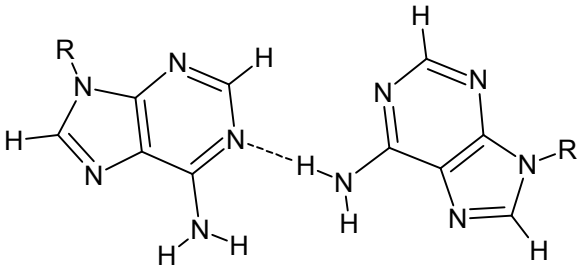
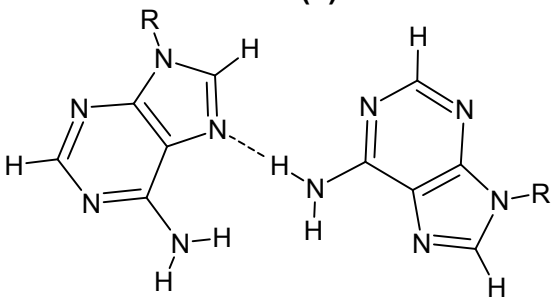
Supporting information for "Coupled nucleotide substitution experiments and molecular dynamics simulation indicate the nature of A-A base pairing and putative structures of coralyne-induced homo-adenine duplexes" by Joung, Çetinkol, Hud and Cheatham (2009), *Nucleic Acids Research*.

The set of possible AA base pairs:

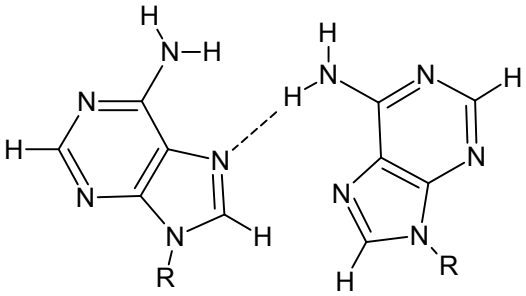
A cisWW



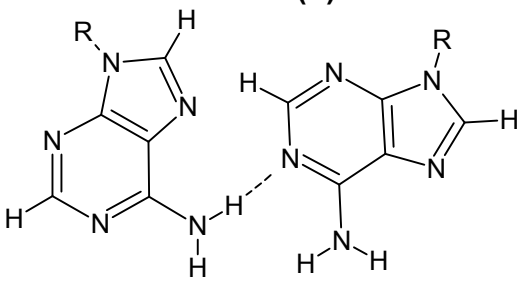
F cisWH(a)



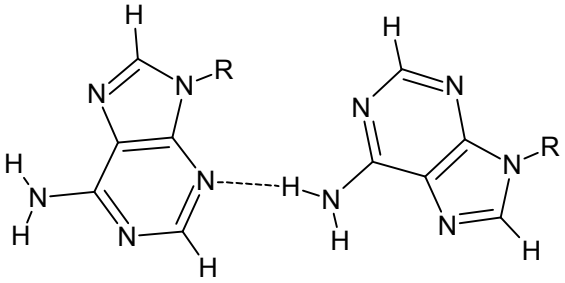
B cisHH



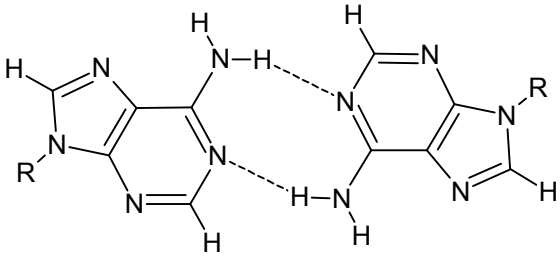
G cisWH(b)



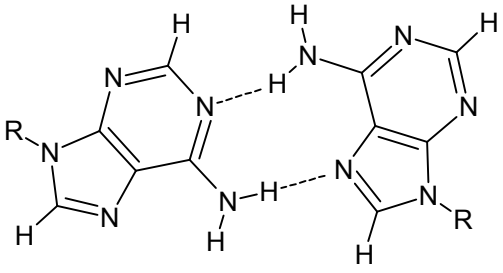
C cisWS



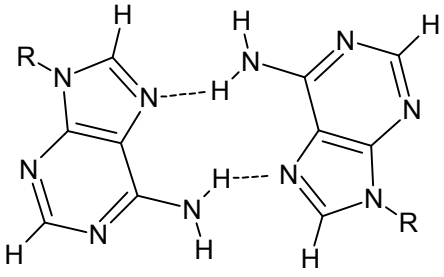
H transWW



D transWH



I transHH



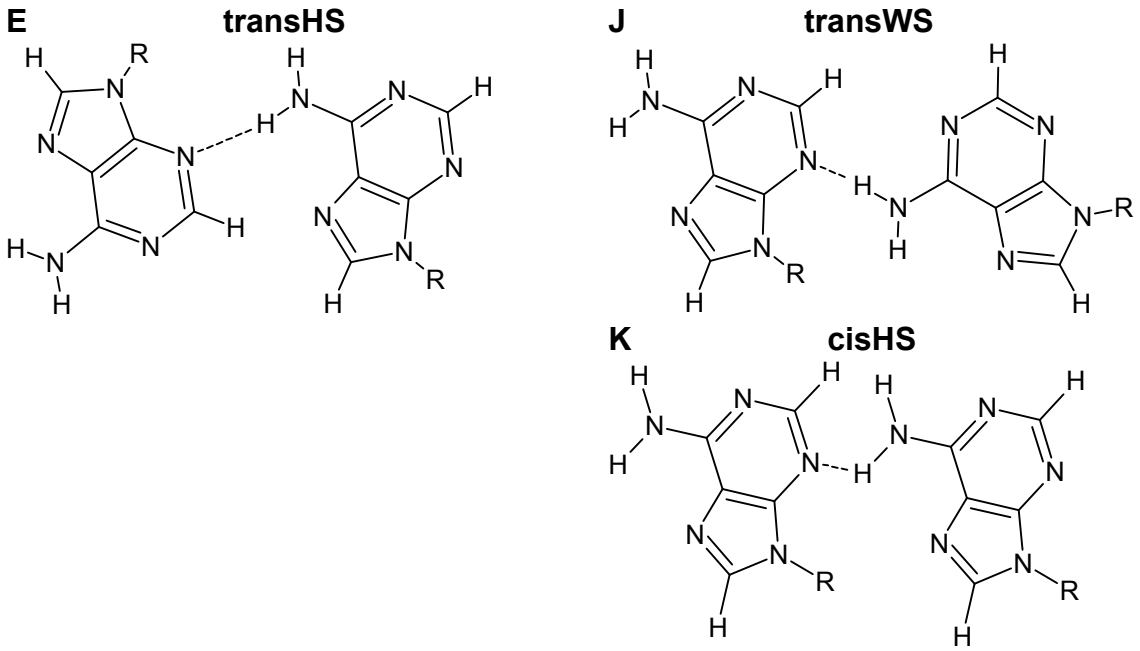


Figure S1: The complete set of adenine base pairs between two adenine bases. Eleven possible base pairs can be constructed from two adenine bases if only regular H-bonds between adenine bases are considered. For convenience, the names of the structures are given with the glycosidic bond orientation followed by the two interacting edges: A: cisWW, B: cisHH, C: cisWS, D: transWH, E: transHS, F: cisWH(a), G: cisWH(b), H: transWW, I: transHH, J: transWS, K: cisHS.

DETAILED METHODS

Parameterization of coralyne molecule

The initial structure of coralyne was built with GaussView v3.0 (1). Geometry optimization using the Gaussian03 program (2) was performed at the Hartree-Fock level with the 6-31G* basis set. The electrostatic potential of the optimized structure was also calculated to allow determination of the restrained electrostatic potential (RESP) charges (3). The rest of the force field parameters for the coralyne molecule were obtained from the *general AMBER force field* (GAFF) (4) using the **`antechamber`** program from the AMBER 8 suite. A minor modification was applied to the parameters; briefly, hydrogen atoms connected to the aromatic carbons were reassigned to be an aromatic hydrogen type. Missing force field parameters were deduced using the **`parmchk`** program from the AMBER 8 suite. The final force field parameters, RESP charges, and optimized coordinates are shown in Table S1.

Generating coordinates

Four distinct types of anti-parallel 12 base-paired homo-(dA) duplex models were initially built, and the models were relaxed in MD simulations with base-pair restraints.

Schematic diagrams of the possible duplexes are displayed in Figure S2. The duplex type I had a non-alternating *anti-anti* glycosidic conformations, the duplex type II had an alternating *anti-syn* conformations, the duplex type III had a non-alternating *anti-syn* conformations, and the duplex type IV had a non-alternating *syn-syn* conformations.

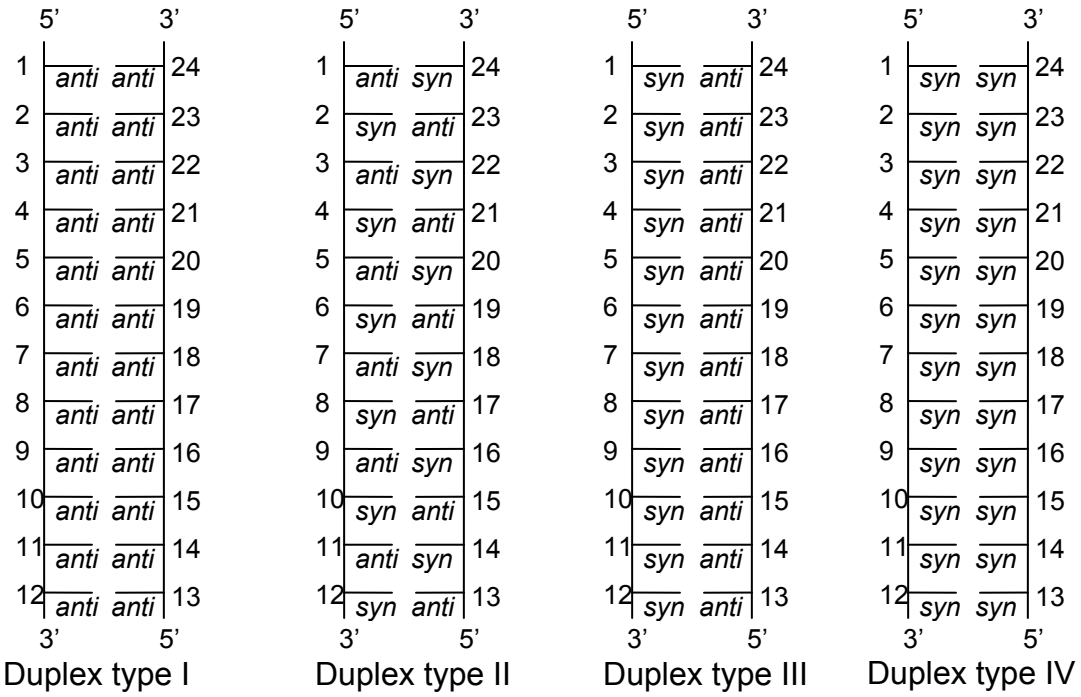


Figure S2: Schematic diagrams of the 12-mer homo-(dA) duplexes that were used as initial structures. The numbers next to the each adenine bases denote their residue numbers. The base-sugar conformation (or the orientations of the glycosidic bond(χ)) are also indicated as either *anti* or *syn*.

Coralyne Docking

The coralyne was docked into the homo-(dA) duplex structure in a systematic manner. Essentially this involved initially docking the coralyne into the intercalation plane between the bases in two orientations (standard and "flipped"). For each of these two coralyne orientations, a further series of models were built with different rotations (referred to as the "coralyne angle") of the coralyne relative to the base pair step. In addition to binding via intercalation, it is possible the coralyne could bind into the grooves of the poly(dA) duplexes. Therefore, various groove-binding models were also built. For binding free energy estimates coralyne self-aggregates were built by stacking (in either of the two orientations) and randomly rotated along the stacking plane. During simulation, the stacking planes of the coralyne aggregates were consistently parallel with a distance between the planes of ~3-4 Å.

Base-pair restraints

If the models are built correctly, the DNA structures will be at or near an equilibrium state. However, when the models are tentative, judicious application of base pair restraints can force the structure to more rapidly obtain a regularized structure with the putative hydrogen bonding patterns. The coordinates of the initial duplex structures (duplex I, II, III, and IV) were initially crude (i.e. not properly adjusted for individual base pairing) and were built. For each base pair, restraints were applied to maintain the appropriate base pair spacing and planarity with a force constant of 5 kcal/mol·Å^2 for distance restraints and 5 kcal/mol·rad^2 for dihedral restraints (Table S2). Note that when only a single H-bond is involved in the base pair, a second distance restraint was assigned in an *ad hoc* manner to promote formation of regular base pairing.

Force field

All the MD simulations were carried out with AMBER 8-10 program suites (5,6) using either the *ff99* (7) or *ff99-bsc0* (8) parameters for nucleic acids and ions and the GAFF (4) parameters for the coralyne molecules. The initial model building and structure generation was performed with the *ff99* force field with subsequent analysis using both force fields. Unless otherwise stated, assume that the results presented were generated with the *ff99* force field. For the explicit water simulations, the TIP3P (9) water model was used. To perform MD simulations in implicit water, the Hawkins, Crammer and Truhlar pairwise generalized Born model (10,11) with the parameters of Tsui and Case (12) was used.

MD simulation

-Implicit solvent

Energy minimization of the system was carried out by performing 1000 steps of steepest descent (except systems containing only homo-(dA) duplexes where 500 steps were used). The non-bonded cutoff was 25 Å and the maximum distance for calculating pair-wise summation for effective Born radii was 20 Å. The MD simulation integration time step was 2 fs and the SHAKE algorithm (13) was employed with a 10^{-5} Å tolerance. At the beginning of the simulation, the temperature of the system was set 250 K and the temperature was maintained at 300 K during simulation using Langevin dynamics (14,15) with a collision frequency of 1.0 ps^{-1}.

-Explicit solvent

Net-neutralizing Na^+ or Cl^- ions were added and the whole system was solvated in a truncated octahedron box filled with TIP3P water molecules. The edges of the periodic box were at least 10 Å away (or 15 Å for the larger complexes with 5 and 7 coralynes) from the boundaries of the solute molecules. The lengths of edges of the periodic boxes were initially 38.3 Å for a single coralyne molecule, 40.8(±2.8) Å for three coralyne molecules, 43.8(±1.4) Å for five coralyne molecules, 50.3(±1.2) Å for seven

coralyne molecules, 61.2 Å for a single homo-(dA), 65.2(±2.4) Å for homo-(dA) duplexes, 69.8(± 2.1) Å for coralyne-homo-(dA) complexes with one coralyne molecule, 84.4(± 2.3) Å for coralyne-homo-(dA) complexes with five coralyne molecules, and 92.9(±0.7) Å for coralyne-homo-(dA) complexes with seven coralyne molecules. For the initial equilibration, harmonic positional restraints with a force constant of 500 kcal/mol·$Å^2$ were applied to the solute molecules and the system was minimized for 1000 steps of steepest descent minimization. Additional minimization was performed to relax the whole system by running another 1000 steps of steepest descent after removing the restraint on the solute molecules. Regarding the MD simulation, the SHAKE algorithm (13) was applied with a tolerance of 10^{-7} Å with a 2 fs time step. The non-bond cutoff was 9 Å and the particle mesh Ewald method (16) with default parameters was used to calculate long-range electrostatic interactions. The temperature of the system was set initially to 250 K and 300 K was maintained during MD simulation using Berendsen coupling algorithm (17) with 1 ps time constant. Initially, the system was equilibrated for 20 ps at constant volume and temperature applying harmonic positional restraints on the solute with a force constant of 10 kcal/mol·$Å^2$. In the production MD simulations, the Berendsen coupling constant was readjusted to 5 ps.

Analyses of the MD trajectories

The hydrogen bonds (H-bonds) between the bases were analyzed based on the following definition: the distance between the H-bond acceptor and the electronegative atom bonded to the hydrogen atom is less than 3.1 Å and the angle of the three atoms, specifically the H-bond acceptor, the hydrogen atom and the electronegative atom, should be between 120°-180°. Although this definition is not generally rigorous, it was defined for comparative purposes to obtain statistical results regarding optimum H-bonds between the adenine bases. Only the H-bonds indicated in Figure S1 were traced. The occupying percentage of each H-bond during the assigned period was calculated based upon the number of snapshots in which the definition of the H-bond are satisfied and the total number of snapshots. The snapshots were produced every 1 ps in the MD simulation. All the occupying percentages of the H-bonds calculated from individual pairs in a system were averaged to obtain the average of the occupying percentage, which is referred as “occupancy of H-bond” in the text.

The molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) method is an efficient and approximate way to calculate relative free energies of bio-molecules (18,19) Although the accuracy of the MM-PBSA method is lower than more computationally expensive methods, such as thermodynamic integration or free energy perturbation (20,21), it is still very useful in ranking the free energies of different conformations of a given molecular structure. In this work, snapshots were taken every 10 ps and the explicit water molecules and ions were stripped out if present. The calculation was assisted by MM-PBSA script in AMBER 8 suite which uses the **`sander`** program in AMBER 8 suite to calculate molecular mechanics (MM) energies, the Delphi V program (22,23) to calculate the electronic contribution of the hydration free energy via Poisson-Boltzmann (PB) equation, and **`molsurf`** in Amber 8 to measure the solvent accessible surface area (SASA). Dielectric constants for the environment of the implicit

solvent were assumed to be 1.0 for solutes and 80.0 for solvent and the solvent probe radius was assumed to be 1.4 Å which is generally accepted for water molecules. The nonpolar contribution in the hydration free energy was estimated using the equation, $\Delta G_{nonpolar} = 0.00542$ kcal·Å^{-2}mol^{-1} × (SASA) + 0.92 kcal·mol^{-1}. The number of grids per Ångstrom was 2 and the number of iterations with linear PB equation was 500. The entropic contribution was estimated using the **nmode** program in AMBER 8 assuming the rigid rotor and harmonic approximation to the normal modes at 300 K. For the calculation of the entropy, the stripped snapshots were prepared in the same way as described above but collected every 100 ps. These structures were minimized with generalized Born method (10,11) for 50,000 steps and then minimized further with Newton-Raphson method up to 10 steps until the root-mean-square of the elements of the gradient vector reached below 10^{-4} kcal·mol^{-1}Å^{-1}. The calculated entropy was averaged for every 1 ns. Accordingly, the free energy difference of a binding reaction (ΔG) equals to $\Delta E_{MM} + \Delta G_{polar} + \Delta G_{nonpolar} - T\Delta S$, where $\Delta S = \Delta S_{trans} + \Delta S_{rot} + \Delta S_{vib}$.

Homo-(dA) duplexes bound with 5 coralyne molecules simulated using the *f99bsc0* force field were used for the cluster analysis. Cluster analysis based on RMSD was applied to two parts of the molecules, specifically the whole molecule including the homo-(dA) duplex and the 5 coralynes and also the minimal unit of coralyne intercalation which is composed of a single coralyne molecule and the four adenine bases surrounding the coralyne molecule. In the latter case, five units could be obtained per snapshot. For the both cases snapshots were taken every 20 ps from the MD trajectories. Clustering using **ptraj** from AmberTools 1.0 isolated 10 clusters with the K-means algorithm for each case (24).

Building the initial coralyne-free homo-(dA) duplex models

The first duplex (duplex type I from Figure S2 in the main text) was generated as a canonical B-DNA duplex (25) by the **nucgen** program from AMBER 8 suite (5). The second duplex (duplex type II) was reconstructed using the anti-parallel Hoogsteen duplex (PDB ID: 1GQU) (26). Specifically, the middle 4-mer of the DNA duplex was taken from the original structure and was extended to 12-mer DNA using Chimera ver.1 (27) by sequentially superimposing two base pairs over each other to extend the helix (deleting the extra nucleotides). Then, all the bases were converted into adenine using the same program. The third duplex (duplex type III) was built by manipulating the duplex type II. All the χ dihedral angles of the adenine bases residing in the strand with lower residue number (i.e. 1-12) were adjusted to create *syn* conformations. Likewise, all the χ angles of the other strand were modified into *anti* conformation. The duplex type IV was also created by modifying the duplex type II. That is, all the χ angles of the duplex type II were simply adjusted to create *syn* conformations. Schematic diagrams of the four structures are illustrated in Figure S2.

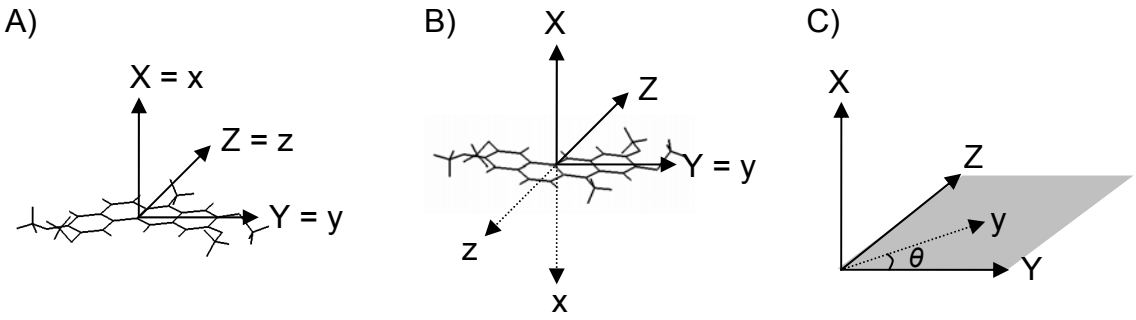


Figure S3: Axes for insertion of a coralyne molecule into a plane of intercalation. Axes indicated by upper-case characters are the axes of the plane of intercalation, axes with lower-case characters are those of the coralyne. A) The three internal axes of the coralyne (xyz axes) match the corresponding axes of the plane of intercalation (XYZ axes). B) After the initial placement, the coralyne is rotated 180 degrees about its y-axis (flipped coralyne). Thereupon the x-axis and z-axis of the coralyne are aligned to the opposite direction of the X-axis and Z-axis of the plane of intercalation, respectively. C) The coralyne positioned as either in A) or B) could be rotated about the x-axis of the plane of intercalation with the degree of θ (angle from Y-axis to y-axis counterclockwise). The coralyne still belongs to the plane indicated by gray color which passes through the Y- and Z-axes of the plane of intercalation.

Defining an internal coordinate system for the coralyne molecule

First, the normal axis of the plane which passes the four atoms, C11, C12, C17 and C20 in the coralyne molecule (see Figure 1 in the main text) was calculated based upon least-squares fitting. This axis will be referred as the x-axis of coralyne molecule. The vector that has the direction from the geometric center of C11 and C12 to the geometric center of C17 and C20 was then determined. Note that this vector is not always orthogonal to the x-axis of coralyne. The cross product of the two vectors was employed and it was defined as the z-axis of coralyne. Finally, the cross product of the z-axis and the x-axis were defined as the y-axis of coralyne, giving three orthogonal axes. Finally, the position of the C3 atom of coralyne was defined as the origin of the coralyne coordinate system. (Figure S3A)

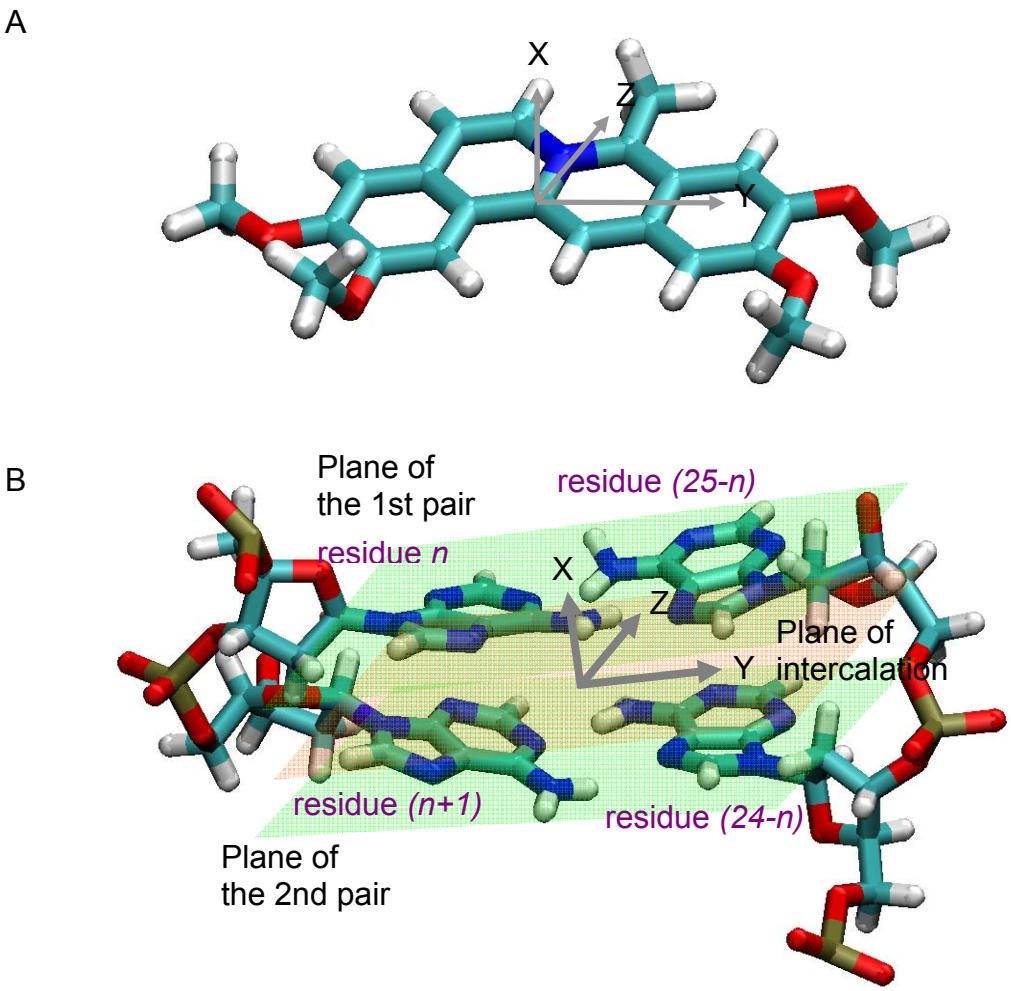


Figure S4: The internal axes for coralyne molecule (A) and the internal axes of the plane of intercalation (B). *n*'s indicate the residue numbers of the adenine bases (*n* = 1 to 11). Note that panels A and B are not drawn to the same scale. See the text for details on the definitions of the internal axes.

