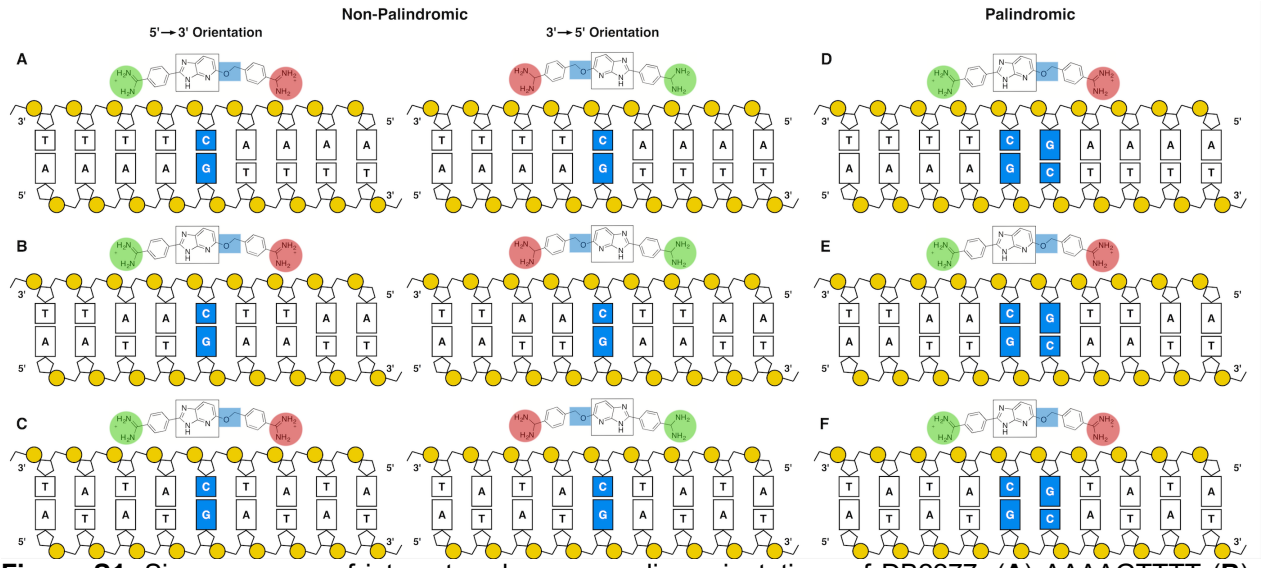

Figure S1. Six sequences of interest and corresponding orientations of DB2277. (**A**) AAAAGTTTT (**B**) AATTGAATT (**C**) ATATGATAT (**D**) AAAAGCTTTT (**E**) AATTGCAATT (**F**) ATATAGCATAT. (**A-C**) The single G•C sequences are non-palindromic and the asymmetric small molecule DB2277 was positioned in two orientations in the minor groove. (**5' → 3' Orientation**) In the 5'→ 3' orientation, amidine-1 (green) is closest to the 5' end of the target sequence. (**3' → 5' Orientation**) In the 3'→5' orientation, amidine-1 is closest to the 3' end of the target sequence. (**D-F**) The GpC sequences are palindromic, therefore, only one orientation of DB2277 was required.

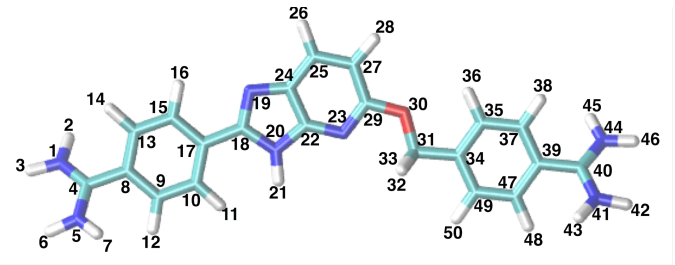

Figure S2. DB2277 atom numbers

Table S1. RESP charges for DB2277

	Atom Name	*Atom Type*	*Charge*		*Atom Name*	*Atom Type*	*Charge*
1	N4	n2	-0.8408	**26**	H2	ha	0.1927
2	H8	hn	0.4461	**27**	C3	ca	-0.3945
3	H9	hn	0.4461	**28**	H1	ha	0.1843
4	C12	ca	0.7536	**29**	C13	ca	0.5710
5	N5	n2	-0.8408	**30**	O1	os	-0.2454
6	H10	hn	0.4461	**31**	C14	c3	-0.0063
7	H11	hn	0.4461	**32**	H12	h1	0.0832
8	C11	cc	-0.1912	**33**	H13	h1	0.0832
9	C10	cc	-0.1008	**34**	C15	ca	0.2101
10	C8	cd	-0.0635	**35**	C16	ca	-0.1700
11	H5	ha	0.1213	**36**	H14	ha	0.1474
12	H7	ha	0.1375	**37**	C18	ca	-0.0480
13	C9	cc	-0.1008	**38**	H16	ha	0.1324
14	H6	ha	0.1375	**39**	C20	ca	-0.2510
15	C7	cd	-0.0635	**40**	C21	ca	0.7810
16	H4	ha	0.1213	**41**	N6	n2	-0.8413
17	C6	cd	0.0221	**42**	H18	hn	0.44690
18	C5	cc	0.4458	**43**	H21	hn	0.44690
19	N3	nc	-0.4495	**44**	N7	n2	-0.8413
20	N2	na	-0.7654	**45**	H19	hn	0.44690
21	H3	hn	0.4364	**46**	H20	hn	0.44690
22	C2	ca	0.8120	**47**	C19	ca	-0.0480
23	N1	nb	-0.7281	**48**	H17	ha	0.13240
24	C1	ca	0.0201	**49**	C17	ca	-0.1700
25	C4	ca	-0.0922	**50**	H15	ha	0.14740

Table S2. AMBER force field modification file (.frcmod). Parameters not found in the GAFF force field were added by ANTECHAMBER or calculated using *ab initio* methods in Gaussian09

	DB2277 Parameterization				
MASS					
BOND					
ANGLE					
cc-nc-ca	70.56	107.15			Added by ANTECHAMBER
os-c3-ca	100.14	109.18			Calculated using Gaussian09
c3-os-ca	70.35	119.79			Calculated using Gaussian09
ca-ca-n2	54.0	119.66			Cheatham, T.E. DOI: 10.1021/ja025660d
cc-ca-n2	54.0	119.66			Cheatham, T.E. DOI: 10.1021/ja025660d
n2-ca-n2	70.0	122.00			sp2 geometry
DIHE					
c3-os-ca-ca	2	3.24	180.0	2.0	Calculated using Gaussian09
na-cc-cd-cd	4	3.39	180.0	2.0	Calculated using Gaussian09
cc-cc-ca-n2	4	1.99	180.0	2.0	Calculated using Gaussian09
n2-ca-ca-ca	4	1.70	180.0	2.0	Calculated using Gaussian09
ca-c3-os-ca	3	9.00	0.0	1.0	Calculated using Gaussian09
ca-ca-ca-n2	4	0.79	327.0	-4.0	Cheatham, T.E. DOI: 10.1021/ja025660d
ca-ca-ca-n2	4	-3.12	0.0	-4.0	Cheatham, T.E. DOI: 10.1021/ja025660d
ca-ca-ca-n2	4	0.610	90.0	1.0	Cheatham, T.E. DOI: 10.1021/ja025660d
ca-ca-n2-hn	4	9.60	180.0	2.0	Added from parm99.dat
cc-ca-n2-hn	4	9.60	180.0	2.0	Added from parm99.dat
IMPROPER					
cc-n2-ca-n2	1.1	180.0	2.0		Using default value
ca-cc-cc-cc	1.1	180.0	2.0		Using default value
cc-cd-cc-ha	1.1	180.0	2.0		Using default value
cc-cd-cd-ha	1.1	180.0	2.0		Using default value
cc-cd-cd-cd	1.1	180.0	2.0		Using default value
cd-na-cc-nc	1.1	180.0	2.0		Using default value
ca-cc-na-hn	1.1	180.0	2.0		General improper torsional angle (2 general atom types)
ca-na-ca-nb	1.1	180.0	2.0		Using default value
ca-ca-ca-nc	1.1	180.0	2.0		Using default value
ca-ca-ca-ha	1.1	180.0	2.0		General improper torsional angle (2 general atom types)
ca-nb-ca-os	1.1	180.0	2.0		Using default value
ca-ca-ca-ca	1.1	180.0	2.0		Using default value
ca-n2-ca-n2	1.1	180.0	2.0		Using default value
NONBON					

