PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Physical Chemistry Chemical Physics
CREATED USING THE RSC LaTeX PCCP ARTICLE TEMPLATE – SEE www.rsc.org/electronicfiles FOR DETAILS
ARTICLE TYPE
www. rsc.org/xxxxxx | XXXXXXXX

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]

Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
First published on the web Xth XXXXXXXXXX 200X
DOI: 10.1039/b000000x

Protonated nucleobases have significant roles in facilitating catalytic functions of RNA, and in stabilizing different structural motifs. Reported pK_a values of nucleobase protonation suggest that the population of neutral nucleobases is 10^3 – 10^4 times higher than that of protonated nucleobases under physiological conditions (pH $\sim$ 7.4). Therefore, a molecular level understanding of various putative roles of protonated nucleobases cannot be achieved without addressing the question of how their occurrence propensities and stabilities are related to the free energy costs associated with the process of protonation under physiological conditions. With water as proton donor, we use advanced QM methods to evaluate the site specific protonation propensities of nucleobases in terms of their associated free energy changes (ΔG_{prot}). Quantitative follow up on the energetics of base pair formation and database search for evaluating their occurrence frequencies, reveal a lack of correlation between base pair stability and occurrence propensities on the one hand, and ease of protonation on the other. For example, although N7 protonated Adenine (ΔG_{prot} = 40.0 kcal/mol) is found to participate in stable base pairing, base pairs involving N7 protonated Guanine (ΔG_{prot} = 36.8 kcal/mol), on geometry optimization, converge to a minima where Guanine transfers its extra proton to its partner base. Such observations, along with examples of weak base pairs involving N3 protonation of Cytosine (ΔG_{prot} = 37.0 kcal/mol) are rationalized by analysing the protonation induced charge redistributions which are found to significantly influence, both positively and negatively, the hydrogen bonding potentials of different functional sites of individual nucloebases. Protonation induced charge redistribution is also found to strongly influence (i) the aromatic character of the rings of the participating bases and (ii) hydrogen bonding potential of the free edges of the protonated base pair. Comprehensive analysis of a non-redundant RNA crystal structure dataset further reveals that, while availability of stabilization possibilities determine the feasibility of occurrence of protonated bases, their occurrence context and specific functional roles are important factors determining their occurrence propensities.

Introduction

While reports of RNA molecule, with newer functionalities, continue to appear in increasing numbers,[1–7] the basic question of how these molecules can display such complex structural and functional diversity continues to demand a satisfactory and comprehensive answer.[8] Just as in the case of proteins, where the presence of charged amino acids with acidic or basic side chains have been associated with catalysis related functionalities[9] (*e.g.*, lysine and arginine in oxyanion hole formation, hystidine in general acid-base catalysis, *etc.*), charged (protonated or deprotonated) nucleobases are also expected to be associated with RNA functionalities. A detailed understanding of (de/)protonation propensities of RNA bases, and their stabilities, is therefore one of the necessary requirements for investigating the functional diversity of RNA molecules.

A major issue that needs to be addressed in this connection is the fact that, unlike the charged amino acids which, because of their pKa values, are expected to be protonated or deprotonated, the same is not true for nucleotides under physiological conditions (pH $\sim$ 7.4).[10,11] For nucleobases in single stranded unfolded RNA, the pK_{a1} values of Adenine, Guanine and Cytosine are 3-4 units and pK_{a2} values of Guanine and Uracil are $\sim$2 units away from neutrality.‡ Though, they have been occasionally inferred on the basis of circumstantial evidences, instances of participation of deprotonated nucleobases are rare in nucleic acid literature.[13–16] However, reported instances of the involvement of protonated RNA bases are more numerous in the literature. For example, a noncanon-

† Electronic Supplementary Information (ESI) available: Initial geometries, thermodynamic cycle for $\triangle\triangle G_{prot,sol}$ calculation, interaction energy calculation procedure, vibrational frequency analysis, ESP and Mulliken partial charge analysis, NUPARM parameters, NICS values and HOMO-LUMO for protonated base pairs. See DOI: 10.1039/b000000x/

