SUPPORTING INFORMATION FILE “supp-1” FOR THE PAPER

On the Role of cis Hoogsteen: Sugar Edge Family of Base Pairs in Platforms and Triplets – Quantum Chemical Insights into RNA Structural Biology

Purshotam Sharma,[±] Judit E. Šponer,[†] Jiří Šponer,[†] Sitansh Sharma,[±] Dhananjay Bhattacharyya*,[‡] and Abhijit Mitra*,[±]

[±]*Center for Computational Natural Sciences and Bioinformatics (CCNSB), International Institute of Information Technology (IIIT-H) Gachibowli, Hyderabad 500032, India*

[†]*Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic*

[‡]*Biophysics Division, Saha Institute of Nuclear Physics (SINP), 1/AF, Bidhannagar, Kolkata 700064, India*

* Corresponding authors.

dhananjay.bhattacharyya@saha.ac.in (DB) or abi_chem@iiit.ac.in (AM)

In addition to this file, the supporting information consists of one additional file “supp-2” containing PDB ids of analyzed RNA X-ray crystal structures, occurrences of analyzed T3 structures using BPFIND and coordinates of optimized geometries of thirty base pairs studied in this work.

Table S1: Comparison of different approaches for automated detection of analyzed H:S cis base pairs

Base pair	PDB ids; interaction	X3DNA	NDB tool	BPFind
A:A(T1)	1GID; A225(A):A226(A) ¹	-	+	-
A:C(T1)	1JJ2; A1106(0):C1105(0) ¹	-	-	-
A:G(T1)	1JJ2; A1236(0):G1235(0) ¹	+	-	-
A:U(T1)	1J5E; A1499:U1498 ¹	-	-	-
C:A(T1)	1JJ2; C2533(0):A2532(0) ¹	-	+	-
C:C(T1)	1FJG; C748:C749 ¹	-	+	-
C:G(T1)	1J5E; C596(A):G595(A) ¹	-	+	-
C:U(T1)	1DRZ; C156:U155 ¹	-	+	-
G:A(T1)	2J01; G2052:A2051 ¹	-	-	+
G:G(T1)	1JJ2; G2093(0):G2092(0) ¹	-	-	-
U:C(T1)	1ET4; U223:C222 ¹	-	-	-
U:G(T1)	430D; U11:G10 ¹	-	-	-
U:U(T1)	1JJ2; U831:U832 ¹	-	-	-
A:A(T2)	1JJ2; A56(9):A54(9) ¹	-	+	-
C:C(T2)	1JJ2; C2787(0):C2785(0) ¹	+	+	-
C:U(T2)	1JJ2; C2475(0):U2473(0) ¹	-	+	-
U:A(T2)	1GRZ; U412:A410 ¹	-	-	-
U:G(T2)	1JJ2; U2527(0):G2525(0) ¹	-	-	-
A:C(T3)	1NWX; A1143:C2008	-	+	+
A:G(T3)	1NWX; A2246:G963	-	-	+
A:U(T3)	1JJ2; A923(0):U2107(0)	-	+	+
C:C(T3)	1J5E; C899(A):C810(A)	-	+	+
U:G(T3)	1FJG; U173:G198	-	+	+

¹These base pairs are taken from Ref. 2 and 3.

Table S2: Geometrical characteristics of the base pair structures obtained on full optimization of the crystal geometries of H:S cis structures.

Base pair	Edge orientation	Geometrical characteristics
Adjacent base pairs (T1)		
A:rA	H:S cis	Viable H:S cis geometry.
A:rC	H:S cis	Viable H:S cis geometry.
A:rG	H:S cis	Geometry not consistent with the intended H:S cis pattern.
A:rU	H:S cis	Highly nonplanar geometry with perpendicular mutual orientation of the bases. Cannot be regarded as a viable base pair for RNA structures.
C:rA	W:S cis	No base-base hydrogen bond. Spurious interaction with 3'-OH group. Cannot be regarded as a viable base pair for RNA structures.
C:rC	H:S cis	Viable H:S cis geometry.
C:rG	H:S cis	Nonplanar geometry.
C:rU	H:S cis	Viable H:S cis geometry.
G:rA	-	Viable H:S cis geometry.
G:rG	H:S cis	Viable H:S cis geometry.
U:rC	H:S cis	Spurious interaction of N1 of U with O2 of rC.
U:rG	H:S cis	Viable H:S cis geometry.
U:rU	W:H cis	Highly nonplanar (buckled) geometry. Cannot be regarded as a viable base pair for RNA structures.
Near- adjacent base pairs (T2)		
A:rA	H:S cis	Viable H:S cis geometry.
C:rC	H:S cis	Highly nonplanar geometry with perpendicular mutual orientation of the bases. Cannot be regarded as a viable base pair for RNA structures.
C:rU	H:S cis	Viable H:S cis geometry.
U:rA	-	Spurious interaction of N1 of U with N3 of rA.
U:rG	H:S cis	Viable H:S cis geometry.
Distant base pairs (T3)		
A:rC	H:S cis	Viable H:S cis geometry.
A:rG	H:S cis	Viable H:S cis geometry.
A:rU	H:S cis	Viable H:S cis geometry.
C:rC	H:S cis	Viable H:S cis geometry.
U:rG	H:S cis	Viable H:S cis geometry.
Predicted structures (P)		
A:rU	H:S cis	Viable H:S cis geometry.
A:rU	H:S cis	Viable H:S cis geometry.
A:rG	H:S cis	Viable H:S cis geometry.
G:rA	H:S cis	Highly nonplanar partially stacked geometry.
G:rC	H:S cis	Viable H:S cis geometry.
G:rU	H:S cis	Viable H:S cis geometry.
U:rC	-	Highly nonplanar (buckled) geometry. Cannot be regarded as a viable base pair for RNA structures.

Table S3: Results of energy decomposition analysis based on Morokuma scheme at the HF/6-31G(d,p) level of theory.

