Electronic Supplementary Information for

Feasibility of occurrence of different types of protonated base pairs in RNA: a quantum chemical study

Antarip Halder,[a] Sukanya Halder,[b] Dhananjay Bhattacharyya,*[b] and Abhijit Mitra*[a]

[a]Center for Computational Natural Sciences and Bioinformatics(CCNSB), International Institute of Information Technology (IIIT-H), Gachibowli, Hyderabad 500032, India
[b]Computational Science Division, Saha Institute of Nuclear Physics(SINP), 1/AF, Bidhannagar, Kolkata 700064, India

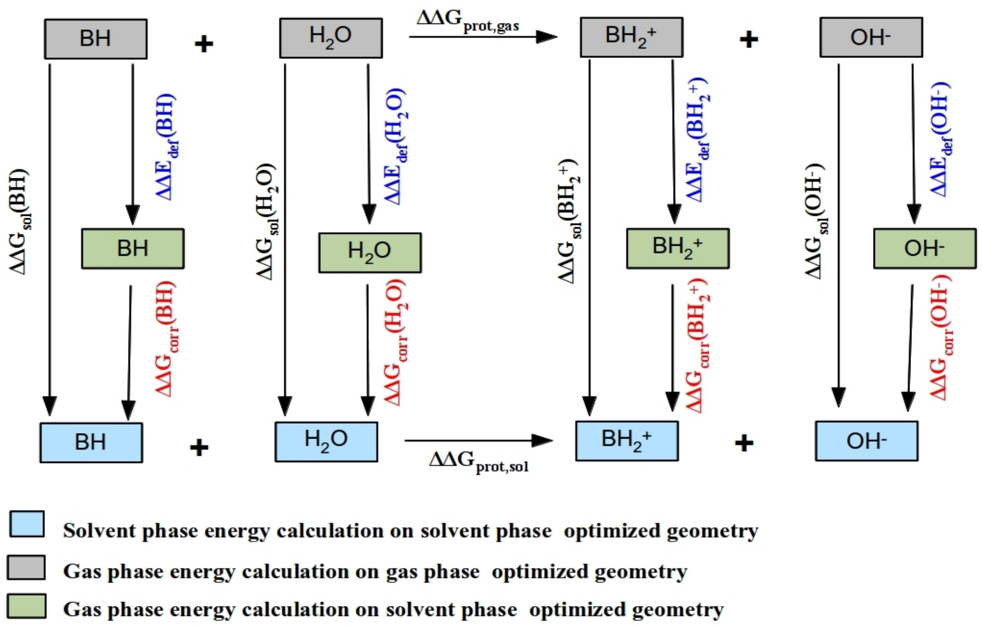
---

Contents:

1. Details of PDB files used in modeling different base pairs.

BP Geometry	**PDB Id**	**Base Pair Id**	**Resolution**
A+:C W:W Cis	402D	A105:C112	X-RAY DIFFRACTION (1.90 Å)
A+:C W:W Trans	1NKW	A2371:C2403	X-RAY DIFFRACTION (3.10 Å)
A+:G W:H Cis	1NJP	A1061:G2731	X-RAY DIFFRACTION (3.50 Å)
A+:G W:H Trans	2TRA	A46:G22	X-RAY DIFFRACTION (3.00 Å)
A+:G S:H Cis	1FJG	A1001:G1002	X-RAY DIFFRACTION (3.00 Å)
C+:C W:W Cis	1MME	C170d:C30c	X-RAY DIFFRACTION (3.10 Å)
C+:C W:W Trans	1B23	C16:C59	X-RAY DIFFRACTION (2.60 Å)
C+:U W:W Cis	1J5A	C2554:U2490	X-RAY DIFFRACTION (3.50 Å)
C+:A W:W Trans	1NKW	C2403:A2371	X-RAY DIFFRACTION (3.10 Å)
C+:A W:H Trans	1FFK	C505:A509	X-RAY DIFFRACTION (4.20 Å)
C+:G W:H Cis	1GID	C260:G108	X-RAY DIFFRACTION (2.50 Å)
C+:G W:H Trans	1DRZ	C141:G161	X-RAY DIFFRACTION (2.30 Å)
G+:C S:W Cis	1ET4	G326:C224	X-RAY DIFFRACTION (2.30 Å)
G+:G S:H Trans	1DK1	G18:G51	X-RAY DIFFRACTION (2.80 Å)
A+:G H:H Cis	4ABS	G1062:A1088	X-RAY DIFFRACTION (3.10 Å)
A+:G H:H Trans	3UXQ	A2169:G2112	X-RAY DIFFRACTION (3.20 Å)
A:T W:W Cis	1BNA	A18(A) : T7(A)	X-RAY DIFFRATCION (1.9Å)
G:C W:W Cis	1BNA	G16(A) : C9(A)	X-RAY DIFFRATCION (1.9Å)