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

‡ Adenine (pK_{a1} = $\sim$4.1), Cytosine(pK_{a1} = $\sim$4.4) and Guanine(pK_{a1} = $\sim$3.2, pK_{a2} = $\sim$9.2), Uracil (pK_{a2} = $\sim$9.2)[10,12]

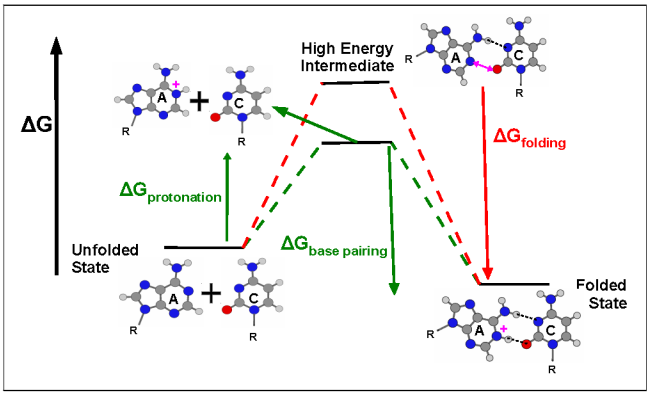

Fig. 1 Free energy diagram showing coupling between RNA folding and RNA base protonation. The model is conceptual and the actual values of free energy changes (ΔG) are unknown.

ical base pairing between N3 protonated Cytosine and Hoogsteen edge of Guanine has been invoked in the formation of DNA triple helices,[17,18] stacked C+:C base pairs are involved in the stabilization of i-DNA quadruplex motif,[19,20] the tuning of adenosine deaminase (ADAR) mediated RNA editing process has been explained via A+:C wooble base pairs involving N1 protonation of Adenine,[21] the protonation of C75 residue has been implied in the active site of HDV ribozyme,[22] etc. Among the protonated nucleobases, Class I nucleobases form base pairs via the loaded proton and participate in catalysis *e.g.* by participating in oxyanion hole formation (lysine-arginine type role). On the other hand, Class II nucleobases are not paired and therefore, can participate in general acid base catalysis (hystidine like role).[10] The importance of charged nucleic acids in enzymatic activities has also been demonstrated via studies on DNAzymes by Perrin *et al.* which suggest that rates of RNA enzyme mediated catalysis are comparable to those of protein enzymes when the functional groups of the bases involved are protonated.[22]

Such examples imply significant changes in the relative population of neutral and protonated nucleobases, which cannot be explained in terms of their normal pK_a values as reported for these bases in unfolded single stranded RNAs and, therefore, implies a shift of pK_{a1} values towards neutrality. A major paradigm invoked in explaining such pK_a shifts relates to variation in the local environment associated with RNA folding. When an RNA folds, certain nucleobases may enter into pockets where the neighboring electrostatics provides the basis for substantial pK_a shifts which may lead to protonation under physiological condition (red line in Fig. 1). Be that as it may, we are interested in assessing molecular level factors, such as proton mediated hydrogen bonding and other factors including stacking, electrostatics, *etc.*, which can explain not only the occurrence of certain protonated bases, but also the relative abundances in known structures. Of the different driving forces, which may lead to the formation of globally stable folded structures involving protonated bases, base pairing (as in Class I) constitutes possibly the most important factor (green line in Fig. 1). In other words, $\Delta G_{folding}$ is dominated by $\Delta G_{base-pairing}$. It is therefore expected that, the less positive the value of $\Delta G_{protonation}$ and the more negative the value of $\Delta G_{base-pairing}$, the more probable will be the occurrence of the corresponding protonated nucleobases in RNA crystal structures.