Base pair	fragment	E _{int} (kcal/mol)	BSSE (kcal/mol)	E _{elec} (kcal/mol)	E _{EX} (kcal/mol)	E _{pol} (kcal/mol)	E _{CT} (kcal/mol)	E _{HOC} (kcal/mol)
T1								
A:rA	B-B'	-3.91	1.11	-9.42	7.92	-1.35	-2.33	0.15
	B-S	-5.13	2.18	-14.21	14.10	-2.82	-4.46	0.47
A:rC	B-B'	-4.55	1.19	-9.25	8.08	-1.91	-2.46	-0.20
	B-S	-4.10	2.06	-13.46	13.49	-2.19	-4.34	0.34
A:rG	B-B'	-4.29	1.29	-9.83	8.60	-1.70	-2.84	0.19
	B-S	-3.07	0.94	-5.72	3.81	-0.53	-1.47	-0.10
A:rU	B-B'	-2.29	1.20	-5.29	4.60	-0.88	-1.85	-0.07
	B-S	-1.43	2.03	-4.02	3.48	-0.65	-2.15	-0.12
C:rA	B-B'	-2.56	0.60	-2.55	0.71	-0.40	-0.92	-0.01
	B-S	1.96	1.98	-0.74	3.04	-0.27	-2.06	0.01
C:rC	B-B'	-8.47	1.37	-12.92	8.28	-2.30	-2.61	-0.30
	B-S	0.74	0.84	-0.94	3.03	-0.63	-1.78	0.22
C:rG	B-B'	-1.10	0.90	-2.55	2.08	-0.44	-1.13	0.04
	B-S	-0.02	1.23	-1.27	1.50	-0.19	-1.31	0.01
C:rU	B-B'	-4.40	1.09	-6.00	3.16	-0.92	-1.66	-0.08
	B-S	0.41	1.00	-1.94	3.06	-0.39	-2.06	-0.08
G:rA	B-B'	1.47	1.24	0.67	1.80	-0.89	-1.20	-0.15
	B-S	-4.35	1.19	-6.28	2.76	-0.55	-1.45	-0.02
G:rG	B-B'	-4.50	1.16	-7.70	5.80	-1.47	-2.22	-0.07
	B-S	-5.32	1.84	-12.05	10.39	-2.04	-3.44	-0.02
U:rC	B-B'	-1.89	0.64	-2.50	1.67	-0.69	-0.95	-0.07
	B-S	0.44	1.32	-2.03	3.26	-0.36	-1.86	0.10
U:rG	B-B'	-5.99	1.48	-9.88	6.68	-1.46	-2.56	-0.26
	B-S	-0.84	2.01	-3.84	3.81	-0.44	-2.41	0.02
U:rU	B-W-B'	-7.10	2.72	-13.91	9.74	-1.48	-4.31	0.14
	B-S	-0.90	1.89	-3.25	3.04	-0.47	-2.15	0.01
T2								
A:rA	B-B'	-0.73	0.63	-1.14	0.41	-0.16	-0.49	0.01
	B-S	-5.53	2.67	-20.17	21.44	-3.05	-6.74	0.31
C:rC	B-B'	-3.99	1.13	-4.29	1.00	-0.62	-1.18	-0.04
	B-S	0.75	2.72	-3.79	5.98	-0.58	-2.81	-0.17
C:rU	B-B'	-4.68	0.90	-7.20	4.49	-1.09	-1.70	-0.08
	B-S	-1.22	0.78	-1.80	0.97	-0.24	-0.90	-0.03
U:rA	B-B'	-0.99	1.24	-2.83	2.82	-0.65	-1.54	-0.03
	B-S	-2.58	0.90	-3.67	1.79	-0.45	-1.13	0.00
U:rG	B-B'	-5.80	0.90	-9.55	6.35	-1.41	-1.95	-0.13
	B-S	-1.69	0.77	-3.88	2.82	-0.24	-1.16	0.00
P								
A:rU	B-B'	-2.22	0.81	-2.85	0.80	-0.30	-0.68	-0.01
	B-S	-7.07	2.63	-21.41	22.14	-3.88	-7.27	0.71
A:rU	B-B'	-1.87	0.77	-2.75	1.13	-0.33	-0.69	0.01
	B-S	-5.30	2.68	-19.53	20.89	-2.97	-6.66	-0.29
A:rG	B-B'	-2.81	1.81	-10.33	10.11	-1.45	-3.05	0.11
	B-S	-4.01	2.61	-14.96	15.10	-1.89	-4.96	0.09

G:rA	B-B'	1.91	0.90	1.16	1.45	-0.66	-0.90	-0.04
	B-S	-4.29	1.77	-8.69	6.64	-1.19	-2.74	-0.09
G:rC	B-B'	-	-	-	-	-	-	-
	B-S	-3.43	2.29	-18.08	22.32	-4.15	-6.69	0.89
G:rU	B-B'	-	-	-	-	-	-	-
	B-S	-4.15	2.00	-14.39	15.33	-3.02	-4.56	0.49
U:rC	B-W-B'	-5.39	3.40	-18.51	17.00	-1.71	-5.75	0.18
	B-S	-1.34	0.96	-3.13	2.12	-0.21	-1.09	0.02
T3								
A:rC	B-B'	-3.40	0.99	-4.67	2.05	-0.78	-0.97	-0.02
	B-S	-5.06	2.67	-18.51	19.69	-2.79	-6.33	0.21
A:rG	B-B'	-4.81	1.31	-14.97	14.64	-2.10	-4.13	0.51
	B-S	-0.75	0.86	-1.35	0.70	-0.13	-0.83	0.01
A:rU	B-B'	-3.04	0.69	-6.04	4.50	-0.84	-1.27	-0.08
	B-S	-5.67	1.65	-16.18	16.64	-3.45	-4.96	0.63
C:rC	B-B'	-10.02	1.59	-12.53	6.09	-2.16	-2.54	-0.47
	B-S	0.06	1.47	1.58	0.01	-0.10	-0.08	0.00
U:rG	B-B'	-6.07	1.56	-12.00	9.78	-1.84	-3.23	-0.34
	B-S	-2.88	0.92	-4.57	2.61	-0.57	-1.26	-0.01

E^{elec} term describes the classical coulombic interaction between the occupied orbitals of one fragment with those of the other. E^{EX} represents the interaction between occupied orbitals of one fragment with the occupied orbitals of the other fragment that causes electron delocalization between the occupied orbitals of two fragments. The polarization term E^{pol} causes the mixing of occupied and virtual orbitals within the same fragment. The E^{CT} component results from the interfragment delocalization by mixing of occupied orbitals of one fragment with the virtual orbitals of the other and vice versa. The difference between the total energy and sum of individual components is the higher order coupling term, E^{HOC} .

Table S4: Root mean square deviation obtained after superposition of one of the bases in crystal and optimized geometries of the base pairs of H:S cis family.