2. Thermodynamic cycle used to calculate $\Delta\Delta G_{prot,sol}$



Free energy change of the protonation process ($\Delta\Delta G_{prot,sol}$) is calculated as follows:

$$\Delta\Delta G_{prot,sol} = \Delta\Delta G_{sol}(BH_2^+) + \Delta\Delta G_{sol}(OH^-) - \Delta\Delta G_{sol}(BH) - \Delta\Delta G_{sol}(H_2O)$$

Where, standard free energy of solvation ($\Delta\Delta G_{sol}$) has been partitioned into two parts,

$$\Delta\Delta G_{sol} = \Delta\Delta E_{def} + \Delta\Delta G_{corr}$$

$\Delta\Delta E_{def}$ is the deformation energy, which is calculated as electronic energy difference (obtained from gas phase calculations) between solvent phase optimized geometry and gas phase optimized geometry.

$\Delta\Delta G_{corr}$ is the correction term, which is the difference between the free energies calculated in gas phase and solvent phase for the solvent phase optimized geometry.

3. Details of calculation of interaction energy.

For the base pairs optimized at B3LYP/6-31G++(2d,2p) level of theory we calculated the single point interaction energy at MP2/aug-ccpVDZ level. The interaction energy (ΔE_{AB}) of a base pair AB formed by the individual bases A and B is defined as,

$$\Delta E_{AB} = E_{AB} - E^0_A - E^0_B$$

where E_{AB} is the total energy of the optimized base pair AB and E^0_A and E^0_B are the total energies of the individual bases A and B, in their optimized geometries respectively. This interaction energy was further corrected for Basis Set Superposition Error (BSSE) and deformation energy ($E_{def(AB)}$). BSSE correction of the interaction energy (E^{BSSE}) was done by using standard counterpoise calculations. The deformation energy is represented as,

$$(E_{def(AB)}) = (E^{AB}_A - E^0_A) + (E^{AB}_B - E^0_B)$$

where, E^{AB}_A and E^{AB}_B are the energies of the bases A and B respectively in the optimized geometry of AB. So the total corrected interaction energy (E^{gas}_{int}) is calculated as,

$$E^{gas}_{int} = \Delta E_{AB} + E^{BSSE} + E_{def(AB)}$$

4. Vibrational frequency analysis of hydrogen bond stretching

Compared to N—H bonds in canonical base pairs (A:T W:W Cis and G:C W:W Cis), in protonated base pairs, red shift in corresponding N(+)—H bond stretching frequency due to hydrogen bond formation is 3-4 times higher for both the N-H --- O and N-H --- N type hydrogen bonds. Influence of protonation (at a ring nitrogen atom) on the other functional groups of the base may be realized from the following observations:

1. N1 protonation of Adenine has enhanced the red shift in N6-H --- O bonds in system 4 and system 5 compared to the N6-H --- O4 bond in canonical AT pair, implying the fact that protonation at ring nitrogen atoms positively influences the hydrogen bond donor sites.
2. N3 protonation of Cytosine positively influences the hydrogen bond donor potential of N4 site (system 9, system 10 and system 13 show remarkable increase in red shift of N4-H --- O bond stretching frequencies and corresponding intensities compared to N4-H --- O6 bond in canonical GC pair) and negatively influences hydrogen bond acceptor potential at O2 site (red shift in N-H --- O2 bond stretching frequencies undergoes a significant decrease in the systems 9,10,11 and 12 compared to N2-H --- O2 bond in canonical GC pair).