Table S3 Computed pKa values for azabenzimidazole nitrogens N2 and N3

Protonation	*pKa₁*	*pKa₂*
N3	4.02	10.97
N2	4.37	13.12

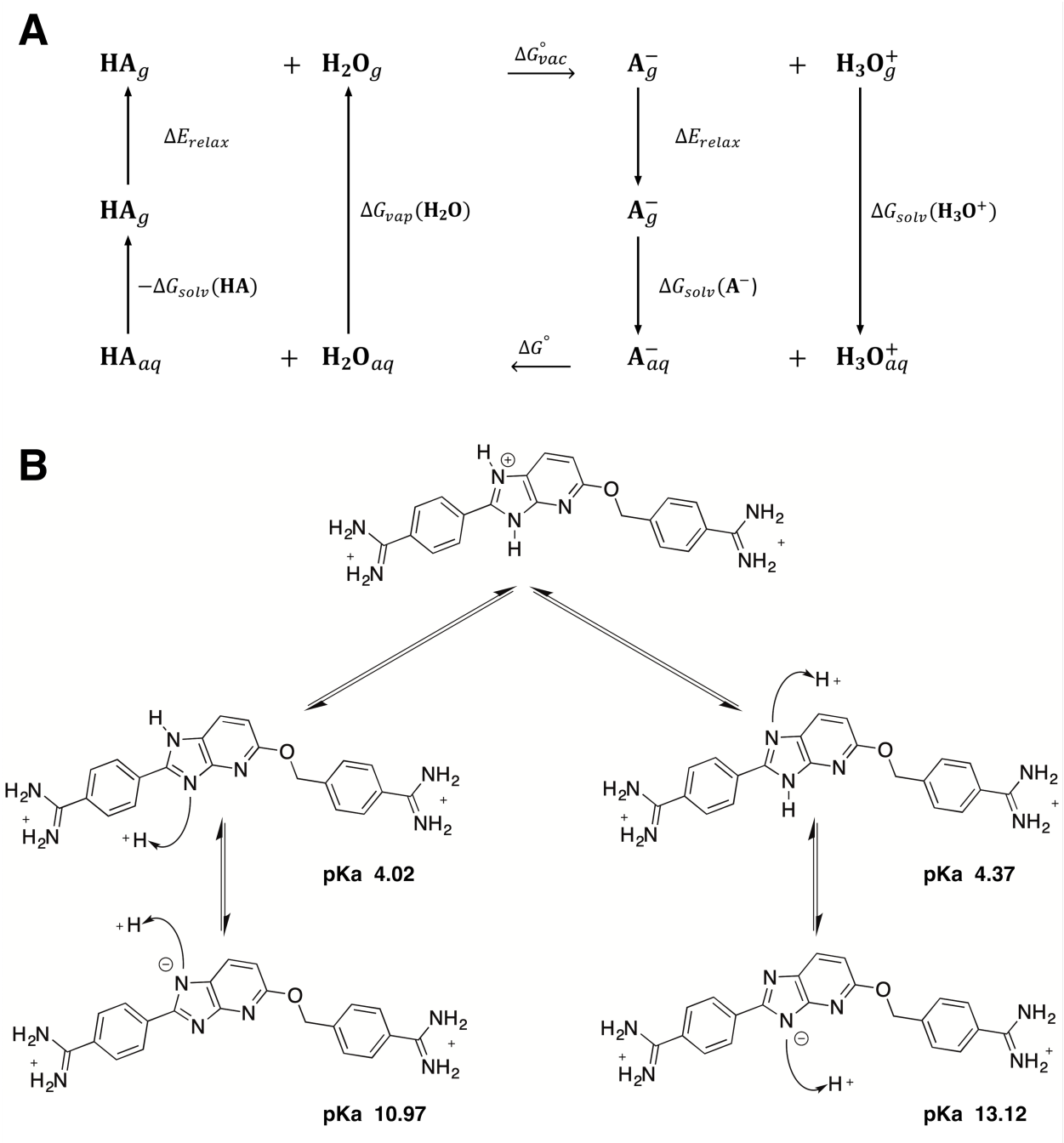

Figure S3 Theoretically determined pKa values for the imidazole ring of DB2277. (**A**) Thermodynamic cycle used in the determination of the pKa's based on [1]. (**B**) Depiction of the deprotonation and resultant pKa at each site.

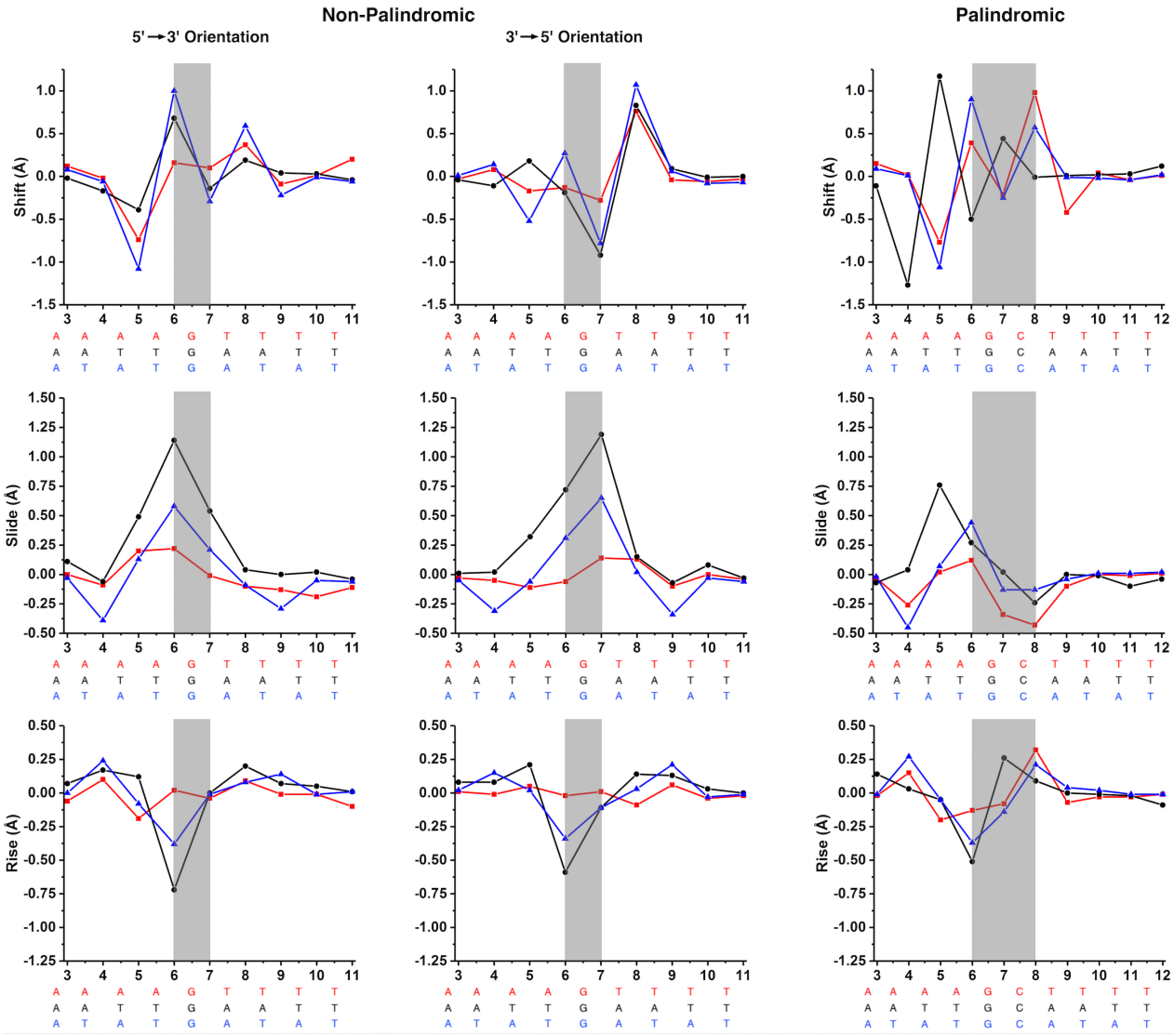

Figure S4. Change in translational inter-base pair parameters upon binding. Translational parameters (shift, slide, rise) for sequences AAAAGTTTT and AAAAGCTTTT are labeled in red, AATTGAATT and AATTGCAATT are in black and ATATGATAT ATATGCATAT are in blue.