Base pair (B-B'S)	RMSD when B is superposed			RMSD when B' of B'S is superposed	
	B'-B' RMSD (Å)	S-S RMSD (Å)	B'S-B'S RMSD (Å)	B-B RMSD (Å)	S-S RMSD (Å)
Adjacent base pairs (T1)					
A:rA	3.39	4.16	3.83	3.08	0.62
A:rC	1.70	1.95	1.85	4.18	0.77
A:rG	1.43	2.93	2.35	1.05	1.26
A:rU	1.91	1.59	1.72	4.04	1.20
C:rA	1.32	3.30	2.61	0.85	2.05
C:rC	3.82	4.37	4.15	1.65	0.60
C:rG	0.49	1.24	0.97	0.37	0.78
C:rU	4.37	5.70	4.94	3.95	1.52
G:rA	1.66	2.04	1.88	3.01	1.03
G:rG	4.38	2.64	3.56	2.84	0.88
U:rC	6.65	5.26	5.88	2.85	0.28
U:rG	2.61	1.82	2.22	3.16	1.78
U:rU	1.31	1.31	1.31	1.42	0.88
Near- adjacent base pairs (T2)					
A:rA	5.93	6.05	6.00	4.65	0.34
C:rC	5.68	5.24	5.43	6.20	1.34
C:rU	2.59	1.85	2.17	0.94	0.52
U:rA	6.19	6.99	6.64	3.21	0.93
U:rG	2.64	4.07	3.47	2.80	2.09
Distant base pairs (T3)					
A:rC	5.56	4.48	5.17	3.19	0.62
A:rG	7.73	9.19	8.52	5.02	1.52
A:rU	4.67	3.14	3.81	3.66	0.74
C:rC	3.47	3.91	3.73	6.34	1.21
U:rG	3.56	1.34	2.63	1.97	0.43

Table S5: Main Geometrical Parameters of the Interbase and Ribose-Base Contacts of the adjacent and non-adjacent H:S cis Base Pairs optimized at the B3LYP/6-31G(d,p) level of theory.

System No.	base pair	Donor atom(D)	Acceptor atom (A)	r(D-A) (Å)	r(H-A) (Å)	∠D-H-A (deg)
Adjacent base pairs (T1)						
1	A:rA	N6 (A)	N3(rA)	3.00	2.02	159.97
		O2' (rA)	N7(A)	2.79	1.89	153.43
2	A:rC	N6 (A)	O2(rC)	2.86	1.85	172.42
		O2' (rA)	N7(A)	2.83	1.87	161.49
3	A:rG	C8(A)	O2' (rG)	3.17	2.15	155.21
		N2(rG)	N7(A)	3.01	1.99	176.95
4	A:rU	N6 (A)	O2 (rU)	3.01	2.00	177.74
		C1' (rU)	N7 (A)	3.65	2.63	155.43
5	C:rA	N4(C)	N3(rA)	3.59	2.67	151.50
6	C:rC	N4 (C)	O2(rC)	2.88	1.93	154.46
		C5(C)	O2(rC)	3.23	2.49	124.40
7	C:rG	N4(C)	N3(rG)	3.28	2.48	136.25
8	C:rU	N4 (C)	O2 (rU)	3.52	2.46	162.87
		C5 (C)	O2' (rU)	3.10	2.13	167.03
9	G:rA	C2(rA)	O6(G)	3.55	2.48	167.26
10	G:rG	N2 (rG)	O6 (G)	2.95	1.95	164.74
		O2'(rG)	N7 (G)	3.01	2.03	177.81
11	U:rC	C5(U)	O2(rC)	3.39	2.35	161.46
12	U:rG	N2 (rG)	O4 (U)	2.93	1.94	161.31
		C5 (U)	O2' (rG)	3.42	2.37	163.24
13	U:rU	O (W)	O2 (rU)	2.79	1.89	154.61
		O (W)	O4 (U)	3.23	2.57	126.00
		C5(U)	O2'(rU)	3.53	2.56	148.78
Non-adjacent base pairs separated by one or two nucleotides in between (T2)						
14	A:rA	N6 (A)	O2'(rA)	2.99	2.01	161.05
		O2' (rA)	N7(A)	2.75	1.81	156.38
15	C:rC	N4(C)	O2(rC)	3.07	2.06	170.29
16	C:rU	N4(C)	O2(rU)	3.01	2.01	172.47
		C5(C)	O2'(rU)	3.55	2.57	150.66
17	U:rA	C6 (U)	O2'(rA)	3.27	2.43	133.55
		C5(U)	N3(rA)	3.43	2.46	149.25
18	U:rG	N2(rG)	O4(U)	2.88	1.92	156.24
		C5(U)	O2' (rG)	3.30	2.24	166.74

Table S6: Main Geometrical Parameters of the Interbase and Ribose-Base Contacts of the modeled as well as newly identified H:S cis Base Pairs, optimzied at B3LYP/6-31G(d,p) level of theory.

System No.	base pair	Donor atom(D)	Acceptor atom (A)	r(D-A) (Å)	r(H-A) (Å)	∠D-H-A (deg)
Predicted structures (P)						
19	A:rU	O2'(rU)	N7(A)	2.74	1.78	160.71
		N6(A)	O2'(rU)	3.06	2.10	156.78
20	A:rU	O2'(rU)	N7(A)	2.75	1.81	157.12
		N6(A)	O2'(rU)	3.01	2.04	159.87
21	A:rG	N7(A)	O2'(rG)	2.84	1.94	150.48
		O2'(rU)	N3(A)	3.02	2.04	161.18
		N1(A)	N2(rG)	3.05	2.05	164.85
22	G:rA	O2' (rA)	N7(G)	2.98	2.15	142.65
		C2(rA)	O6(G)	3.38	2.43	145.21
23	G:rC	O2' (rC)	N7(G)	2.72	1.72	173.80
24	G:rU	O2' (rU)	N7(G)	2.81	1.83	176.23
25	U:rC	C5 (U)	O2' (rC)	3.30	2.39	140.51
		O (W)	O4 (U)	2.88	1.91	169.52
		O (W)	O2 (rC)	2.69	1.83	145.03
		O2' (rC)	O(W)	2.66	1.66	173.47
Base pairs between two nucleotides far in the sequence (T3)						
26	A:rC	N6 (A)	O2'(rC)	3.03	2.06	160.27
		O2' (rC)	N7(A)	2.77	1.83	157.18
27	A:rG	N2(rG)	N3(A)	3.23	2.25	160.66
		O2'(rG)	N7(A)	2.93	1.93	162.54
28	A:rU	N6 (A)	O2 (rU)	2.95	1.99	158.67
		O2' (rU)	N7 (A)	2.75	1.78	164.57
29	C:rC	N4(C)	O2(rC)	3.00	2.00	168.71
30	U:rG	N2 (rG)	O4 (U)	2.90	1.89	166.59
		C5 (U)	N2 (rG)	3.54	2.51	158.22
		C6(U)	O2' (rG)	3.20	2.31	138.08

Table S7: Atoms in molecules (AIM) and Natural bond orbital (NBO) analysis for weak C-H...O and C-H...N hydrogen bonding interactions in base pairs of trans H:S family.^a