Table S1: Shifts in IR frequency corresponding to symmetric stretch of D−H bonds which participate in D−H···A type hydrogen bond formation.*

Sl. No.	Base Pair Geometry	Prot. Site	Hydrogen Bond	ν in cm-1	ν0 in cm-1	Red shift in cm-1	I KM/mol	I0 KM/mol	I/I0
1	A:T W:W Cis		N6-H --- O4	3497.04	3680.70	183.66	948.63	132.62	7.15
			C2-H --- O2	3233.71	3251.29	17.58	1.47	18.64	0.08
			N1 --- H-N3	2972.63	3663.74	691.11	2543.28	93.32	27.55
2	G:C W:W Cis		N1-H --- N3	3301.90	3669.03	367.13	2193.20	78.10	28.08
			N2-H --- O2	3479.75	3655.71	175.96	1235.34	71.89	17.18
			O6 --- H-N4	3259.52	3685.51	425.99	183.91	121.55	1.51
3	A+:C W:W Cis	N1	N6-H --- N3	3254.54	3577.24	322.70	1757.78	186.63	9.42
			N1-H(+) --- O2	2872.62	3585.52	712.90	2565.32	153.31	16.73
4	A+:C W:W Trans	N1	N6-H --- O2	3207.63	3577.24	369.61	448.96	186.63	2.41
			N1-H(+) --- N3	3045.54	3585.52	539.98	1395.79	153.31	9.10
5	A+:G W:H Cis	N1	N6-H --- O6	3196.42	3577.24	380.82	2727.51	186.63	14.61
			N1-H(+) --- N7	2925.36	3585.52	660.16	2336.12	153.31	15.24
6	A+:G W:H Trans	N1	N6-H --- N7	3140.34	3577.24	436.90	3522.90	186.63	18.88
			N1-H(+) --- O6	2997.17	3585.52	588.35	1647.68	153.31	10.75
7	A+:G S:H Cis	N3	N3-H(+) --- N7	2877.40	3594.67	717.27	2860.26	93.82	30.49
			C2-H --- O6	3188.30	3210.40	22.10	217.20	5.21	41.66
8	C+:C W:H Cis	N3	N4-H --- N3	3192.43	3587.37	394.94	2262.80	266.97	8.48
			N3-H(+) --- O2	2931.25	3570.44	639.19	1746.80	82.77	21.10
9	C+:C W:H Trans	N3	N4-H --- O2	2917.29	3587.37	670.08	3335.95	266.97	12.50
			N3-H(+) --- N3	2709.51	3570.44	860.93	1303.96	82.77	15.75

			O2 --- H-N4	3451.90	3599.04	147.14	769.83	70.05	10.99
10	C+:U W:W Cis	N3	N3-H(+) --- O4	2556.38	3570.44	1014.06	3438.81	82.77	41.55
			O2 --- H-N3	3438.63	3605.27	166.64	547.94	66.51	8.24
11	C+:A W:W Trans	N3	N3-H(+) --- N1	1943.64	3570.44	1626.80	4876.60	82.77	58.92
			O2 --- H-N6	3507.71	3605.48	97.77	444.19	85.40	5.20
12	C+:A W:H Trans	N3	N3-H(+) --- N7	2368.29	3570.44	1202.15	3776.14	82.77	45.62
			O2 --- H-N6	3523.52	3605.48	81.96	259.15	85.40	3.03
13	C+:G W:H Cis	N3	N4-H --- O6	3110.70	3587.37	476.67	2489.66	266.97	9.33
			N3-H(+) --- N7	2652.92	3570.44	917.52	2856.17	82.77	34.51
14	C+:G W:H Trans	N3	N4-H --- N7	3077.51	3587.37	509.86	3718.52	266.97	13.93
			N3-H(+) --- O6	3077.51	3570.44	492.93	3718.52	82.77	44.93
15	G+:C S:W Cis	N3	N2-H --- N3	3185.54	3597.93	412.39	2061.78	251.76	8.19
			N3-H(+) --- O2	2869.71	3617.23	747.52	1949.11	110.86	17.58
16	G+:G S:H Trans	N3	N2-H --- N7	3087.76	3597.93	510.17	4309.20	251.76	17.12
			N3-H(+) --- O6	3087.76	3617.23	529.47	4309.20	110.86	38.87