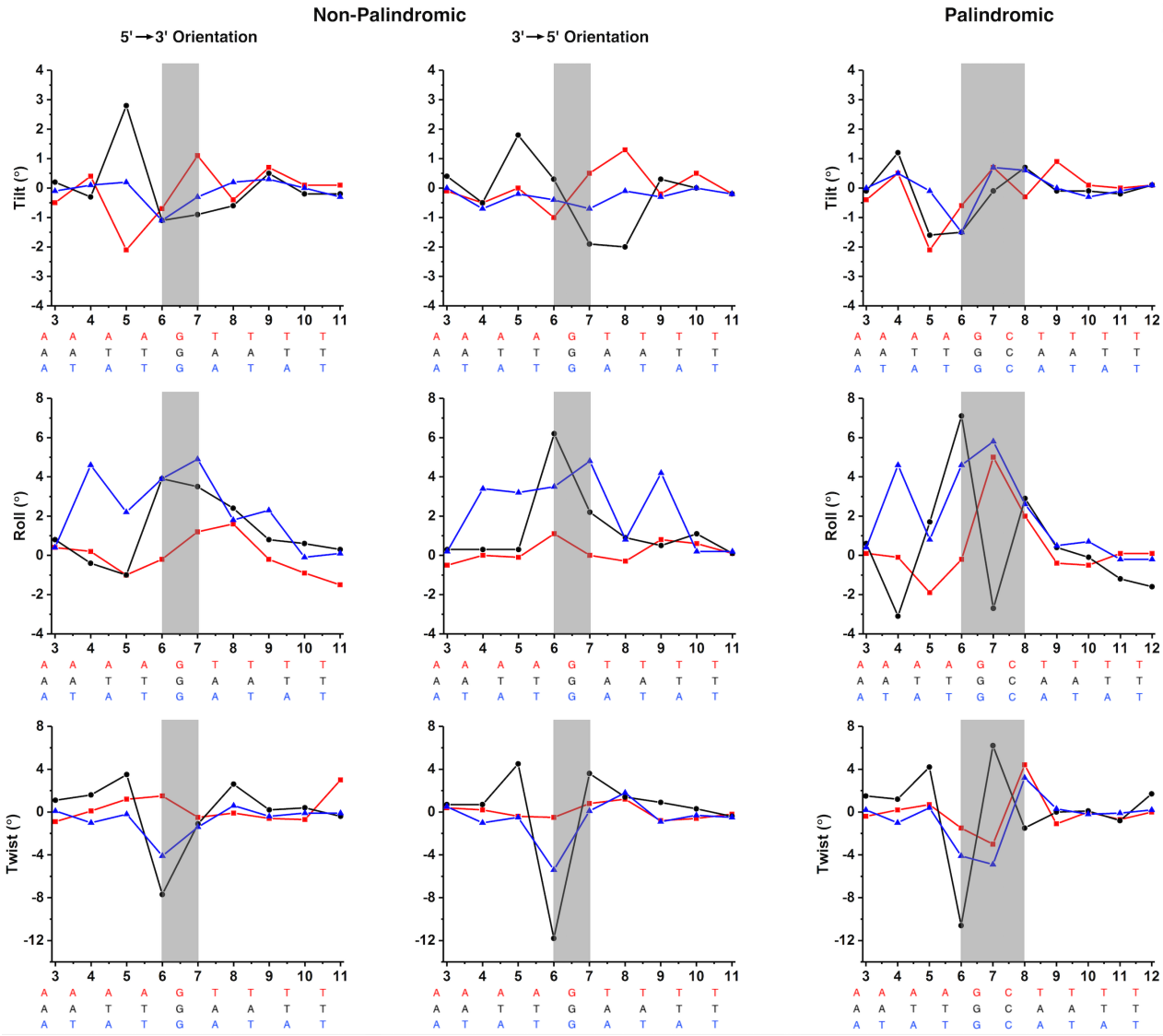

Figure S5. Change in rotational inter-base pair parameters upon binding. Rotational parameters (tilt, roll, twist) for sequences AAAAGTTTT and AAAAGCTTTT are labeled in red, AATTGAATT and AATTGCAATT are in black and ATATGATAT ATATGCATAT are in blue.

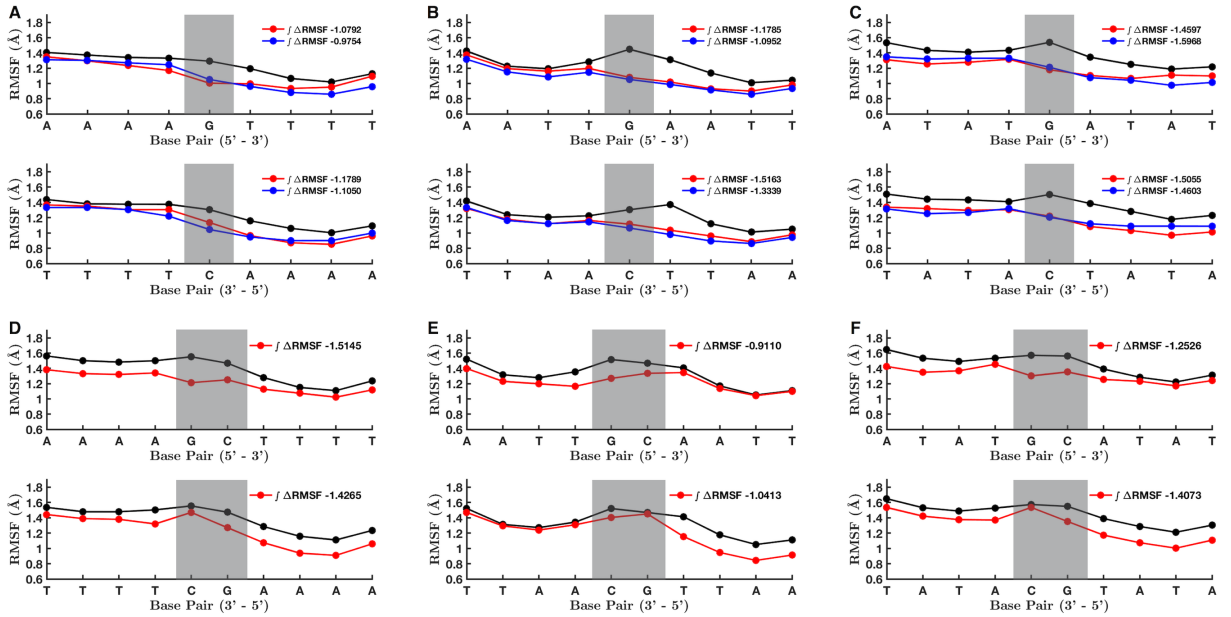

Figure S6. Root mean square fluctuation (RMSF) per base pairs shows increased rigidity (decreased RMSF) upon binding for all six sequences. (**A**) AAAAGTTTT, (**B)** AATTGAATT, (**C**) ATATGATAT, (**D**) AAAAGCTTTT, (**E**) AATTGCAATT, (**F**) ATATAGCATAT. The RMSF for the unbound structures are shown in black for all sequences. The nonpalindromic G•C sequences have two binding orientations (5'→3', red) (3'→5', blue). The palindromic (GpC) sequences only have one binding orientation (red). The target binding sites, G•C and GpC, are highlighted in gray. Integrated change in RMSF for each bound state are shown on the top right for each strand