System No.	Base pair	Bond	AIM analysis		NBO analysis
			ρ (e/a ₀ ³)	$\nabla^2\rho$ (e/a ₀ ³)	E(2) (kcal/mol)
Adjacent base pairs (T1)					
3	A:rG(I)	C8-H(A)...O2'(rG)	0.0153	0.0593	3.22
4	A:rU(I)	C1'-H(rU)...N7(A)	0.0079	0.0261	1.92
8	C:rU	C5-H(C)...O2'(rU)	0.0101	0.0289	1.95
11	U:rC(I)	C5(U)...O2(rC)	0.0099	0.0381	2.21
12	U:rG(I)	C5-H(U)...O2'(rG)	0.0110	0.0407	2.56
13	U:rU	C5-H(U)...O2'(rU)	0.0083	0.0281	0.97
Non-adjacent base pairs separated by one or two nucleotides in between (T2)					
16	C:rU(II)	C5-H(C)... O2'(rU)	0.0074	0.0291	1.20
17	U:rA	C5-H(U)...N3(rA)	0.0118	0.0323	3.33
		C6-H(U)...O2'(rA)	0.0105	0.0347	1.52
18	U:rG(II)	C5-H(U)...O2'(rU)	0.0136	0.0492	2.05
Predicted Structures (P)					
22	G:rA(II)	C2-H(rA)...O6(G)	0.0089	0.0340	1.32
25	U:rC(II)	C5-H(U)...O2'(rU)	0.0102	0.0395	1.12
Base pairs between two nucleotides far in the sequence (T3)					
29	U:rG(III)	C5-H(U)...N3(rG)	0.0100	0.0323	3.12

^aThe values for weak hydrogen bonds of G:rA(T1) are given separately in Table S17.

Table S8: Description of the D-A network and relevant geometrical parameters of sugar edge base in crystal and optimized geometries. Interaction energies in optimized geometries, and occurrence frequencies are also displayed for comparison.

(a)

	rA	O4'	O2'	O2'	C1'	N3	C2	geomtry	r(O2'-N3)	r(O2'-C2)	r(O4'-N3)	Pucker	χ (deg)	E _{RIMP2} (kcal/mol)	Occurrence ¹
A	ArA (T1)					N6		Crystal	3.83	4.96	4.21	C2' endo	-137.9	-16.9	15.6%
			N7			N6		Optimized	3.84	4.95	4.27	C2' endo	-142.6		
	ArA (T2)		N7			N6		Crystal	4.20	5.45	4.35	C3' endo	-156.9	-13.6	
			N7	N6				Optimized	4.02	5.27	4.39	C3' endo	-162.1		
G	GrA (T1)						O6	Crystal	3.22	4.39	4.36	C2' endo	-160.6	-6.4	1.6%
							O6	Optimized	2.76	3.90	4.42	C2' endo	165.4		
U	UrA (T2)						O4	Crystal	3.79	4.99	4.18	C3' endo	-146.5	-6.3	1.0%
				C6			C5		Optimized	4.14	5.41	4.45	C3' endo		
C	CrA (T1)	C5				N4		Crystal	4.56	5.74	3.85	C2' endo	-105.9	-5.2	4.9%
						N4		Optimized	2.82	3.99	4.41	C2' endo	174.4		

(b)

(2)

	rC	O2'	O2'	C1'	O2	geometry	r(O2'-O2) (Å)	Pucker	χ (deg)	E _{RIMP2} (kcal/mol)	Occurrence frequency
A	ArC T1			N7	N6	Crystal	3.84	C2' endo	-123.3	-15.6	1.0%
		N7			N6	Optimized	3.31	C2' endo	-148.5		
	ArC T3		N6			Crystal	3.89	C3' endo	-154.7	-15.6	
N7		N6			Optimized	3.45	C3' endo	-164.0			
U	UrC T1				C5	Crystal	3.61	C2' endo	-129.7	-5.1	1.4%
					C5	Optimized	3.63	C2' endo	-137.0		
C	CrC T1		C6		C5,N4	Crystal	3.67	C3' endo	-150.6	-11.7	0.6%
					C5,N4	Optimized	3.37	C3' endo	-170.9		
	CrC T2		N4		N4	Crystal	3.80	C2' endo	-123.9	-7.7	
					N4	Optimized	2.70	C2' endo	-170.8		
CrC T3		C5,N4		N4	Crystal	3.38	C3' endo	-163.5	-9.5		
				N4	Optimized	2.70	C2' endo	-171.3			

(c)

	rG	O4'	O2'	O2'	C1'	N3	N2	geometry	r(O2'-N2) (Å)	r(O2'-N3) (Å)	r(O4'-N3) (Å)	Pucker	χ (deg)	E _{RIMP2} (kcal/mol)	Occurrence ¹	
A	ArG (T1)						N7	Crystal	5.55	3.75	4.11	C2' endo	-132.9	-12.1	1.5%	
				C8			N7	Optimized	4.20	2.78	4.33	C2' endo	-177.9			
	ArG (T3)						N7	Crystal	5.63	3.96	4.37	C3' endo	-165.7	-15.8		
							N7(D),N6(A)	Optimized	4.25	2.82	4.37	C2' endo	173.6			
G	GrG (T1)						O6	Crystal	5.60	3.80	4.16	C2' endo	-134.3	-14.6	0.2%	
			N7				O6	Optimized	4.90	3.33	4.33	C2' endo	-170.0			
U	UrG (T1)						O4	Crystal	6.66	4.62	3.75	C2' endo	-98.0	-11.8	46.9%	
				C5			O4	Optimized	4.29	2.83	4.32	C2' endo	-175.1			
	UrG (T2)						O4	Crystal	6.41	4.44	3.89	C2' endo	-109.6	-10.0		
				C5			O4	Optimized	4.19	2.74	4.37	C2' endo	-171.4			
	UrG (T3)				C6		C5	O4	Crystal	5.77	4.07	4.39	C3' endo	-171.5		-12.4
				C6		C5	O4	Optimized	5.91	4.18	4.37	C3' endo	-164.9			
C	CrG (T1)			C5		N4		Crystal	6.16	4.26	3.95	C2' endo	-166.7	-6.3	3.0%	
						N4		Optimizied	6.03	4.16	4.06	C2' endo	-128.2			

(d)

(a)

A	rU	O2'	O2'	C1'	O2	O(w)	geometry	r(O2'-O2) (Å)	Pucker	χ (deg)	E _{RIMP2} (kcal/mol)	Occurrence ¹
	ArU T1			N7	N6		Crystal	3.64	C2' endo	-127.1	-8.4	10.1%
				N7	N6		Optimized	2.80	C2' endo	-165.3		
	ArU T3	N7			N6		Crystal	3.75	C3' endo	-158.0	-15.2	
		N7			N6		Optimized	3.60	C3' endo	-166.5		
U	UrU T1		C6		C5	O2/O4	Crystal	3.64	C2' endo	-129.2	-11.5	3.0%
			C5			O2/O4	Optimized	2.78	C2' endo	-162.1		
C	CrU T1				N6/C5		Crystal	3.71	C2' endo	-120.0	-7.9	8.6%
			C5		N4		Optimized	2.75	C2' endo	-172.2		
	CrU T2				N4		Crystal	3.86	C3' endo	-153.3	-7.9	
		C5		N4		Optimized	3.58	C3' endo	-166.7			

Table S9: Description of the D-A network Hoogsteen edge base in crystal and optimized geometries.

Base pair	A-Hoogsteen					
	N7-acceptor		N6-donor		N6-acceptor	
	Crystal	Optimized	Crystal	Optimized	Crystal	Optimized
A:rA(T1)	-	O2'	N3	N3	-	-
A:rA(T2)	O2'	O2'	N3	O2'	-	-
A:rC(T1)	C1'	O2'	O2	O2	-	-
A:rC(T3)	-	O2'	O2'	O2'	-	-
A:rG(T1)	N2	N2	-	-	-	-
A:rG(T3)	N2	N2	-	-	-	N2
A:rU(T1)	C1'	C1'	O2	O2	-	-
A:rU(T3)	O2'	O2'	O2	O2	-	-
	C-Hoogsteen					
	C5-donor		N4-donor		C6-donor	
	Crystal	Optimized	Crystal	Optimized	Crystal	Optimized
C:rA(T1)	O4'	-	N3	N3	-	-
C:rC(T1)	O2	O2	O2	O2	-	-
C:rC(T2)	-	-	O2'/O2	O2	-	-
C:rC(T3)	O2'	-	O2'	O2	-	-
C:rG(T1)	O2'	-	N3	N3	-	-
C:rU(T1)	-	O2'	-	O2	O2	-
C:rU(T2)	-	O2'	O2	O2	-	-

Base pair	G-Hoogsteen			
	N7-acceptor		O6-acceptor	
	Crystal	Optimized	Crystal	Optimized
G:rA(T1)	-	-	C2	C2
G:rG(T1)	-	O2'	N2	N2

Base pair	U-Hoogsteen							
	C6-donor		C5-donor		O2-acceptor		O4-acceptor	
	Crystal	Optimized	Crystal	Optimized	Crystal	Optimized	Crystal	Optimized
U:rA(T2)	-	O2'	-	N3	-	-	O2	-
U:rC(T1)	-	-	O2	O2	-	-	-	-
UrG(T1)	-	-	-	O2'	-	-	N2	N2
U:rG(T2)	-	-	-	O2'	-	-	N2	N2
U:rG(T3)	O2'	O2'	N3	N3	-	-	N2	N2
U:rU(T1)	O2'	-	O2	O2'	O(water)	O(water)	O(water)	O(water)

Table S10: Details regarding the extent of planarity of base pairs within and without the crystal context, their participation in triplet formation in crystal, and the requirement of constraints

Base pair	Buckle (deg)	Propel (deg)	Characteristic in crystal	Change on optimization.	Constraint?
Adjacent base pairs (T1)					
A:rA	12.10 (-18.34)	-65.48 (-6.19)	Planar (DNPT, MG, no interaction of 3 rd base with SE base)	High propeller; (no constraints, binding pattern changes)	No
A:rC	13.80 (-25.79)	-6.59 (-11.28)	Slightly buckled (DNP)	Decrease in buckle and increase in propeller (binding pattern changes)	No
A:rG	-2.46 (-18.08)	-31.90 (-38.31)	Propelled (DNP)	Decrease in buckle and slight decrease in propeller	Yes
A:rU	-37.40 (-17.56)	-6.53 (7.92)	Planar (DNPT, mg)	Increase in buckle	Yes
C:rA	-4.68 (-3.45)	10.34 (5.31)	Highly planar(DNPT, MG, no interaction of 3 rd base with SE base)	Remains planar	Yes
C:rC	28.01 (4.90)	32.24 (11.47)	Highly planar(DNPT, MG, no interaction of 3 rd base with SE base)	Increase in buckle, open, propeller	No
C:rG	-3.25 (-2.06)	8.47 (5.73)	Highly planar(DNPT, MG, has interaction of 3 rd base with SE base)	Remains planar	Yes
C:rU	-74.78 (4.28)	-22.69 (12.65)	Planar (DNP).	Sharp increase in buckle	No
G:rA	12.93 (-11.96)	-10.48 (-5.90)	Planar (DNPT, MG, no interaction of 3 rd base with SE base)	Remains planar	No
G:rG	-40.10 (8.74)	10.10 (7.27)	Highly planar(DNPT, MG, has interaction of 3 rd base with SE base)	Sharp increase in buckle	No
U:rC	23.36 (-37.34)	42.78 (6.71)	Relatively high buckle (DNPT, mg)	Decrease in buckle but sharp increase in propeller	Yes
U:rG	-52.64 (-6.67)	1.86 (-2.69)	Highly planar (DNPT, no interaction of 3 rd base with SE base)	Sharp increase in buckle propeller	No
U:rU	-34.07 (-16.85)	9.92 (3.29)	Planar (DNPT, wMG)	Increase in buckle	Yes
Near-adjacent base pairs (T2)					
A:rA	56.50 (-31.39)	-30.30 (-54.54)	Buckled and propelled (mgT)	Increase in buckle, change of sign of buckle (binding pattern changes)	No
C:rC	-53.22 (-3.44)	2.67 (-33.87)	Propelled (B)	Sharp increase in buckle, sharp decrease in open and propeller (binding pattern changes).	Yes
C:rU	8.06 (-23.81)	0.56 (-5.51)	Slightly buckled (mgT)	Decrease in buckle and propeller	No
U:rA	-27.78 (-18.81)	-4.85 (-36.33)	Propelled (MGT)	Sharp decrease in open and propeller (change in binding pattern).	Yes
U:rG	17.03 (-15.49)	-16.00 (0.61)	Planar (MGT, interaction of 3 rd base with SE base)	Increase in propeller	No
Distant base pairs (T3)					
A:rC	-57.29 (7.25)	-16.62 (-29.70)	Propelled (mgT, no interaction of 3 rd base with SE base)	Sharp increase in buckle	No
A:rG	34.92 (-53.38)	60.84 (-9.78)	Buckled (MGT)	Sharp increase in propeller. Change in sign of buckle (binding pattern changes)	No
A:rU	-44.84 (22.42)	-39.49 (-29.94)	Propelled (mgT, no interaction of 3 rd base with SE base)	Increase in buckle, open, propeller. change in sign of buckle	No
C:rC	-50.12 (-39.81)	28.88 (-36.09)	Buckled and propelled (Q, mg)	Increase in buckle Sharp increase in open.	No
U:rG	15.55 (55.78)	-12.97 (9.03)	Buckled (mgT)	Decrease in buckle	No

Table S11: Variation in base pair geometry and hydrogen bonding patterns in the context of base pair occurrence in a platform and/or triplet, and the nature of interactions with the third base

(a) **BASE PAIRS WHICH HAVE PLANAR CRYSTAL GEOMETRIES^a**

	Base pair	Buckle	Propellar	environment	H-bonding pattern changes?	Open changes significantly?	3 rd nucleotide interacts with sugar base?
1	A:rA(I)	→	↑	DNPT, MG	Yes	No →	No
2	A:rU(I)	↑	→	DNPT, mg	No	No →	Yes
3	C:rA(I)	→	→	DNPT, MG	Only weak C5-O4' is broken	No →	No
4	C:rC(I)	↑	↑	DNPT, MG	weak C6-O2' bond formed.	Yes(~29°)	No
5	C:rG(I)	→	→	DNPT, MG	Only weak C5-O2' is broken	No →	Yes
6	C:rU(I)	↑	→	DNP	Only weak C5-O2' is formed	No →	3 rd base absent
7	G:rA(I)	→	→	DNPT, MG	No	No →	No
8	G:rG(I)	↑	→	DNPT, MG	Yes	No →	Yes
9	U:rG(I)	↑	→	DNPT	Only weak C5-O2' is formed	No →	Yes
10	U:rU(I)	↑	→	DNPT, wMG	Only C-H mediated bonds change	No →	Yes
11	U:rG(II)	→	↗	MGT	Only weak C5-O2' is broken	No →	Yes

^aRed cases: Only buckle increases on optimization; Green cases: Remain planar on optimization; Blue case: Only propeller increase on optimization; Yellow case: Both buckle and propeller increase on optimization.

(b) **BASE PAIRS WHICH HAVE BUCKLED CRYSTAL GEOMETRIES^a**

	Base pair	Buckle	Propellar	environment	H-bonding pattern changes?	Open changes significantly?	3 rd nucleotide interacts with 1 st base?
12	A:rC(I)	↓	→	DNP	Yes	No	3 rd base absent
13	U:rC(I)	↓	↑	DNPT, mg	No	No	Yes
14	C:rU(II)	↓	→	mgT	Only weak C5-O2' is formed	No	3 rd base absent
15	A:rG(III)	↓	↑	MGT	Yes	Yes(~29°)	No
16	U:rG(III)	↓	→	mgT	No	No	No

^aYellow cases: Which undergo decrease in buckle and no change in propeller on optimization; Violet cases: Which undergo decrease in buckle and increase in propeller on optimization.

(c) **BASE PAIRS WHICH HAVE PROPELLED CRYSTAL GEOMETRIES^a**

	Base pair	Buckle	Propellar	environment	H-bonding pattern changes?	Open changes significantly?	3 rd nucleotide interacts with sugar base?
17	A:rG(I)	↓	→	DNP	Only weak C8-O2' is formed	No	No 3 rd base
18	C:rC(II)	↑	↓	B	Yes	Yes(~29°)	No 3 rd base
19	U:rA(II)	→	↓	MGT	Yes	Yes(~50°)	Yes
20	A:rC(III)	↑	↗	mgT	Yes	No	No

^aYellow case: Only buckle decreases on optimization; Pink case: Only propeller decreases optimization; Green case: Buckle increases and propeller decreases on optimization; Blue case: Buckle increase and propeller slightly increases on optimization.

(d) **BASE PAIRS WHICH HAVE BUCKLED AND PROPELLED CRYSTAL GEOMETRIES^a**

	Base pair	Buckle	Propellar	environment	H-bonding pattern changes?	Open changes significantly?	3 rd nucleotide interacts with 1st base?
21	A:rA(II)	↑	↓	mgT	Yes	No	No
22	A:rU(III)	↑	↑	mgT	No	No	No
23	C:rC(III)	↑	↓	Q, mg	Yes	Yes	No

^aYellow case: Buckle and propeller further increases on optimization; Green cases: Buckle increases and propeller increases on optimization.

Table S12: Examples of isolated DNPs and DNPs in a triplet, as identified by different sources.

Base pair	Example of Isolated DNP	Example of DNP in a triplet
A:rA	A219:A218; 1GID (from SCOR)	A226:A225:U249; 1GID (from Leontis et al) ¹
A:rC	A1106:C1105; 1JJ2* (from Leontis et al) ¹	A1130:C1129:G1143; 2I2P (from BPFind)
A:rG	A1236:G1235; 1JJ2* (from Leontis et al) ¹	-
A:rU	A625:U624; 1I94 (from BPFind)	A1499:U1498:C1403; 1J5E (from Leontis et al) ¹
C:rA	C149:A148; 1NBS (from BPFind)	C2533:A2532:G2489; 1JJ2 (from Leontis et al) ¹
C:rC	-	C749:C748:G657; 1FJG (from Leontis et al) ¹
C:rG	-	C596:G595:G644; 1J5E (from Leontis et al) ¹
C:rU	C156:U155; 1DRZ (from Leontis et al) ¹	-
G:rA	-	G2052:A2051:C2617; 2J01 (from Leontis et al) ¹
G:rG	-	G2093:G2092:U2652; 1JJ2 (from Leontis et al) ¹
U:rC	-	U223:C222:G207; 1ET4 (from Leontis et al) ¹
U:rG	U920:G919; 1EUQ (from BPFind)	U11:G10:A20; 430D (from Leontis et al) ¹
U:rU	-	U832:U831:U850; 1JJ2 (from Leontis et al) ¹

Table S13. Range of interaction energies and types of hydrogen bonds present in different H:S cis base pairs.

Interaction energy	Type of H-bonds	Category			
		T1	T2	T3	Predicted
Above 17 kcal/mol	Water-mediated	-	-	-	U:rC
-13 to -17 kcal/mol	Three N-H...O/N	-	-	-	A:rG
	Two N-H...O/N	A:rA, A:rC, G:rG	A:rA	A:rC, A:rG, A:rU	A:rU(P1), A:rU(P2)
	One N-H...O/N	-	-	-	G:rC, G:rU
	Water mediated	U:rU	-	-	-
-9 to -13 kcal/mol	One C-H...O/N one N-H...O/N	A:rG, C:rC, U:rG	U:rG	-	-
	Two C-H...O/N one N-H...O/N	-	-	U:rG	-
	One N-H...O/N	-	-	C:rC	-
Below -9 kcal/mol	One C-H...O/N one N-H...O/N	A:rU, C:rU	C:rU	-	G:rA
	One N-H...O/N	C:rA, C:rG, G:rA, U:rC	C:rC	-	-
	Two C-H...O/N	-	U:rA	-	-