*ν0 and I0 represent harmonic frequencies (in cm^{-1}) and IR intensities (in KM/mol), respectively, of the unpaired monomers. ν and I represent harmonic frequencies and IR intensities, respectively, of the paired systems.

5. ESP and Mulliken partial charges analysis

Table S2: ESP charges at different functional sites of the four RNA bases for their neutral and charged states.

Base	**Charged State**	**Functional Sites of Nucleobases**				
		N1	**N3**	**N7**	**N6**	
Adenine	Neutral	-0.744	-0.75	-0.558	-0.795	
	N1+	-0.541	-0.532	-0.513	-0.794	
	N3+	-0.588	-0.513	-0.518	-0.767	
	N7+	-0.725	-0.712	-0.221	-0.984	
		O2	**N3**	**N4**		
Cytosine	Neutral	-0.624	-0.796	-0.839		
	N3+	-0.472	-0.588	-0.943		
		N3	**N7**	**O6**	**N1**	**N2**
Guanine	Neutral	-0.713	-0.494	-0.578	-0.762	-0.773
	N3+	-0.533	-0.439	-0.426	-0.621	-0.934
	N7+	-0.748	-0.169	-0.506	-0.719	-0.976
	N1-	-0.88	-0.537	-0.738	-0.913	-0.842
		O2	**O4**	**N3**		
Uracil	Neutral	-0.573	-0.584	-0.614		
	N3-	-0.743	-0.783	-0.963		

Table S3: Mulliken charges at different functional sites of the four RNA bases for their neutral and charged states.

Base	Charged State	Functional Sites of Nucleobases				
		N1	**N3**	**N7**	**N6**	
Adenine	Neutral	-0.294	-0.324	-0.343	-0.250	
	N1+	-0.066	-0.197	-0.295	-0.165	
	N3+	-0.196	-0.101	-0.264	-0.135	
	N7+	-0.206	-0.240	-0.075	-0.208	
		O2	**N3**	**N4**		
Cytosine	Neutral	-0.403	-0.380	-0.236		
	N3+	-0.272	-0.147	-0.151		
		N3	**N7**	**O6**	**N1**	**N2**
Guanine	Neutral	-0.446	-0.275	-0.386	-0.206	-0.309
	N3+	-0.214	-0.200	-0.240	-0.189	-0.204
	N7+	-0.390	-0.014	-0.330	-0.201	-0.219
	N1-	-0.569	-0.300	-0.496	-0.435	-0.331
		O2	**O4**	**N3**		
Uracil	Neutral	-0.388	-0.356	-0.210		
	N3-	-0.517	-0.495	-0.454		

6. NUPARM parameters for optimized geometry of protonated base pairs.

Table S4: NUPARM parameters calculated over B3LYP/6-31G++(2d,2p) level optimized geometries of the protonated base pairs.