Earlier, we have carried out database analysis of occurrence of Class I protonated base pairs and have reported their optimized geometries and intrinsic interaction energies using quantum chemical computations.[23] Given the lack of hydrogen atom coordinates in X-ray crystal structures, and the known issues associated with unambiguous characterization of exchangeable protons from NMR structures, detection of protonated base pairs from structures constitute a major challenge. We had used BPFind software[24] (modified for detecting all possible protonated base pairs), which adopts an *in silico* hypothesis driven approach for analysing crystal structures, to address this challenge and had reported 18 distinct protonated base pairs[§] involving two or more hydrogen bonds, one of which involves the extra proton of the protonated base.[23]

In this study, we have used advanced QM methods to calculate the protonation propensities at different sites of the four RNA bases and to assess the stabilities of base pairs involving them. Stability of a base pair has thus been determined by calculating the intrinsic interaction energy of the base pair which is the difference between the ground state electronic energies of the base pair and its "infinitely separated" monomers. Further, we have calculated the free energy of formation of the neutral nucleobases and corresponding charged species. Finally, the protonation propensity at specific sites of individual nucleobases was determined by the free energy changes of the modeled protonation processes involving nucleobases protonated at that particular site, respectively. Comparison of these energetics data with the occurrence frequencies of protonated base and base pairs reveals that occurrence frequency of protonated base pairs are not simply determined by the ease of protonation at specific sites of the nucleobases, nor 'only' by the stabilities of the base pairs that they may form. Rather, we have found that, protonation induced charge redistribution increases the potential of protonated base pairs for participating in higher order structures (e.g., base triples, stacked base pairs, *etc.*). Of course, apart from these considerations which are related to the thermodynamic feasibility of the protonation of bases, one may not undermine the importance of the

§ A larger collection of structures of protonated base pairs, both naturally occurring and modeled, are available in the "RNA Base Pair Database" (http://www.saha.ac.in/biop/www/db/local/BP/rnabasepair.html) and "RNABP COGEST" database (http://bioinf.iiit.ac.in/RNABPCOGEST/)

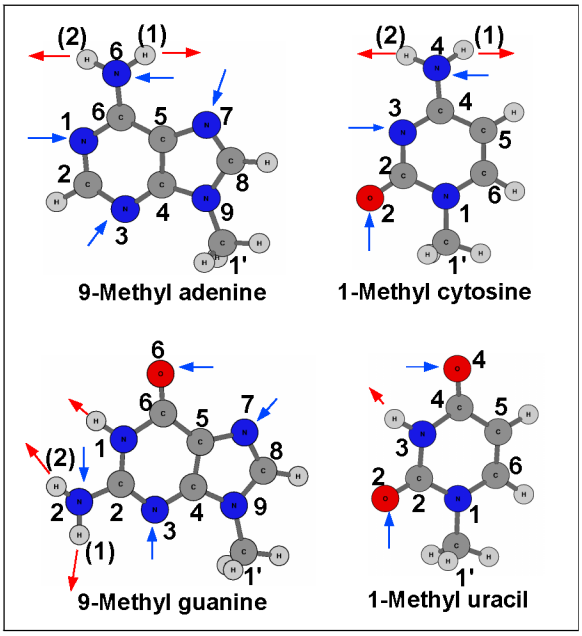

Fig. 2 Modelled bases with sugar and backbone replaced by methyl group. The blue (red) arrows indicate the possible sites of protonation (deprotonation) considered in this study.

context of occurrence and functional role of structural motifs involving protonated bases, as a major determinant of their occurrence propensities in RNA structure.

Methods

Modelling

Coordinates of four different nucleotides of RNA were extracted from appropriate PDB structures. Initial structural models of RNA bases were made by substituting the sugar-phosphate moiety by a methyl group corresponding to C1′ of sugar. Omission of sugar phosphate backbone has been shown to reduce the computational effort without significantly affecting the properties of the real systems.[25,26] Moreover, *ab initio* calculations as well as experimental results have established that for a nucleotide, the nucleobase is the origin of the lowest energy ionization pathway.[27,28] Fig. 2 shows the modelled bases in their neutral state and possible sites of protonation and deprotonation which were considered in this study. Charged bases were modelled by adding/removing a hydrogen atom at/from the site of protonation/deprotonation of the neutral bases respectively. Similarly, while studying the base pairs, the coordinates of the interacting nucleobases were extracted from appropriate PDB files[†] and modelled accordingly. GaussView5[29] package was used for editing the molecules.