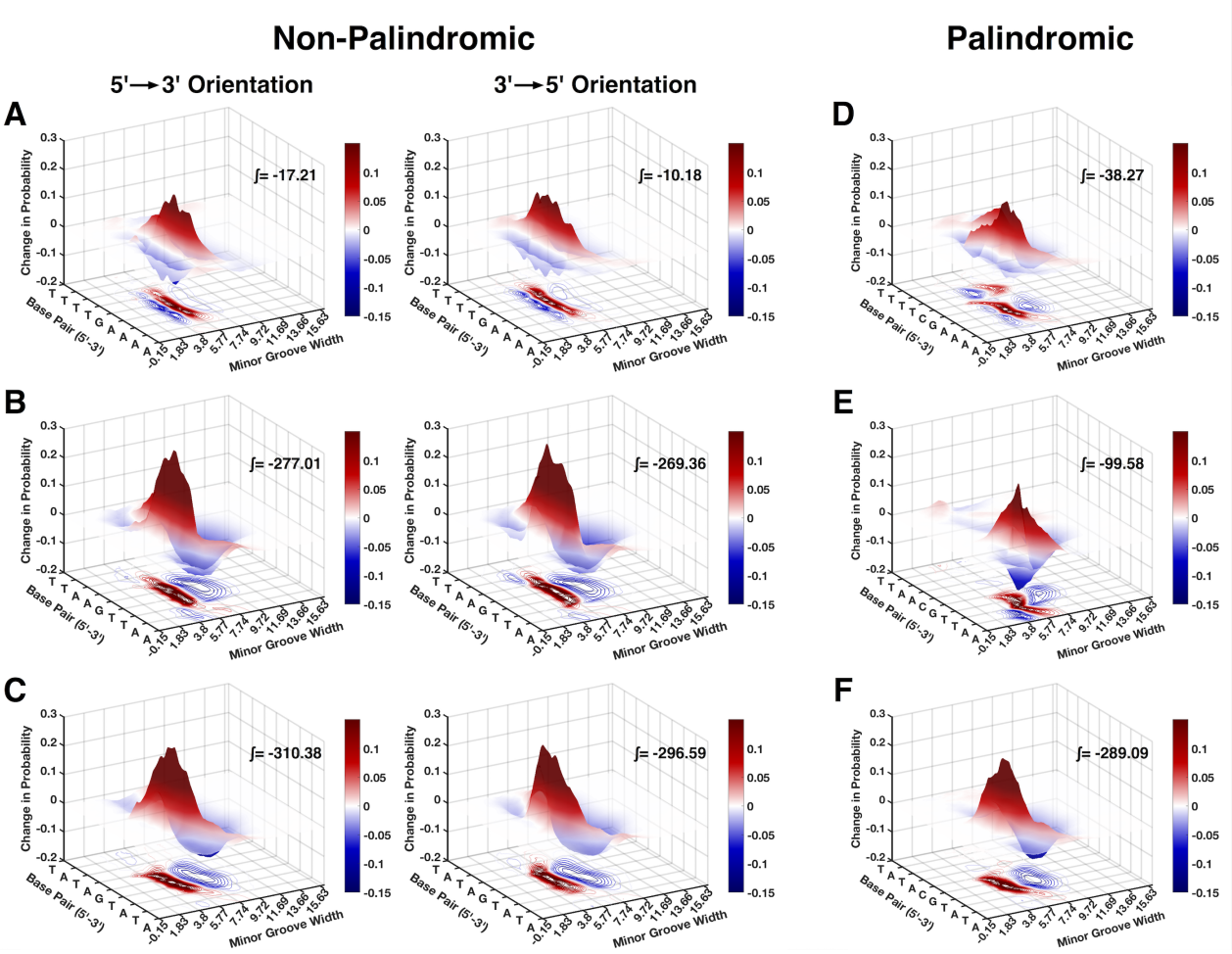

Figure S7. Change in probability of minor groove width upon binding. (**A**) AAAAGTTTT, (**B**) AATTGAATT (**C**) ATATGATAT (**D**) AAAAGCTTTT (**E**) AATTGCAATT (**F**) ATATAGCATAT. The nonpalindromic G•C sequences have two binding orientations (5'→3' and 3'→5'). The palindromic (GpC) sequences only have one binding orientation. The color gradient indicates increasing probability (blue to red) of width in Angstroms (Å). Surfaces were integrated to get a total change in width. Values are reported alongside figures.

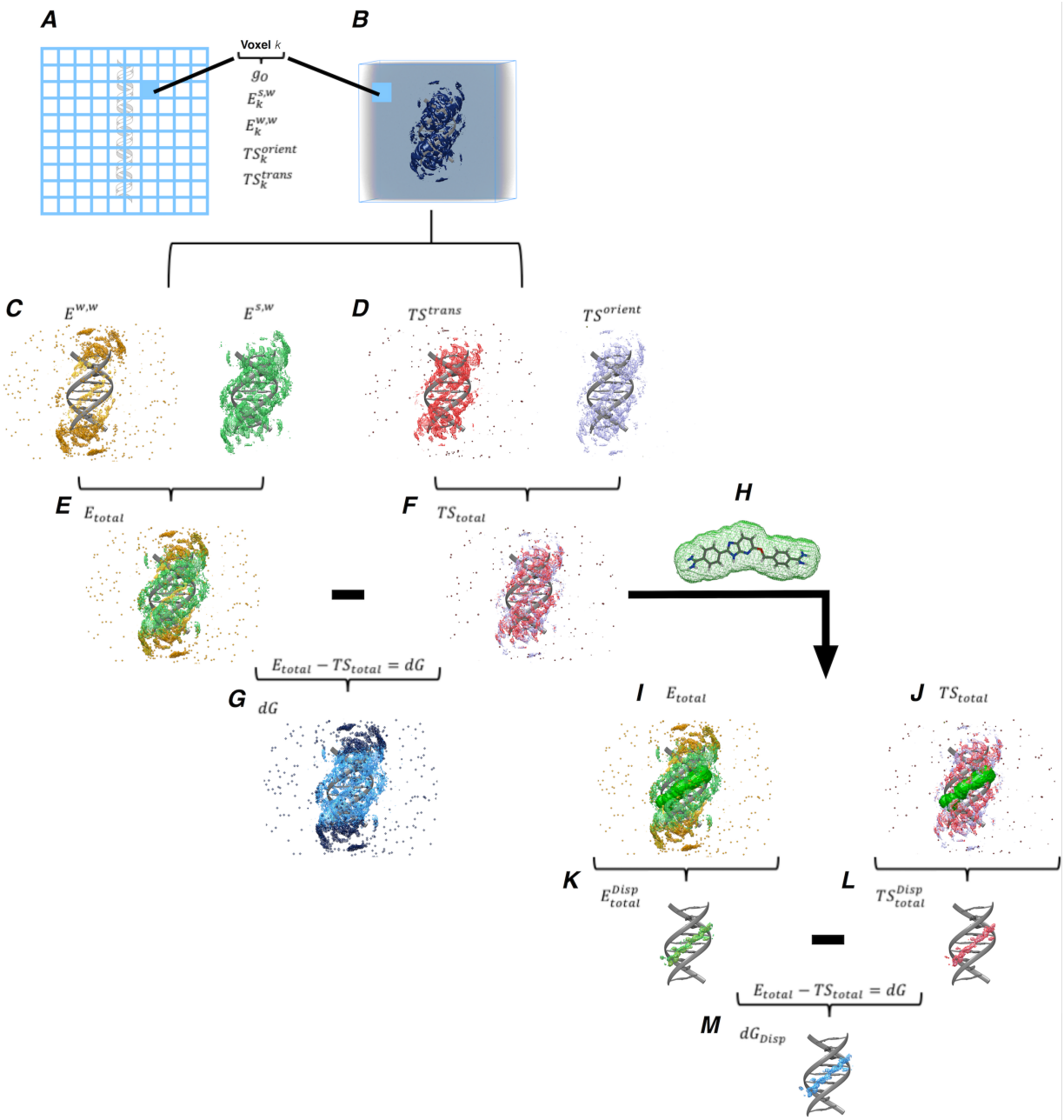

Figure S8. Implementation of grid inhomogenous solvation theory (GIST). (**A**) The system is discretized in to equal volume voxels. From this, density of the solvent (g_O), solute-water enthalpy ($E_k^{s,w}$), water-water enthalpy ($E_k^{w,w}$), translations entropy (TS_k^{trans}) and orientational entropy (TS_k^{orient}) are reported per voxel. (**B, C, D**) Using the reported solvent density, the energetic calculations can be restricted to the structured waters near the DNA duplex. (**E, F)** The solute-water and water-water enthalpy terms are summed to get total enthalpy ($E_{total} = E^{s,w} + E^{w,w}$) and the translational and orientational entropy terms are summed to get total entropy ($TS_{total} = TS^{trans} + TS^{orient}$). (**G**) Subtracting the total entropy map from the total enthalpy map results an isodensity of free energy ($\Delta G \quad = E_{total} - TS_{total}$). (**C-G**) The isosurfaces can be visualized in order to diagnose solvation hotspots. (**H**) A binary density map is created to represent the volume of water displaced by DB2277. (**I, J**) The binary map is multiplied the total entropy isodensity and the total enthalpy isodensity. (**K, L)** This produces maps of the enthalpy and entropy of the water displaced by DB2277 (E_{totaal}^{Disp} , TS_{total}^{Disp}, respectively). (**M**) The entropy of the expelled water can be subtracted from the enthalpy of the expelled water to determine the free energy of the displaced waters $\Delta G_{Disp} = E_{total}^{Disp} - TS_{total}^{Disp}$.

Table S4. Thermodynamic data for GIST analysis of the waters displaced from the minor groove.

	Enthalpy (ΔE) (kcal/mol)	**Entropy (TΔS)** (kcal/mol)	**Free Energy (ΔG)** (kcal/mol)
AAAAGTTTT			
Displaced Water (5'—>3')	-10.13 ± 0.0003	-1.04 ± 3.46e-05	-9.09 ± 0.0003
Displaced Water (3'—>5')	-10.26 ± 0.0003	-1.08 ± 3.58e-05	-9.18 ± 0.0003
AATTGAATT			
Displaced Water (5'—>3')	-6.14 ± 0.0002	-0.66 ± 2.24e-05	-5.45 ± 0.0002
Displaced Water (3'—>5')	-6.57 ± 0.0002	-0.71 ± 2.40e-05	-5.86 ± 0.0002
ATATGATAT (5'—>3')			
Displaced Water (5'—>3')	-3.50 ± 0.0001	-0.30 ± 1.21e-05	-3.20 ± 0.0001
Displaced Water (3'—>5')	-3.33 ± 0.0001	-0.29 ±1.17e-05	-3.04 ± 0.0001
AAAAGCTTTT			
Displaced Water	-6.24 ± 0.0002	-0.57 ±2.17e-05	-5.67 ± 0.0002
AATTGCAATT			
Displaced Water	-3.97 ± 0.0002	-0.56 ± 2.00e-05	-3.41 ± 0.0002
ATATGCATAT			
Displaced Water	-2.83 ± 0.0001	-0.25 ± 9.66e-06	-2.59 ± 0.0001

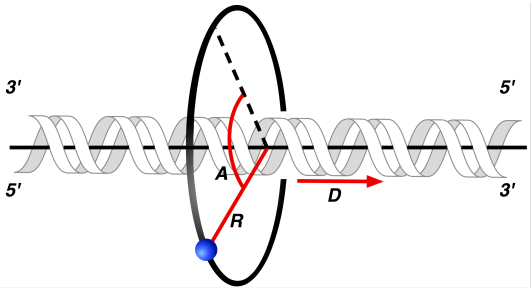

Figure S9. Curvilinear helicoidal coordinates system of Canion (***R***) radial distance in Å (***A***), angle from reference point (dotted line) which tracks the helical twist of the nucleic acid, and (***D***) longitudinal distance in Å along the base pair sequence.

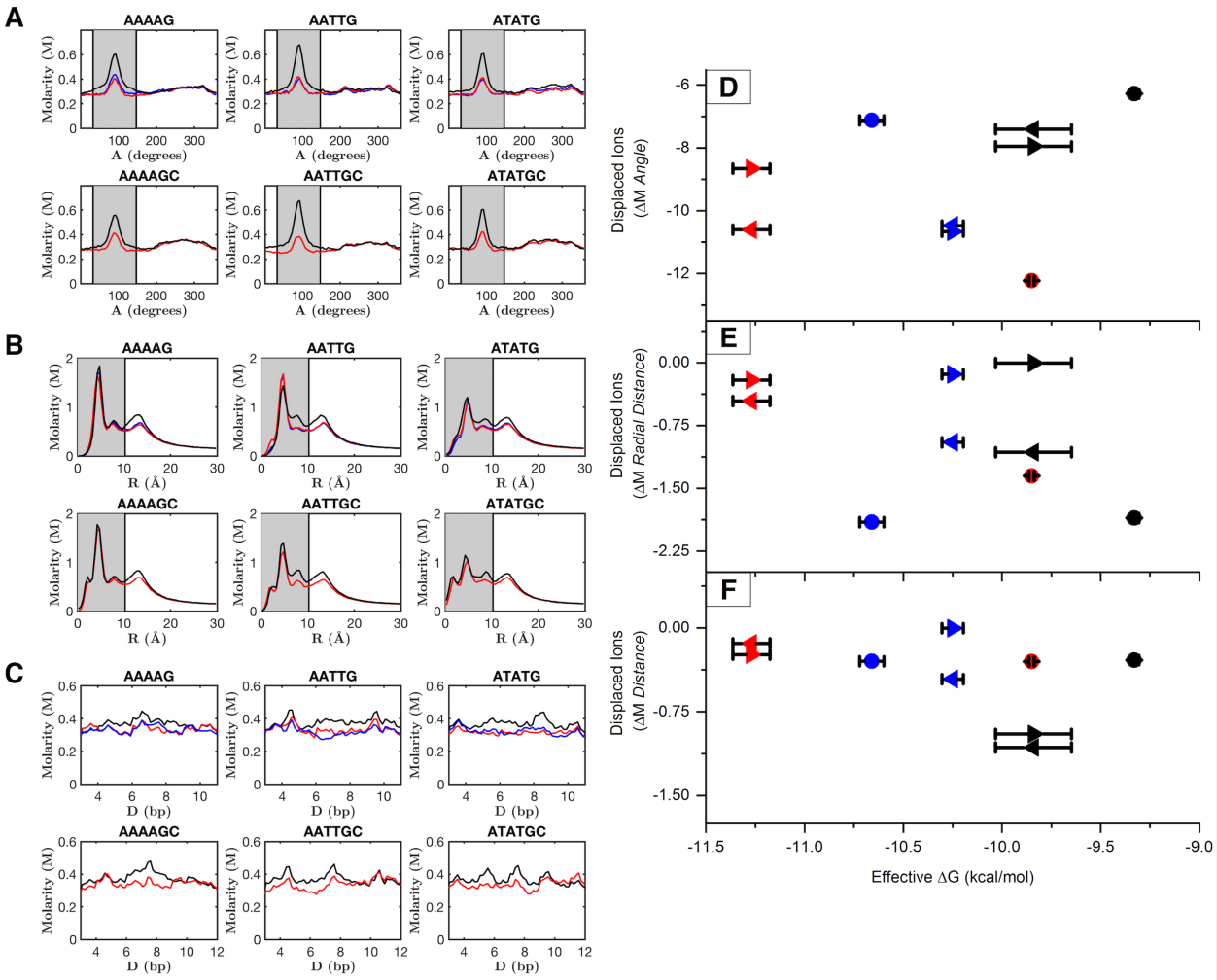

Figure S10. Change in distribution of positive ions upon binding. (**A**) Ion distribution in the free and bound states in the angular dimension (***A***). (**B**) Ion distribution in the free and bound states in the radial distance dimension (***R***). Ion distribution in the longitudinal distance dimension (***D***). (**A,C,B**) The free structures are shown as black lines. The nonpalindromic G•C sequences have two binding orientations (5'→3', red) (3'→5', blue). The palindromic (GpC) sequences only have one binding orientation (red). Integrated change in molarity is shown on each plot. (**D**) Integrated change in molarity in the minor groove (33° < ***A*** < 147°) plotted against effective ΔG. (**E**) Integrated change in molarity in the duplex (***R*** < 10.25 Å) plotted against effective ΔG. (**F**) Integrated change in molarity along the binding sequence (***D***) plotted against effective ΔG.

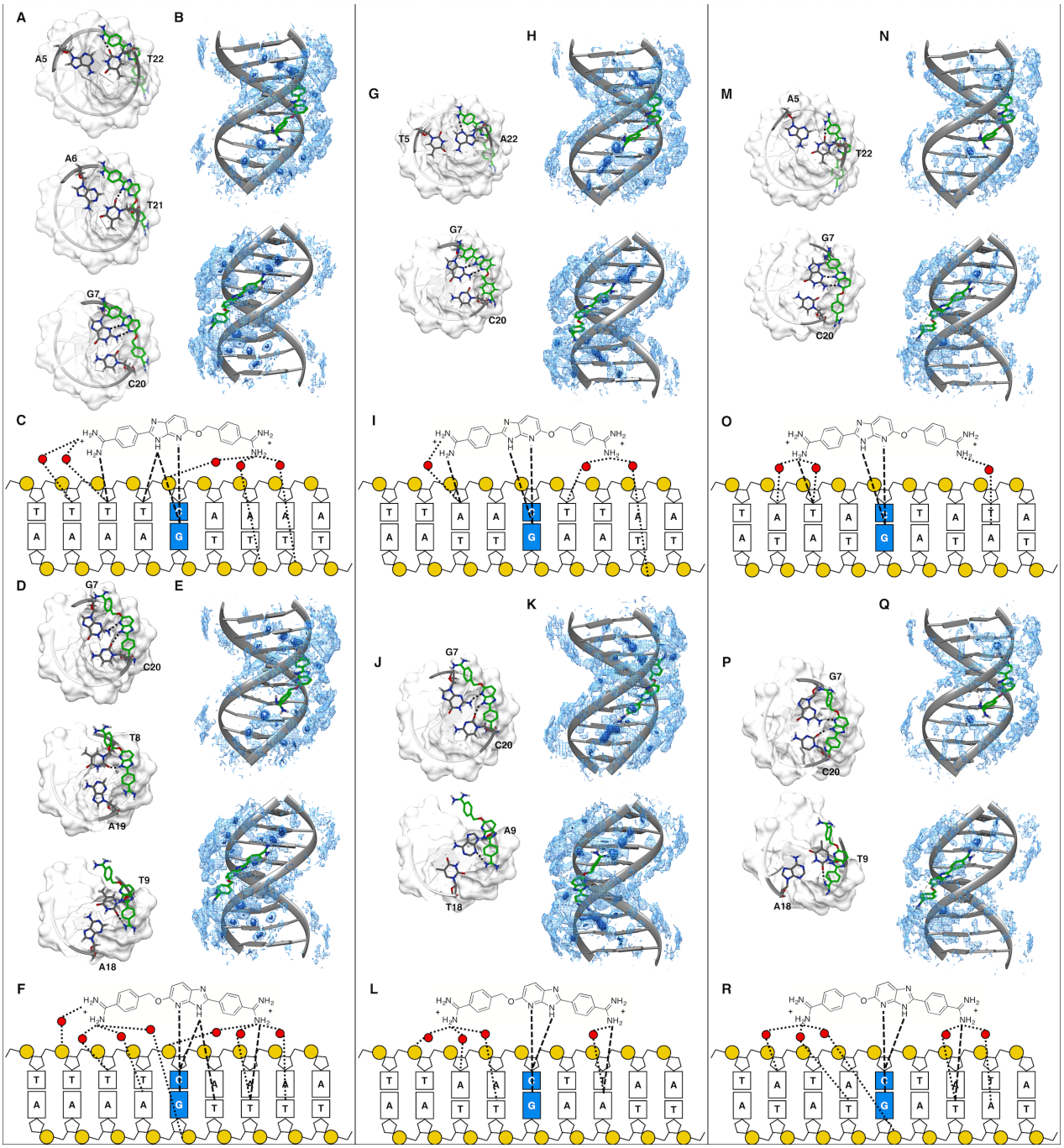

Figure S11. DB2277 contacts the DNA via direct hydrogen bonds and water mediated contacts in the non-palindromic sequences. Contact with persistence > 15% are drawn between the DB2277 and the DNA bases or backbone where the hydrogen bond is formed. Symmetric hydrogen bonds, like those formed by the amidines, are resented one hydrogen bond from that site. (**A-F**) In sequence AAAAGTTTT, DB2277 forms persistent contacts to three base pairs of the sequence with amidine-1 and the core azabenzimidiazole in both the 5'→3' (**A-C**) and 3'→5' (**D-F**) orientations. (**G-L**) Contacts between DB2277 and AATTGAATT are less numerous but several water-mediated contacts can be seen at amidine-1 and amindine-2. (**M-R)** The persistent contacts in the ATATGATAT sequence are highly similar to those of the AATTGAATT sequence. Water mediated contacts visualized as isodensity (isovalues blue mesh: -0.009, blue surface: -0.02) in panels **B, E, H, K, N,** and **Q** and as red dots in panels **C, F, I, L, O,** and **R**.