Defining the internal coordinate system for intercalating base pair planes

The geometric center of the base pairs of the homo-(dA) duplex and the xyz-axes of intercalation plane were defined similarly to that of the coralyne (Figure S4B). All the absolute coordinates of the six atoms (N1, C2, N3, C4, C5, and C6) from the four bases involved in the intercalation were measured. The origin (center) of the plane of intercalation was determined as the geometric center of those 24 atoms. The normal vectors of the two planes of base pairs were determined using 12 atoms of each pair by least-squares fitting. The normal vector of the plane of intercalation (x-axis) was defined as the average of the two normal vectors calculated above. Note that the normal vectors were designed to point upwards with respect to the geometry of the four bases, as shown in Figure S4. Another vector having the direction from the geometric

center of the 12 atoms from the two bases with lower residue numbers (*n* and *n*+1 in Figure S4B) to the other geometric center of the 12 atoms from the two bases with higher residue numbers ((24-*n*) and (25-*n*) in Figure S4B) was determined (n = 1-11). The cross product of the x-axis and this newly determined vector was defined as the z-axis of the plane of intercalation. The cross product of the z-axis and the x-axis was defined as the y-axis of the plane of intercalation, giving three orthogonal axes. In the same way, the plane of the base pairs were also defined as it is visualized as the green planes in Figure S4B.

The intercalation of a coralyne molecule into an intercalation pocket was carried out by moving the internal origin of the coralyne to the internal origin of the intercalation plane and matching their internal xyz axes. The coralyne could also be inserted in a flipped geometry. In this case, the x-axis of coralyne was aligned to the opposite direction of the x-axis and z-axis of the plane of intercalation (Figure S3A and S2B). A coralyne in the "flipped" state will be referred to as a flipped coralyne. Notice that the z-axis of the flipped coralyne also points in the opposite direction of the z-axis of the plane of intercalation, although the y-axes are still aligned. Coralyne could also be rotated with respect to the x-axis of the plane of the intercalation (Figure S3C). This angle will be referred as "coralyne angle" in the remainder of the text. Since the planes of the coralyne and the intercalation site are not always parallel during MD, the y-axis of the coralyne was projected on the plane of intercalation and the coralyne angle was measured from the y-axis of the plane of the intercalation to the projected y-axis of the coralyne counterclockwise (in the view from the positive direction of the x-axis of the plane of intercalation). This angle provides a means to see how the coralyne rotates during the simulation and to assess distinct minima along the rotation surface.

Docking coralyne molecules on groove-sides of poly(dA) duplexes

DNA groove-binding molecules, such as distamycin and netropsin, typically have crescent shapes that match the convex surface of the minor groove (28). Coralyne has a similar curved shape, suggesting the possibility that this molecules could interact with the coralyne-induced homo-(dA) structures via groove binding. Therefore, we built models of coralyne interacting with the homo-(dA) grooves where the convex edges of the coralyne molecules were facing the surfaces of the grooves. One complication encountered relates to the definition of the major and minor grooves as the homo-(dA) duplex structures are not canonical. We made the assumption that coralyne molecules would bind more favorably in the narrower groove, consistent with most DNA groove binding agents. For convenience, the narrower groves will be denoted as "minor grooves" in this report. As coralyne is an asymmetric molecule, it can bind in two distinct orientations, unless the homo-(dA) duplex has a center of inversion. For instance, all homo-(dA) duplexes with alternating base pairs have centers of inversion. In the other cases, coralyne molecules were placed either "head-up" (with the C20 atom in coralyne located closer to the first residue of adenine than the C11 atom) or "head-down" (where the C11 atom in coralyne is located closer to the first residue of adenine than it is to the C20 atom). When the glycosidic bond orientations of the bases are *trans*, the narrower groove is not obvious because both grooves have similar widths. In

this case, complexes for both grooves were built. In order to place a coralyne molecule onto the homo-(dA) duplex structure grooves, the coralyne molecules were first placed between the base pairs without adjustment of the coralyne angle in the same way described in the previous paragraph. Secondly, the coralyne was moved along to the z-axis 5-6 Å in either the positive or negative direction, depending on the location of the groove. Finally, the coralyne was rotated properly along the z-axis of the coralyne in the range of -60 to 60 degree to align the coralyne with the groove.

Table S1: Force field parameters for coralyne. Bond, angle, torsion and van der Waals parameters were determined based on the 'atom type' shown in Table S1 and the corresponding parameters from the GAFF force field.(4) Note that to build a complete residue/topology file, two missing angle parameters in GAFF were needed ('ha-ca-na' and 'na-ca-c3'). They were deduced using the **`parmchk`** program. Harmonic force constants chosen were 51.20 and 64.57 kcal $mol^{-1}rad^{-2}$ and the equilibrium angles were 112.42 and 115.06 degrees, respectively.

	Optimized Coordinates			Atom Type	RESP
Atom Name	x	y	z	in GAFF	Charge
C1	2.288	-0.988	-0.144	ca	-0.1056
C2	1.360	-2.017	-0.182	ca	0.1756
C3	-0.455	-0.446	-0.019	ca	0.1560
C4	0.454	0.579	0.014	ca	-0.3361
C5	1.825	0.352	-0.044	ca	0.2408
C6	-0.902	-2.809	-0.138	ca	-0.2565
C7	-1.896	-0.220	0.043	ca	-0.0583
C8	-2.774	-1.295	0.020	ca	0.1871
C9	-2.211	-2.616	-0.074	ca	-0.1546
C10	-4.160	-1.101	0.093	ca	-0.4521
H1	-4.802	-1.960	0.082	ha	0.2103
C11	-4.682	0.166	0.187	ca	0.3930
C12	-3.793	1.275	0.195	ca	0.1599
C13	-2.446	1.076	0.127	ca	-0.2444
H2	-0.492	-3.787	-0.207	ha	0.2058
H3	0.105	1.585	0.090	ha	0.2049
H4	-2.855	-3.474	-0.094	ha	0.1886
H5	-1.839	1.956	0.157	ha	0.2080
N1	0.024	-1.743	-0.115	na	0.1760
C14	1.773	-3.463	-0.297	c3	-0.3196
H6	1.450	-4.032	0.567	hc	0.1312
H7	1.356	-3.917	-1.188	hc	0.1312
H8	2.841	-3.556	-0.362	hc	0.1312
O1	-4.266	2.527	0.338	os	-0.3535
O2	-5.973	0.461	0.276	os	-0.2557
C15	-4.893	3.118	-0.787	c3	-0.0384
H9	-4.200	3.171	-1.619	h1	0.0907
H10	-5.775	2.566	-1.075	h1	0.0907
H11	-5.171	4.115	-0.483	h1	0.0907
C16	-6.937	-0.565	0.342	c3	-0.1338
H12	-6.933	-1.161	-0.563	h1	0.1151

H13	-6.766	-1.198	1.205	h1	0.1151
H14	-7.888	-0.069	0.442	h1	0.1151
C17	3.713	-1.195	-0.199	ca	-0.2349
H15	4.141	-2.170	-0.287	ha	0.2237
C18	2.746	1.432	-0.011	ca	-0.4717
H16	2.359	2.428	0.055	ha	0.2056
O3	5.027	2.118	-0.065	os	-0.2615
O4	5.893	-0.380	-0.283	os	-0.3620
C19	4.084	1.199	-0.073	ca	0.4547
C20	4.574	-0.165	-0.163	ca	0.1662
C21	4.695	3.492	-0.032	c3	-0.1362
H17	4.101	3.760	-0.896	h1	0.1199
H18	4.163	3.735	0.879	h1	0.1199
H19	5.633	4.020	-0.059	h1	0.1199
C22	6.682	-0.280	0.893	c3	-0.0064
H20	6.634	0.715	1.310	h1	0.0848
H21	6.353	-1.007	1.627	h1	0.0848
H22	7.693	-0.501	0.592	h1	0.0848

Table S2: List of restraints applied between adenine bases during equilibration phases of the MD simulation. The penalty on the potential energy function varies as a function of the restraint distance or angle (R) as follows: When $r_2 \leq R < r_3$, there is no penalty. When $r_1 \leq R < r_2$ or $r_3 \leq R < r_4$, a harmonic penalty with a force constant of 5 kcal mol^{-1}Å^{-2} or 5 kcal mol^{-1}rad^{-2} was applied. When $R < r_1$ or $r_4 \leq R$, a linear penalty is applied starting from the current restraint penalty value at r_1 and r_4.

Base pair	[a]	Atoms involved in the restraints[b]	r_1	r_2	r_3	r_4
cisWW cisWW-rev	D	N6(1-12) - N1(13-24)	0.00	2.60	3.10	5.00
	D	N6(1-12) - N1(13-24)	0.00	3.61	4.31	6.94
	T	N1(1-12) – C6(1-12) – N1(13-24) – C2(13-24)	-45	-18	18	45
	T	C2(1-12) – C5(1-12) – C6(13-24) – N3(13-24)	-45	-18	18	45
cisWW-alt cisWW-alt-rev	D	N6(even) - N1(odd)	0.00	2.60	3.10	5.00
	D	N6(even) - N1(odd)	0.00	3.61	4.31	6.94
	T	N1(even) – C6(even) – N1(odd) – C2(odd)	-45	-18	18	45
	T	C2(even) – C5(even) – C6(odd) – N3(odd)	-45	-18	18	45
cisHH cisHH-rev	D	N6(1-12) - N7(13-24)	0.00	2.60	3.10	5.00
	D	N7(1-12) - C8(13-24)	0.00	2.54	3.03	4.89
	T	C4(1-12) – N1(1-12) – C5(13-24) – N9(13-24)	-45	-18	18	45
	T	C5(1-12) – C6(1-12) – N7(13-24) – C8(13-24)	-45	-18	18	45
cisHH-alt cisHH-alt-rev	D	N6(even) - N7(odd)	0.00	2.60	3.10	5.00
	D	N7(even) - C8(odd)	0.00	2.54	3.03	4.89
	T	C4(even) – N1(even) – C5(odd) – N9(odd)	-45	-18	18	45
	T	C5(even) – C6(even) – N7(odd) – C8(odd)	-45	-18	18	45
cisWS cisWS-rev	D	N6(1-12) - N3(13-24)	0.00	2.60	3.10	5.00
	D	N1(1-12) - N9(13-24)	0.00	3.0	3.58	5.77
	T	N1(1-12) – C6(1-12) – N3(13-24) – C4(13-24)	-45	-18	18	45
	T	C2(1-12) – C5(1-12) – C2(13-24) – C5(13-24)	-45	-18	18	45
cisWS-alt cisWS-alt-rev	D	N6(even) - N3(odd)	0.00	2.60	3.10	5.00
	D	N1(even) - N9(odd)	0.00	3.0	3.58	5.77
	T	N1(even) – C6(even) – N3(odd) – C4(odd)	-45	-18	18	45
	T	C2(even) – C5(even) – C2(odd) – C5(odd)	-45	-18	18	45
transWH transWH-rev	D	N1(1-12) - N6(13-24)	0.00	2.60	3.10	5.00
	D	N6(1-12) - N7(13-24)	0.00	2.60	3.10	5.00

	T	C6(1-12) – N1(1-12) – C6(13-24) – C5(13-24)	-45	-18	18	45
	T	C5(1-12) – C2(1-12) – N1(13-24) – C4(13-24)	-45	-18	18	45
transWH-alt	D	N1(even) - N6(odd)	0.00	2.60	3.10	5.00
transWH-alt-rev	D	N6(even) - N7(odd)	0.00	2.60	3.10	5.00
	T	C6(even) – N1(even) – C6(odd) – C5(odd)	-45	-18	18	45
	T	C5(even) – C2(even) – N1(odd) – C4(odd)	-45	-18	18	45
transHS	D	N3(1-12) - N6(13-24)	0.00	2.60	3.10	5.00
transHS-rev	D	C2(1-12) - N9(13-24)	0.00	2.39	2.85	4.60
	T	C2(1-12) – N3(1-12) – C6(13-24) – C5(13-24)	-45	-18	18	45
	T	N1(1-12) – C4(1-12) – N1(13-24) – C4(13-24)	-45	-18	18	45
transHS-alt	D	N3(even) - N6(odd)	0.00	2.60	3.10	5.00
transHS-alt-rev	D	C2(even) - N9(odd)	0.00	2.39	2.85	4.60
	T	C2(even) – N3(even) – C6(odd) – C5(odd)	-45	-18	18	45
	T	N1(even) – C4(even) – N1(odd) – C4(odd)	-45	-18	18	45
cisWH(a)	D	N7(1-12) - N6(13-24)	0.00	2.60	3.10	5.00
	D	C8(1-12) - N1(13-24)	0.00	3.74	4.46	7.19
	T	C8(1-12) – N7(1-12) – C6(13-24) – C1(13-24)	-45	-18	18	45
	T	N9(1-12) – C5(1-12) – C5(13-24) – C2(13-24)	-45	-18	18	45
cisWH(a)-alt	D	N7(even) - N6(odd)	0.00	2.60	3.10	5.00
	D	C8(even) - N1(odd)	0.00	3.74	4.46	7.19
	T	C8(even) – N7(even) – C6(odd) – C1(odd)	-45	-18	18	45
	T	N9(even) – C5(even) – C5(odd) – C2(odd)	-45	-18	18	45
cisWH(a)-rev	D	N7(13-24) - N6(1-12)	0.00	2.60	3.10	5.00
	D	C8(13-24) - N1(1-12)	0.00	3.74	4.46	7.19
	T	C8(13-24) – N7(13-24) – C6(1-12) – C1(1-12)	-45	-18	18	45
	T	N9(13-24) – C5(13-24) – C5(1-12) – C2(1-12)	-45	-18	18	45
cisWH(a)-alt-rev	D	N7(odd) - N6(even)	0.00	2.60	3.10	5.00
	D	C8(odd) - N1(even)	0.00	3.74	4.46	7.19
	T	C8(odd) – N7(odd) – C6(even) – C1(even)	-45	-18	18	45
	T	N9(odd) – C5(odd) – C5(even) – C2(even)	-45	-18	18	45
cisWH(b)	D	N6(1-12) - N1(13-24)	0.00	2.60	3.10	5.00
	D	N7(1-12) - C2(13-24)	0.00	2.39	2.85	4.60
	T	C5(1-12) – C6(1-12) – N1(13-24) – C2(13-24)	-45	-18	18	45
	T	C4(1-12) – N1(1-12) – C6(13-24) – N3(13-24)	-45	-18	18	45
cisWH(b)-alt	D	N6(even) - N1(odd)	0.00	2.60	3.10	5.00
	D	N7(even) - C2(odd)	0.00	2.39	2.85	4.60
	T	C5(even) – C6(even) – N1(odd) – C2(odd)	-45	-18	18	45
	T	C4(even) – N1(even) – C6(odd) – N3(odd)	-45	-18	18	45
cisWH(b)-rev	D	N6(13-24) - N1(1-12)	0.00	2.60	3.10	5.00
	D	N7(13-24) - C2(1-12)	0.00	2.39	2.85	4.60
	T	C5(13-24) – C6(13-24) – N1(1-12) – C2(1-12)	-45	-18	18	45
	T	C4(13-24) – N1(13-24) – C6(1-12) – N3(1-12)	-45	-18	18	45
cisWH(b)-alt-rev	D	N6(odd) - N1(even)	0.00	2.60	3.10	5.00
	D	N7(odd) - C2(even)	0.00	2.39	2.85	4.60
	T	C5(odd) – C6(odd) – N1(even) – C2(even)	-45	-18	18	45
	T	C4(odd) – N1(odd) – C6(even) – N3(even)	-45	-18	18	45
transWW	D	N6(1-12) - N1(13-24)	0.00	2.60	3.10	5.00
	D	N1(1-12) - N6(13-24)	0.00	2.60	3.10	5.00
	T	N1(1-12) – C6(1-12) – N1(13-24) – C6(13-24)	-45	-18	18	45
	T	C2(1-12) – C5(1-12) – C2(13-24) – C5(13-24)	-45	-18	18	45
transWW-alt	D	N6(even) - N1(odd)	0.00	2.60	3.10	5.00
	D	N1(even) - N6(odd)	0.00	2.60	3.10	5.00
	T	N1(even) – C6(even) – N1(odd) – C6(odd)	-45	-18	18	45
	T	C2(even) – C5(even) – C2(odd) – C5(odd)	-45	-18	18	45
transHH	D	N6(1-12) - N7(13-24)	0.00	2.60	3.10	5.00
	D	N7(1-12) - N6(13-24)	0.00	2.60	3.10	5.00
	T	C5(1-12) – C6(1-12) – C5(13-24) – C6(13-24)	-45	-18	18	45
	T	C4(1-12) – N1(1-12) – C4(13-24) – N1(13-24)	-45	-18	18	45
transHH-alt	D	N6(even) - N7(odd)	0.00	2.60	3.10	5.00

	D	N7(even) - N6(odd)	0.00	2.60	3.10	5.00
	T	C5(even) – C6(even) – C5(odd) – C6(odd)	-45	-18	18	45
	T	C4(even) – N1(even) – C4(odd) – N1(odd)	-45	-18	18	45
transWS	D	N6(1-12) - N3(13-24)	0.00	2.60	3.10	5.00
	D	N1(1-12) - C2(13-24)	0.00	3.61	4.30	6.93
	T	N1(1-12) – C6(1-12) – N3(13-24) – C2(13-24)	-45	-18	18	45
	T	C2(1-12) – C5(1-12) – C4(13-24) – N1(13-24)	-45	-18	18	45
transWS-alt	D	N6(even) - N3(odd)	0.00	2.60	3.10	5.00
	D	N1(even) - C2(odd)	0.00	3.61	4.30	6.93
	T	N1(even) – C6(even) – N3(odd) – C2(odd)	-45	-18	18	45
	T	C2(even) – C5(even) – C4(odd) – N1(odd)	-45	-18	18	45
transWS-rev	D	N6(13-24) - N3(1-12)	0.00	2.60	3.10	5.00
	D	N1(13-24) - C2(1-12)	0.00	3.61	4.30	6.93
	T	N1(13-24) – C6(13-24) – N3(1-12) – C2(1-12)	-45	-18	18	45
	T	C2(13-24) – C5(13-24) – C4(1-12) – N1(1-12)	-45	-18	18	45
transWS-alt-rev	D	N6(odd) - N3(even)	0.00	2.60	3.10	5.00
	D	N1(odd) - C2(even)	0.00	3.61	4.30	6.93
	T	N1(odd) – C6(odd) – N3(even) – C2(even)	-45	-18	18	45
	T	C2(odd) – C5(odd) – C4(even) – N1(even)	-45	-18	18	45

[a] 'D' and 'T' denote a distance restraint and a torsional (dihedral) restraint, respectively.
[b] Two atoms and four atoms are listed for distance restraints and dihedral restraints. Numbers in the parentheses denote the residue numbers of adenine bases.

Simulation and selection of corlyane-free homo-(dA) duplexes

Although an explicit solvent model usually gives more reliable results, implicit solvent models are significantly faster and were assumed to be adequate for ranking the structural stability of duplexes for the initial modeling effort. In an attempt to stabilize the coralyne-free homo-(dA) model structure, restraints between the base pairs were applied. Even with restraints, a number of the structures were not stable in simulation. Therefore, we assumed that any instability of homo-(dA) duplexes with the restraints implied structural problems with the homo-(dA) duplex models. For the first 1 ns, the MD simulations for all 40 homo-(dA) duplexes were performed with the full set of base-pair restraints listed in Table S2. After the initial 1 ns simulation, MD simulations continued for another 1 ns with a partial modification of the initial restraints to remove all the distance restraints that were not involved in H-bonds. The H-bond occupancy was measured every 1 ps (Table S3). Base pairs with two H-bonds, such as transHH, transWH and transWW maintained the H-bond occupancy over the two periods because their restraints were not modified in the second period. In the case of transHS and transHS-alt, the occupancies increased after the removal of one distance restraint while most of the occupancies decreased or maintained similar levels. With the exception of the transHS and transHS-alt model structures, the restraints on the base pairs maintained the homo-(dA) duplex H-bonding.

Based on the H-bond occupancies during the initial MD simulation on the coralyne-free homo-(dA) duplexes, some of the poor structures listed in Table S3 were removed from consideration. In order to avoid discarding potentially reasonable candidate duplexes, a rather generous criterion for remaining in the pool of plausible structures was employed, specifically only duplexes with less than 50% H-bonding occupancy during the final 1 ns

were considered unreasonable and removed. The retained model structures are indicated in bold face in Table S3. Although there are a number of reasons for low occupancy, including the inappropriateness of the initial model structure, steric hindrance, or high tension in the backbone chains, the low occupancy also potentially indicates that the specific base pairs are not suitable for building longer homo-(dA) duplexes.

Note also that some of the structures which had an H-bond occupancy > 50% encountered problems when building complexes with coralynes. To insert coralynes, the gaps between the adjacent base pairs or the groove binding sites must be maintained and form repetitive structures that are similar to one another during simulations. One of the general modeling criteria applied was that all the planes of the base pairs in the duplex need to be nearly parallel. Otherwise, the overall shape of the poly(dA) will be distorted and such small segments of poly(dA) structures cannot form long-chained poly(dA) duplexes, as were observed experimentally (29). As we could not ignore the possibility that the insertion of the coralyne would stabilize the partially distorted poly(dA) structures, we also applied rather generous criterion. The angles between the individual x-axes of the base pair planes and the average vector of all the x-axes was measured, and all the angles were averaged. If the base pair planes are perfectly parallel the average angle would be zero. This angle can also be interpreted as the tilt of the base pairs from the longitudinal direction of the duplex. This tilting angle is only 1° in normal B-DNA (30). Therefore, it is reasonable to omit structures with significantly higher tilting. Based on the tilting angles from the last 1 ns, structures with angles above 20° were discarded. The duplexes in bold face in Table S4 indicate the selected structures.

Table S3: The average occupancy of the hydrogen bonds. For the first 1 ns, the full restraints (Table S2) between base pairs were applied and in the next 1 ns, the distance restraints which restrain non-H-bond interactions were omitted. The homo-(dA) duplexes typed with bold face have over 50% average occupancy of the H-bonds for the last 1 ns.

	Initial 1 ns (Full restraints)		Next 1ns (Partial restraints)	
Homo-(dA) duplex	Mean(%)	stdev	Mean(%)	Stdev
cisHH	**68.87**	**8.06**	**52.43**	**16.12**
cisHH-alt	64.56	10.05	31.35	24.82
cisHH-alt-rev	**74.40**	**10.94**	**70.88**	**15.13**
cisHH-rev	**93.69**	**5.45**	**88.38**	**14.02**
cisWH(a)	**52.00**	**13.73**	**54.45**	**14.43**
cisWH(a)-alt	**52.52**	**7.14**	**50.79**	**7.71**
cisWH(a)-alt-rev	41.87	24.07	18.95	20.76
cisWH(a)-rev	42.63	21.22	38.68	26.65
cisWH(b)	50.21	4.36	43.77	6.52
cisWH(b)-alt	**54.64**	**9.82**	**53.39**	**7.65**
cisWH(b)-alt-rev	26.32	23.04	17.24	17.87
cisWH(b)-rev	35.81	26.34	15.70	18.67
cisWS	51.14	14.52	5.09	7.28
cisWS-alt	44.12	31.90	29.72	23.45
cisWS-alt-rev	35.98	26.55	14.80	24.81
cisWS-rev	**81.72**	**15.69**	**70.89**	**36.95**
cisWW	55.38	3.77	46.25	7.39
cisWW-alt	49.93	10.07	26.00	23.18
cisWW-alt-rev	82.85	12.92	46.84	42.95
cisWW-rev	64.31	15.44	31.57	24.23
transHH	**65.45**	**3.08**	**65.15**	**2.98**
transHH-alt	**67.85**	**8.34**	**65.95**	**3.61**
transHS	0.04	0.08	49.45	11.53
transHS-alt	0.00	0.00	39.98	20.85
transHS-alt-rev	0.03	0.09	23.34	24.95
transHS-rev	0.17	0.28	32.27	32.76
transWH	**67.02**	**2.92**	**66.69**	**3.19**
transWH-alt	**62.67**	**6.76**	**62.40**	**8.97**
transWH-alt-rev	**87.44**	**11.59**	**89.19**	**9.34**
transWH-rev	**91.26**	**3.32**	**90.65**	**4.08**
transWS	**67.31**	**9.09**	**66.93**	**7.22**
transWS-alt	47.16	14.07	48.85	15.43
transWS-alt-rev	28.35	35.27	23.17	23.99
transWS-rev	12.12	15.93	12.99	15.98
transWW	**68.41**	**2.76**	**68.49**	**2.79**
transWW-alt	**66.63**	**4.79**	**67.86**	**4.82**
cisHS	61.10	16.36	34.96	25.13
cisHS-alt	45.18	7.47	47.90	7.96
cisHS-alt-rev	26.22	24.96	13.63	24.62
cisHS-rev	32.23	18.09	5.85	8.99

Table S4: Angle deviation from the principal axis. For the first 1 ns, the full restraints (Table S2) between base pairs were applied and in the next 1 ns, the distance restraints which restrain the non-H-bond interactions were omitted. The homo-(dA) duplexes typed with bold face have less than 20° average tilting angles for the last 1 ns.