Sl. No.	Base Pair	Buckle	Open	Propel	Stagger	Shear	Stretch
1	A+:C W:W Cis	0.05	10.40	0.06	0.00	-2.62	2.80
2	C+:C W:W Cis	3.42	9.49	-7.04	0.02	-2.76	2.77
3	A+:G W:H Cis	0.04	-1.91	0.00	0.00	0.41	2.83
4	A+:G S:H Cis	1.85	93.00	1.46	-0.03	4.91	0.81
5	C+:G W:H Trans	0.23	-2.00	0.06	0.00	2.93	2.77
6	C+U W:W Cis	0.00	-15.33	0.00	0.00	2.30	2.91
7	C+:C W:W Trans	-0.02	-4.79	0.05	0.00	-0.14	2.83
8	G+:G S:H Trans	4.12	115.55	-5.03	0.12	-3.09	4.06
9	C+:G W:H Cis	-0.15	-1.02	-0.02	0.00	0.26	2.75
10	A+:G W:H Trans	-0.70	-4.34	1.89	0.00	2.75	2.80

7. Influence of protonated base pairing on the local aromaticity of the base rings.

Table S5: NICS(1) values of the pyrimidine (Pyr.) and imidazole (Imi.) rings of the bases participating in protonated base pairing have been compared with their NICS(1) values when they participate in formation of canonical GC and AT/U base pairs. Intrinsic interaction energies of the protonated base pairs calculated at MP2/aug-cc-pVDZ//B3LYP/6-31G++(2d,2p) level of theory are also given.

Sl. No.	Base Pair	Prot. Site	Base	Ring	NICS(1) in		Difference
					Protonated BP	Canonical BP	
1	A(+):C W:W Cis (-32.1 kcal/mol)	N1	A(+)	Pyr.	-6.80	-7.50	-0.70
			A(+)	Imi.	-9.87	-9.54	0.33
			C	Pyr.	-3.46	-2.65	0.81
2	C(+):C W:W Cis (-35.2 kcal/mol)	N3	C(+)	Pyr.	-2.24	-2.65	-0.41
			C	Pyr.	-3.96	-2.65	1.31
3	A(+):G W:H Cis (-34.5 kcal/mol)	N1	A(+)	Pyr.	-6.32	-7.50	-1.18
			A(+)	Imi.	-9.81	-9.54	0.27
			G	Pyr.	-3.75	-3.50	0.25
			G	Imi.	-8.75	-8.62	0.13
4	A(+):G S:H Cis (-30.1 kcal/mol)	N3	A(+)	Pyr.	-6.96	-7.50	-0.54
			A(+)	Imi.	-9.75	-9.54	0.21
			G	Pyr.	-3.37	-3.50	-0.13
			G	Imi.	-8.69	-8.62	0.07
5	C(+):G W:H Trans (-36.4 kcal/mol)	N3	C(+)	Pyr.	-2.12	-2.65	-0.53
			G	Pyr.	-3.89	-3.50	0.39
			G	Imi.	-8.65	-8.62	0.03
6	C(+):U W:W Cis (-26.3 kcal/mol)	N3	C(+)	Pyr.	-2.71	-2.65	0.06
			U	Pyr.	-2.19	-2.14	0.05
7	C(+):C W:W Trans (-42.5 kcal/mol)	N3	C(+)	Pyr.	-2.35	-2.65	-0.30
			C	Pyr.	-3.18	-2.65	0.53
8	G(+):G S:H Trans (-40.7 kcal/mol)	N3	G(+)	Pyr.	-2.30	-3.50	-1.20
			G(+)	Imi.	-9.28	-8.62	0.66
			G	Pyr.	-4.08	-3.50	0.58
			G	Imi.	-8.65	-8.62	0.03
9	C(+):G W:H Cis (-39.9 kcal/mol)	N3	C(+)	Pyr.	-2.35	-2.65	-0.30
			G	Pyr.	-3.95	-3.50	0.45
			G	Imi.	-8.68	-8.62	0.06
10	A(+):G W:H Trans (-37.6 kcal/mol)	N1	A(+)	Pyr.	-6.83	-7.50	-0.67
			A(+)	Imi.	-9.31	-9.54	-0.23
			G	Pyr.	-4.03	-3.50	0.53
			G	Imi.	-8.58	-8.62	-0.04

8. **Analysis HOMO and LUMO orbitals of protonated base pairs.**

Fig. S1: Distribution of HOMO and LUMO orbitals of protonated base pairs as obtained from NBO analysis of B3LYP/6-31G++(2d,2p) optimized geometry.