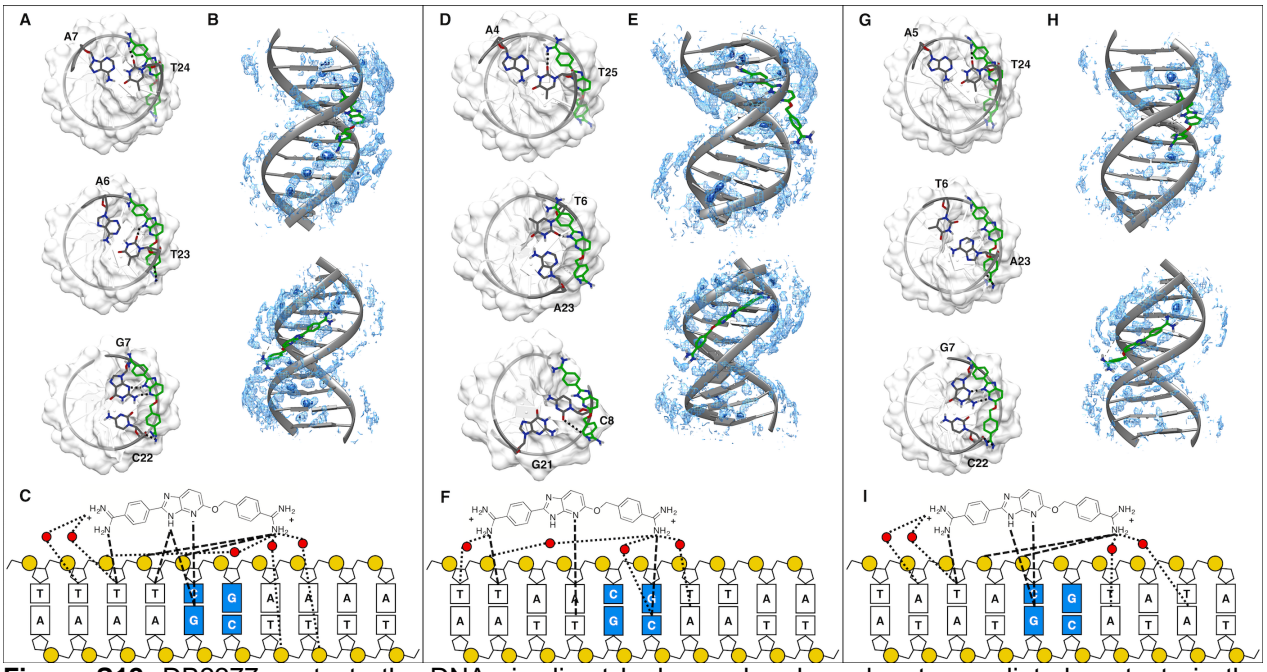

Figure S12. DB2277 contacts the DNA via direct hydrogen bonds and water mediated contacts in the palindromic sequences. Contact with persistence > 15% are drawn between the DB2277 and the DNA bases or backbone where the hydrogen bond is formed. Symmetric hydrogen bonds, like those formed by the amidines, are resented one hydrogen bond from that site. (**A-C**) In sequence AAAAGCTTTT, DB2277 forms persistent contacts to three base pairs of the sequence with amidine-1, the core azabenzimidiazole and amidine-2. (**D-F**) To optimize binding, DB2277 shifts one step upstream in the AATTGCAATT sequence and does not make contacts with the target guanine. (**G-I)** The persistent contacts in the ATATGCATAT sequence are highly similar to those of the AAAAGCTTTT sequence. Water mediated contacts visualized as isodensity (isovalues blue mesh: -0.009, blue surface: -0.02) in panels **B, E,** and **H** and as red dots in panels **C, F,** and **I**.

Table S5. Contact persistence between DB2277 and DNA sequences for the nonpalindromic sequences.

	Acceptor Residue	Residue Number	Acceptor	Donor Residue	Residue Number	Donor	Donor Hydrogen	Persistence
AAAAGTTTT (5'—>3')								
	DT	*22*	*O2*	*DB1*	*27*	*N4/N5*	*H9/H10/H11*	*0.9183*
	DB1	*27*	*N1*	*DG*	*7*	*N2*	*H22*	*0.8930*
	DT	*21*	*O2*	*DB1*	*27*	*N2*	*H3*	*0.5227*
	DG	*7*	*N3*	*DB1*	*27*	*N2*	*H3*	*0.3569*
AAAAGTTTT (3'—>5')								
	DB1	*27*	*N1*	*DG*	*7*	*N2*	*H22*	*0.8767*
	DT	*9*	*O2*	*DB1*	*27*	*N4/N5*	*H8/H9/H10/H11*	*0.7175*
	DC	*20*	*O2*	*DB1*	*27*	*N2*	*H3*	*0.5194*
	DT	*8*	*O2*	*DB1*	*27*	*N2*	*H3*	*0.3206*
AATTGAATT (5'—>3')								
	DB1	*27*	*N1*	*DG*	*7*	*N2*	*H22*	*0.8682*
	DA	*22*	*N3*	*DB1*	*27*	*N4/N5*	*H8/H10/H11*	*0.8312*
	DG	*7*	*N3*	*DB1*	*27*	*N2*	*H3*	*0.6869*
AATTGAATT (3'—>5')								
	DB1	*27*	*N1*	*DG*	*7*	*N2*	*H22*	*0.8941*
	DA	*9*	*N3*	*DB1*	*27*	*N4/N5*	*H8/H9/H10/H11*	*0.7905*
	DC	*20*	*O2*	*DB1*	*27*	*N2*	*H3*	*0.7208*
ATATGATAT (5'—>3')								
	DT	*22*	*O2*	*DB1*	*27*	*N4/N5*	*H8/H9/H10*	*0.9711*
	DB1	*27*	*N1*	*DG*	*7*	*N2*	*H22*	*0.9539*
	DG	*7*	*N3*	*DB1*	*27*	*N2*	*H3*	*0.7112*
ATATGATAT (3'—>5')								
	DT	*9*	*O2*	*DB1*	*27*	*N4/N5*	*H8/H9/H10/H11*	*0.9081*
	DB1	*27*	*N1*	*DG*	*7*	*N2*	*H22*	*0.8444*
	DC	*20*	*O2*	*DB1*	*27*	*N2*	*H3*	*0.7806*

Table S6. Contact persistence between DB2277 and DNA sequences for the palindromic sequences.

	Acceptor Residue	Residue Number	Acceptor	Donor Residue	Residue Number	Donor	Donor Hydrogen	Persistence
AAAAGCTTTT								
	DB1	*29*	*N1*	*DG*	*7*	*N2*	*H22*	*0.9270*
	DT	*24*	*O2*	*DB1*	*29*	*N4/N5*	*H8/H9/H10/H11*	*0.9255*
	DT	*23*	*O2*	*DB1*	*29*	*N2*	*H3*	*0.4822*
	DT	*23*	*O1P*	*DB1*	*29*	*N6/N7*	*H18/H19/H20/H21*	*0.3847*
	DG	*7*	*N3*	*DB1*	*29*	*N2*	*H3*	*0.3647*
	DC	*22*	*O3'*	*DB1*	*29*	*N6/N7*	*H18/H19/H20/H21*	*0.2289*
AATTGCAATT								
	DT	*25*	*O2*	*DB1*	*29*	*N4/N5*	*H8/H9/H10/H11*	*0.9490*
	DT	*6*	*O2*	*DB1*	*29*	*N2*	*H3*	*0.8364*
	DC	*8*	*O2*	*DB1*	*29*	*N6/N7*	*H18/H19/H20/H21*	*0.1762*
ATATGCATAT								
	DT	*24*	*O2*	*DB1*	*29*	*N4/N5*	*H8/H9/H10/H11*	*0.9797*
	DB1	*29*	*N1*	*DG*	*7*	*N2*	*H22*	*0.9682*
	DG	*7*	*N3*	*DB1*	*29*	*N2*	*H3*	*0.6821*
	DA	*23*	*O1P*	*DB1*	*29*	*N6/N7*	*H18/H19/H20/H21*	*0.3996*
	DC	*22*	*O3'*	*DB1*	*29*	*N6/N7*	*H18/H19/H20/H21*	*0.1490*

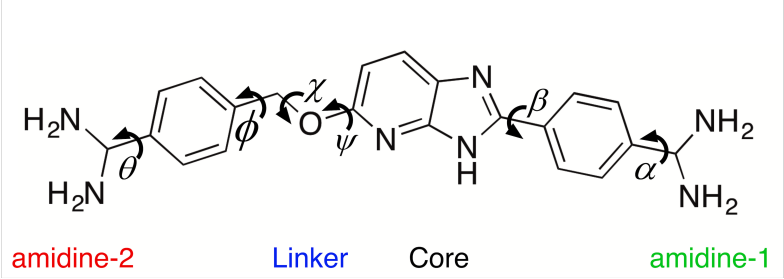

Figure S13 Torsional angles of DB2277

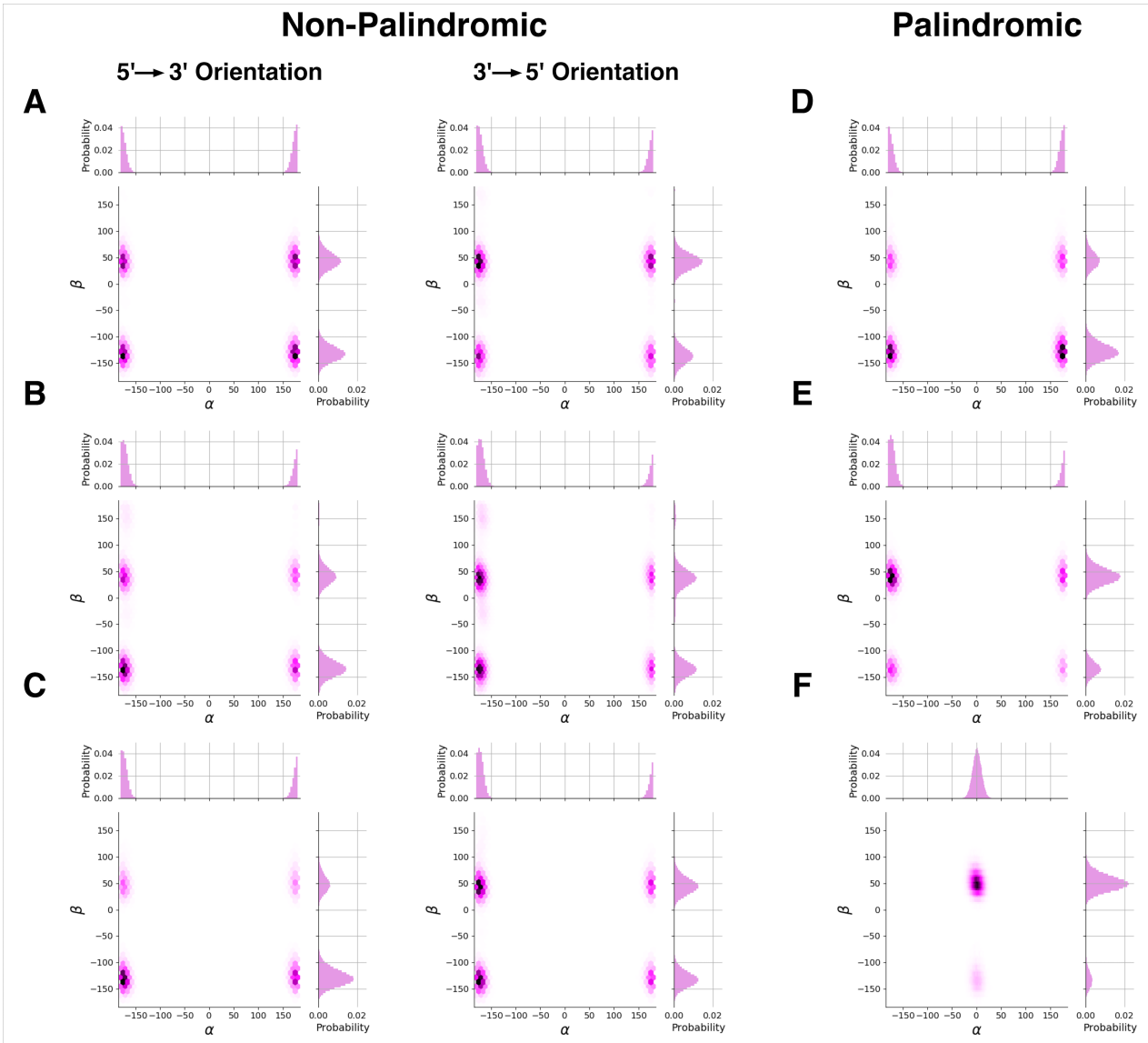

Figure S14 Restricted rotation of amidine-1 due to hydrogen bond formation with the floor of the minor groove. It is interesting to note that the phenyl group of the ATATGCATAT sequence is rotated 180° relative to other conformations. In this wide minor groove, the planar conformation of DB2277 is altered to better accommodate hydrogen bond formation.

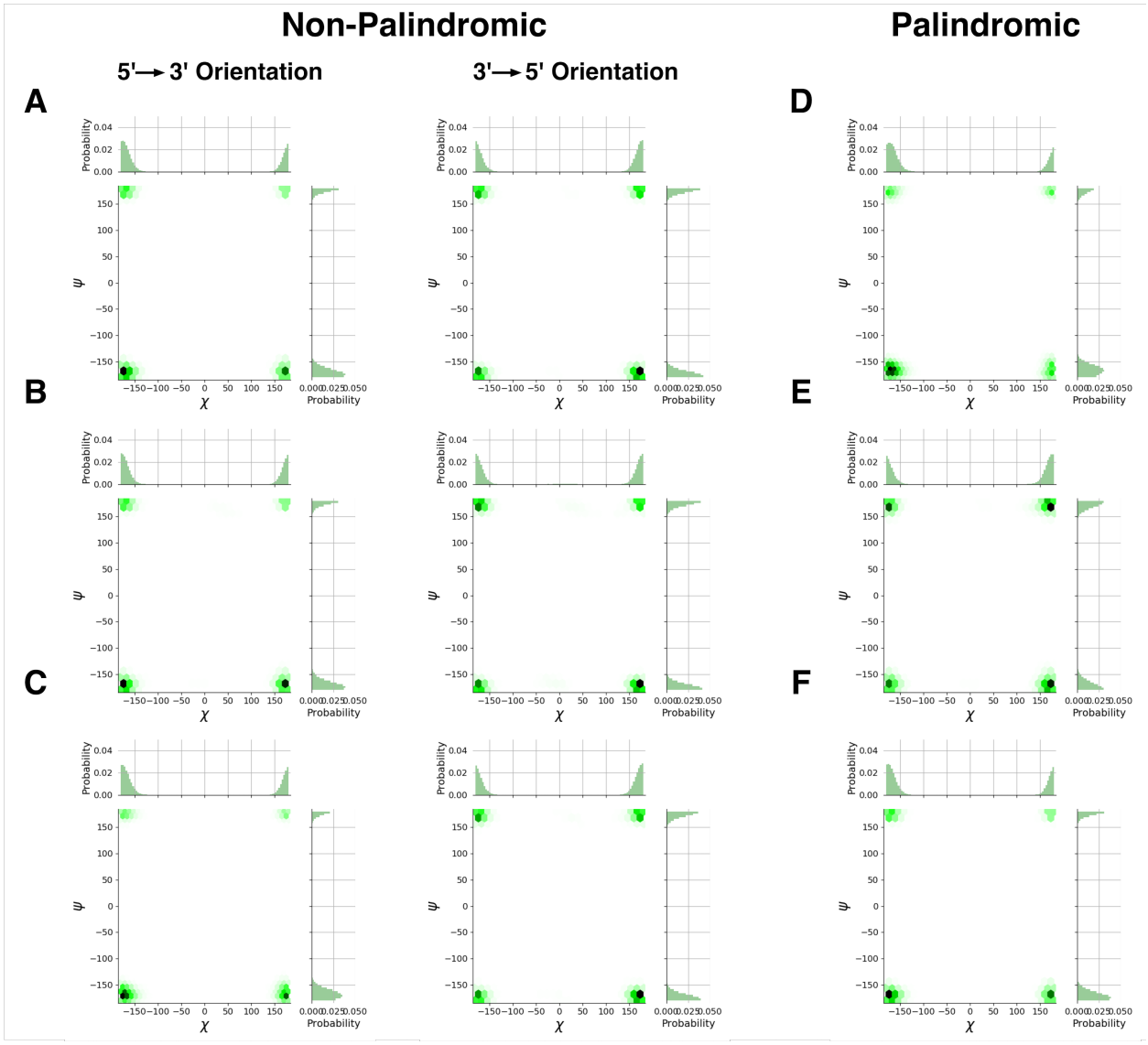

Figure S15 Restricted bond rotation on either side of the ether functional group is due to electrostatic repulsion between O1 and N1. See Table S1 for further details.

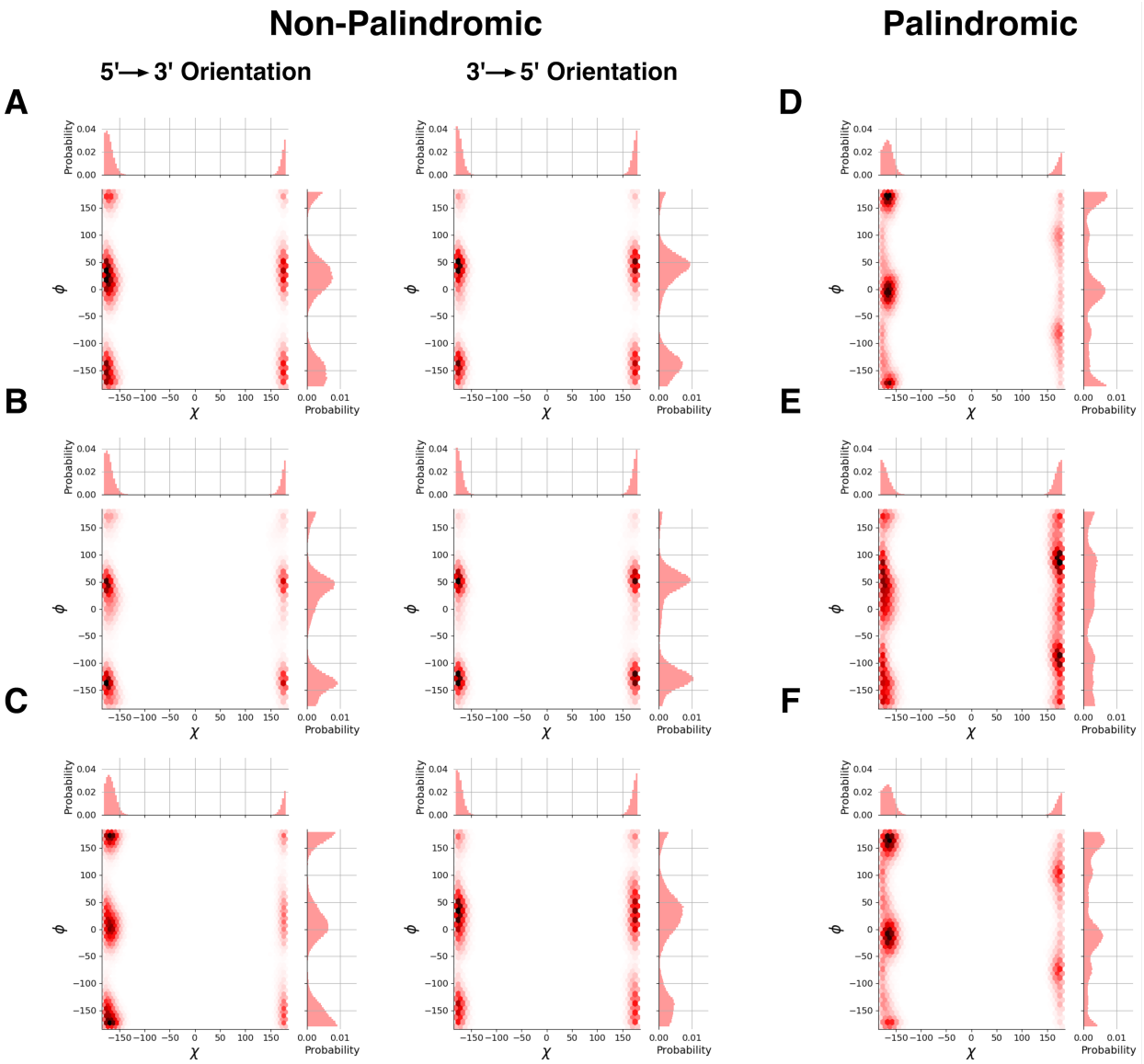

Figure S16 Rotational freedom of the phenyl group relative to the methyl-ether bond. In the non-palindromic sequences, the minor groove collapses around the small molecule and the amidine group of amdine-2 forms numerous water-mediated contacts. This restricts the rotation of the phenyl group to preferred orientations (**A**, **B**, **C**). Amidine-2 in the palindromic sequences is more dynamic as it is more solvent exposed compared to in the non-palindromic sequences. The minor groove is unable to collapse around the small molecule and it forms less persistent direct and indirect bonds. This results in greater rotational freedom (**D**, **E**, **F**).

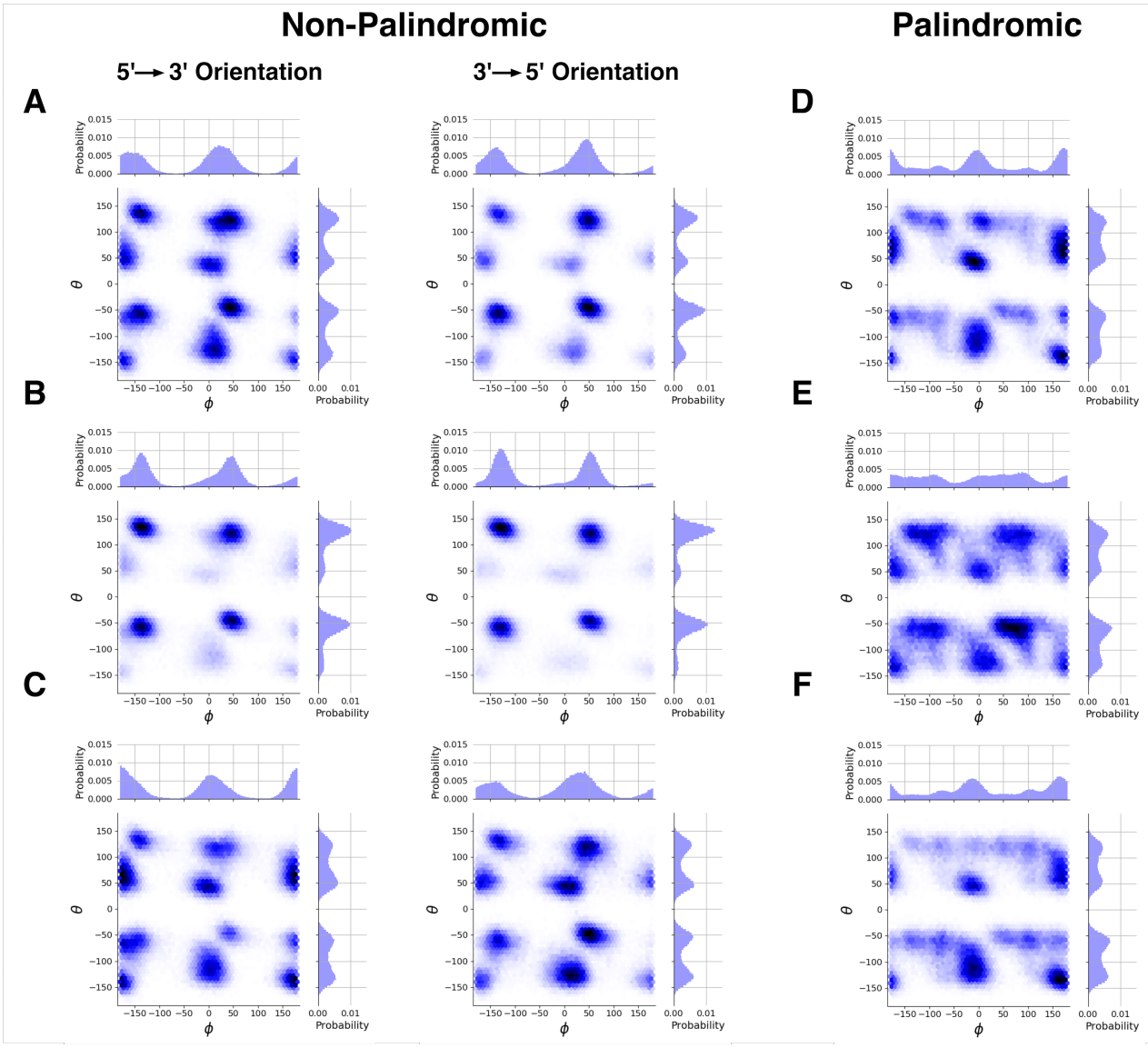

Figure S17 Rotational angles of aminide-2 show influence of water mediate contacts on small molecule rigidity. In the non-palindromic sequences, the minor groove collapses around the small molecule, aiding in contact formation and decreases the rotational freedom (**A**, **B**, **C**). In the palindromic sequences, the minor groove does not collapse around the small molecule and the contact formation is more transient resulting in greater rotational freedom (**D**, **E**, **F**).

References

1. C. O. Silva, E. C. da Silva and M. A. C. Nascimento, *The Journal of Physical Chemistry A*, 2000, **104**, 2402-2409.
2. R. Cammi and J. Tomasi, *J Comput Chem*, 1995, **16**, 1449-1458.
3. M. Cossi, V. Barone, R. Cammi and J. Tomasi, *Chemical Physics Letters*, 1996, **255**, 327-335.