	Initial 1 ns (Full restraints)		Next 1ns (Partial restraints)	
	Mean(degree)	stdev	Mean(degree)	stdev
cisHH	27.04	8.74	26.11	7.50
cisHH-alt	36.17	15.33	30.71	13.14
cisHH-alt-rev	35.71	4.86	30.32	11.05
cisHH-rev	24.85	3.25	23.16	9.09
cisWH(a)	**14.56**	**8.12**	**15.09**	**6.92**
cisWH(a)-alt	21.83	4.73	24.17	6.87
cisWH(a)-alt-rev	37.64	10.84	35.66	12.81
cisWH(a)-rev	**23.43**	**12.41**	**16.47**	**6.96**
cisWH(b)	**12.18**	**2.24**	**13.11**	**1.61**
cisWH(b)-alt	**14.93**	**3.81**	**18.15**	**4.70**
cisWH(b)-alt-rev	**24.36**	**8.33**	**18.68**	**5.95**
cisWH(b)-rev	37.68	8.68	34.20	12.16
cisWS	25.13	11.04	25.65	11.24
cisWS-alt	27.01	12.74	26.75	8.71
cisWS-alt-rev	41.81	11.37	44.35	10.97
cisWS-rev	47.82	10.32	52.51	11.56
cisWW	12.63	4.04	22.37	2.81
cisWW-alt	16.84	9.56	21.22	9.38
cisWW-alt-rev	31.64	14.13	32.46	5.16
cisWW-rev	45.54	11.00	43.04	11.00
transHH	**20.28**	**7.34**	**19.37**	**8.41**
transHH-alt	22.86	9.87	35.33	12.60
transHS	25.09	7.56	25.35	6.25
transHS-alt	31.00	8.47	22.29	11.16
transHS-alt-rev	53.32	6.88	48.17	11.32
transHS-rev	44.92	15.83	51.60	9.66
transWH	**12.33**	**5.80**	**11.75**	**4.91**
transWH-alt	33.66	16.73	40.85	20.31
transWH-alt-rev	41.21	12.09	39.18	6.91
transWH-rev	35.20	6.83	35.92	9.15
transWS	**17.45**	**3.04**	**19.13**	**5.30**
transWS-alt	**27.37**	**12.94**	**18.67**	**6.66**
transWS-alt-rev	23.62	11.22	29.48	8.41
transWS-rev	33.44	12.82	28.23	4.25
transWW	**12.44**	**5.49**	**12.77**	**6.00**
transWW-alt	32.35	15.06	29.53	7.49
cisHS	16.81	11.95	24.57	4.10
cisHS-alt	**11.23**	**5.80**	**10.12**	**2.98**
cisHS-alt-rev	26.72	5.26	26.26	12.41
cisHS-rev	34.14	7.89	30.89	9.89

Six structures were considered reasonable, those which satisfy both the criteria from the average occupancy of the H-bonds and the average tilting angles: cisWH(a), cisWH(b)-alt, transHH, transWH, transWS, and transWW.

Table S5: The binding free energies (Δ*G*) of the groove binding models calculated using MM-PBSA method. To verify that coralyne prefers to intercalate rather than bind in the major groove, MD simulations were performed on groove-binding models compared to intercalative models and MM-PBSA performed to estimate the binding energies. Only the polar and non-polar contributions of the free energies were calculated. Solute entropic contributions were not calculated. The free energies of the complex (homo-(dA) duplex and coralyne), receptor (homo-(dA) duplex) and ligand (coralyne) were calculated using the partial structures extracted from the same snapshot. Δ*G* was averaged over the first 1 ns period. In the names of the complexes, the information describing the complexes is linked with hyphens. 'c1' means one coralyne molecules, 'm1' and 'm2' indicate the orientations of the coralyne ('m1': head-up and 'm2': head-down, see the method section for the details). The additional last digit (-2) specified for some of the complexes indicates the other groove. This is only applicable to 'trans' duplexes with no center of inversion. The unit of the energy is kcal/mol.

Complex	Δ*G*	*stdev*
cisWH(a)-c1-m1	-6.96	3.40
cisWH(a)-c1-m2	-6.78	3.46
cisWH(b)-alt-c1-m1	-6.54	5.19
transHH-c1-m1	-5.27	4.78
transHH-c1-m2	-5.26	5.58
transWH-c1-m1	3.80	4.26
transWH-c1-m1-2	-7.22	3.47
transWH-c1-m2	2.63	5.50
transWH-c1-m2-2	-6.62	3.72
transWS-c1-m1	-7.49	4.40
transWS-c1-m1-2	8.32	5.88
transWS-c1-m2	-11.1	3.98
transWS-c1-m2-2	8.17	7.02
transWW-c1-m1	-0.91	5.76
transWW-c1-m2	-1.95	4.81

Table S6: The binding free energies (Δ*G*) of the intercalation binding models calculated using MM-PBSA method. Only the polar and non-polar contributions of the free energies were calculated. The free energies of the complex (homo-(dA) duplex and coralyne), receptor (homo-(dA) duplex), and ligand (coralyne) were calculated using the partial structures extracted from the same snapshots. Δ*G* was averaged over the final 1 ns period. The naming convention of the complexes is explained in the text. The occupancies of the H-bonds were calculated separately for the two base pairs (between residue 6 and 19 and between residue 7 and 18). The structures listed in bold were selected for further study. As the coralyne could reorient, many of the models converged to similar structures. The cisWH(a)-c1-135d and cisWH(a)-c1-180d model structures both had low binding energies but their coralyne angles were very similar to the cisWH(a)-c1-270d model structure, so only the latter was retained. In the group of cisWH(b) model structures, only the cisWH(b)-alt-c1-45d model structure showed average H-bond occupancy > 40%. For the transHH models, two structures with similar binding energies were selected from a set of five models that sampled two different coralyne angles. Three structures were selected for transWH type because of similar binding free energies: transWH-c1-180d, transWH-c1-270d-f, and transWH-c1-45d. The transWH-c1-180d-f and transWH-c1-225d-f were the same as transWH-c1-270d-f, and transWH-c1-270d was the same as transWH-c1-180d as far as the coralyne angle is concerned. TransWS-c1-315d-f and transWS-c1-45d-f had similar coralyne angles and transWS-c1-45d-f was chosen due to the slightly lower binding free energies. We chose transWW-c1-0d, transWW-c1-135d-f, and transWW-c1-315d-f as they had similar binding free energies, while transWW-c1-225d-f was considered to have the same coralyne angle as transWW-c1-315d-f

	H-bond occupancy (%)		Δ*G* (kcal/mol)		Coralyne angle (°)	
Complex	Res 6-19	Res 7-18	*Avg*	*Dev*	*Avg*	*Dev*
cisWH(a)-c1-0d	0.0	0.0	-18.97	7.17	331.76	24.70
cisWH(a)-c1-45d	15.2	0.0	-20.25	4.98	43.79	20.66
cisWH(a)-c1-90d	0.0	0.0	-17.33	4.94	57.11	28.52
cisWH(a)-c1-135d	45.1	63.9	-25.41	3.51	216.96	8.25
cisWH(a)-c1-180d	54.7	31.0	-23.97	4.60	237.87	25.36
cisWH(a)-c1-225d	8.9	55.1	-22.68	3.05	237.96	12.87
cisWH(a)-c1-270d	**42.3**	**63.8**	**-25.33**	**3.42**	**216.78**	**6.73**
cisWH(a)-c1-315d	22.7	35.6	-28.19	3.51	2.29	8.01
cisWH(a)-c1-0d-f	0.0	46.7	-25.01	3.83	29.91	9.89
cisWH(a)-c1-45d-f	20.0	0.0	-19.99	4.15	60.53	13.25
cisWH(a)-c1-90d-f	0.0	5.8	-22.17	3.91	60.67	26.85
cisWH(a)-c1-135d-f	32.2	30.3	-27.16	3.11	158.03	6.07
cisWH(a)-c1-180d-f	0.0	1.7	-19.19	3.42	254.35	19.31
cisWH(a)-c1-225d-f	6.3	36.6	-18.11	3.28	264.04	12.66
cisWH(a)-c1-270d-f	14.2	60.6	-20.84	4.07	260.28	16.03
cisWH(b)-alt-c1-0d	73.6	0.0	-21.49	4.38	339.79	6.81
cisWH(b)-alt-c1-45d	**75.5**	**82.0**	**-28.01**	**2.75**	**10.03**	**5.05**
cisWH(b)-alt-c1-90d	38.4	0.0	-22.20	4.42	53.21	11.84
cisWH(b)-alt-c1-135d	0.0	0.0	-18.10	4.17	90.07	13.76
cisWH(b)-alt-c1-180d	0.0	85.5	-18.67	4.64	105.18	15.48

cisWH(b)-alt-c1-225d	11.6	2.3	-19.42	4.84	278.53	30.65
cisWH(b)-alt-c1-270d	7.9	0.0	-21.85	3.94	230.89	12.00
cisWH(b)-alt-c1-315d	0.0	68.8	-15.89	3.87	260.58	12.76
transHH-c1-0d	**56.8**	**58.9**	**-26.15**	**3.47**	**22.25**	**7.03**
transHH-c1-45d	57.2	61.8	-26.84	3.48	21.14	5.17
transHH-c1-90d	61.9	61.2	-26.49	3.59	21.89	7.04
transHH-c1-135d	38.1	59.8	-23.00	3.47	95.71	10.94
transHH-c1-180d	59.7	56.0	-22.87	4.02	249.48	28.15
transHH-c1-225d	**58.3**	**59.6**	**-25.46**	**3.66**	**219.93**	**15.55**
transHH-c1-270d	58.7	58.6	-24.29	3.86	220.32	14.97
transHH-c1-0d-f	64.6	58.3	-20.83	3.19	54.62	11.79
transHH-c1-45d-f	62.8	59.4	-21.95	3.09	54.91	10.26
transHH-c1-90d-f	41.8	57.3	-19.82	3.42	52.82	20.99
transHH-c1-135d-f	0.0	52.8	-22.70	3.76	213.96	20.36
transHH-c1-180d-f	62.0	51.0	-21.55	3.43	224.24	13.69
transHH-c1-225d-f	0.0	55.3	-18.45	4.12	231.07	27.44
transHH-c1-270d-f	59.8	0.0	-21.23	3.35	219.67	12.68
transHH-c1-315d-f	60.4	61.2	-21.48	3.17	55.07	12.26
transWH-c1-0d	62.3	65.8	-26.60	4.04	39.17	12.34
transWH-c1-45d	**59.0**	**66.8**	**-26.53**	**3.72**	**29.78**	**17.06**
transWH-c1-90d	0.0	0.0	-22.70	2.95	32.89	8.58
transWH-c1-135d	54.8	0.0	-25.77	2.93	36.61	7.39
transWH-c1-180d	**61.2**	**59.8**	**-32.48**	**3.12**	**193.57**	**5.81**
transWH-c1-225d	47.8	38.6	-21.93	4.07	229.15	27.43
transWH-c1-270d	56.7	56.1	-31.79	2.56	194.81	5.82
transWH-c1-0d-f	63.4	60.7	-21.77	3.54	74.01	25.65
transWH-c1-45d-f	63.3	64.1	-21.79	3.71	66.07	14.09
transWH-c1-90d-f	13.7	0.0	-26.00	2.84	11.47	12.96
transWH-c1-180d-f	56.3	62.5	-24.62	3.71	203.75	25.35
transWH-c1-225d-f	28.0	61.9	-24.26	4.61	184.82	17.78
transWH-c1-270d-f	**53.9**	**59.7**	**-27.39**	**3.77**	**184.24**	**13.07**
transWH-c1-315d-f	0.0	0.3	-23.46	3.67	180.42	11.78
transWS-c1-0d	65.7	0.0	-16.57	6.48	34.01	11.51
transWS-c1-45d	59.8	44.7	-12.14	3.56	47.81	8.75
transWS-c1-90d	0.0	20.5	-14.87	4.38	346.22	23.89
transWS-c1-135d	0.0	32.8	-25.28	4.23	181.88	8.89
transWS-c1-180d	0.0	37.7	-23.23	3.93	186.60	10.50
transWS-c1-225d	57.1	59.0	-18.50	5.07	219.51	4.80
transWS-c1-270d	0.0	22.3	-14.29	4.98	300.67	23.24
transWS-c1-315d	0.0	50.7	-16.26	4.72	312.08	16.99
transWS-c1-0d-f	41.7	58.0	-24.71	4.24	336.06	6.26
transWS-c1-45d-f	**45.4**	**57.9**	**-25.02**	**4.28**	**336.51**	**7.25**
transWS-c1-90d-f	2.8	0.0	-17.34	4.88	96.79	18.10
transWS-c1-135d-f	1.5	39.2	-13.76	4.77	98.22	17.49
transWS-c1-180d-f	0.0	0.0	-18.51	6.43	145.19	18.12
transWS-c1-225d-f	0.0	47.9	-16.95	4.84	127.18	10.36
transWS-c1-270d-f	37.4	57.7	-24.60	3.95	336.01	5.88
transWS-c1-315d-f	41.5	58.2	-24.97	4.06	335.20	6.50
transWW-c1-0d	**63.8**	**62.9**	**-28.73**	**2.95**	**344.21**	**3.85**
transWW-c1-45d	0.0	52.2	-18.74	4.10	97.73	21.73

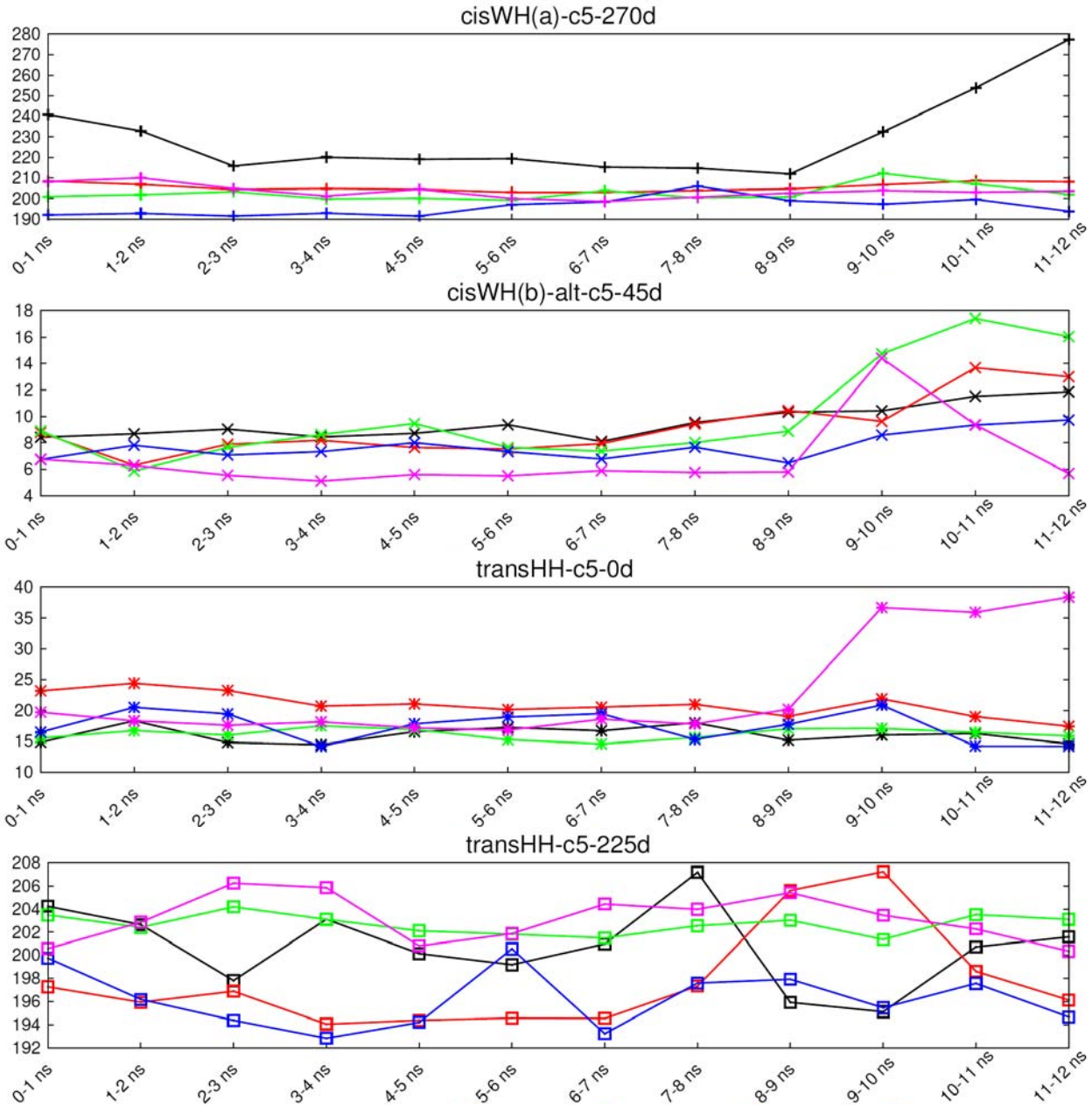
transWW-c1-90d	0.0	51.9	-17.71	3.17	91.46	11.39
transWW-c1-135d	29.2	60.6	-21.16	4.83	102.48	36.54
transWW-c1-180d	59.7	0.0	-28.48	2.82	171.50	5.04
transWW-c1-225d	62.7	53.7	-13.28	5.67	244.33	8.06
transWW-c1-270d	63.5	59.7	-21.95	2.95	242.12	7.66
transWW-c1-315d	60.9	0.0	-19.20	2.95	247.02	6.34
transWW-c1-0d-f	0.0	64.5	-24.72	3.36	322.04	8.97
transWW-c1-45d-f	61.7	61.4	-21.01	2.79	82.33	9.34
transWW-c1-135d-f	**57.3**	**64.3**	**-31.06**	**2.73**	**148.89**	**4.93**
transWW-c1-180d-f	0.0	51.4	-18.64	3.38	121.49	15.56
transWW-c1-225d-f	52.5	60.2	-30.13	3.10	326.35	5.28
transWW-c1-270d-f	0.0	58.5	-15.79	5.35	288.57	23.03
transWW-c1-315d-f	**61.1**	**64.4**	**-29.68**	**2.74**	**325.29**	**4.09**

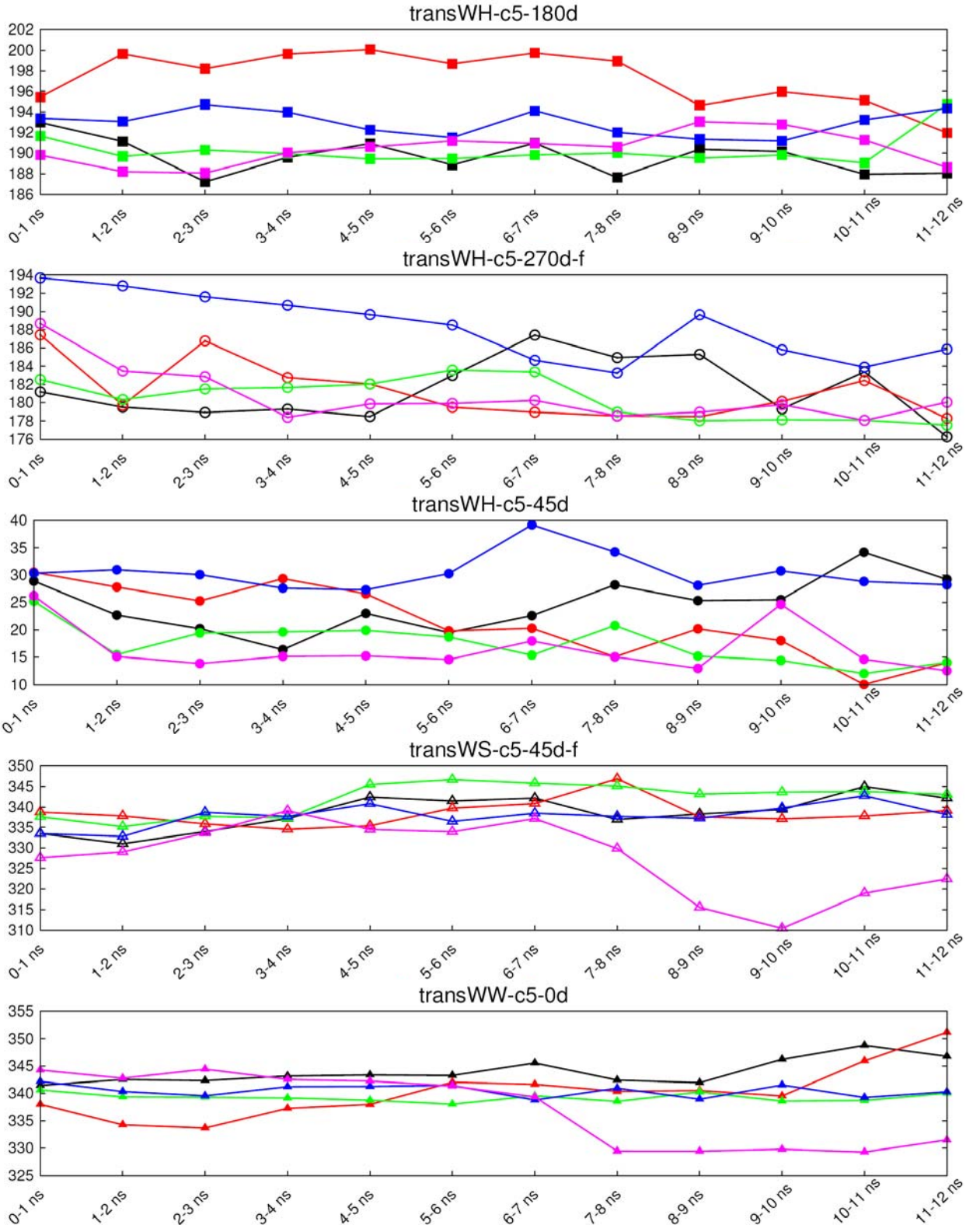
Coralyne angles of homo-(dA)-coralyne complexes in explicit solvent

From the simulations of homo-(dA)-coralyne complexes in explicit water, coralyne angles were also measured to confirm that every coralyne molecule was in the same orientation with respect to their surrounding adenine residues (Figure S5-7). The results of the initial 6 ns simulations are displayed in Figure S5. Except for the first coralyne in cisWH(a)-c5-270d, the coralyne angles of the other coralynes ranged from 190° to 210°. The coralyne angles in transHH-c5-0d were maintained between 15° and 25°, but coralyne 5 deviated a little from the average. All the coralyne molecules in each of the transHH-c5-225d, transWH-c5-180d, and transWH-c5-270d-f models shared similar values of their respective coralyne angle and they were packed within 20° around the average. The coralyne angles of transWH-c5-45d did not deviate significantly, but the values are rather scattered. The coralyne angles in the transWS-c5-45d-f model shifted a little after the first removal of the restraints (after 3 ns). However, they were well maintained after that, except for coralyne 5. Coralyne angles were scattered more in the transWW-c5-0d and transWW-c5-135d-f models after the complete removal of the base pair restraints (after 6 ns). The coralyne angle of coralyne 4 in the transWW-c5-315d-f model deviated significantly initially, but it merged into the average after 10 ns. Generally, coralyne molecules were aligned well within the same complex. Some deviations of the coralyne angles were observed especially for the coralyne molecules placed near the terminal of the duplex (coralyne 1 and 5 in Figure S5). Their deviations imply that the terminal coralynes are less stable in the structure than the others.

In Figure S6-7, coralyne angles from extended periods are displayed. The coralyne angles were confined within a narrow range for both the transWH-c5-180d and transWW-c5-315d-f models over the extended period in *ff99* (Figure S6). The coralyne angles of transWH-c5-270d-f did not fluctuate greatly, but do deviate with respect to each other. TransWH-c5-45d and transWW-c5-135d-f suggest instability in the coralyne angles due to relatively large positional variation. Similar patterns were seen with *ff99-bsc0* (Figure S7). The coralyne angles of transWH-c5-180d and transWW-315d-f populated a narrow range, although some minor excursions were occasionally

observed. Unlike with *ff99*, transWH-c5-270d-f seemed less stable judging by the deviation in coralyne angle observed with *ff99-bsc0*. Overall, the stability of the coralyne angle follows the order: transWH-c5-180d, transWW-315d-f > transWH-c5-270d-f > transWH-c5-45d, transWW-c5-135d-f. This order corresponds to the order of the free energy observed with *ff99-bsc0*.




transWH-c5-180d
transWH-c5-270d-f
transWH-c5-45d
transWS-c5-45d-f
transWW-c5-0d
0-1 ns
1-2 ns
2-3 ns
3-4 ns
4-5 ns
5-6 ns
6-7 ns
7-8 ns
8-9 ns
9-10 ns
10-11 ns
11-12 ns

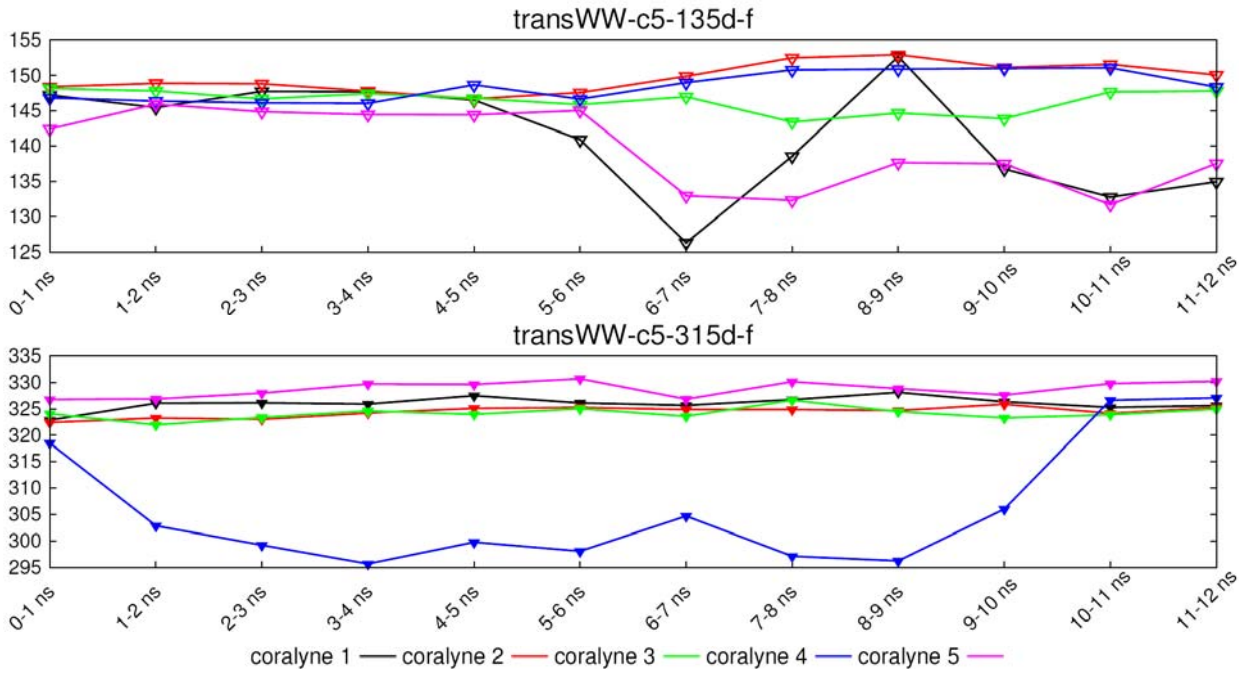


Figure S5: The change of coralyne angles of the 10 complexes during the 12 ns MD simulations. Y-axis indicates the angle in degrees, each with different scales. Standard deviations at the data points were 5.8 degrees on average. Coralyne *n* is the one surrounded by the adenine residues: *2n*, *25-2n*, *2n+1*, and *24-2n*.

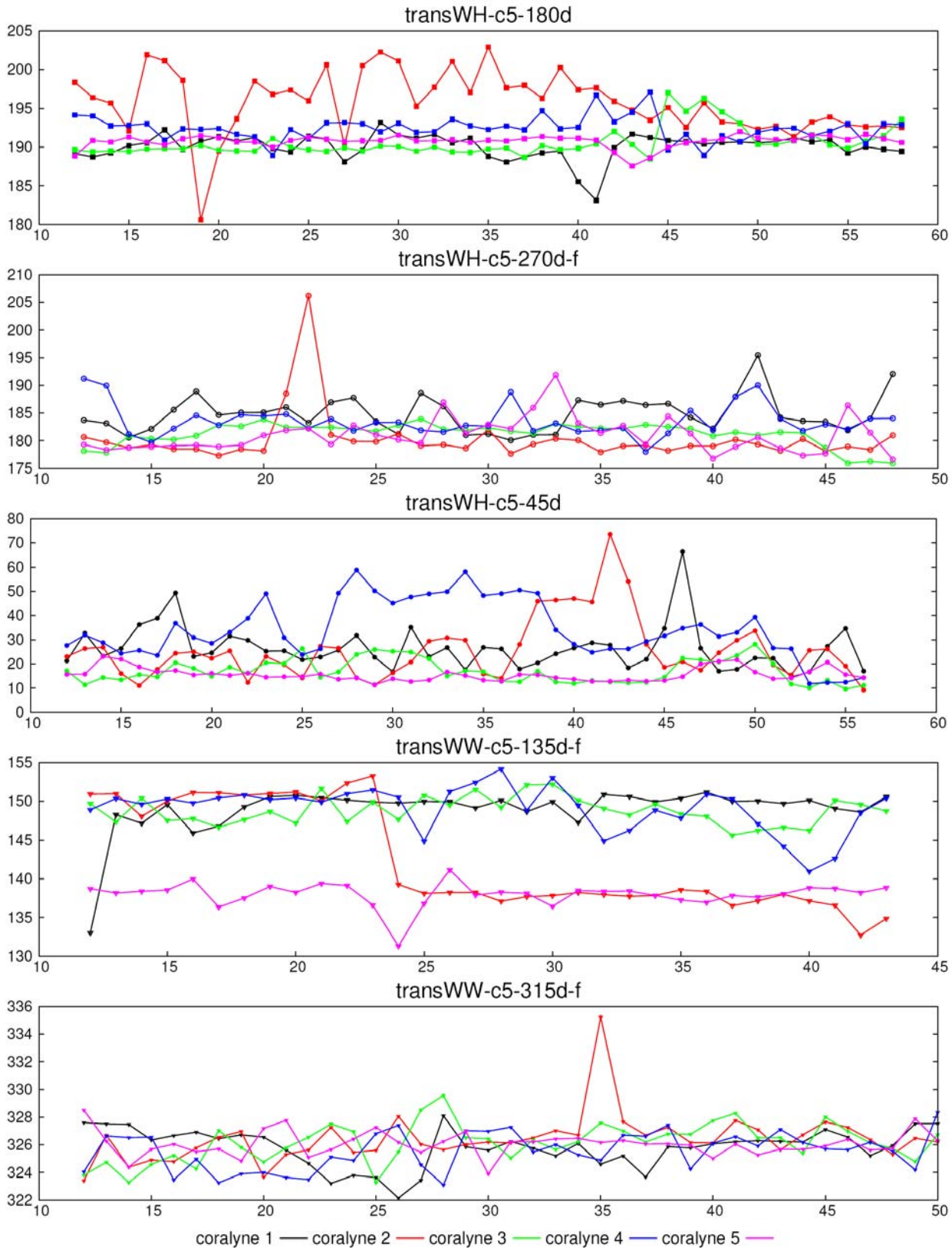


Figure S6: Coralyne angles over the course of extended MD simulations with the *ff99* force field. The y-axis is the angle in degrees and the x-axis is time in ns.

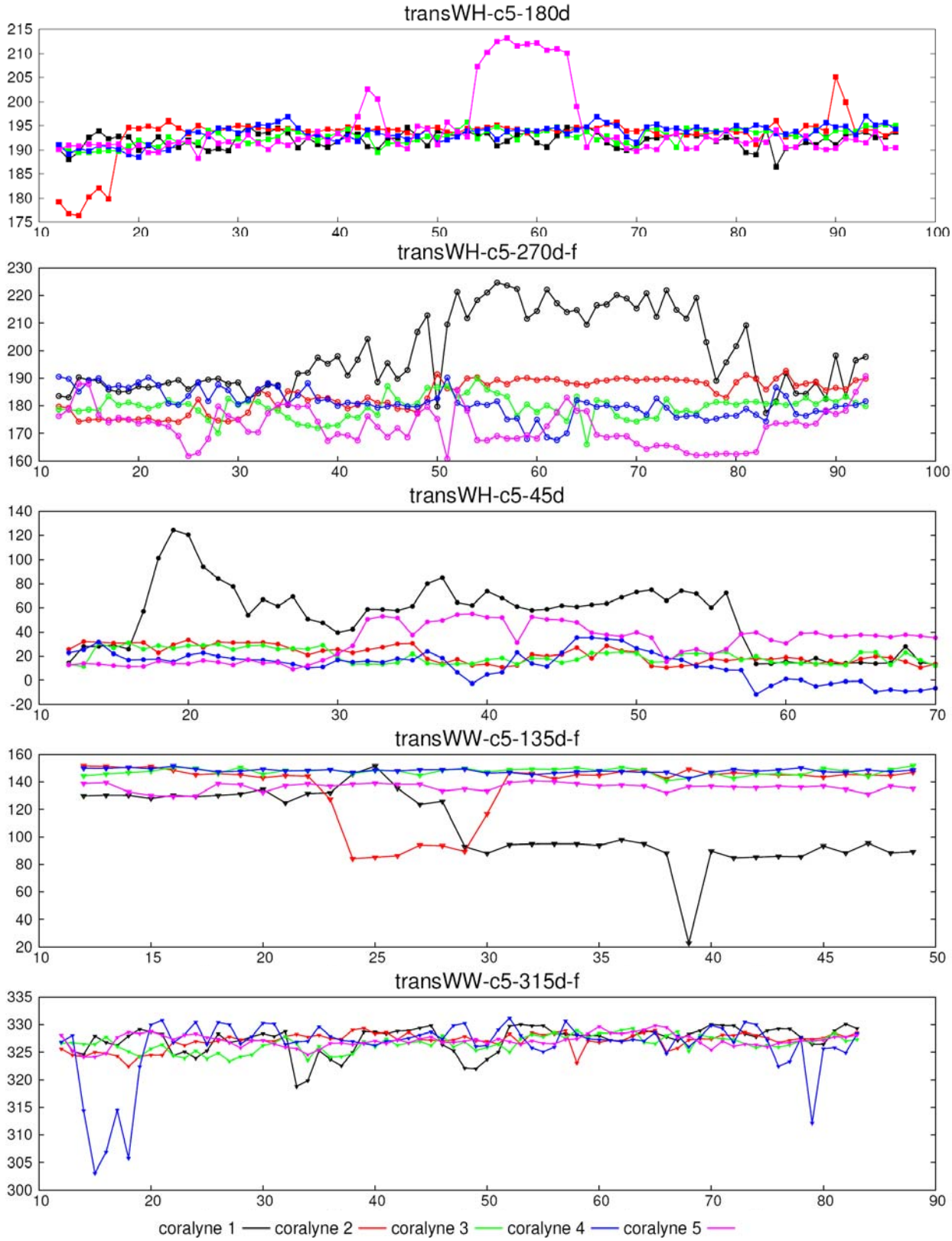


Figure S7: Coralyne angles from MD simulations with the *ff99-bsc0* force field modifications to *ff99*. The y-axis is the angle in degrees and the x-axis is time in ns.

In the simulations, the average occupancy of the hydrogen bonds tend to decrease as the simulation elapsed.

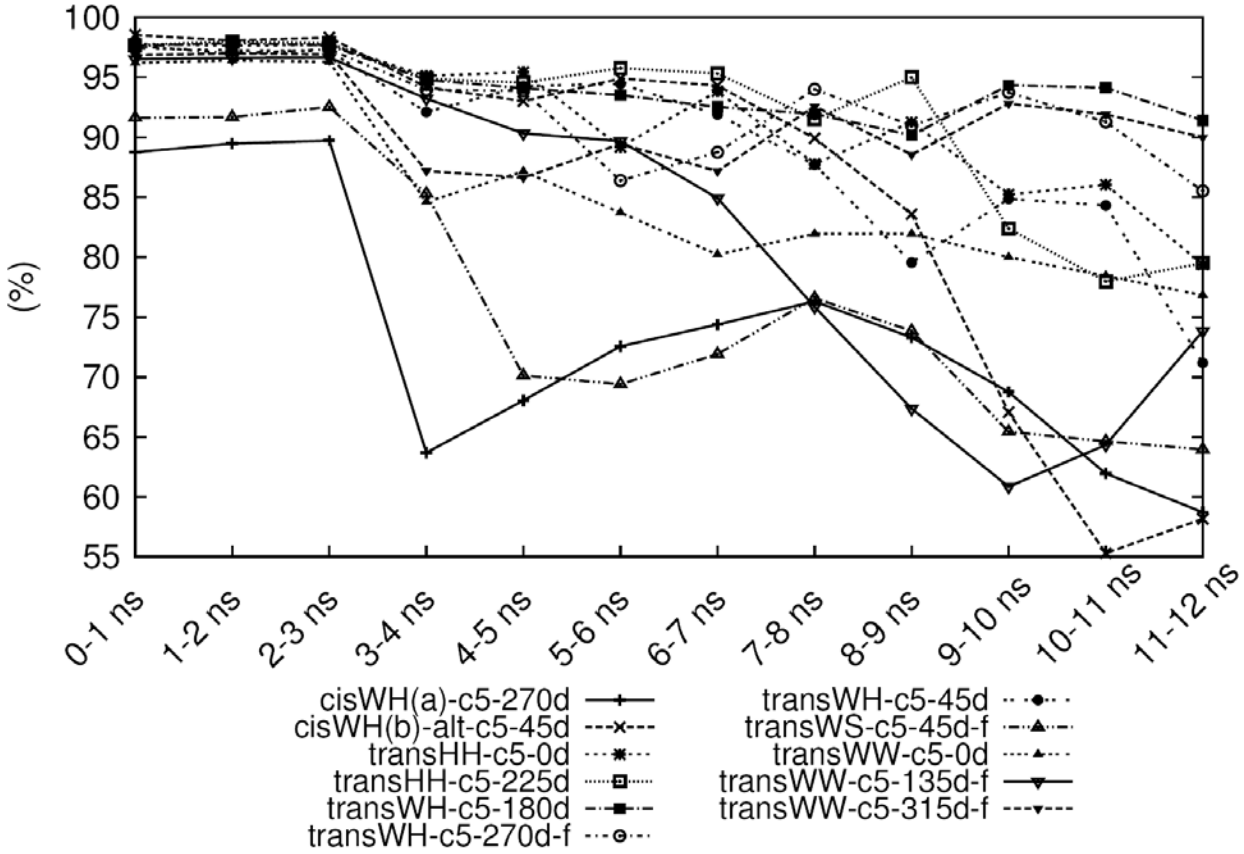


Figure S8: Average occupancy of the H-bonds in 5-coralyne-homo-(dA)-duplex's. The terminal base pairs (between residue 1 and 24 and between 12 and 13) were excluded in the calculation because they do not have neighboring coralynes intercalated.

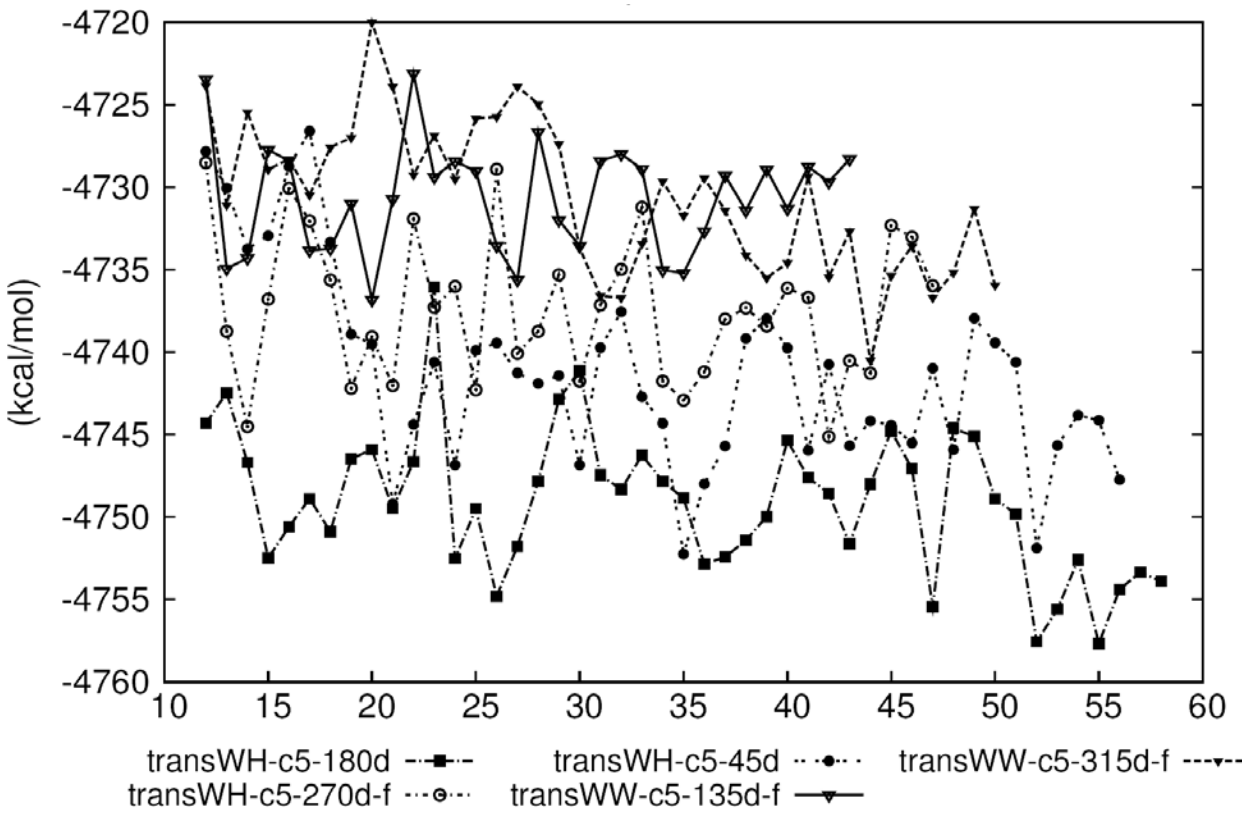


Figure S9: The free energies of the selected model structures during the extended MD period. The units of the x-axis are nanoseconds of MD simulation with the *ff99* force field.

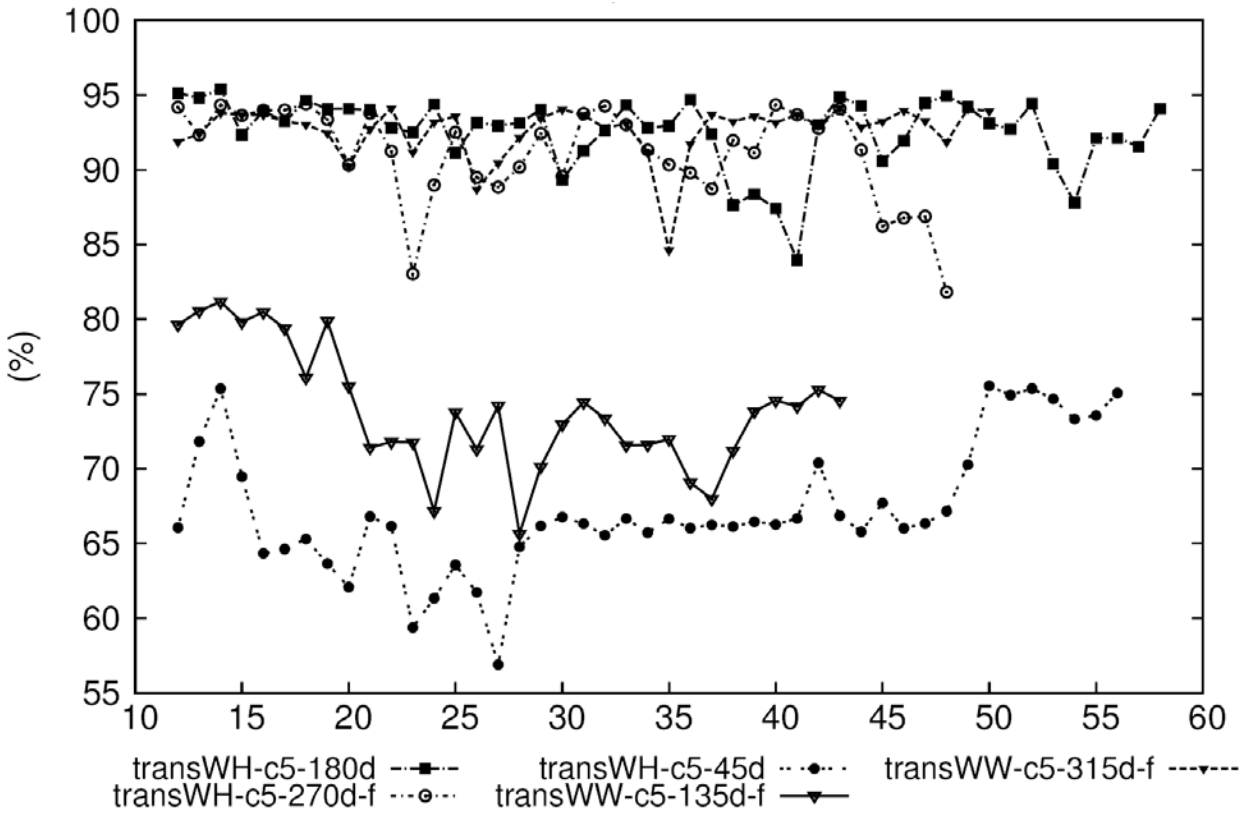


Figure S10: The average occupancies of the H-bonds of selected model structures during the extended MD period. The H-bond occupancy (%) is on the y-axis and the time in nanoseconds is on the x-axis. The data is with the *ff99* force field.

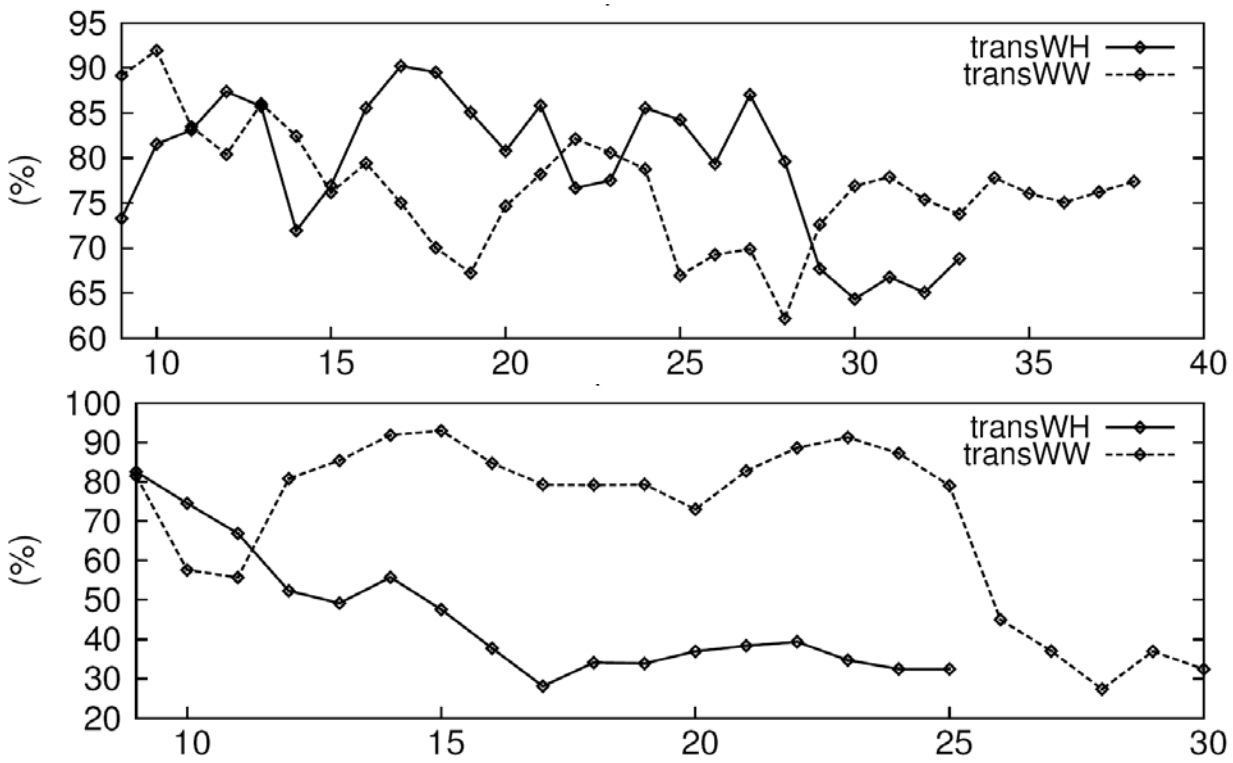


Figure S11: H-bond occupancies (%) versus time (ns) during control MD simulations of coralyne-free homo-(dA) duplex model structures with the two different force fields. The MD simulations of the two homo-(dA) duplexes (transWW and transWH) without coralyne molecules were carried out under the same conditions as the complexes. The occupancies fluctuated for both structures in simulations with both force fields. With the *ff99-bsc0* force field (bottom), the occupancies dropped around 30% within 25-30 ns. The duplexes in *ff99* (above) did not show such drastic decreases in the occupancy. With intercalated coralyne, the H-bond occupancy generally remained > 90%, while the occupancies of the coralyne-free duplexes fluctuated near 75-80%. Although we did not run longer simulations for the coralyne-free duplexes, we could observe local base pair breakage, collapse and base stacking changes in the duplex-only structures. These structural changes could account for the larger fluctuations and slow decreases in the occupancy as the local structures drifted away from 1:1 base pairing.

Estimation of absolute binding free energies

Although the putative model structures for the homo-(dA) and coralyne complexes have been successfully narrowed down, the absolute binding free energies of the structures are unknown. To calculate this requires an estimate of the free energies of single coralyne molecules and single homo-(dA) molecules. To estimate these values, the free energy of various numbers of coralyne molecules alone in TIP3P water (post MD

simulation) were estimated. Like most DNA-binding drugs, coralyne is also relatively hydrophobic and may be expected to aggregate to some extent in water. Therefore, the free energy of a single coralyne molecule in water is expected to be higher than the free energy of coralyne aggregates. This was tested. Coralyne aggregates with various numbers of coralyne molecules in TIP3P water were simulated for 400 ps and MM-PBSA contributions were calculated from the last 300 ps. Figure S12 shows how the free energy of coralyne molecules changes as the number of coralyne molecules increases in an aggregate. The MM, polar and non-polar contributions decreased as the size of the aggregate increased. The solute entropic contribution also decreased. However, these entropic contributions were significantly smaller than the other contributions. Overall, the total free energy per coralyne molecule decreased as the size of the aggregate increased. Therefore the free energy of coralyne molecules in water would be estimated to be more negative than the free energy of 7-coralyne aggregates (= 36.9 kcal/mol = 75.1 kcal/mol - 300 K $\times$ 127.1 cal/mol/K). The estimation of the free energy of single-stranded homo-(dA) is harder to estimate because of the larger number of degrees of freedom. A single 12-mer homo-(dA) of the duplex type I model structure was solvated in TIP3P water and simulated for 6 ns. The free energy was estimated from the initial 6 ns simulations by the MM-PBSA method (Table S6). The duplex type I was built using regular B-DNA parameters. Therefore the single homo-(dA) was also shaped like B-DNA and 6 ns is not sufficient for the molecule to reorganize itself and explore the most stable structure. Therefore the real free energy of the single-stranded homo-(dA) would be lower than the free energy estimated in Table S6. The absolute binding energy of the complexes with 5 coralyne molecules could be estimated as shown in Table S7.

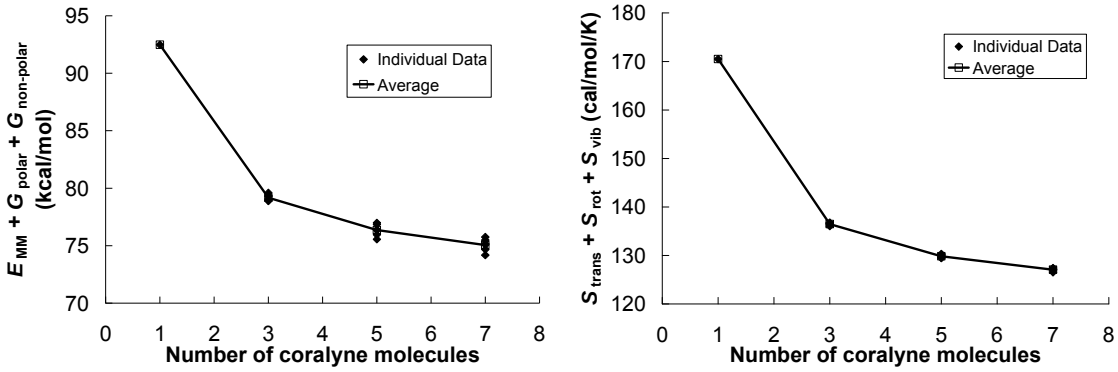


Figure S12: MM-PBSA contribution per coralyne molecule depending on the number of coralyne molecules in an aggregate. The left graph shows the MM, polar and non-polar contributions versus coralyne aggregation number and the right graph shows the solute entropic contributions. Except for the single coralyne case, measurements were from 6 separate simulations which had different initial coordinates.

Table S6: Estimated free energies at 300 K calculated by MM-PBSA method. The snapshots for the calculations were taken from the following period: the first 6 ns for the single-stranded homo-(dA), the last 10 ns for the complexes with 5 coralyne molecules, and the last 3 ns for the complexes with 7 coralyne molecules. Column A: the number of coralyne molecules in the complex, Column B: $E_{MM} + G_{polar} + G_{non\text{-}polar}$ (in kcal/mol), Column C: $S_{trans} + S_{rot} + S_{vib}$ (in cal/mol/K), Column D: col. B - (300 K) x col. C (in kcal/mol)

Force Field		*ff99*			*ff99-bsc0*		
	A	B	C	D	B	C	D
Single-strand		-2139.6	1085.8	**-2465.3**	-2144.3	1085.3	**-2469.8**
transWH-180d	5	-3973.1	2598.2	**-4752.5**	-3959.5	2605.9	**-4741.3**
	7	-3834.4	2830.8	**-4683.6**	-3824.9	2929.6	**-4703.7**
transWW-315d-f	5	-3946.7	2626.0	**-4734.5**	-3953.7	2605.4	**-4735.3**
	7	-3820.6	2854.2	**-4676.8**	-3802.9	2830.6	**-4652.0**

In order to measure the binding free energies of the complexes with 7 coralyne molecules, transWH-c7-180d and transWW-c7-315d-f were simulated. These complexes violate the neighbor exclusion principle and, therefore, they are expected to have positive binding free energy. The initial structures were built by adding two more coralyne molecules into the each complex with 5 coralyne molecules. The new coralyne molecules were inserted into the pockets next to the central intercalation pocket. After equilibration, they were simulated for 3 ns with the full base pair restraints. After that, the restraints were completely removed and another 3 ns simulations were carried out. The free energies were measured with the snapshots collected from the last 3 ns and the results are also displayed in Table S6. The absolute binding energy of the complexes with 7 coralyne molecules could also be estimated (Table S7).

If the neighbor exclusion principle is strictly obeyed, the average binding free energies per coralyne molecule of the 7-coralyne complexes should be slightly positive while those of the 5-coralyne complexes should be slightly negative. Most of the differences of the estimated average binding free energies between 5- and 7-coralyne molecule complexes were negative (Table S7) and this implies that the formation of 7-coralyne complexes is more favorable than that of 5-coralyne complexes. This is likely not correct in a practical sense because the binding free energy should gradually decrease as the number of ligand molecules increases (especially when the neighbor-exclusion principle is violated). Moreover, intercalation reactions likely are not cooperative. The experimentally estimated binding free energy at 300K, approximately -6.9 kcal/mol calculated from the binding constant (1.05×10^5 M^{-1}) is a few kcal/mol more negative than the average binding free energies in Table S7. The discrepancy is likely to be caused by the approximations made with the MM-PBSA treatment. Any shift of the free energies of the complexes or single homo-(dA) molecules will change the absolute values of the binding free energies, but such shifts cannot affect the differences of the average binding free energies between 5- and 7-coralyne molecule complexes in Table S7. However, the shift of the free energy of the coralyne molecules can affect the differences as well as the absolute binding free energies. If the free energy of the coralyne molecules becomes more favorable, for example, the differences in Table S7

become more positive. More aggregation of the coralyne molecules could also lower the free energy and this also partially offsets the observation that the binding energy of a 5-coralyne complex is less negative than that of a 7-coralyne complex.

Table S7: Estimated binding free energies at 300 K in kcal/mol. The 'Total' column shows the binding free energies of the various complexes based on the data in Table S6. The second column presents the number of coralyne molecules in each complex. The average binding free energies presents the binding free energies per coralyne molecule. The differences between the average binding free energies of 7-coralyne molecule complex and 5-coralyne molecule are also displayed.

Force Field	*n*	*ff99*		*ff99-bsc0*	
		Total	Average	Total	Average
transWH-180d	5	-6.4	-1.3	13.8	2.8
	7	-11.3	-1.6	-22.4	-3.2
Difference			-0.3		-6.0
transWW-315d-f	5	11.6	2.3	19.8	4.0
	7	-4.5	-0.6	29.3	4.2
Difference			-3.0		0.2

Overall, although approximate, the results in Table S7 are not completely out of bounds and suggest the greater relative stability of the transWH model structure. However, clearly the estimation of the absolute binding free energy was not very successful, and this is likely due to the approximations of the force field, approximations of the MM-PBSA method, and most likely the rather simplified representation of the free coralyne and free homo-(dA) duplex structural states.

Cluster analysis of the transWH and transWW model structures.

Cluster analyses based on RMSd for the transWH-c5-180d (12,021 ps – 96,354 ps) and transWW-c5-315d-f (16,286 ps – 83,590 ps) MD trajectories with the *ff99-bsc0* force field were performed. The cluster analysis was applied to two parts of the molecules, specifically the whole molecule including the homo-(dA) duplex and the 5 coralynes and also the minimal unit of coralyne intercalation which is composed of a single coralyne molecule and the four adenine bases surrounding the coralyne molecule. In the latter case, five units could be obtained per snapshot. For the both cases snapshots were taken every 20 ps from the MD trajectories. Clustering using **`ptraj`** from AmberTools 1.0 isolated 10 clusters with the K-means algorithm for each case(31). For the whole molecule of transWH-c5-180d, cluster 1 was the most populated cluster (Table S8). The RMS deviations between any pair of representative structures from cluster 1, 3, 8, and 9 were less than 3 Å as shown in Figure S13A. Thus, their centroids are close each other. If all of their populations are summed up, the total population becomes 52.4%. Although cluster 5 is slightly further away from cluster 1 (3.08 Å) and 3 (3.47 Å), it was within 3 Å far from cluster 8 and 9. If the population of cluster 5 is also added, the total population reached up to 68.5%. For the binding unit of transWH-c5-180d, cluster

6 was most populated. Figure S13B shows that the representative structures of cluster 3 and 6 were within 1 Å RMSd. Their total population was 42.5%. It is noticeable that cluster 6 was also very close to cluster 7 (0.98 Å) and 9 (1.04 Å). The whole molecule of transWW-c5-315d-f had major population at cluster 0, 4, 5, 7, and 9. All their representative structures are within 3 Å far each other and the total population of them was 49.4%. Cluster 3 was also within 3 Å far to cluster 0, 4, 5, and 9 although it was a little farther from cluster 7 (3.11 Å). The binding unit of transWW-c5-315d-f was dominated by cluster 7. The cluster was also close to cluster 0 and 2 within 1 Å and their total population was 56.9%. The representative structures of the most dominant clusters are illustrated in Figure 11 in the main text. In conclusion, all the cluster analyses revealed that there is a dominant population occupying about 50% of the simulation time.

The average coralyne angle of the binding unit of transWH-c5-180d in cluster 6 was 193.9 (±3.8) and that of the binding unit of transWW-c5-315d-f in cluster 7 was 327.3(±3.9). It suggests that the coralyne angles are highly preserved in the dominant clusters.

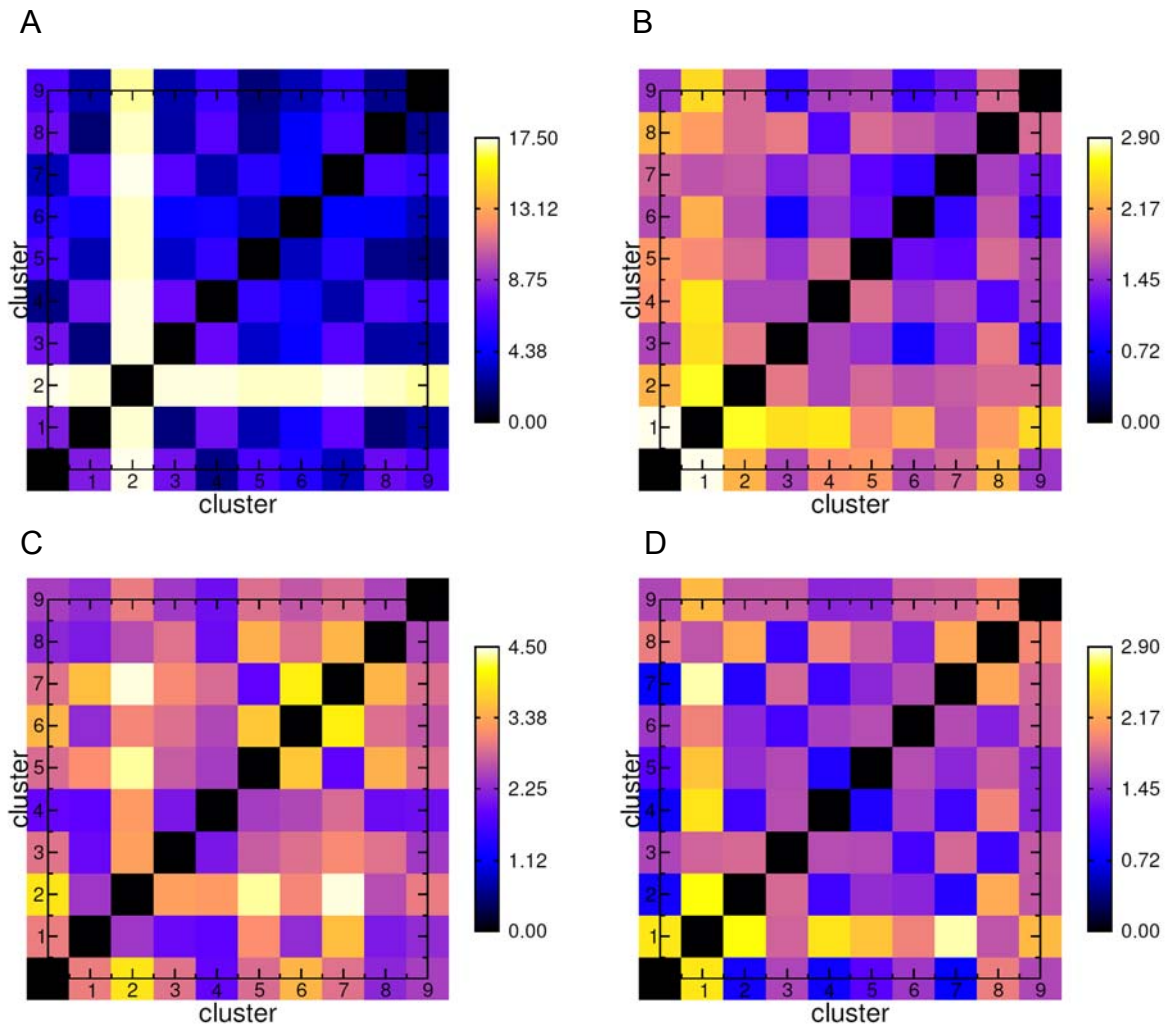
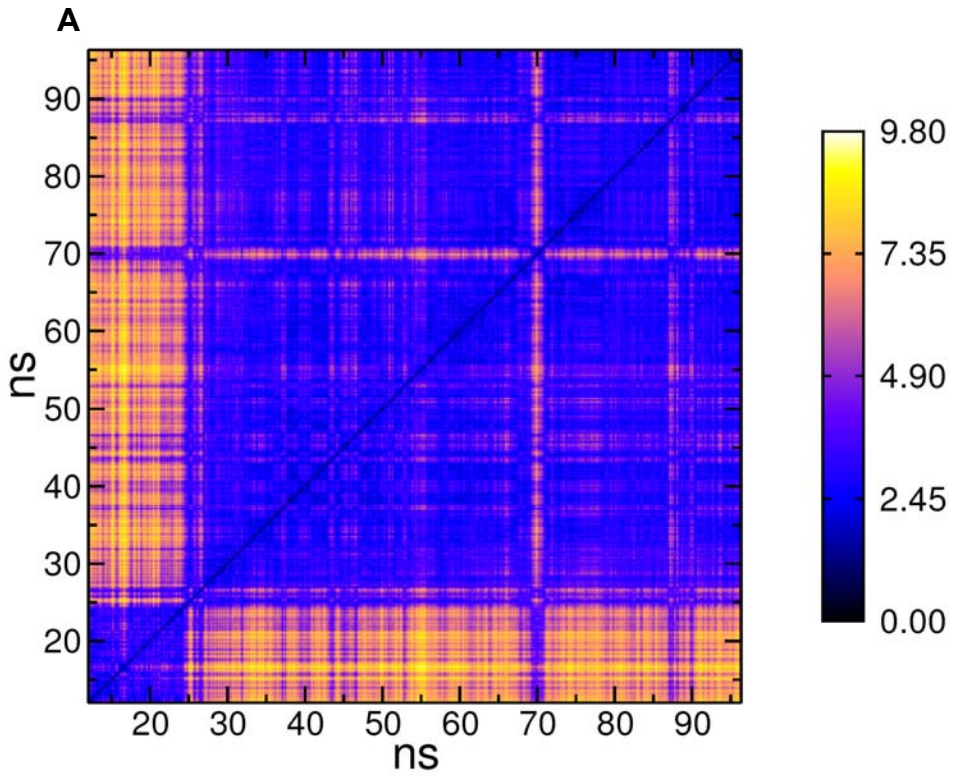
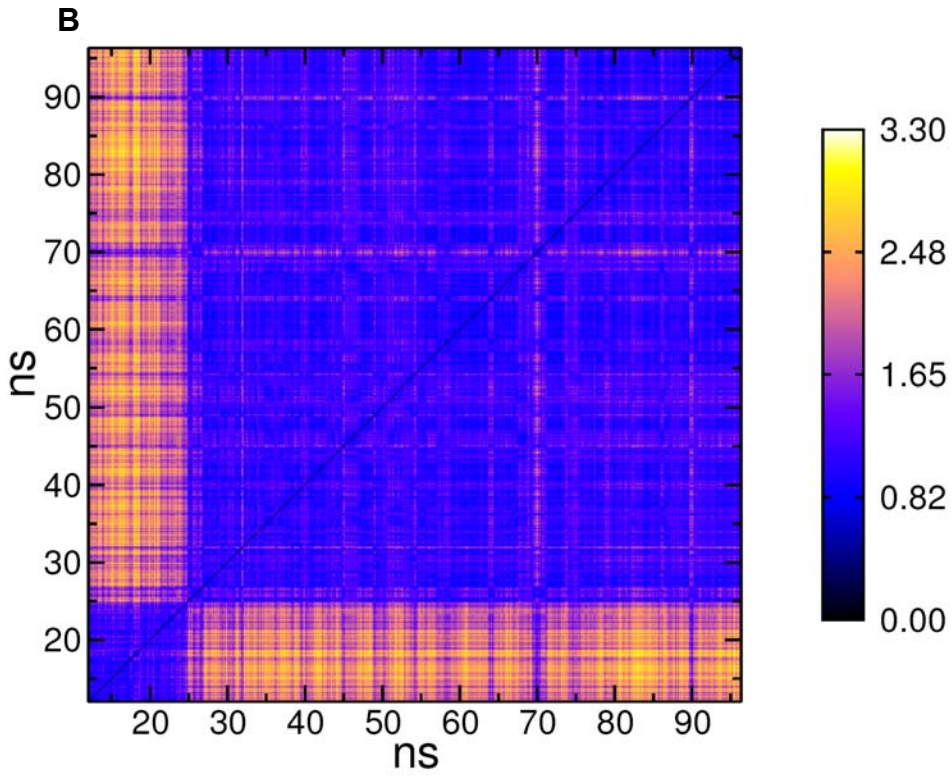


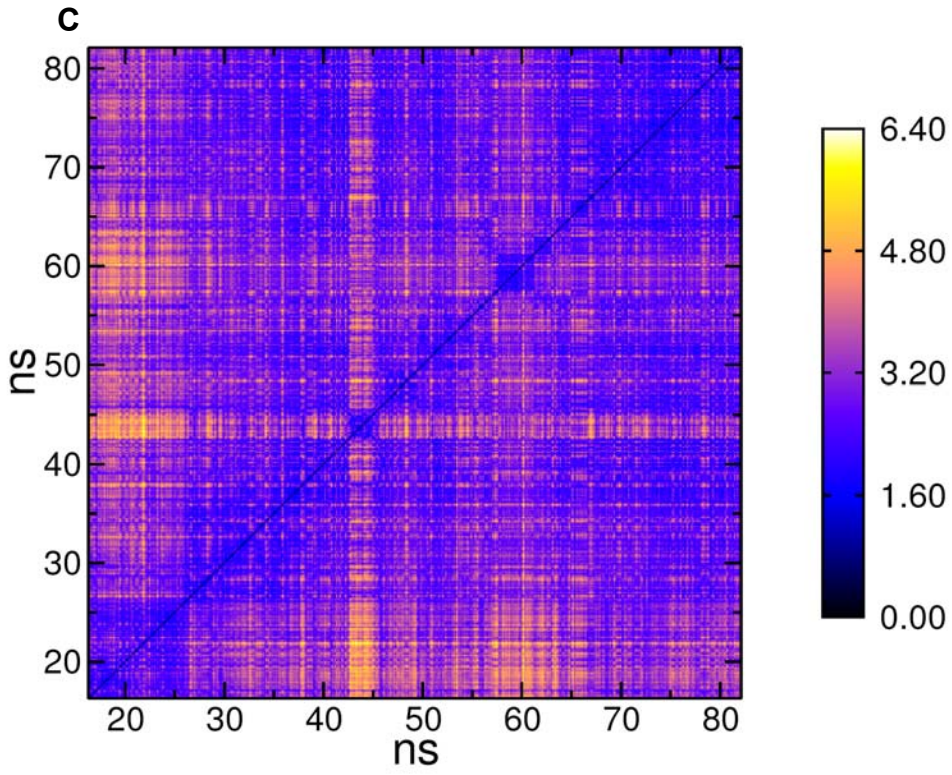
Figure S13: 2D-RMSd plots among the representative structures from individual clusters. Cluster numbers corresponds to the numbers in Table S7 and the unit of RMS deviation is Å. (A: whole molecule of transWH-c5-180d, B: binding unit of transWH-c5-180d, C: whole molecule of transWW-c5-315d-f, D: binding unit of transWW-c5-315d-f)

Table S8: Cluster populations in percentage. Populations indicated in bold face shows neighboring clusters within 3 Å (whole molecule) or 1 Å (binding unit) in RMSd. The total populations summed up from the dominant clusters are listed in the last row.

	transWH-c5-180d		transWW-c5-315d-f	
Cluster Number	Whole Molecule	Binding Unit	Whole Molecule	Binding Unit
0	8.9	1.9	**6.0**	**20.5**
1	**17.2**	14.0	13.7	1.0
2	2.2	0.4	4.2	**11.8**
3	**13.1**	**14.0**	14.0	9.8
4	7.8	1.1	**18.3**	14.7
5	16.1	0.6	**6.4**	9.1
6	3.0	**28.5**	5.7	2.2
7	9.7	22.7	**7.2**	**24.6**
8	**11.1**	1.1	13.0	6.2
9	**11.0**	15.6	**11.5**	0.0
Dominant	52.4	42.5	49.4	56.9




B
90
80
70
60
50
40
30
20
ns
20 30 40 50 60 70 80 90
ns
3.30
2.48
1.65
0.82
0.00


C
80
70
60
50
40
30
20
ns
20 30 40 50 60 70 80
ns
6.40
4.80
3.20
1.60
0.00

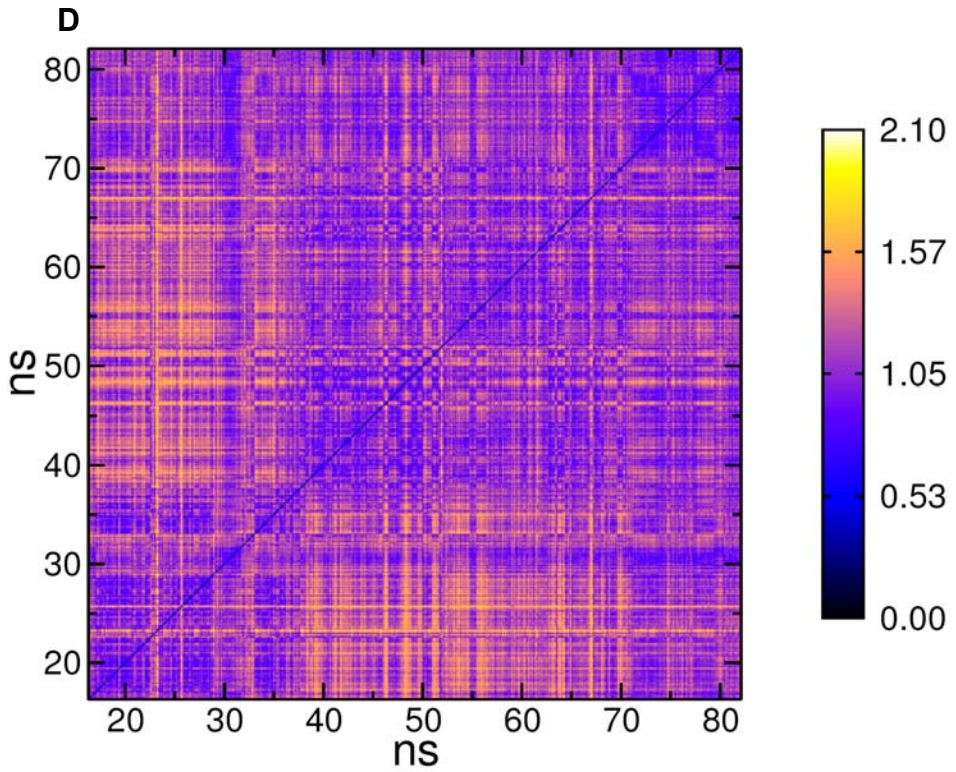


Figure S14: 2D-RMSd plots of the extended MD simulations of the transWH and transWW model structures of the coralyne-induced homo-d(A) duplex. 2D-RMSd was measured from the trajectories of simulations using *ff99-bsc0* force field. The xy-axes show simulation times on the nanosecond scale and the color codes denote RMSd in Å. A: whole molecule of transWH-c5-180d, B: the central binding unit of transWH-c5-180d, C: whole molecule of transWW-c5-315d-f, D: the central binding unit of transWW-c5-315d-f.

The 2D RMSd plots show that a dominant set of similar conformations is found in extended MD simulation. In the case of the transWH-c5-180d trajectories, a distinct conformational switch occurs at ~25 ns. This corresponds to a greater twisting of the structure relative to the initial geometry and is likely related to the α/γ substate problem as the initial coordinates were taken from the simulations with *ff99*. See the forthcoming discussion on the twisting of the structure for more information.

Table S9: Geometries of the H-bonds. The first two columns indicate residue numbers (n = 2, 4, …, 10) for the involved in the H-bonds. Residue numbers and atom names from the columns of distance and angle are in the same order in a row. The geometry of the H-bonds in the binding unit of transWH-c5-180d and transWW-c5-315d-f were measured from their most dominant clusters above (Table S8). The results showed that there is no statistical difference between neighboring H-bonds in the aspects of distance and angle.

transWH-c5-180d									
Res	Res	Distance (N1-N6)		Angle (N1-H-N6)		Distance (N6-N7)		Angle (N6-H-N7)	
n	25-*n*	3.01	± 0.17	160.0	± 8.9	3.06	± 0.20	163.5	± 8.4
n+1	24-*n*	3.02	± 0.27	157.9	± 10.1	3.07	± 0.27	159.5	± 11.1
transWW-c5-315d-f									
Res	Res	Distance (N6-N1)		Angle (N6-H-N1)		Distance (N1-N6)		Angle (N1-H-N6)	
n	25-*n*	3.05	± 0.28	161.6	± 9.7	3.10	± 0.24	161.8	± 11.6
n+1	24-*n*	2.99	± 0.14	164.0	± 8.5	3.05	± 0.17	165.5	± 7.6

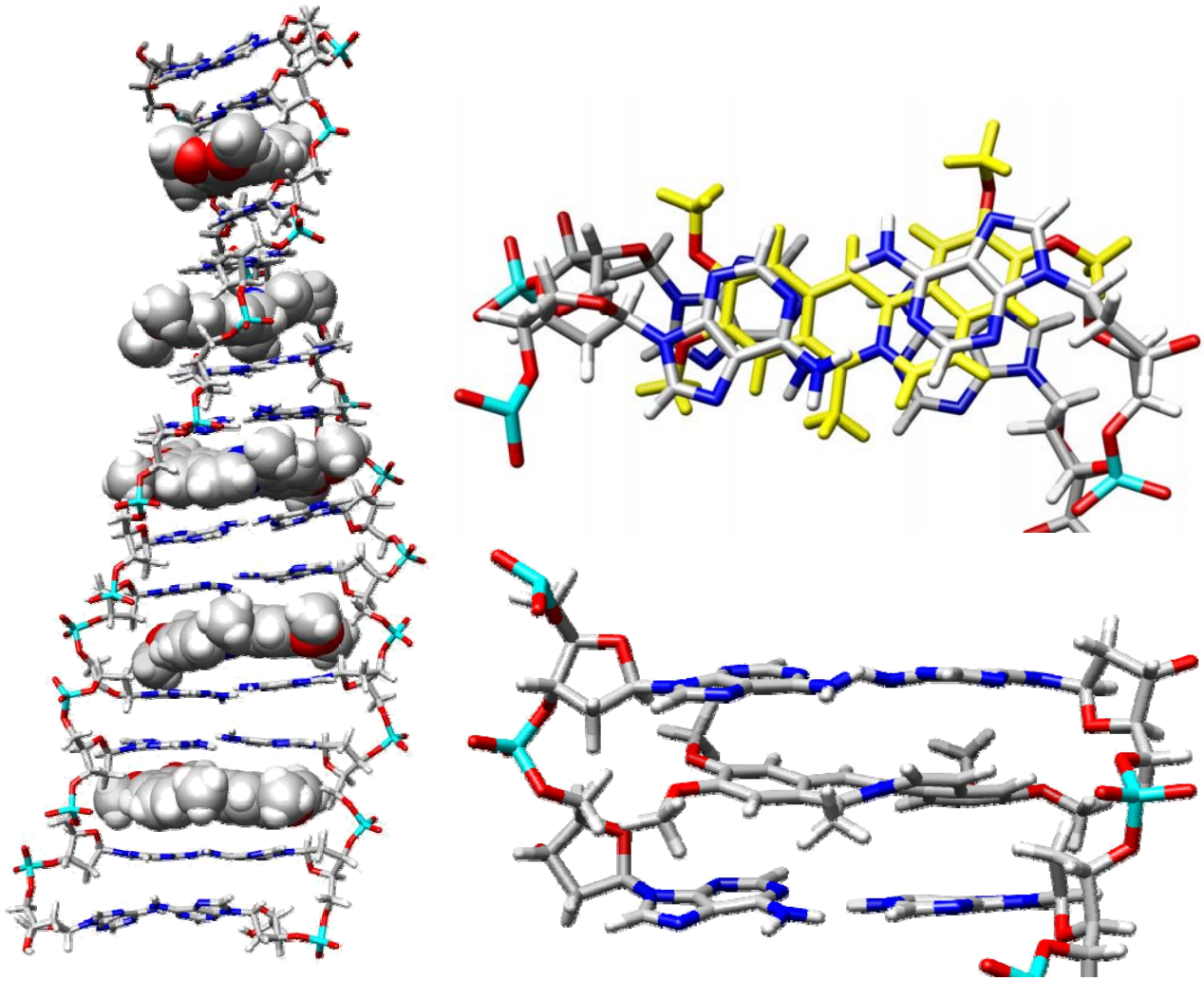


Figure S15: Graphical illustrations of representative structures of the transWW geometry from cluster analysis. All of the molecular graphics represent representative configurations from dominant clusters sampled over the course of MD simulation on the transWW-c5-135d-f model structure. The clustering was performed over the MD trajectory from ~12-96 ns with **`ptraj`** either over the entire molecule or over the 4 base-coralyne binding unit as described previously. A: The entire coralyne-induced homo-(dA) duplex with the DNA shown as a stick figure and the coralyne as solid spheres. Note that the structure is rather extended and under-twisted as compared to canonical B-DNA. B. View of the stacking with coralyne (yellow) for the minimal binding unit. C. A close up view of the minimal binding unit. The kink in the coralyne structure at the nitrogen imparted by the nearby methyl leads to a distinct and slight buckle in the W-W base pairs.

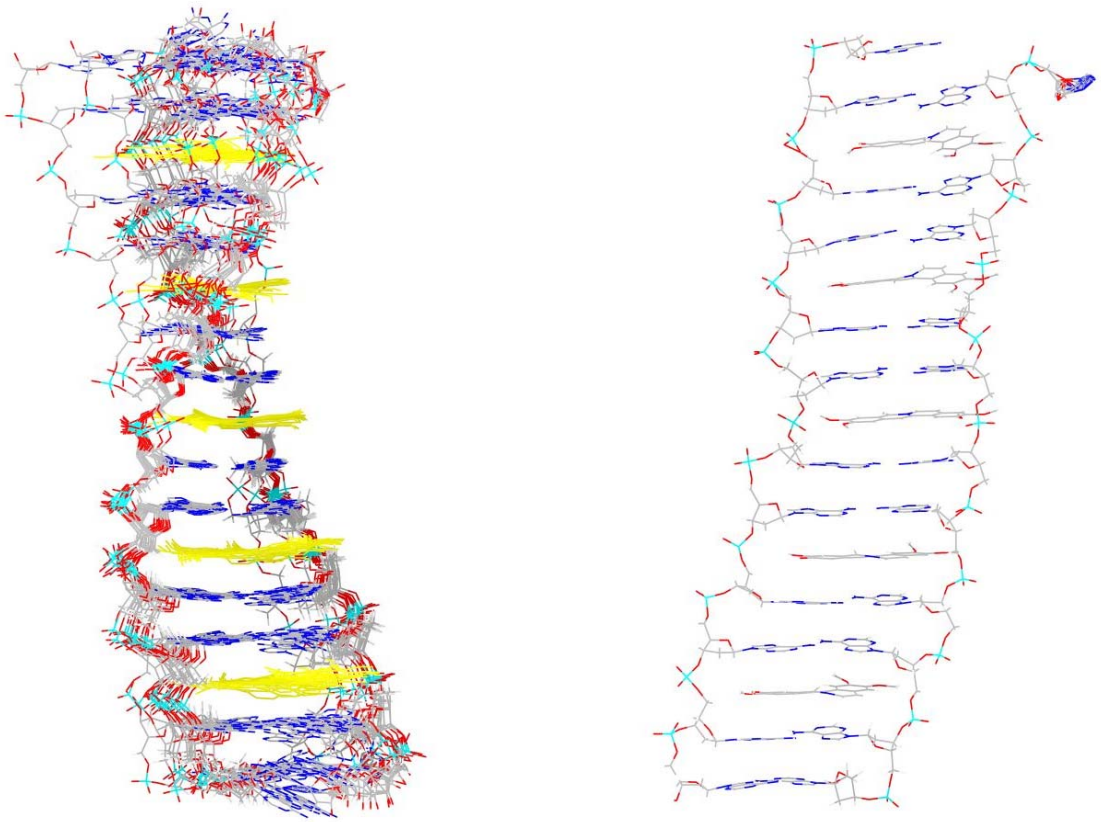


Figure S16: Twisting and structural variation in the transWH model over the course of MD simulation. Structures were taken at 5 ns intervals after straight coordinate average based smoothing over 100 ps windows for the transWH-c5-180d trajectory with *ff99-bsc0*. The structures are aligned to the central two adenine base pairs. Significantly greater twisting / opening is observed for the Watson-Crick strand, under-twisting to such an extent that the structure becomes essentially linear and ladder-like at some points during the MD simulation (shown in the figure on the right).

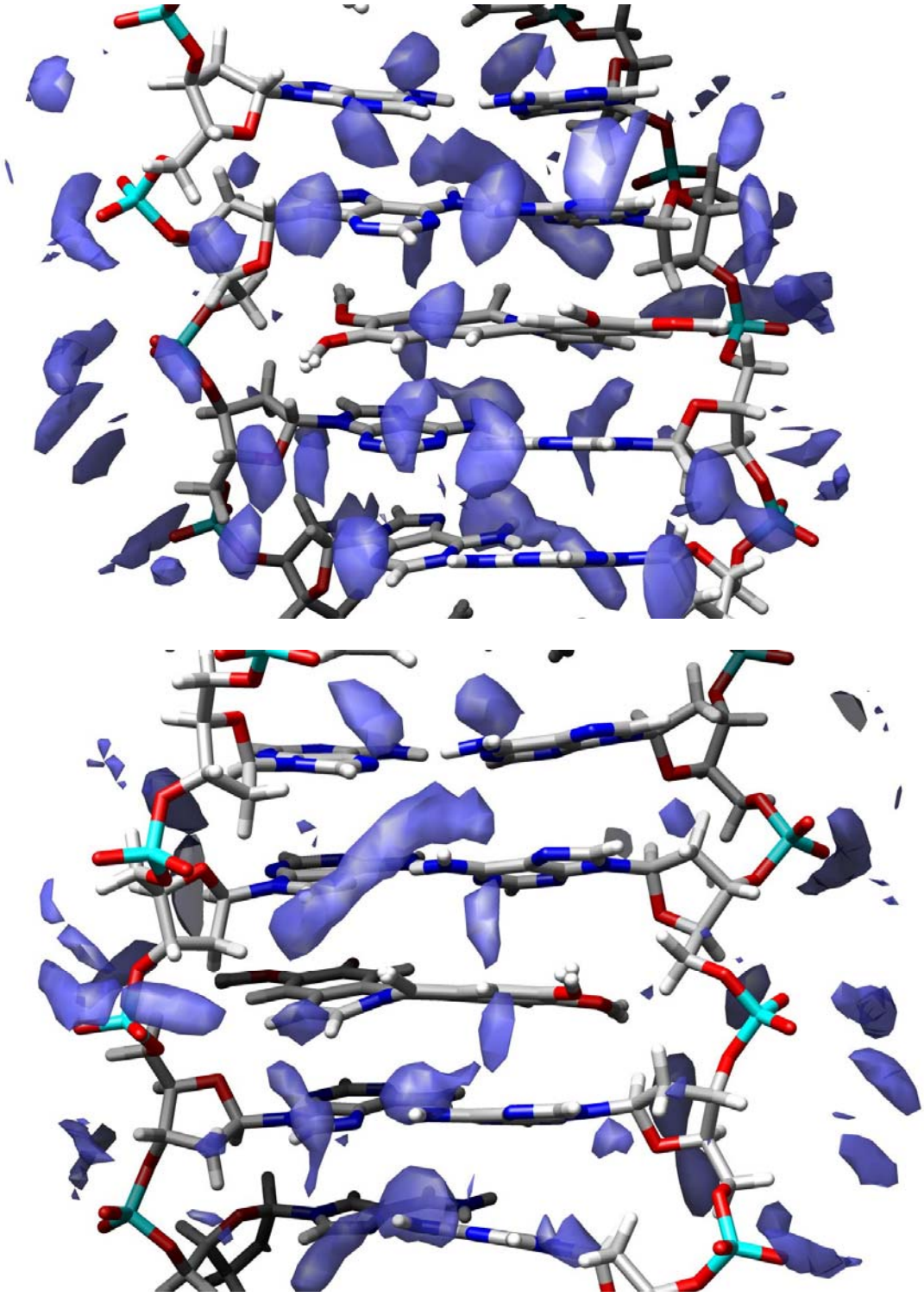


Figure S17: Hydration sites around the central coralyne binding site over the transWH-c5-180d MD trajectory with the *ff99-bsc0* force field. The top figure shows a view into the minor groove and the bottom figure shows a view into the major groove. The structure shown is an average structure from 75-76 ns with hydration density from the entire trajectory contoured at 3x bulk hydration.

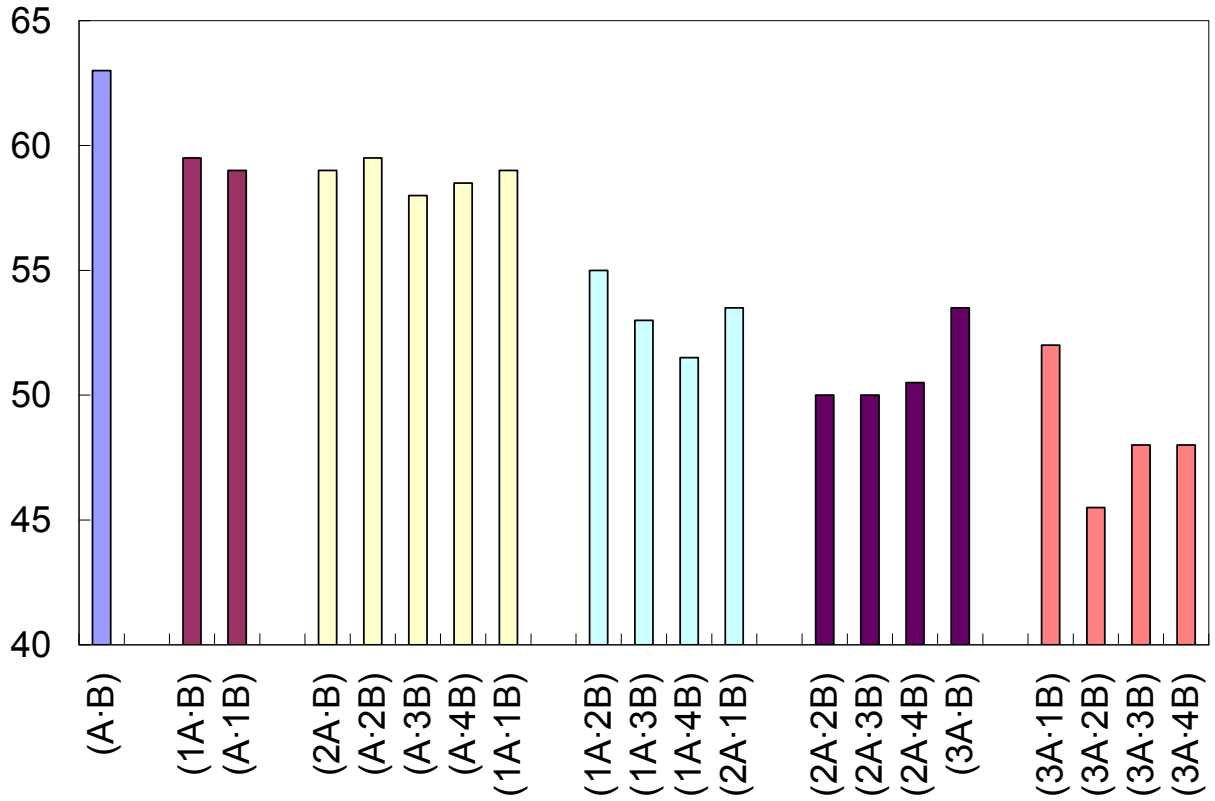
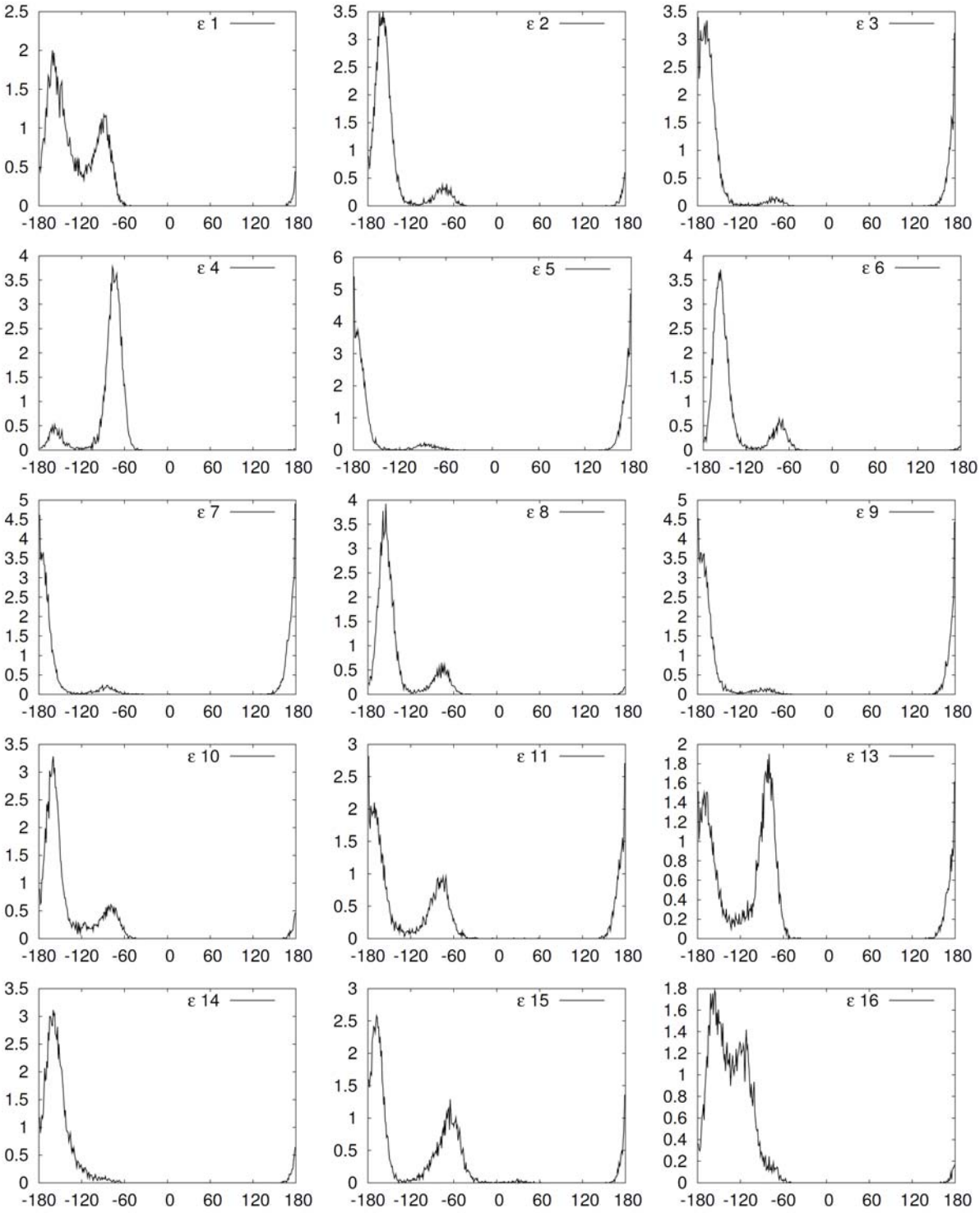
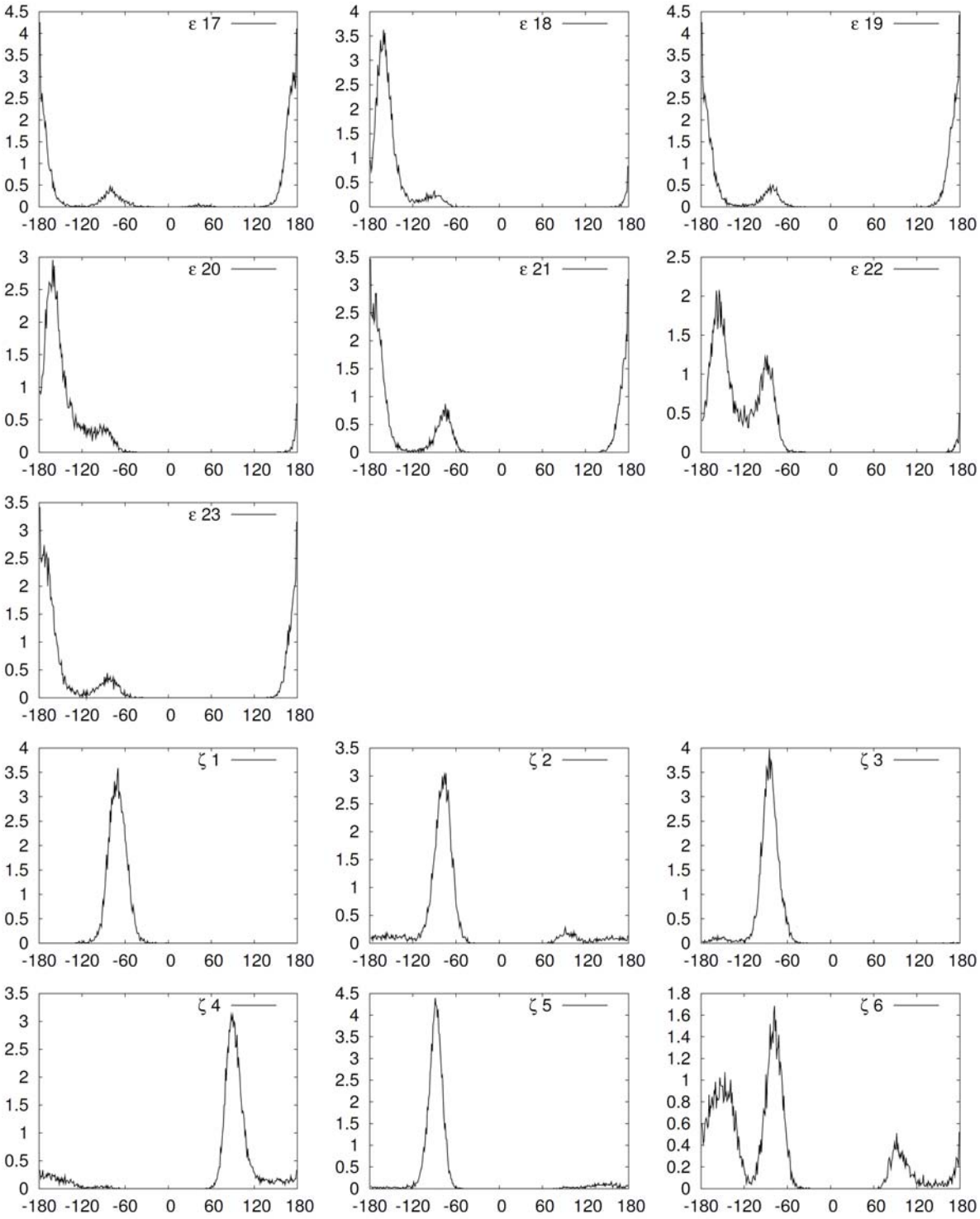
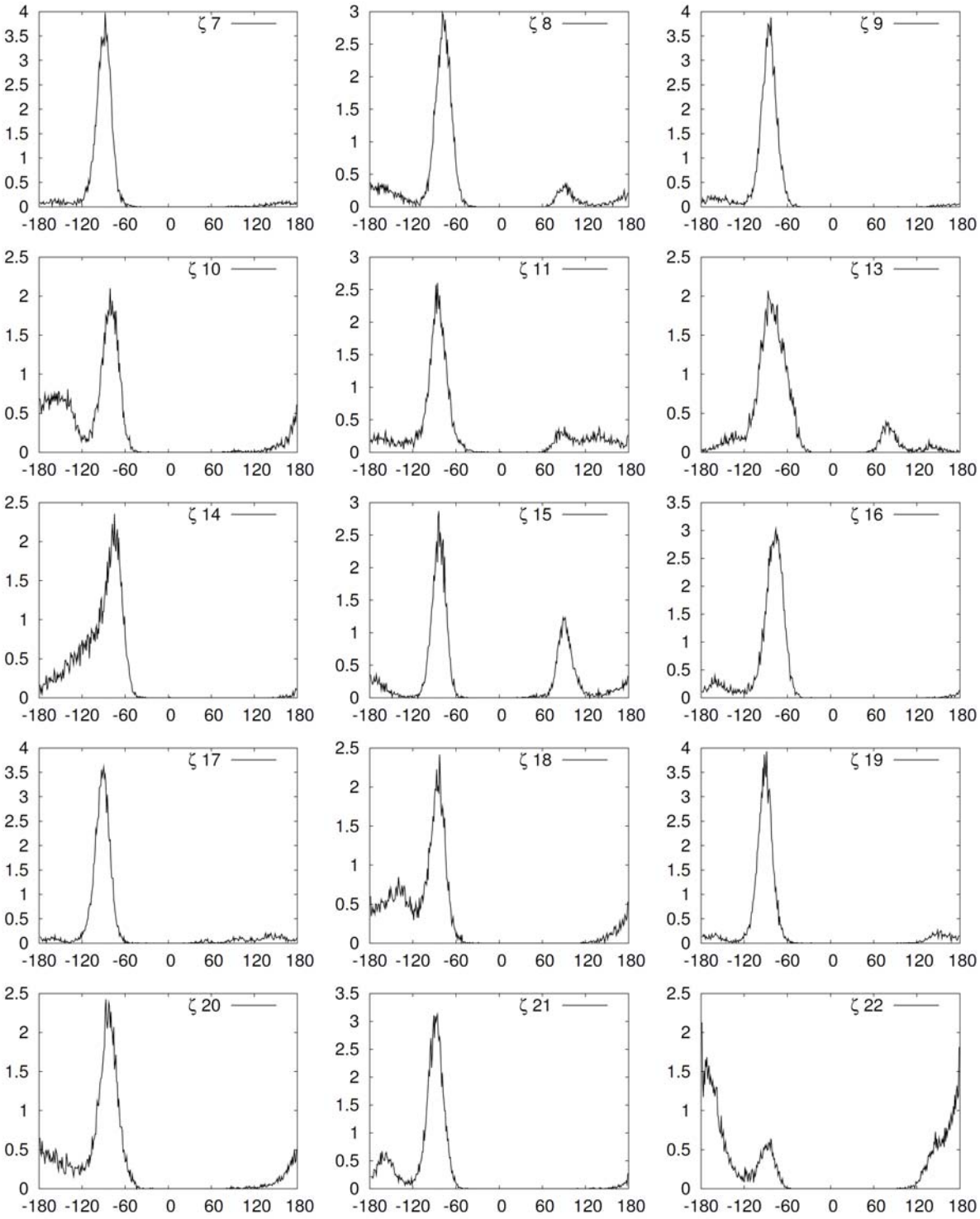
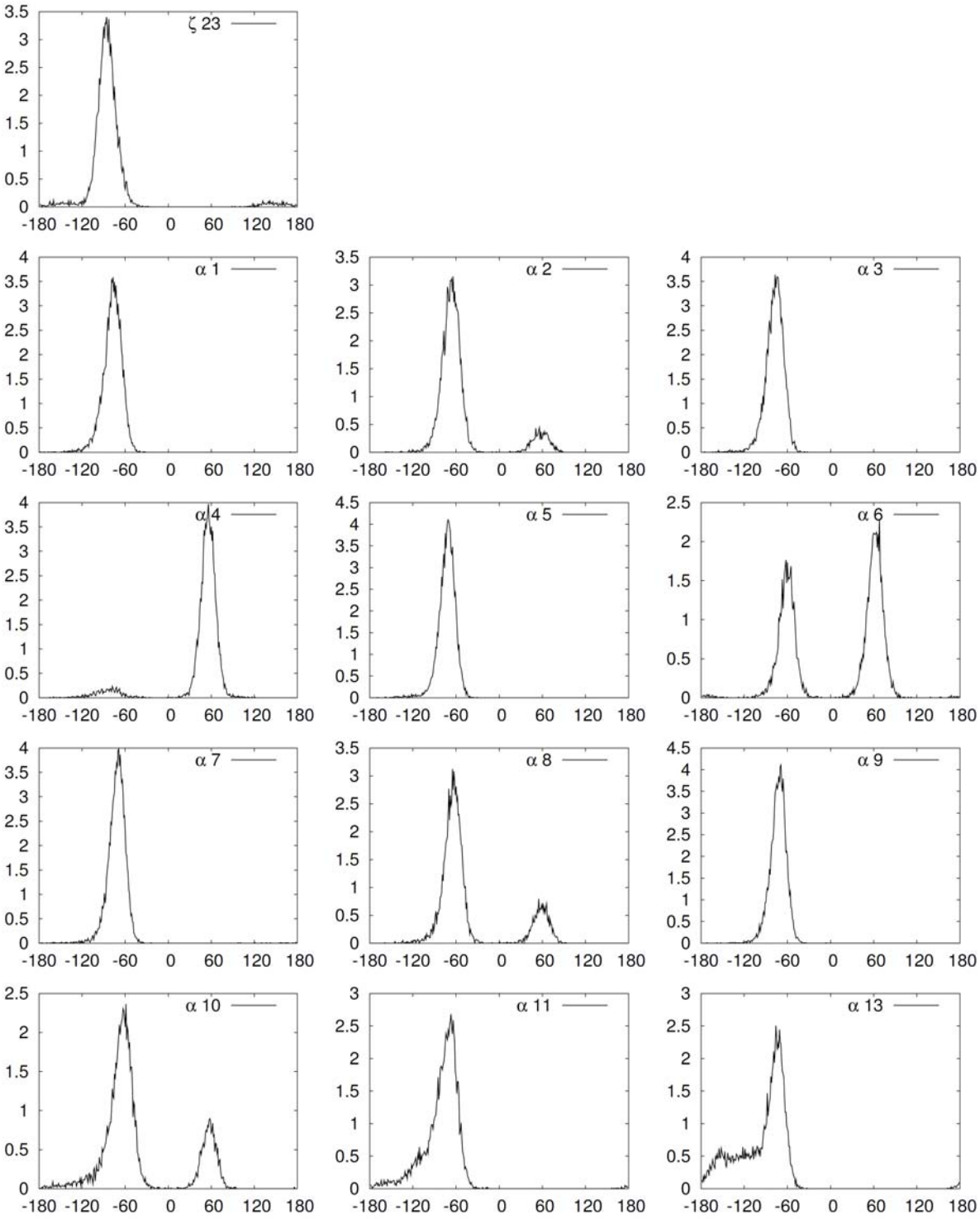