Computational Details

Quantum mechanical calculations were mainly carried out using Gaussian03[30] package. The Gaussian09[31] package was used in some selected cases. Considering the recent report[32] on success of the nonlocal hybrid Becke three-parameter Lee-Yang-Parr (B3LYP)[33,34] functional in estimating (with good correlation with experimental values) the intrinsic interaction energy of protonated base pairs, we used the same with 6-31G++(2d,2p) basis set for ground state geometry optimization of the modelled systems (neutral and charged bases and base pairs) in gas phase. Earlier it has been shown that,[35] results of computations using B3LYP with a split valance double ζ augmented with (i) a d type polarization function for all non-hydrogen atoms and (ii) a p type polarization function for hydrogen atoms, and also including an s-p and p-d diffused orbitals compare very well with reference RIMP2/cc-pVTZ data.[36] Considering the fact that B3LYP functional has issues in tackling long range correlations,[81] we have calculated the dispersion corrections for the B3LYP/6-31G++(2d,2p) optimized geometry for some of the systems, where the middle range dispersion interactions might play a significant role. We have used Grimme's DFT-D3 formalism[82] with B3LYP functional and zero damping force to calculate the dispersion energy (E_{disp}). DFT generated Kohn-Sham orbitals sometimes do not give the real picture of the molecular orbitals. Hence, for analysis of ground state molecular orbitals, the second order Møller-Plesset perturbation theory was used with a larger, augmented correlation-consistent polarized valence only double ζ (aug-cc-pVDZ), basis set[37]

To have a better correspondence with experimental environment we have incorporated the solvent effects via the computationally efficient method of implicit solvation which involves the representation of solvent as a continuous medium instead of individual explicit solvent molecules. Bulk solvation was included in the calculations through the Conductor-Like Polarizable Continuum Model (CPCM)[38,39] which is an implementation of COSMO in the polarizable continuum model (PCM) framework as implemented in Gaussian03. This was found to be more appropriate for polar liquids, where the electrostatic potential goes to zero on the surface. It is considered to be computationally extremely efficient and robust, and to be less sensitive to outlying charge error.[40] Vibrational frequency analysis on optimized geometries in both gas and solvent phase showed 3N-6 real frequencies in all cases.

Analysis of charge distribution in the neutral and charged systems were carried out based on (i) NBO charges, obtained by performing a Natural Bond Orbital (NBO) analysis[41] on the optimized geometry in gas phase, (ii) Electrostatic Potential (ESP) surfaces (mapped over the total electron density) and (iii) Mulliken partial charges. Following earlier studies,[42,43] we quantified the effect of (de/)protonation on the local aro-

maticity of the six member and five member rings of the nucleobases, by calculating the Nuclear Independent Chemical Shift (NICS)[44] values (the negative of the absolute magnetic shielding in ppm unit) of the rings at 1Å above the center of the ring (NICS(1)), using the GIAO (Gauge Invariant Atomic Orbital)[45,46] method at the B3LYP/6-31G++(2d,2p) level. Negative value of NICS is quantitatively related to aromaticity: more negative the value greater the aromatic character of the ring. Since, NICS(0) values (calculated at the geometrical center of the ring) have been shown to be affected by local contributions of the σ framework,[47] we rely on NICS(1) values to study the local π aromaticity of the 6 member (pyrimidine)and 5 member (imidazole) rings.

Intrinsic stability of protonated base pairs (optimized in gas phase at DFT level of theory with B3LYP functional) were calculated using MP2/aug-cc-pVDZ level of theory, using a well established technique,[23,35] as the difference between the energies of the complex minus the energies of the individual interacting bases. The interaction energies were also corrected for Basis Set Superposition Error (BSSE) and deformation energy correction at the same level of theory. Details of the procedure is explained under section 3 of Supplementary Information.†

Intra base pair parameters are those parameters which determine the relative spatial orientation of the constituent bases in a base pair. According to the IUPAC-IUB convention, there are three rotational and three translational intra-base pair parameters - Buckle, