Single GC Base Pair Recognition by a Heterocyclic Diamidine: Structures, Affinities, and Dynamics

Edwin N. Ogbonna,[a,†] J. Ross Terrell,[a†] Ananya Paul, [a] Abdelbasset Faharat, [a,b,c] Gregory M.K. Poon, [a] David W Boykin,[a] W. David Wilson[a*]

[a]Department of Chemistry and Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, GA 30303-3083, USA. [b] Department of Pharmaceutical Organic Chemistry, Faculty of Pharmacy, Mansoura University, Mansoura 35516, Egypt. [c] Master of Pharmaceutical Sciences Program, California North State University, 9700 W Taron Dr., Elk Grove, CA 95757, U.S.A

[†]These authors contributed equally to this work

Address correspondence to this author

***W. David Wilson Tel: 404-413-5503; Fax: 404-413-5505; Email: wdw@gsu.edu

MATERIALS AND METHODS

Synthesis of DB2447

For the detailed synthesis scheme of Structure of pyridyl bis-methoxybenzamidine (DB2447)**.** [1]

Preparation of Singel GC base pair DNA

The oligonucleotide duplexes F-5'- AATAAGAGGAAGTGGG-3'/R-5'-TCCCACTTCCTCTTAT-3' and F-5'-AATAGAAGGAAGTGGG-3'/R-5'-TCCCACTTCCTTCTAT-3' from Integrated DNA Technologies, Coralville, IA, were dissolved in 10 mM HEPES, pH=7.4, 150 mM NaCl buffer and mixed in a 1:1 stochiometric ratio with appropriate strand pairs. Annealing was carried out by placing duplex solutions in a 2L water bath heated to 95°C followed by slow cooling to room temperature.

Purification of PU.1 Protein

Human PU.1 ETS domain was purified as previously described. [2] In brief, Human PU.1 ETS domain (residues 165-270) was cloned into the pET-28b (+) vector and expressed in BL21-Gold (DE3) pLysS cells. Cells were grown at 37°C to $OD600_{nm}$ of 0.6 and expression induced by the addition of IPTG to a final concentration of 0.5 mM. The culture incubation temperature was decreased to 23°C and cells were allowed to shake for an additional 16h. Cells were pelleted by centrifugation and suspended/lysed in 10 mM HEPES, pH=7.4, 500 mM NaCl, 1 mM PMSF buffer. Cell lysate was clarified by centrifugation and loaded onto a 1 mL HiTrap SP HP column (Cytiva) and eluted

against a 10 mM HEPES, pH=7.4, 2M NaCl buffer gradient with fractionation. Fractions containing PU.1 ETS domain were pooled, concentrated, and loaded onto a BioRad SEC 70 column pre-equilibrated with 10 mM HEPES, pH=7.5, 150 mM NaCl buffer. Highly pure fractions from the SEC purification were pooled and concentrated for use in crystallography and biophysics experiments.

Crystallization of Protein-DNA complex with DB2447

DNA complexes were prepared by mixing PU.1 and duplex DNA at a 1:1 stoichiometric ratio and diluting to a final concentration of 220 μM complex mixed. Complex crystals were grown in hanging drops with a 1:1 mixture of stock solution with well solution containing 100 mM sodium acetate, pH=4.6, 2% PEG 3350. Crystals were cryoprotected by transfer into 100 mM sodium acetate, 2% PEG 3350, 20% glycerol solution without (native) and with (ligand-bound) DB2447 at a concentration of 500 μM for 24h prior to looping and flash-freezing.

Data Collection and Structure Determination

X-ray diffraction data sets were collected at 100K at SER-CAT at the Argonne National Laboratory Advanced Photon Source (APS) (Lemont, IL), and the National Synchrotron Light Source II (NSLS-II) at Brookhaven National Laboratory (BNL) (Upton, NY). See Table S1 for crystal sample collection sources and details. Data sets were initially auto processed in XDS and were further truncated in the data reduction module of the CCP4i2 software using Aimless. [3] Molecular replacement was conducted in the PHASER-MR module of the PHENIX suite using maximum-likelihood search procedures and structures prepared via iterative cycles of refinement, model building,

and ligand fitting using both phenix.refine and the Crystallographic Object-Oriented Toolkit (COOT) software respectively. [4,5] The Chimera X software was used to generate all the figures containing crystal structures and models. [6] Subsequently, the atomic structure and coordinate factors for all structures pertaining to this study have been deposited to the RCSB Protein and Nucleic Acid Data Bank.

Molecular Dynamics (MD) Simulations

Structure optimization of DB2447 was performed by using DFT/B3LYP theory with the 6-31+G* basis set in Gaussian 09 (Gaussian, Inc., 2009, Wallingford, CT) with Gauss-view 5.09. [7] Partial charges were derived using the RESP fitting method (restrained electrostatic potential). [8] AMBER 16 (Assisted Model Building with Energy Refinement) software suite was used to perform molecular dynamics (MD) simulations. [9] The X-ray structures of DB2447-PU1-DNA complexes were taken as a initial coordinate to do the MD simulations. The ANTECHAMBER Tools were used to create *LEaP* Input topology files for the ligand DB2831 which was used in the AMBER simulation programs. [9] Specific atom types assigned for the DB2447 molecule were adapted from the *ff99* force field. Most of the force field parameters for DB2447 molecule were derived from the existing set of bonds, angles, and dihedrals for similar atom types in *parm99* and *GAFF* force fields. Some dihedral angle parameters were obtained from previously reported parametrized data. [10] The molecular structure with specific atom types used for the DB2447 molecule is shown in Figure S7. Parameters of DB2447 in .frcmod file are listed in Table S3.

The AMBER16 package was used to equilibrate the DB2447-PU1−DNA complex system using OL15 force field modifications for DNA. MD simulations were performed in explicit solvation conditions where ligand-PU1-DNA complexes were solvated in a truncated octahedron box[11] molecules by using *TLeap* [12] program in AMBER16. To reach physiological salt concentration, 150 mM Na+ and Cl-, were added to the systems. Na+ and Cl-, an appropriate number of ions were added to the systems. This is a higher salt concentration than required to achieve electrical neutrality but is more biologically relevant. The particle mesh Ewald (PME) [13] method was used to handle Coulombic interactions, and a 10 Å cutoff was applied on all van der Waals interactions. The MD simulations were performed by using Sander module with the SHAKE [14] algorithm applied to constrain all bonds involving hydrogen atoms with an integration time step of 2 fs. In the multistage equilibration protocol, the system was relaxed with 500 steps of steepest-descent energy minimization. The temperature of the system was increased from 0 K to 310 K for over 10 ps under constant-volume conditions. In the final step, the production runs on the system was subsequently performed for 600 ns under NPT (constant-pressure) conditions on the PMEMD CUDA module of AMBER16. [9, 12] Trajectories were post-processed using the CPPTRAJ module of AMBERTOOLS16 MD3,MD6 to produce 30000 snapshots for analysis and visualization in UCSF Chimera visualization software. [15] The steepest descent algorithm is useful for quickly removing the largest strains in the system, but it also converges slowly when close to a minimum.

Biosensor-Surface Plasmon Resonance (SPR) Experiments to Determine Ligand-DNA Binding Constant:

Biacore SPR measurements were performed with streptavidin-derivatized (SA) CM5 sensor chips by using four-channel Biacore T200 optical biosensor system (Cytiva, Global life science solutions USA LLC). The procedure of the preparation of SA chips and immobilization of biotinylated-DNAs (**ATAGAAGGAA**:

5'-biotin-CCAAATAGAAGGAAGTGAAACCAAGCTCTCTTGGTTTCACTTCCTTCTATTTGG-3'; **ATAAGAGGAA**:

5'-biotin- CCAAATAAGAGGAAGTGAAACCAAGCTCTCTTGGTTTCACTTCCTTCTATTTGG -3'; **AAAGTGTTT**:

5'-biotin-CC**AAAGTGTTT**GCCTCTGCAAACACTTTGG-3') on-chip surface (cell 3, 4 respectively) were described before. Ligand solutions were prepared with degassed and filtered 50 mM Tris-HCl-buffer with varied NaCl concentrations (100 mM NaCl to 400 mM NaCl) pH 7.4 with 0.05% (v/v) surfactant P20. A series of ligand concentrations (2 nM to 1 uM nM) were injected over the DNA-immobilized sensor chip with the flow rate of 100 µL/min for 180 s, followed by buffer flow for ligand dissociation (600−1800 s). After each sample run, the sensor chip surface was regenerated by injecting acidic 10 mM glycine solution (pH 2.5) for 30 s followed by several buffer injections to establish a stable baseline for the subsequent cycles. The data analysis was followed by a previously described method where reference response from the blank cell (cell 1) was subtracted from the response in each flow cell containing DNA to give a signal (RU_{obs}, response units) directly related to the amount of bound ligand. The expected maximum

response (RU_{max}) per bound ligand in the steady-state region was determined from the molecular weight of the DNA, the ligand molecular weight, and the refractive index gradient ratio of the ligand and DNA. KaleidaGraph 4.0 software was used to plot RU_{obs} versus free ligand concentration (C_{free}). The equilibrium binding constants (K_1) were determined with a one-site binding model. In this model, r = (RU_{obs}/RU_{max}) represents the moles of bound compound/mol of DNA hairpin duplex, and K_1 is macroscopic binding constant.

$$\boldsymbol{r = K_1 * C_{free} / 1 + K_1 * C_{free}} \qquad (1)$$

To evaluate the stoichiometry of ligand-DNA complex, RU_{max} in the equation was used as a fitting parameter and compared that value with predicted maximal response per bound ligand. Kinetic analysis was achieved by globally fitting the ligand-binding sensorgrams by using a standard 1:1 kinetic model with incorporated mass transport-limited binding parameters as described previously. [16]

Supporting Information

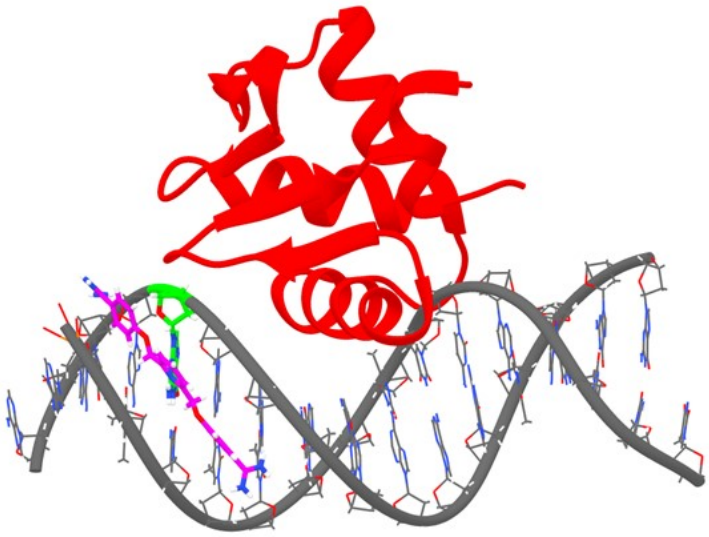

Figure S1: Structure of DB2447 (ball, carbon in magenta) bound to the single G recognition site in the minot groove of PU.1-5'AATAGAAGGAAGTGGG-3' complex (backbone in grey). Guanine 6 residue (ball, carbon in green).

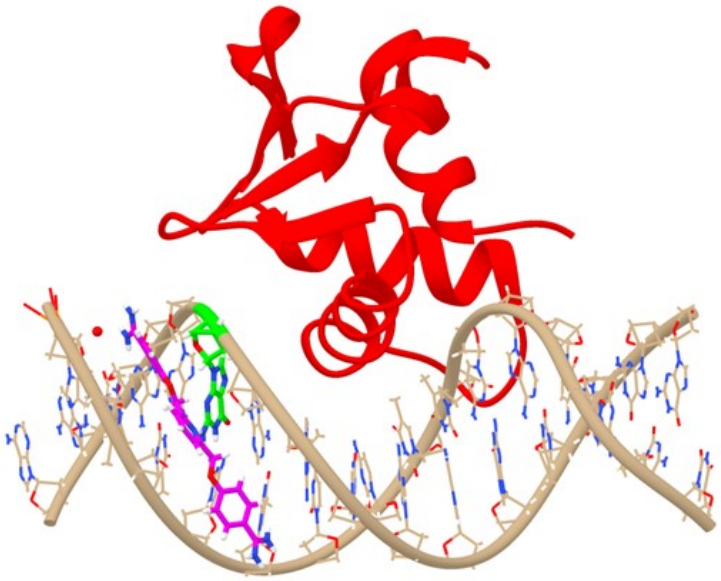

Figure S2: Structure of DB2447 (ball, carbon in magenta) bound to the single G recognition site in the minot groove of PU.1-5'AATAAGAGGAAGTGGG-3' complex (backbone in tan). Guanine 6 residue (ball, carbon in green).

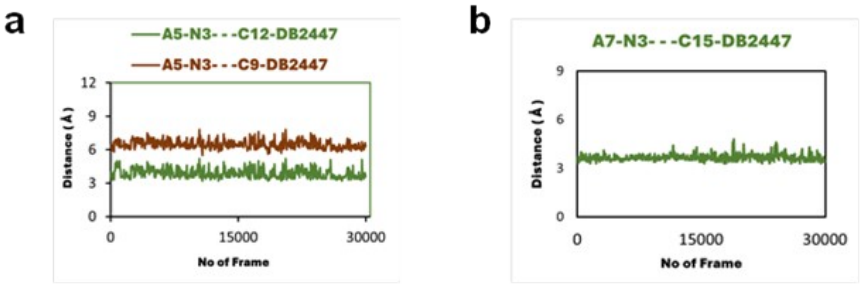

Figure S3: MD simulation study for DB2447 bound to AGAA (a) The MD simulation distance plot to show the interaction between C12 and C9 with N3 of A5 (b)) MD simulation plot showing the interaction between C15 and N3 of A7.

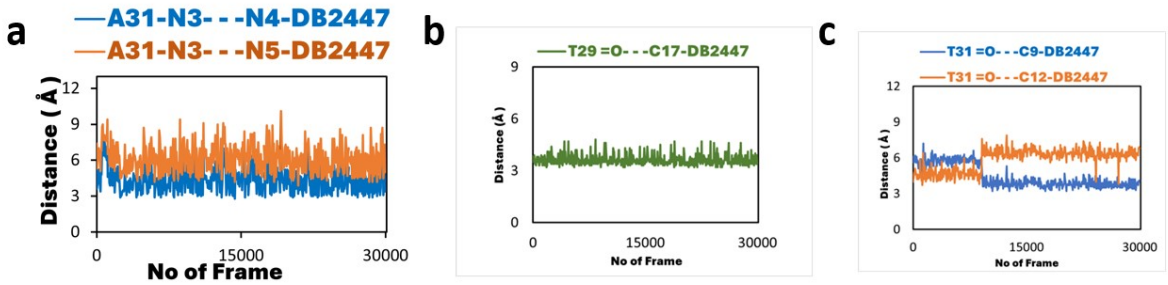

Figure S4: MD simulation study for DB2447 bound to AAGA (a) MD simulation distance plots show the interaction between A31 with N4 and N5 of DB2447 (b)) MD simulation distance plot showing the interaction between C17 and O2 of T29 (C) MD simulation distance plot of C12 and C9 with 02 of T31

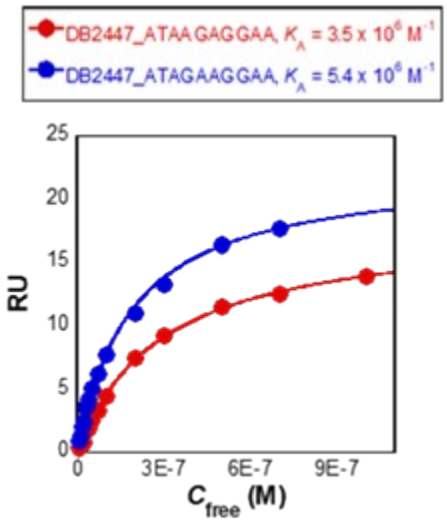

Figure S5: Comparison of the SPR binding affinity for -AATAGAAGGAA-3' and -AATAAGAGGAA- binding sites with DB2447. The solid lines are best fits with a 1:1 binding model.

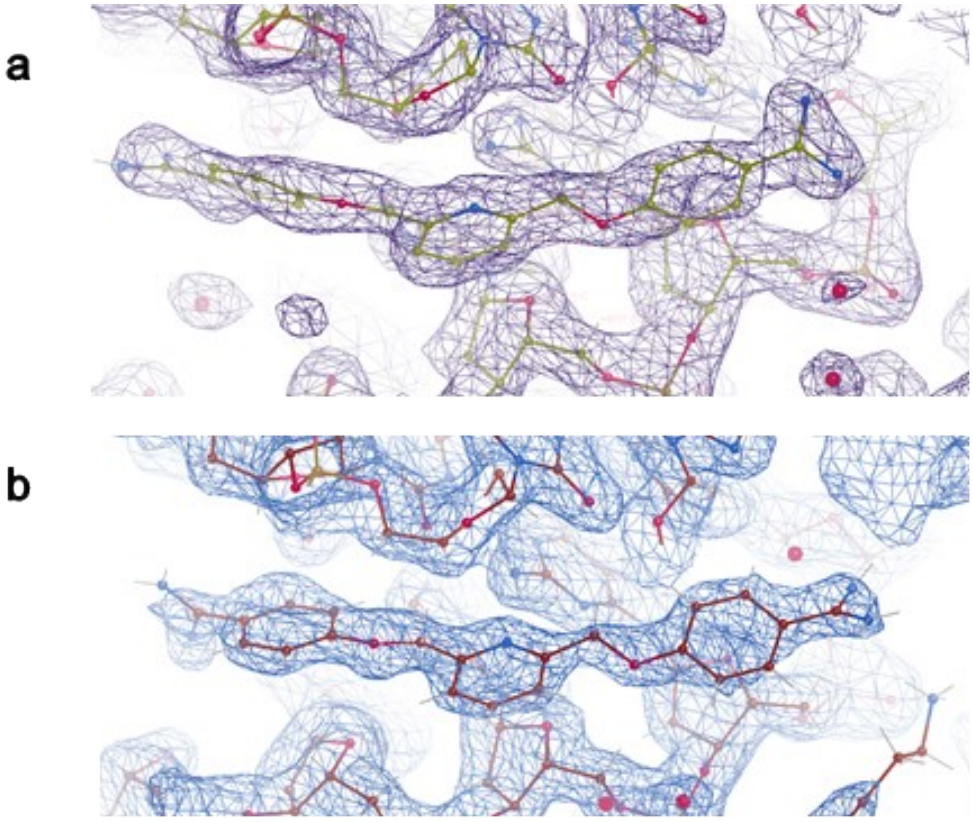

Figure S6: (a) $2F_o$-F_c maps for PU.1-AGAA-DB2447 (in lime green) at 1 σ. (b)) $2F_o$-F_c maps for PU.1-AAGA-DB2447 (in red) at 1 σ.

a

Local base-pair parameters

bp	Shear	Stretch	Stagger	Buckle	Propeller	Opening
1 T-A	0.07	-0.21	-0.05	-1.29	-11.23	2.82
2 A-T	0.14	-0.14	0.06	0.23	-7.42	5.38
3 G-C	-0.35	-0.16	-0.02	4.91	-6.95	0.84
4 A-T	-0.07	-0.21	-0.50	0.09	-12.16	1.80
5 A-T	0.03	-0.20	-0.23	-7.44	-2.04	2.23
6 G-C	-0.26	-0.13	0.12	12.25	4.79	-0.41
7 G-C	-0.28	-0.12	0.09	11.99	-11.90	0.77
8 A-T	0.10	-0.18	-0.03	12.34	-17.88	5.56
9 A-T	0.17	-0.22	0.09	5.20	-18.29	4.20
10 G-C	-0.20	-0.11	0.06	-5.96	-16.57	0.31
11 T-A	-0.14	-0.13	0.17	-5.02	-11.05	5.10
12 G+C	0.61	-3.39	-0.72	16.17	4.97	71.41
13 G-C	-0.10	-0.14	-0.03	-0.49	-11.11	1.01
14 G-C	-0.45	-0.16	-0.08	-11.89	-11.35	0.77
ave.	-0.05	-0.39	-0.08	2.22	-9.16	7.27
s.d.	0.27	0.86	0.25	8.55	7.34	18.57

Local base-pair step parameters

step	Shift	Slide	Rise	Tilt	Roll	Twist
1 TA/TA	0.21	-0.50	3.20	-1.55	7.56	28.93
2 AG/CT	-0.53	-0.42	3.23	-1.62	4.27	32.12
3 GA/TC	0.07	-0.75	3.54	1.51	5.84	30.61
4 AA/TT	0.24	-1.02	3.49	-4.41	4.66	30.96
5 AG/CT	-0.18	-0.54	3.02	-1.14	4.35	24.97
6 GG/CC	-0.96	-0.05	3.35	-2.52	4.01	35.38
7 GA/TC	0.19	-0.35	3.21	1.55	5.51	31.87
8 AA/TT	-0.08	-0.08	3.35	0.08	4.47	33.11
9 AG/CT	-0.20	-0.83	3.47	-3.81	5.36	40.05
10 GT/AC	-0.88	-0.97	3.28	-0.79	1.60	25.08
11 TG/CA	-2.34	3.43	-0.45	174.27	-14.03	-144.38
12 GG/CC	0.36	3.77	-0.90	-135.02	107.03	-178.09
13 GG/CC	0.91	-0.41	3.56	2.42	0.45	29.96

b

Local base-pair parameters

bp	Shear	Stretch	Stagger	Buckle	Propeller	Opening
1 T-A	-0.09	-0.10	0.00	5.39	-13.08	1.95
2 A-T	0.06	-0.14	0.18	6.33	-8.17	4.75
3 G-C	-0.33	-0.06	-0.01	-0.02	-12.07	4.73
4 A-T	0.08	-0.16	-0.27	-4.97	-14.55	6.70
5 A-T	0.14	-0.14	-0.16	-9.40	-5.98	1.52
6 G-C	-0.25	-0.06	0.18	7.82	1.07	0.34
7 G-C	-0.21	-0.12	0.04	9.39	-12.43	1.08
8 A-T	0.15	-0.15	-0.01	10.23	-17.44	4.80
9 A-T	0.14	-0.14	0.06	1.48	-16.63	5.15
10 G-C	-0.21	-0.14	0.08	-7.19	-14.41	-2.09
11 T-A	-0.14	-0.07	0.22	-9.03	-9.78	4.26
12 G+C	0.43	-3.44	-0.41	9.79	7.99	69.47
13 G-C	-0.16	-0.11	-0.14	-4.11	-8.34	-0.34
14 G-C	-0.29	-0.16	-0.21	-10.29	-7.88	1.49
ave.	-0.05	-0.36	-0.03	0.39	-9.41	7.42
s.d.	0.22	0.89	0.18	7.79	6.95	18.03

Local base-pair step parameters

step	Shift	Slide	Rise	Tilt	Roll	Twist
1 TA/TA	0.72	-0.70	3.16	-0.48	6.96	27.62
2 AG/CT	-0.15	-0.70	3.47	-2.07	4.06	35.07
3 GA/TC	-0.04	-0.83	3.46	1.16	3.88	32.25
4 AA/TT	-0.56	-0.88	3.40	-3.29	1.51	31.26
5 AG/CT	0.24	-0.54	3.04	-2.01	3.66	24.46
6 GG/CC	-0.99	-0.16	3.32	-1.33	4.52	34.65
7 GA/TC	0.30	-0.36	3.24	1.44	4.55	31.09
8 AA/TT	-0.08	-0.08	3.37	-0.56	3.38	35.20
9 AG/CT	-0.12	-0.89	3.45	-3.12	4.86	38.55
10 GT/AC	-0.81	-1.06	3.35	-0.89	1.90	25.93
11 TG/CA	-2.18	3.45	-0.42	173.81	-15.24	-141.53
12 GG/CC	-0.37	-3.85	-1.27	136.59	-106.50	169.54
13 GG/CC	0.98	-0.25	3.52	1.96	1.49	30.49

c

Local base-pair parameters

bp	Shear	Stretch	Stagger	Buckle	Propeller	Opening
1 T-A	0.12	-0.16	-0.06	0.63	-10.79	1.56
2 A-T	0.15	-0.16	-0.05	1.38	-5.08	3.30
3 A-T	0.03	-0.15	0.03	6.64	-7.37	2.16
4 G-C	-0.24	-0.18	-0.25	0.84	-8.28	0.22
5 A-T	0.05	-0.16	-0.12	-8.98	-4.02	1.92
6 G-C	-0.21	-0.13	0.14	11.78	3.09	0.14
7 G-C	-0.27	-0.14	0.08	10.90	-10.35	0.55
8 A-T	0.11	-0.17	-0.00	9.36	-16.76	4.52
9 A-T	0.13	-0.20	0.03	3.70	-16.84	4.27
10 G-C	-0.23	-0.14	0.06	-6.43	-15.88	-0.20
11 T-A	-0.04	-0.13	0.08	-4.87	-9.27	4.11
12 G+C	0.52	-3.42	-0.61	13.63	3.71	70.28
13 G-C	-0.24	-0.15	-0.05	-1.00	-9.97	0.10
14 G-C	-0.24	-0.18	-0.18	-12.30	-11.10	-0.17
ave.	-0.03	-0.39	-0.06	1.81	-8.49	6.63
s.d.	0.23	0.87	0.19	8.06	6.37	18.40

Local base-pair step parameters

step	Shift	Slide	Rise	Tilt	Roll	Twist
1 TA/TA	0.39	-0.65	3.27	-0.37	8.95	25.93
2 AA/TT	-0.75	-0.02	3.21	-2.44	2.03	36.24
3 AG/CT	0.25	-0.97	3.51	-0.40	5.43	29.87
4 GA/TC	0.41	-1.20	3.50	-2.03	5.99	30.97
5 AG/CT	-0.18	-0.42	2.96	-0.29	4.56	25.28
6 GG/CC	-0.94	-0.06	3.36	-2.47	4.72	35.56
7 GA/TC	0.14	-0.34	3.28	0.82	5.34	31.65
8 AA/TT	0.01	-0.05	3.32	0.67	4.41	33.35
9 AG/CT	-0.22	-0.81	3.47	-4.39	5.62	39.61
10 GT/AC	-0.98	-0.96	3.26	-0.36	2.46	26.27
11 TG/CA	-2.26	3.48	-0.77	174.04	-15.60	-133.19
12 GG/CC	0.25	3.88	-0.80	-136.57	105.54	-173.59
13 GG/CC	0.85	-0.43	3.60	2.59	1.87	31.37

d

Local base-pair parameters

bp	Shear	Stretch	Stagger	Buckle	Propeller	Opening
1 T-A	-0.09	-0.10	0.00	5.39	-13.08	1.95
2 A-T	0.06	-0.14	0.18	6.33	-8.17	4.75
3 G-C	-0.33	-0.06	-0.01	-0.02	-12.07	4.73
4 A-T	0.08	-0.16	-0.27	-4.97	-14.55	6.70
5 A-T	0.14	-0.14	-0.16	-9.40	-5.98	1.52
6 G-C	-0.25	-0.06	0.18	7.82	1.07	0.34
7 G-C	-0.21	-0.12	0.04	9.39	-12.43	1.08
8 A-T	0.15	-0.15	-0.01	10.23	-17.44	4.80
9 A-T	0.14	-0.14	0.06	1.48	-16.63	5.15
10 G-C	-0.21	-0.14	0.08	-7.19	-14.41	-2.09
11 T-A	-0.14	-0.07	0.22	-9.03	-9.78	4.26
12 G+C	0.43	-3.44	-0.41	9.79	7.99	69.47
13 G-C	-0.16	-0.11	-0.14	-4.11	-8.34	-0.34
14 G-C	-0.29	-0.16	-0.21	-10.29	-7.88	1.49
ave.	-0.05	-0.36	-0.03	0.39	-9.41	7.42
s.d.	0.22	0.89	0.18	7.79	6.95	18.03

Local base-pair step parameters

step	Shift	Slide	Rise	Tilt	Roll	Twist
1 TA/TA	0.72	-0.70	3.16	-0.48	6.96	27.62
2 AG/CT	-0.15	-0.70	3.47	-2.07	4.06	35.07
3 GA/TC	-0.04	-0.83	3.46	1.16	3.88	32.25
4 AA/TT	-0.56	-0.88	3.40	-3.29	1.51	31.26
5 AG/CT	0.24	-0.54	3.04	-2.01	3.66	24.46
6 GG/CC	-0.99	-0.16	3.32	-1.33	4.52	34.65
7 GA/TC	0.30	-0.36	3.24	1.44	4.55	31.09
8 AA/TT	-0.08	-0.08	3.37	-0.56	3.38	35.20
9 AG/CT	-0.12	-0.89	3.45	-3.12	4.86	38.55
10 GT/AC	-0.81	-1.06	3.35	-0.89	1.90	25.93
11 TG/CA	-2.18	3.45	-0.42	173.81	-15.24	-141.53
12 GG/CC	-0.37	-3.85	-1.27	136.59	-106.50	169.54
13 GG/CC	0.98	-0.25	3.52	1.96	1.49	30.49

Figure S7: Results of the helical parameter analysis from 3DNA.[17] (a) AGAA_NAT (b) AGAA_DB2447 (c) AAGA_NAT (d) AAGA_DB2447.

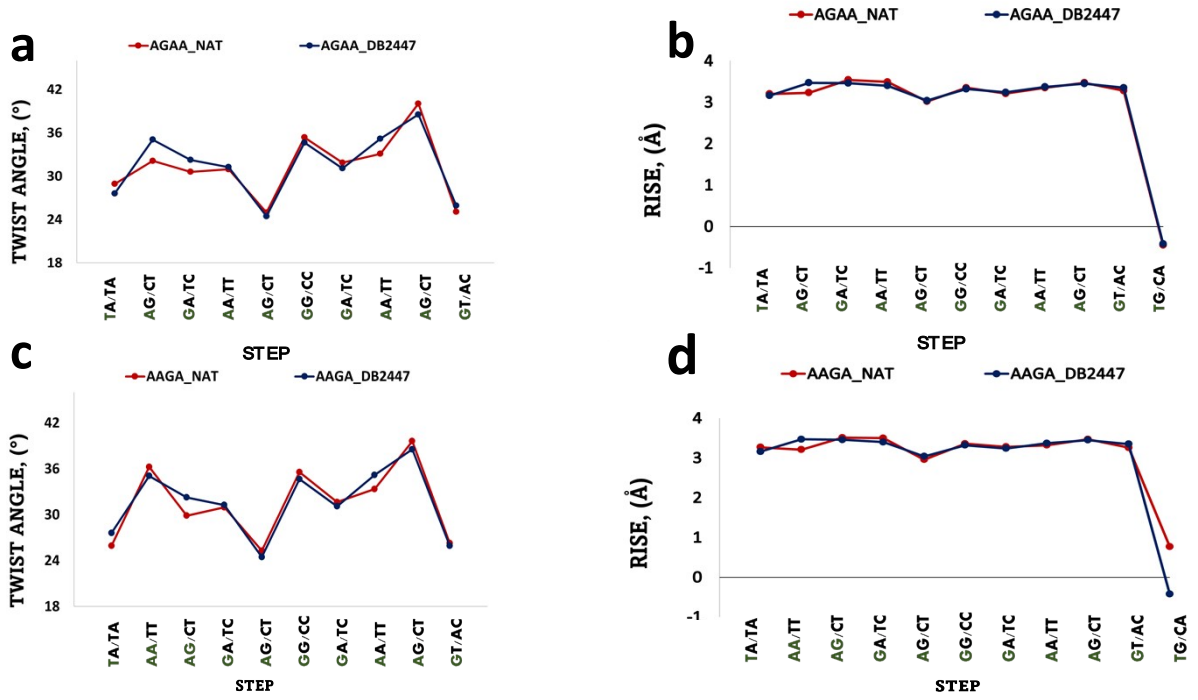

Figure S8: Plots from the helical parameter analysis from 3DNA. (a) Twist for AGAA_NAT and AGAA_DB2447 (b) Rise for AGAA_NAT and AGAA_DB2447 (c) Twist for AAGA_NAT and AAGA_DB2447 (d) Twist for AAGA_NAT and AAGA_DB2447

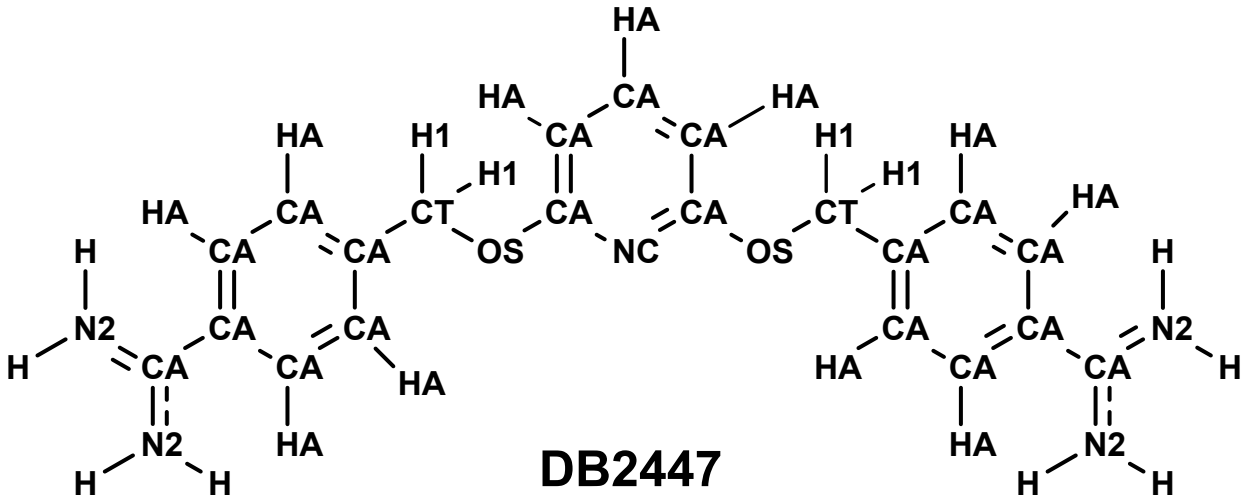

Figure S9: The molecular structure with specific atom types used for the DB2447 molecule.

Table S1: Crystallographic Data

	AAGA_DB2447	AGAA_Native	AGAA_DB2447
PDB ID	**8VDI**	**8V9N**	**8VDH**
Wavelength	1	1	0.9201
Resolution range	33.14 - 1.93 (1.999 - 1.93)	31.72 - 1.78 (1.844 - 1.78)	24.18 - 1.64 (1.699 - 1.64)
Space group	P 1 21 1	P 1 21 1	P 1 21 1
Unit cell	42.996 60.583 44.297 90 116.635 90	42.928 60.363 44.707 90 116.775 90	43.049 61.062 44.178 90 116.306 90
Total reflections	62680 (6378)	75929 (7803)	102715 (10277)
Unique reflections	15325 (1543)	19089 (1900)	25020 (2501)
Multiplicity	4.1 (4.1)	4.0 (4.1)	4.1 (4.1)
Completeness (%)	99.44 (99.87)	97.28 (98.80)	97.21 (98.84)
Mean I/sigma(I)	12.82 (2.36)	17.53 (3.00)	11.37 (2.93)
Wilson B-factor	23.75	19.26	20.87
R-merge	0.08503 (0.6843)	0.05984 (0.5193)	0.06574 (0.304)
R-meas	0.09787 (0.7875)	0.0693 (0.5965)	0.07556 (0.3482)
R-pim	0.04789 (0.3856)	0.03448 (0.2906)	0.0366 (0.1676)
CC1/2	0.998 (0.791)	0.999 (0.874)	0.997 (0.97)
CC*	0.999 (0.94)	1 (0.966)	0.999 (0.992)
Reflections (refinement)	15325 (1541)	19089 (1898)	25029 (2477)
Reflections (R-free)	769 (83)	1897 (196)	1932 (194)
R-work	0.1664 (0.2273)	0.1680 (0.2314)	0.1819 (0.2177)
R-free	0.2018 (0.2819)	0.2076 (0.2760)	0.2076 (0.2656)
CC(work)	0.969 (0.875)	0.974 (0.917)	0.968 (0.935)
CC(free)	0.963 (0.834)	0.971 (0.777)	0.958 (0.847)
No of non-hydrogen atoms	1603	1645	1643
macromolecules	1393	1399	1399
ligands	51	0	51
solvent	182	246	216
Protein residues	91	91	91
RMS(bonds)	0.01	0.01	0.009
RMS(angles)	1.25	1.31	1.23
Ramachandran favored (%)	100	97.75	97.75
Ramachandran allowed (%)	0	2.25	2.25
Ramachandran outliers (%)	0	0	0
Rotamer outliers (%)	1.28	0	1.27
Clashscore	1.17	0.79	1.16

| | AAGA_DB2447 | AGAA_Native | AGAA_DB2447 |

PDB ID	**8VDI**	**8V9N**	**8VDH**
Average B-factor	25.84	23.59	27.28
macromolecules	24.22	21.69	25.49
ligands	36.4		36.43
solvent	33.57	32.49	35.86
Beamline	APS 22-BM	ALS 8.2.2	NSLS-II 17-ID-2
Oscillation Width	1	1	0.2
Frames Collected	200	200	1100

*__AAGA_Native__ (PDB ID: 8E4H) previously published.[36]

Table S2: Types of H-bond Interactions for AGAA and AAGA

AGAA	DISTANCE (Å)	TYPE OF H-BOND INTERACTION
A7-N3---N1-DB2447	2.6 - 3.4	Transient
G5-NH2---N3-DB2447	3.5 - 5.8	Fixed
A4-N3---N4-DB2447	3.5 - 5.8	Transient
AAGA	**DISTANCE (Å)**	**TYPE OF H-BOND INTERACTION**
T27=O---N1-DB2447	2.6 - 3.4	Transient
G6-NH2---N3-DB2447	3.5 - 5.8	Fixed
T30=O---N4-DB2447	3.5 - 5.8	Transient

Table S3: SPR binding information of DB2447 for 5'-AATAGAAGGAAGTGGG-3' and 5'-AATAAGAGGAAGTGGG-3' respectively.

Binding site	Compound	K_A
-ATAAGA-	DB2447	$(3.5 \pm 2) \times 10^6\ M^{-1}$
-ATAGAA-	DB2447	$(5.4 \pm 3) \times 10^6\ M^{-1}$

Table S4: Parameters of DB2447 in .frcmod file

MASS

N2 14.01 0.530 parm99

H 1.008 0.161 parm99

CA 12.01 0.360 parm99

HA 1.008 0.167 parm99

OS 16.00 0.465 parm99

NC 14.01 0.530 parm99

H1 1.008 0.135 parm99

CT 12.01 0.878 parm99

BOND

CA-CA 469.0 1.400 parm99

CA-CT 317.0 1.510 parm99

CA-NC 483.0 1.339 parm99

CA-HA 367.0 1.080 parm99

CA-N2 481.0 1.340 parm99

H -N2 434.0 1.010 parm99

CT-H1 340.0 1.090 parm99

CT-OS 320.0 1.410 parm99

CA-OS 372.4 1.3730 gaff, ca-os

ANGLE

N2-CA-N2 70.0 120.00 parm99

H -N2-H 35.0 120.00 parm99

CA-CA-N2 70.0 119.99 parm99, CM-CA-N2, Gaussian-angle

CA-N2-H 50.0 120.00 parm99

CA-CA-CA 63.0 120.00 parm99

CA-CA-HA 50.0 120.00 parm99

CA-CA-CT 70.0 120.00 parm99

CA-CT-H1 50.0 109.50 parm99, CA-CT-HC

H1-CT-H1 35.0 109.50 parm99

H1-CT-OS 50.0 109.50 parm99

CA-CT-OS 60.0 109.50 parm99, C -CT-OS

CT-OS-CA 62.3 119.76 gaff, c3-os-ca, Gaussian-angle

CA-CA-OS 70.5 115.46 gaff, ce-c2-os forceconst, gaussian O-out

CA-CA-NC 69.2 122.63 gaff, ca-ca-nb

CA-NC-CA 68.59 115.86 SOURCE3 ca-nb-ca GAFF

CT-CA-NC 67.33 116.66 SOURCE4 c3-ca-nb GAFF

DIHE

N2-CA-N2-H 4 9.60 180.0 2.0 parm 99, X -CA-N2-X

CA-CA-N2-H 4 9.60 180.0 2.0 parm 99, X -CA-N2-X

N2-CA-CA-CA 4 0.789 327.000 -4.0 DB921

N2-CA-CA-CA 4 -3.118 0.000 -2.0 DB921

N2-CA-CA-CA 4 0.609 90.000 1.0 DB921

CA-CA-CA-CA 4 14.50 180.0 2.0 parm99, X -CA-CA-X

CA-CA-CA-CT 4 14.50 180.0 2.0 parm99

CA-CA-CT-H1 6 0.72 180.0 2.0 New parameter

CA-CA-CT-OS 6 0.72 180.0 2.0 New parameter

CA-CT-OS-CA 3 6.88 0.0 1.0 New parameter

CT-OS-CA-CA 2 3.29 180.000 2.000 New parameter

OS-CA-CA-CA 4 14.50 180.0 2.0 parm99, X -CA-CA-X

CA-CA-CA-HA 4 14.50 180.0 2.0 parm99, X -CA-CA-X

HA-CA-CA-HA 4 14.50 180.0 2.0 parm99, X -CA-CA-X

HA-CA-CA-CT 4 14.50 180.0 2.0 parm99, X -CA-CA-X

NC-CA-CA-HA 4 14.50 180.0 2.0 parm99, X -CA-CA-X

OS-CA-CA-HA 4 14.50 180.0 2.0 parm99, X -CA-CA-X

CA-CA-CA-HA 4 14.50 180.0 2.0 parm99, X -CA-CA-X

CT-CA-NC-CA 2 9.60 180.0 2.0 Parm99, X -CA-NC-X

NC-CA-CT-OS 6 0.00 0.0 2.0 JCC,7,(1986),23, X -CA-CT-X

NC-CA-CT-H1 6 0.00 0.0 2.0 JCC,7,(1986),23, X -CA-CT-X

CA-OS-CT-H1 3 6.88 0.0 1.0 New parameter, CA-CT-OS-CA db2277

CA-NC-CA-CA 2 9.60 180.0 2.0 Parm99, X -CA-NC-X

NC-CA-CA-CA 4 14.50 180.0 2.0 parm99, X -CA-CA-X

IMPROPER

CA-CA-CA-HA	1.1	180.0	2.0	General improper torsional angle (2 general atom types)
CT-CA-CA-NC	1.1	180.0	2.0	Using default value
CA-CA-CA-OS	1.1	180.0	2.0	Using default value
CA-CA-CA-HA	1.1	180.0	2.0	Using default value
CA-N2-CA-N2	1.1	180.0	2.0	Using default value

NONBON

H1	1.3870	0.0157	parm99
H	0.6000	0.0157	parm99
HA	1.4590	0.0150	parm99
OS	1.6837	0.1700	parm99
CT	1.9080	0.1094	parm99
CA	1.9080	0.0860	parm99 (C*)
NC	1.8240	0.1700	parm99 (N)
N2	1.8240	0.1700	parm99 (N)

References

(1) A. Paul, A.A. Farahat, D.W. Boykin and W.D. Wilson, Life, 2022,**12**, 681.

(2) J.R. Terrell, S.J. Taylor, A.L. Schneider, Y. Lu, T.N. Vernon, S. Xhani, R.H. Gumpper, M. Luo, W.D. Wilson, U. Steidl and G.M.K Poon, *Cell Rep.*, 2023,**42**, 112671.

(3) M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, N.S. Keegan, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin and K.S. Wilson, *Acta. Cryst. D*., 2011, **67**, 235-242.

(4) D. Liebschner, P.V. Afonine, M.L. Baker, G. Bunkóczi, V.B. Chen, T.I. Croll, B. Hintze, L.W. Hung, S. Jain, A.J. McCoy, N.W. Moriarty, R.D. Oeffner, B.K. Poon, M.G. Prisant, R.J. Read, J.S. Richardson, D.C. Richardson, M.D. Sammito, O.V. Sobolev, D.H. Stockwell, T.C. Terwilliger, A.G. Urzhumtsev, L.L. Videau, C.J. Williams and P.D. Adams, *Acta Crystallogra D Struct Biol*., 2019, **75**, 861-877.

(5) P. Emsley and K. Cowtan, *Acta Crystallogr. D*., 2004, **60**, 2126-2132.

(6) E.F. Pettersen, T.D. Goddard, C.C. Huang, E.C. Meng, G.S. Couch, T.I. Croll, J.H. Morris and T.E. Ferrin, *Protein Sci*., 2021, **30**, 70-82.

(7) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.

(8) C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, J. Phys. Chem., 1993, 97, 10269; U. C. Singh, P. A. Kollman*, J. Comput. Chem.*, 1984, **5**, 129.

(9) D. A. Case, V. Babin, J. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, I. A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu and Kollman, P. A. AMBER 14, 2014.

(10) J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graph. Model*, 2006, **25**, 247; N. K. Harika, M. W. Germann and W. D. Wilson, *Chem. Eur. J.*, 2017, **23**, 17612; P. Athri and W. D. Wilson, *J. Am. Chem. Soc*., 2009, **131**, 7618; N. Špačková, T. E. Cheatham, F. Ryjáček, F. Lankaš, L. van Meervelt, P. Hobza and J. Šponer, *J. Am. Chem. Soc*., 2003, **125**, 1759.

(11) W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys*., 1983, **79**, 926.

(12) T. J. Macke, D. A. Case, Molecular Modeling of Nucleic Acids, American Chemical Society, Washington, D.C., USA, 1998, 379.

(13) T. Darden, D. York and L. Pedersen, *J. Chem. Phys*., 1993, **98**, 10089.

(14) J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys*., 1977, **23**, 327.

(15) G. S. Couch, D. K. Hendrix and T. E. Ferrin*, Nucleic Acids Res*., 2006, **34**, e29.
(16) B. Nguyen, F. A. Tanious and W. D. Wilson, *Methods*, 2007, **42**, 150.
(17) L.Xiang-Jun and W.K. Olson, *Nucleic Acid Res*., 2003, 31, 5108-21.