Table S14: Base pair parameters calculated using NUPARM.^a

System No.	Base pair	Buckle (deg)	Open (deg)	Propel (deg)	Stagger (Å)	Shear (Å)	Stretch (Å)
Adjacent base pairs(T1)							
1	A:rA	12.10 (-18.34)	-14.88 (-9.91)	-65.48 (-6.19)	-0.45 (0.12)	2.52 (1.89)	3.17 (3.12)
2	A:rC	13.80 (-25.79)	-30.49 (-19.39)	-6.59 (-11.28)	-1.25 (0.49)	-1.14 (-1.31)	3.70 (3.34)
3	A:rG	-2.46 (-18.08)	2.80 (-6.16)	-31.90 (-38.31)	0.39 (-0.04)	-1.21 (-1.58)	3.49 (3.60)
4	A:rU	-37.40 (-17.56)	-21.73 (-17.36)	-6.53 (7.92)	0.58 (1.29)	-1.08 (-1.41)	3.42 (2.91)
5	C:rA	-4.68 (-3.45)	-20.36 (-12.33)	10.34 (5.31)	1.70 (1.43)	2.94 (2.54)	3.64 (3.04)
6	C:rC	28.01 (4.90)	-40.88 (-11.96)	32.24 (11.47)	1.26 (0.05)	-1.90 (-2.31)	4.09 (3.26)
7	C:rG	-3.25 (-2.06)	-14.66 (-16.09)	8.47 (5.73)	2.48 (2.22)	3.20 (3.28)	2.45 (2.60)
8	C:rU	-74.78 (4.28)	-17.08 (-20.07)	-22.69 (12.65)	-0.50 (0.70)	-0.96 (-1.71)	3.71 (3.58)
9	G:rA	12.93 (-11.96)	-24.00 (-23.46)	-10.48 (-5.90)	-0.31 (0.81)	-0.45 (0.97)	3.44 (3.13)
10	G:rG	-40.10 (8.74)	-3.83 (-19.02)	10.10 (7.27)	-1.35 (-0.15)	1.61 (2.54)	3.31 (3.22)

11	U:rC	23.36 (-37.34)	6.86 (-9.96)	42.78 (6.71)	0.14 (-1.65)	-3.09 (-2.54)	3.07 (3.16)
12	U:rG	-52.64 (-6.67)	-2.10 (-5.78)	1.86 (-2.69)	0.02 (-0.39)	1.04 (1.68)	3.67 (3.63)
13	U:rU	-34.07 (-16.85)	-8.25 (-17.27)	9.92 (3.29)	-0.23 (-0.58)	-1.99 (-1.79)	3.75 (3.71)
Non-adjacent base pairs separated by one or two nucleotides in between							
14	A:rA	56.50 (-31.39)	-26.84 (-18.21)	-30.30 (-54.54)	-0.72 (-1.01)	4.04 (2.82)	3.96 (3.33)
15	C:rC	-53.22 (-3.44)	-22.26 (-65.42)	2.67 (-33.87)	0.58 (-2.92)	-0.38 (1.64)	3.98 (4.14)
16	C:rU	8.06 (-23.81)	-45.28 (-32.67)	0.56 (-5.51)	-0.48 (0.71)	-1.62 (-1.64)	4.61 (4.25)
17	U:rA	-27.78 (-18.81)	-4.02 (-54.15)	-4.85 (-36.33)	-0.26 (-1.72)	0.53 (2.70)	3.38 (3.62)
18	U:rG	17.03 (-15.49)	-30.12 (-45.70)	-16.00 (0.61)	-0.72 (-0.76)	1.28 (0.50)	4.63 (5.33)
Predicted structures							
19	A:rU	55.59	-9.23	-3.59	-0.75	-1.03	3.49
20	A:rU	-58.92	-34.99	-20.34	-1.19	-1.00	4.72
21	A:rG	-44.75	-37.08	1.16	-2.05	3.30	2.31
22	G:rA	37.10	-15.87	1.13	-0.22	1.09	3.66
23	G:rC	-22.39	-60.37	7.64	-2.76	-0.81	2.20
24	G:rU	-5.97	-55.70	9.48	-2.44	-0.69	2.58
25	U:rC	-25.63	-6.97	15.54	0.83	-2.12	4.28
Base pairs between two nucleotides far in the sequence							
26	A:rC	-57.29 (7.25)	-38.93 (-47.54)	-16.62 (-29.70)	-1.08 (-2.25)	-0.91 (-1.46)	4.78 (4.92)
27	A:rG	34.92 (-53.38)	-1.77 (28.91)	60.84 (-9.78)	1.41 (-0.81)	-0.45 (-1.59)	3.39 (2.80)
28	A:rU	-44.84 (22.42)	-33.35 (-29.53)	-39.49 (-29.94)	-1.12 (-1.32)	0.81 (0.23)	4.31 (4.10)
29	C:rC	-50.12 (-39.81)	-70.73 (-45.94)	28.88 (-36.09)	0.79 (-0.70)	-2.34 (1.06)	4.90 (4.94)
30	U:rG	15.55 (55.78)	2.12 (5.98)	-12.97 (9.03)	-0.18 (-1.04)	-0.56 (-0.80)	3.49 (2.90)

^aThe values in parentheses correspond to RNA crystal geometries of base pairs

Table S15. Relationship between intramolecular hydrogen bond and availability of suitable D-A interactions.

Base pair	Pucker		Intramolecular H-bond present in optimized geometry?
T1			
A:rA	C2'-endo	O2' - utilized as donor, N3 utilized as acceptor	No
A:rC	C2'-endo	O2' - utilized as acceptor, O2 utilized as acceptor	No
A:rG	C2'-endo	O2' - utilized as acceptor, N3 free	Yes
A:rU	C2'-endo	O2'-free; O2 utilized as acceptor	Yes
C:rA	C2'-endo	O2'-free, N3 utilized as acceptor	Yes
C:rC	C3'-endo	O2'-free; O2 utilized as acceptor	O2'-O2=3.37Å (doubtful case)
C:rG	C2'-endo	O2'-free, N3 utilized as acceptor	No
C:rU	C2'-endo	O2' - utilized as acceptor, O2- utilized as acceptor	Yes
G:rA	C2'-endo	O2' free, N3 free	Yes
G:rG	C2'-endo	O2' - utilized as donor; N3 free	No
U:rC	C2'-endo	O2' free, O2 utilized as acceptor	No
U:rG	C2'-endo	O2' - utilized as acceptor, N3 free	Yes
U:rU	C2'-endo	O2' - utilized as acceptor, O2 utilized as acceptor	Yes
T2			
A:rA	C3'-endo	O2' - utilized as donor, N3 acceptor free	No
C:rC	C2'-endo	O2'-free; O2 utilized as acceptor	Yes
C:rU	C3'-endo	O2' - utilized as acceptor, O2 utilized as acceptor	No
U:rA	C3'-endo	N3 utilized as acceptor; O2' - utilized as acceptor	No
U:rG	C2'-endo	N3 free, O2' - utilized as acceptor	Yes
T3			
A:rC	C3'-endo	O2' utilized as donor; O2 free	No
A:rG	C2'-endo	O2' free, N3 free	Yes
A:rU	C3'-endo	O2' utilized as donor; O2 blocked	No
C:rC	C2'-endo	O2'-free; O2 utilized as acceptor	Yes
U:rG	C3'-endo	N3- utilized as acceptor; O2' - utilized as acceptor	No

Table S16: Relationship between presence of intramolecular H-bond and planarity for the studied H:S cis base pairs

Base pair	Intramolecular H-bond present in crystal geometry?	Planar in crystal geometry?	Intramolecular H-bond present in optimized geometry?	Planar in optimized geometry?
Adjacent pairs (T1)				
A:rA	No	Yes	No	No
A:rC	No	No	No	Yes
A:rG	No	No	Yes	No
A:rU	No	Yes	Yes	No
C:rA	No	Yes	Yes	Yes
C:rC	No	Yes	O2'-O2=3.37Å (doubtful case)	No
C:rG	No	Yes	No	Yes
C:rU	No	Yes	Yes	No
G:rA	O2'-N3=3.33Å (doubtful case)	Yes	Yes	Yes
G:rG	No	Yes	O2'-N3=3.33Å (doubtful case)	No
U:rC	No	No	No	No
U:rG	No	Yes	Yes	No
U:rU	No	Yes	Yes	No
Near-adjacent pairs (T2)				
A:rA	No	No	No	No
C:rC	No	No	Yes	No
C:rU	No	No	No	Yes
U:rA	No	No	No	No
U:rG	No	Yes	Yes	Yes
Distant base pairs (T3)				
A:rC	No	No	No	No
A:rG	No	No	Yes	No
A:rU	No	No	No	No
C:rC	Doubtful (O2'-O2=3.38Å)	No	Yes	No
U:rG	No	No	No	Yes

^a**Green:** Internal hydrogen bond is not present, but geometry is planar; **Red:** Internal hydrogen bond is not present, and geometry is not planar. **Yellow:** Internal hydrogen bond is present, but geometry is planar. **Blue:** Internal hydrogen bond is present and geometry is planar.

Table S17: NBO and AIM analysis of bifurcated hydrogen bonding interactions in the nonprotonated geometry of G:rA.

Bond	E(2)	ρ (e/a_0^3)	$\nabla^2\rho$ (e/a_0^3)
C8-H(G)...O2'(rA)	0.87	0.0094	0.038
O2'-H(rA)...N3(G)	21.42	0.0253	0.145
C2-H(rA)...N7(G)	0.31	0.0044	0.017
C2-H(rA)...O6(G)	1.47	0.0073	0.038

Table S18: Some occurrences of triplet motif containing A:U H:W trans and U:G H:S cis base pair, as identified by BPFind.

PDB id	Base pair ids
1FJG	A448:U486:G485; A907:U891:G890; A1238:U1301:G1300; A1374:U1348:G1347
1I94	A1219:U1282:G1281; A1356:U1329:G1328
1JBR	A20(F):U11(C):G10(C); A20(D):U11(D):G10(D)
1JJ2	A160:U176:G175; A226:U214:G213; A293:U359:G358; A407:U382:G381; A476:U465:G464; A569:U589:G588; A1013:U954:G953; A912:U1293:G1292; A2054:U1371:G1370; A2010:U1972:G1971; A2702:U2693:G2692; A103(9):U79(9):G78(9)
1NBS	A140:U169:G168;
1NJP	A832:U1202:G1201; A1952:U1202:G1201;
1NKW	A832:U1202:G1201
1NWX	A832:U1202:G1201
1UN6	A100(E):U76(E):G75(E); A100(F):U76(F):G75(F)
1VQO	A160:U176:G175; A226:U214:G213; A293:U359:G358; A407:U382:G381; A476:U465:G464; A569:U589:G588; A1013:U954:G953; A912:U1293:G1292; A2054:U1371:G1370; A2010:U1972:G1971
1YHQ	A160:U176:G175; A226:U214:G213; A293:U359:G358; A407:U382:G381; A476:U465:G464; A569:U589:G588; A1013:U954:G953; A912:U1293:G1292; A2054:U1371:G1370; A2010:U1972:G1971; A2702:U2693:G2692; A103(9):U79(9):G78(9)
1YIT	A160:U176:G175; A226:U214:G213; A293:U359:G358; A407:U382:G381; A476:U465:G464; A569:U589:G588; A1013:U954:G953; A912:U1293:G1292; A2054:U1371:G1370; A2010:U1972:G1971; A2702:U2693:G2692; A103(9):U79(9):G78(9)
2E5L	A907:U891:G890; A1238:U1301:G1300; A1374:U1348:G1347
2I2P	A907:U891:G890; A1238:U1301:G1300; A1374:U1348:G1347
2I2T	A190:U206:G205; A255:U243:G242; A401:U373:G372; A470:U459:G458; A917:U860:G859; A819:U1188:G1187; A2013:A1267:U1266; A1744:U1716:G1715; A1969:U1931:G1930; A2665:U2656:G2655
2J00	A907:U891:G890; A1238:U1301:G1300; A1374:U1348:G1347
2J01	A190:U206:G205; A255:U243:G242; A401:U373:G372; A470:U459:G458; A819:U1188:G1187; A2013:A1267:U1266; A1969:U1931:G1930
2UUB	A448:U486:G485; A907:U891:G890; A1238:U1301:G1300; A1374:U1348:G1347
3BWP	A49:U59:G58
3D0U	A62:U25:G24
3D0X	A62:U25:G24
430D	A20:U11:G10

Figure S1. Base heavy atom Superposition and RMSD evaluation for crystal geometry and optimized geometry (A:A base pair in this example).

PDB id: 1JJ2

PDB id: 1JJ2

PDB id: 1FJG

PDB id: 1J5E

PDB id: 2J01

PDB id: 1JJ2

PDB id: 1JJ2

PDB id: 1GRZ

PDB id: 1JJ2

PDB id: 1NWX

Figure S2. H:S cis triplets involving interaction of sugar edge base with the major groove of the W:W cis pair. In the sixth figure, ball represents the oxygen of the interacting water atom.

Figure S3. H:S cis triplets and quartet involving interaction of Hoogsteen edge base with the minor groove of the W:W cis pair.

A:A(T2)

C:C(T2)

C:U(T2)

U:A(T2)

U:G(T2)

Figure S4. Interaction of flipped out base (in T2 base pairs) with the neighboring residues and water molecules.