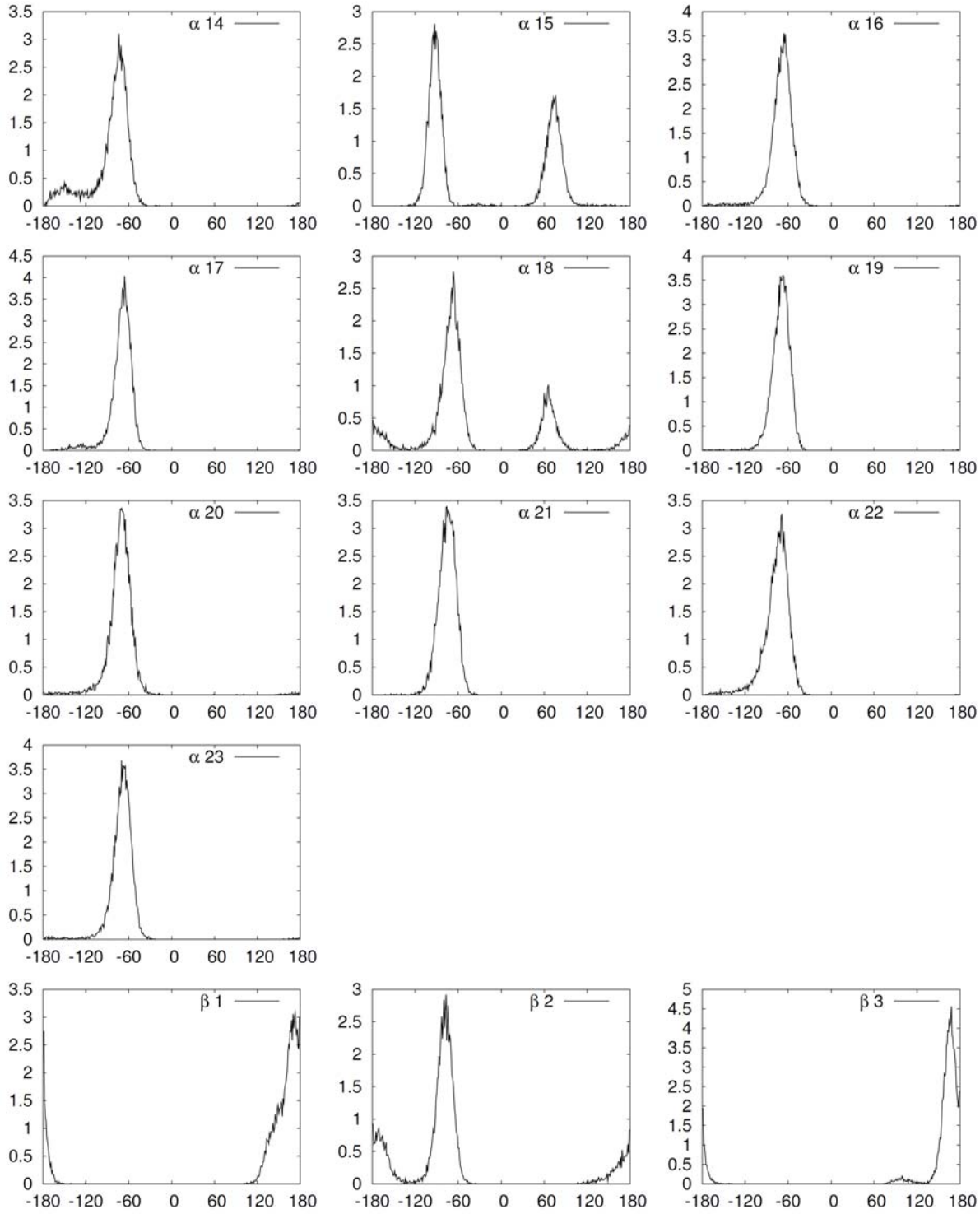
Figure S18: Melting temperatures of DNA duplexes in the presence of coralyne. Numbers in Table 3 is graphically visualized to improve readability. Y-axis is in °C.

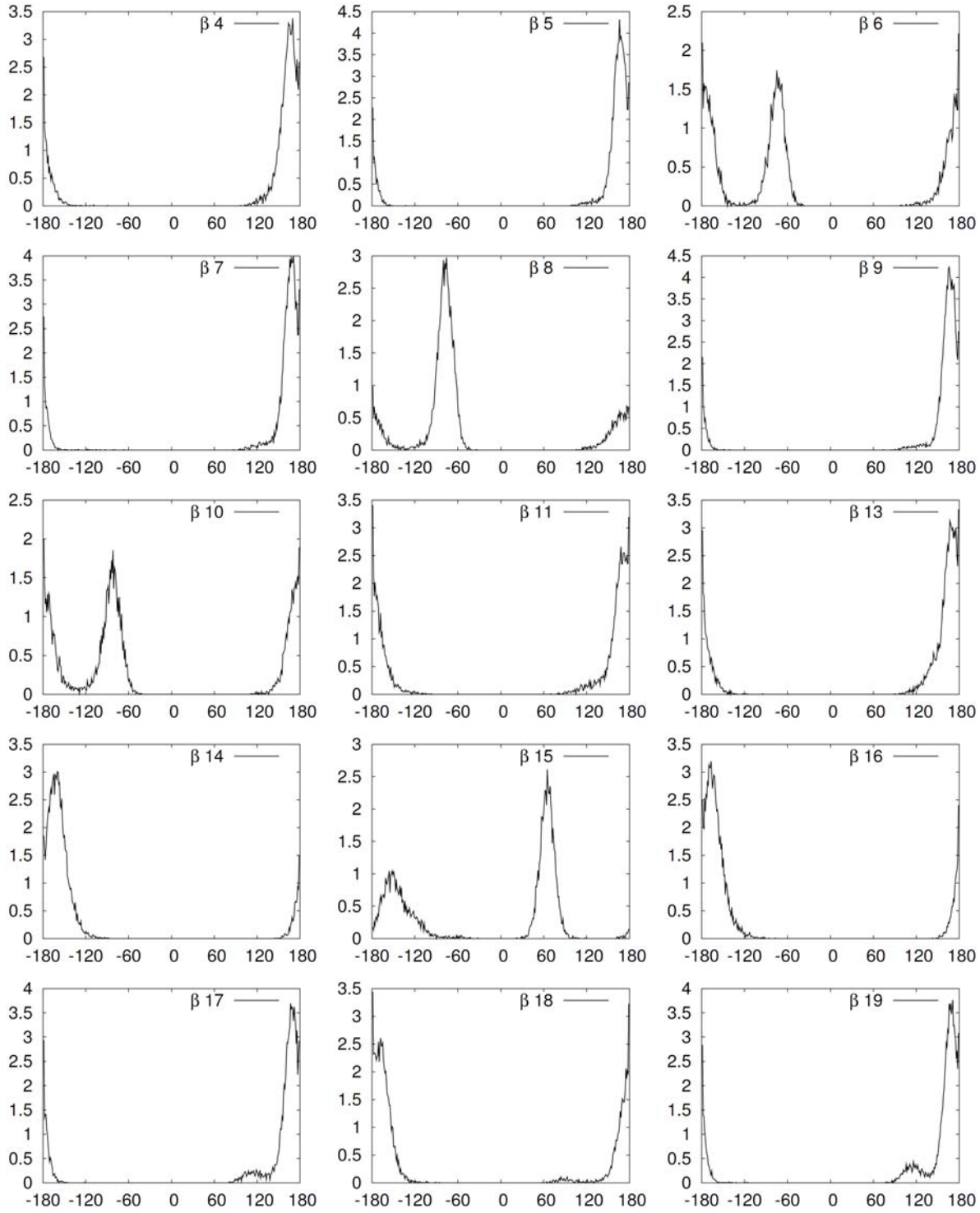

ε 1
ε 2
ε 3
ε 4
ε 5
ε 6
ε 7
ε 8
ε 9
ε 10
ε 11
ε 13
ε 14
ε 15
ε 16
-180 -120 -60 0 60 120 180

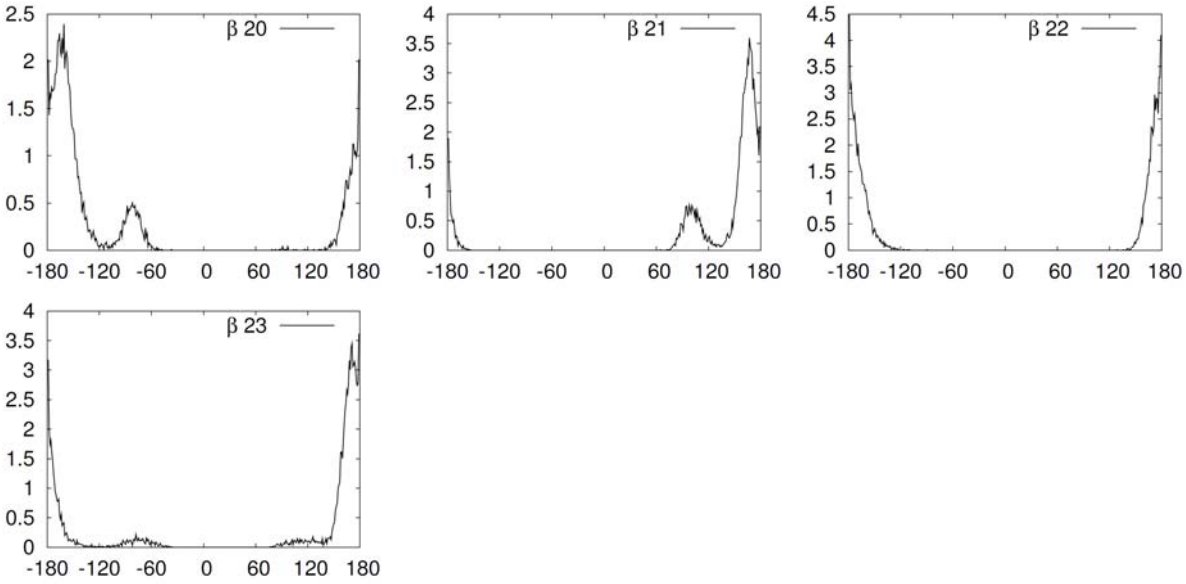

ε 17
ε 18
ε 19
ε 20
ε 21
ε 22
ε 23
ζ 1
ζ 2
ζ 3
ζ 4
ζ 5
ζ 6

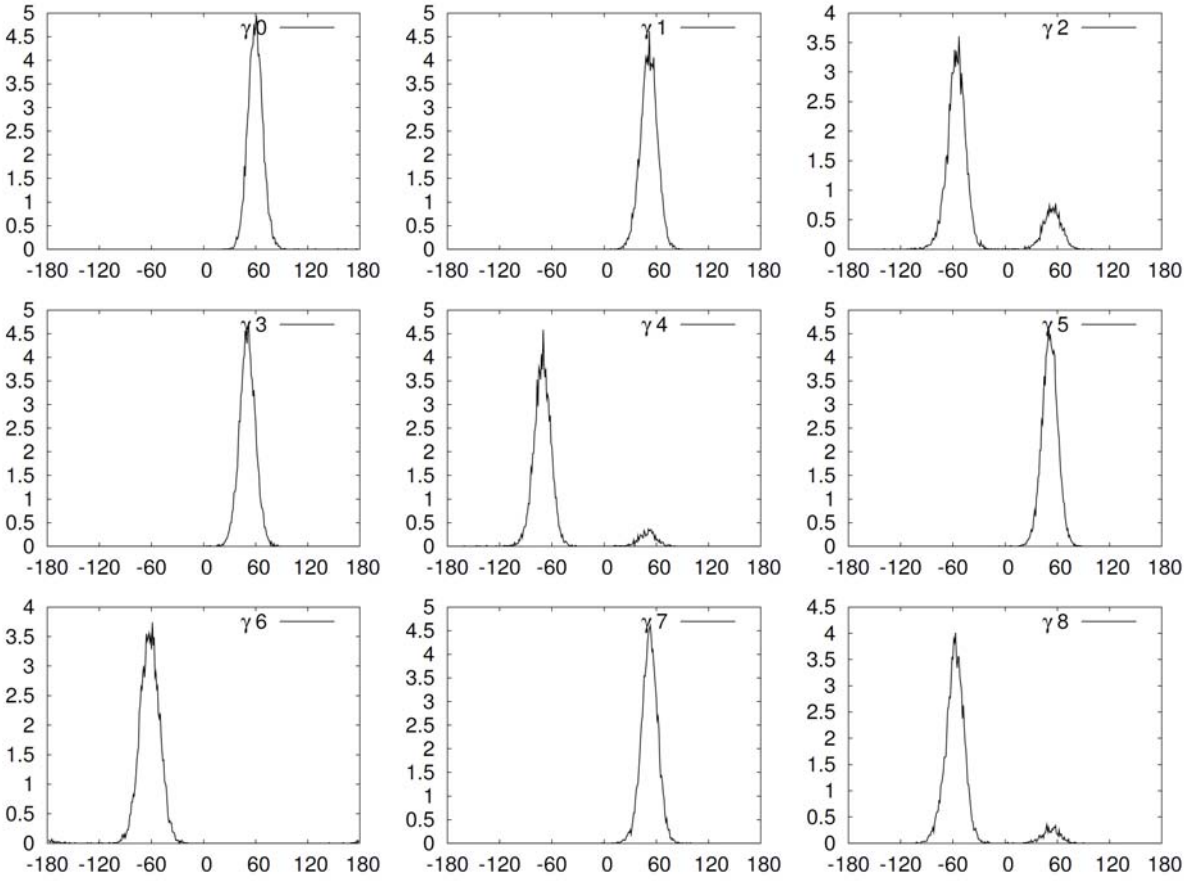

ζ 7
ζ 8
ζ 9
ζ 10
ζ 11
ζ 13
ζ 14
ζ 15
ζ 16
ζ 17
ζ 18
ζ 19
ζ 20
ζ 21
ζ 22

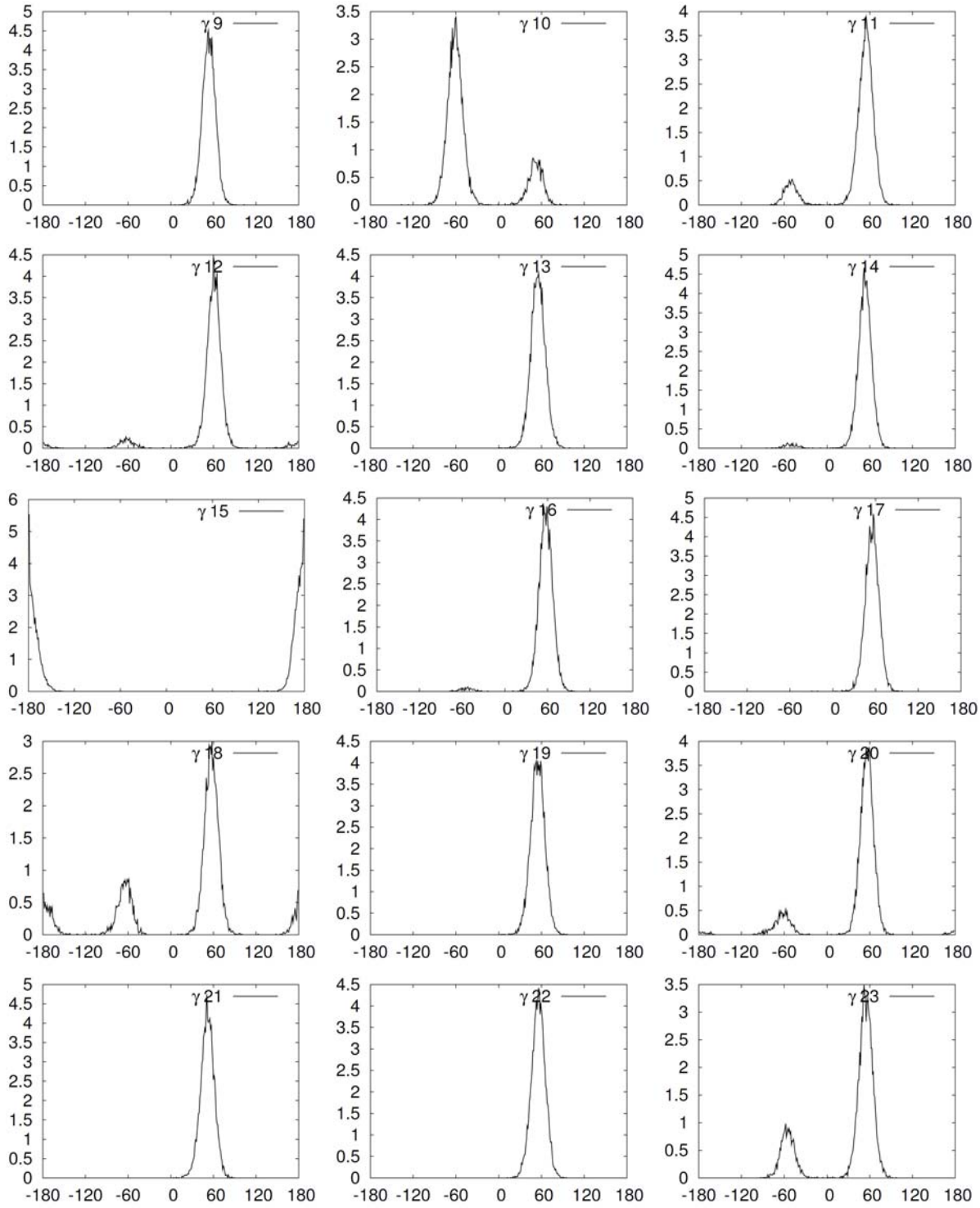

ζ 23
α 1
α 2
α 3
α 4
α 5
α 6
α 7
α 8
α 9
α 10
α 11
α 13
-180 -120 -60 0 60 120 180

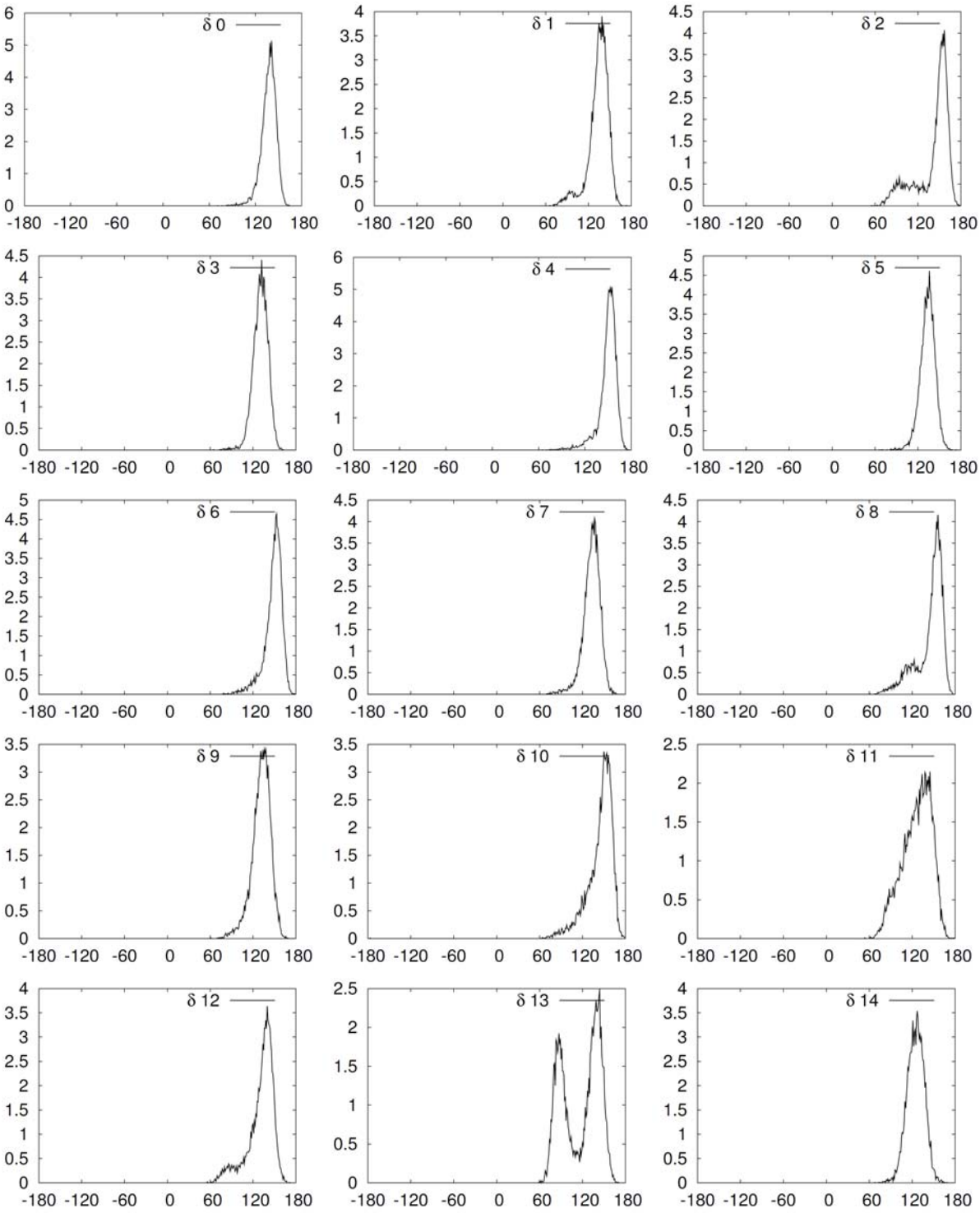

α 14
α 15
α 16
α 17
α 18
α 19
α 20
α 21
α 22
α 23
β 1
β 2
β 3

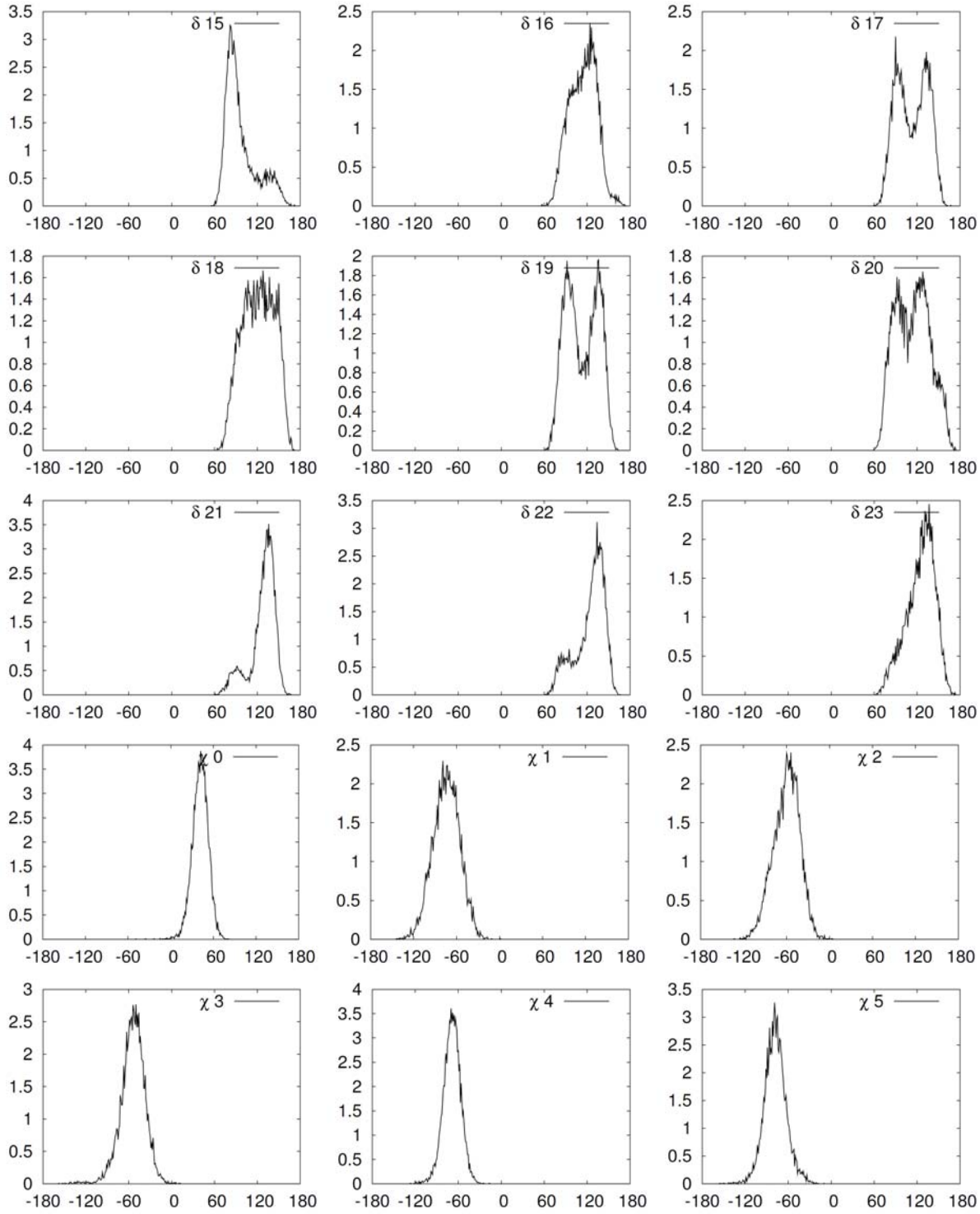

β 4
β 5
β 6
β 7
β 8
β 9
β 10
β 11
β 13
β 14
β 15
β 16
β 17
β 18
β 19
-180 -120 -60 0 60 120 180

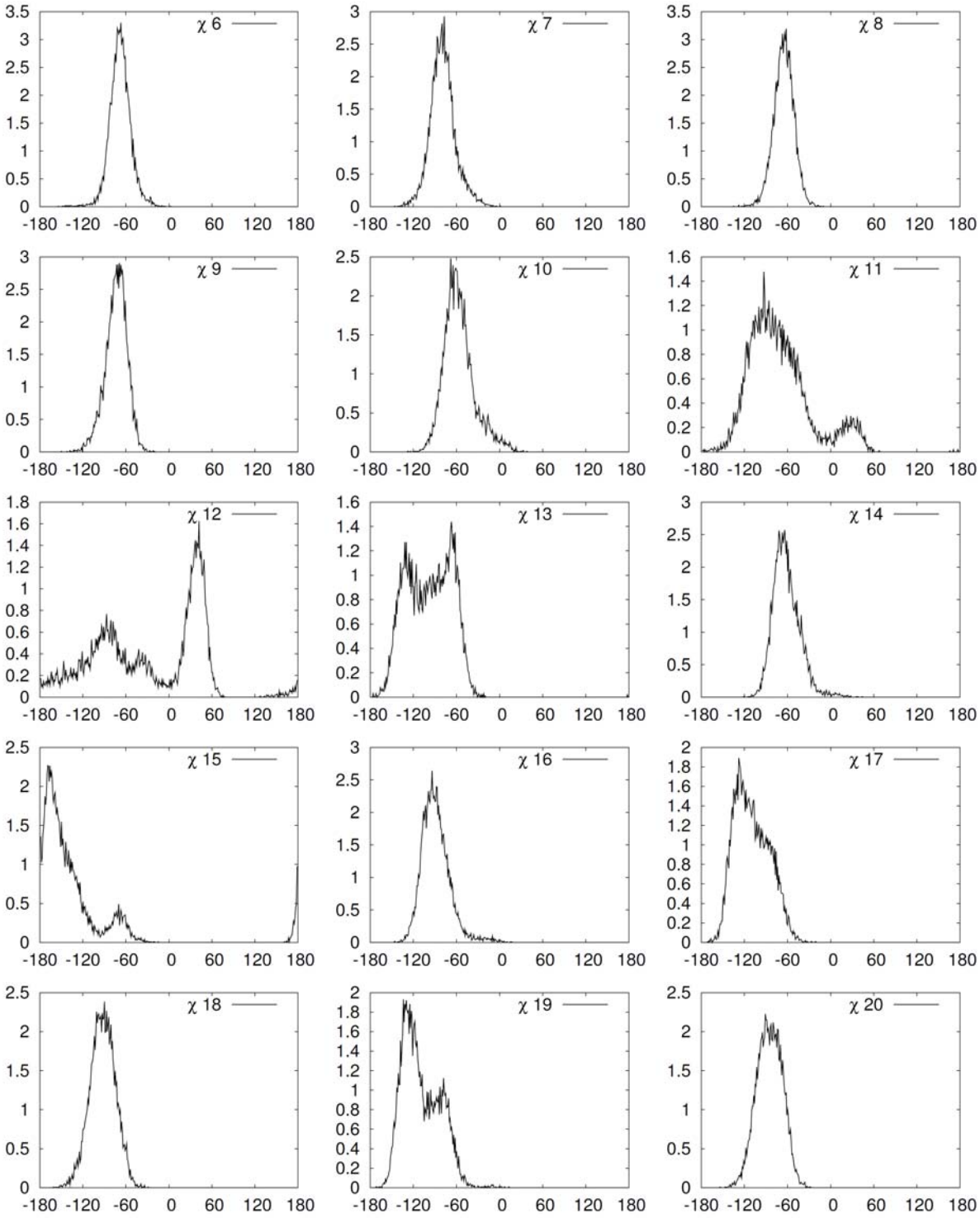

β 20
β 21
β 22
β 23

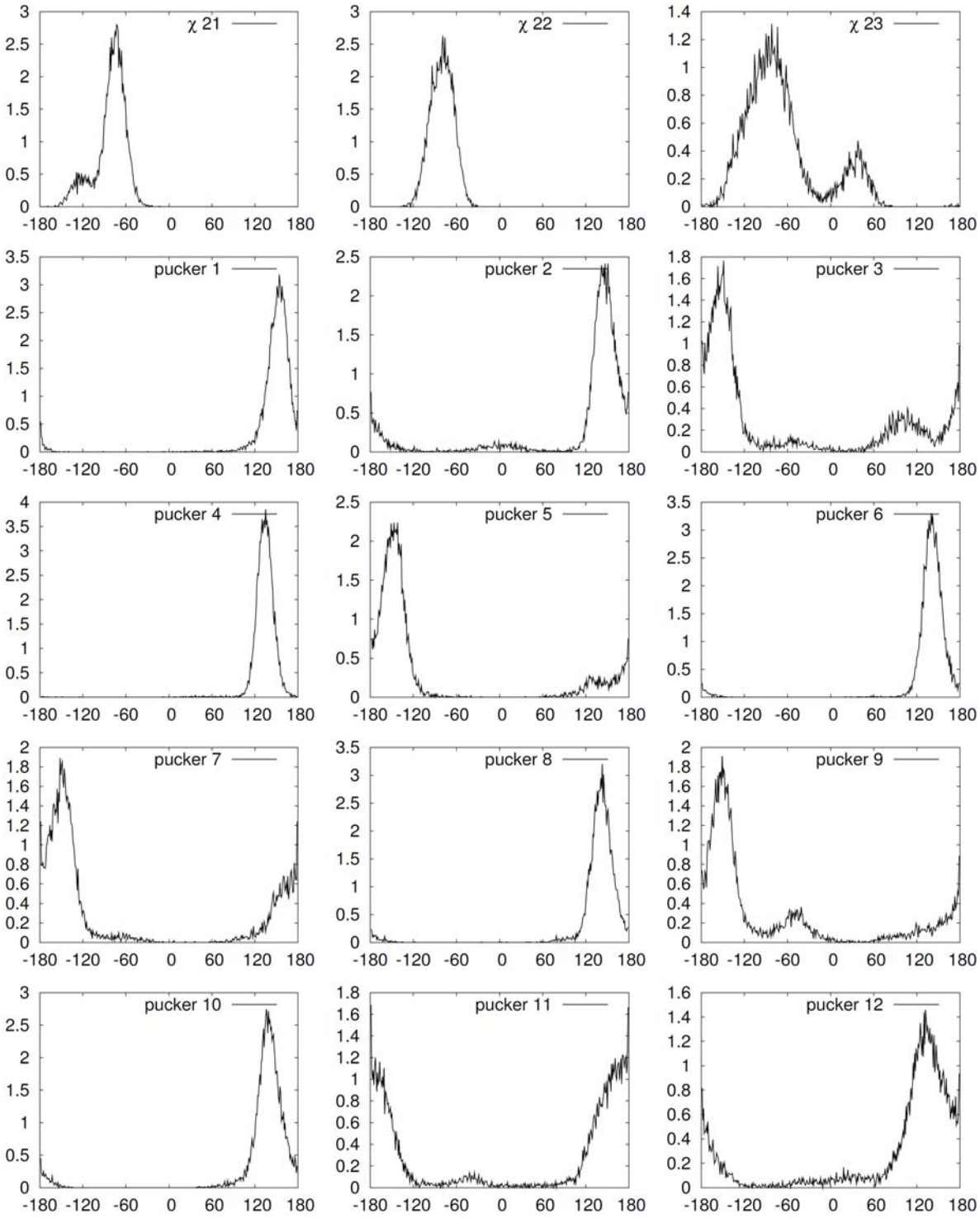

γ 0
γ 1
γ 2
γ 3
γ 4
γ 5
γ 6
γ 7
γ 8


γ 9
γ 10
γ 11
γ 12
γ 13
γ 14
γ 15
γ 16
γ 17
γ 18
γ 19
γ 20
γ 21
γ 22
γ 23


δ 0
δ 1
δ 2
δ 3
δ 4
δ 5
δ 6
δ 7
δ 8
δ 9
δ 10
δ 11
δ 12
δ 13
δ 14
-180 -120 -60 0 60 120 180


δ 15
δ 16
δ 17
δ 18
δ 19
δ 20
δ 21
δ 22
δ 23
χ 0
χ 1
χ 2
χ 3
χ 4
χ 5


χ 6
χ 7
χ 8
χ 9
χ 10
χ 11
χ 12
χ 13
χ 14
χ 15
χ 16
χ 17
χ 18
χ 19
χ 20


χ 21
χ 22
χ 23
pucker 1
pucker 2
pucker 3
pucker 4
pucker 5
pucker 6
pucker 7
pucker 8
pucker 9
pucker 10
pucker 11
pucker 12

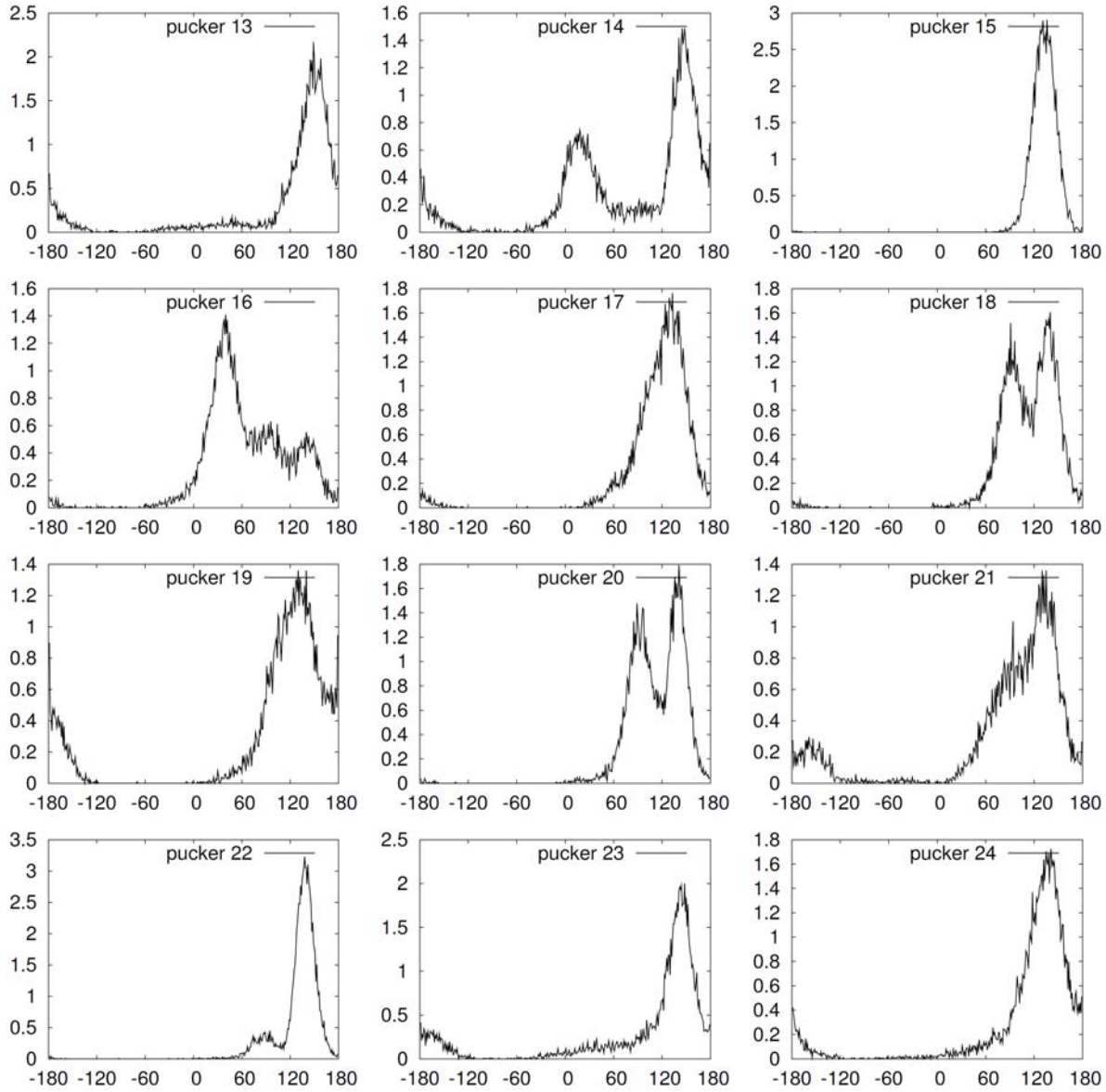


Figure S19: Histograms of the backbone torsions and the sugar pucker of transWH-c5-180d. Backbone torsions and sugar puckers are defined as follows:

ε *n*: C4'(*n*)-C3'(*n*)-O3'(*n*)-P(*n+1*)
ζ *n*: C3'(*n*)-O3'(*n*)-P(*n+1*)-O5'(*n+1*)
α *n*: O3'(*n*)-P(*n+1*)-O5'(*n+1*)-C5'(*n+1*)
β *n*: P(*n+1*)-O5'(*n+1*)-C5'(*n+1*)-C4'(*n+1*)
γ *n*: O5'(*n+1*)-C5'(*n+1*)-C4'(*n+1*)-C3'(*n+1*)
δ *n*: C5'(*n+1*)-C4'(*n+1*)-C3'(*n+1*)-O3'(*n+1*)
χ *n*: O4'(*n+1*)-C1'(*n+1*)-N9(*n+1*)-C4(*n+1*)
pucker *n*: C1'(*n*)-C2'(*n*)-C3'(*n*)-C4'(*n*)-O4'(*n*) (32)

The numbers next to the atom names in parentheses denote residue numbers. The histograms were collected using the data set of transWH-c5-180d with *ff99-bsc0* from

80,210 ps to 90,354 ps. X-axis and y-axis indicate angle (in degree) and arbitrary scale, respectively.

Table S10: The most populated backbone torsions and sugar pucker of transWH-c5-180d. The angles at the highest peaks in Figure S18 are tabulated. Most of the angles are comparable to conventional B-DNA torsions (33):

ε: > 160° / < -90°
ζ: (-130° - -60°) or (> 150° / < -150°)
α: -90° - -30°
β: > 130° / < -160°
γ: 20° - 80°
δ: 70° - 180°
χ: -160° - -60°

n	ε	ζ	α	β	γ	δ	χ	pucker
0	-	-	-	-	61	141	42	-
1	-161	-70	-77	173	52	139	-79	154
2	-159	-74	-64	-77	-53	157	-54	147
3	-179	-85	-77	168	52	132	-50	-149
4	-77	88	56	170	-70	155	-70	135
5	-179	-89	-71	166	50	136	-79	-147
6	-156	-78	68	179	-59	153	-67	140
7	179	-88	-69	171	53	137	-77	-152
8	-155	-79	-65	-76	-57	156	-62	143
9	-179	-83	-69	166	53	136	-69	-151
10	-160	-81	-59	-179	-60	150	-68	136
11	-179	-86	-67	-179	55	145	-93	-179
12	-	-	-	-	60	140	42	132
13	-80	-87	-76	179	56	144	-67	149
14	-160	-75	-74	-160	52	127	-64	145
15	-168	-84	-93	65	-179	82	-169	136
16	-156	-76	-66	-167	55	124	-94	40
17	-179	-90	-66	167	57	90	-128	133
18	-161	-83	-67	-179	58	128	-90	140
19	179	-89	-67	171	53	136	-134	130
20	-161	-87	-70	-161	59	128	-91	141
21	-179	-86	-76	167	51	136	-73	130
22	-155	-179	-69	-179	56	134	-79	138
23	-179	-86	-70	179	52	137	-82	148
24	-	-	-	-	-	-	-	141

REFERENCES

1. Roy Dennington II, Todd Keith, John Millam, Ken Eppinnett, W. Lee Hovell and Gilliland, R. (2003). 3.0 ed.
2. M. J. Frisch, G.W.T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. (2004). C.01 ed.
3. Bayly, C.I., Cieplak, P., Cornell, W. and Kollman, P.A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.*, **97**, 10269-10280.
4. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A. and Case, D.A. (2004) Development and testing of a general amber force field. *J Comput Chem*, **25**, 1157-1174.
5. Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham, T.E., Debolt, S., Ferguson, D., Seibel, G. and Kollman, P. (1995) AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structure and energetic properties of molecules. *Comp. Phys. Comm.*, **91**, 1-41.
6. Case, D.A., Cheatham, T.E., III, Darden, T.A., Gohlker, H., Luo, R., Merz, K.M., Jr., Onufriev, A.V., Simmerling, C., Wang, B. and Woods, R. (2005) The AMBER biomolecular simulation programs. *J. Comp. Chem.*, **26**, 1668-1688.
7. Wang, J., Cieplak, P. and Kollman, P.A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comp. Chem.*, **21**, 1049-1074.
8. Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T.E., 3rd, Laughton, C.A. and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophysical journal*, **92**, 3817-3829.
9. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926-935.
10. Hawkins, G.D., Cramer, C.J. and Truhlar, D.G. (1995) Pairwise solute descreening of solute charges from a dielectric medium. *Chem Phys Lett*, **246**, 122-129.

11. Hawkins, G.D., Cramer, C.J. and Truhlar, D.G. (1996) Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *J. Phys. Chem.*, **100**, 19824-19839.
12. Tsui, V. and Case, D.A. (2001) Theory and applications of the generalized born solvation model in macromolecular simulations. *Biopolymers*, **56**, 275-291.
13. Ryckaert, J.-P., Ciccotti, G. and Berendsen, H.J.C. (1977) Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J Comput Phys*, **23**, 327.
14. Pastor, R.W., Brooks, B.R. and Szabo, A. (1988) An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol. Phys.*, **65**, 1409-1419.
15. Loncharich, R.J., Brooks, B.R. and Pastor, R.W. (1992) Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers*, **32**, 523-535.
16. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H. and Pedersen, L.G. (1995) A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.*, **103**, 8577-8593.
17. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Comp. Phys.*, **81**, 3684-3690.
18. Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W. *et al.* (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res*, **33**, 889-897.
19. Srinivasan, J., Cheatham, T.E., Cieplak, P., Kollman, P.A. and Case, D.A. (1998) Continuum solvent studies of the stability of DNA, RNA and phosphoramidate helices. *Journal of the American Chemical Society*, **120**, 9401-9409.
20. Beveridge, D.L. and DiCapua, F.M. (1989) Free energy via molecular simulation. *Ann. Rev. Biophys. Biophy. Chem.*, **18**, 431-492.
21. Kollman, P.A. (1993) Free energy calculations- Applications to chemical and biochemical phenomena. *Chem. Rev.*, **93**, 2395-2417.
22. Rocchia, W., Alexov, E. and Honig, B. (2001) Extending the applicability of the nonlinear Poisson-Boltzmann equation: Multiple dielectric constants and multivalent ions. *J Phys Chem B*, **105**, 6507-6514.
23. Rocchia, W., Sridharan, S., Nicholls, A., Alexov, E., Chiabrera, A. and Honig, B. (2002) Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J Comput Chem*, **23**, 128-137.
24. Shao, J., Tanner, S.W., Thompson, N. and Cheatham, T.E., III. (2007) Clustering molecular dynamics trajectories: I. Characterizing the performance of different clustering algorithms. *J. Chem. Ther. Comp.*, [pending minor revision].
25. Arnott, S., Campbell-Smith, P.J. and Chandrasekaran, R. (1976), *Nucleic Acids*. 3rd ed, Vol. II, pp. 411-422.
26. Cubero, E., Abrescia, N.G., Subirana, J.A., Luque, F.J. and Orozco, M. (2003) Theoretical study of a new DNA structure: the antiparallel Hoogsteen duplex. *Journal of the American Chemical Society*, **125**, 14603-14612.

27. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*, **25**, 1605-1612.
28. Nelson, S.M., Ferguson, L.R. and Denny, W.A. (2007) Non-covalent ligand/DNA interactions: minor groove binding agents. *Mutat Res*, **623**, 24-40.
29. Persil, O., Santai, C.T., Jain, S.S. and Hud, N.V. (2004) Assembly of an antiparallel homo-adenine DNA duplex by small-molecule binding. *Journal of the American Chemical Society*, **126**, 8644-8645.
30. Berg, J.M., Tymoczko, J.L. and Stryer, L. (2006) *Biochemistry*. 6th ed. W. H. Freeman and Company, New York.
31. Shao, J.Y., Tanner, S.W., Thompson, N. and Cheatham, T.E. (2007) Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *Journal of Chemical Theory and Computation*, **3**, 2312-2334.
32. Harvey, S.C. and Prabhakaran, M. (1986) Ribose puckering: structure, dynamics, energetics, and the pseudorotation cycle. *Journal of the American Chemical Society*, **108**, 6128-6136.
33. Schneider, B., Neidle, S. and Berman, H.M. (1998) Conformations of the sugar-phosphate backbone in helical DNA crystal structures. *Biopolymers*, **42**, 113-124.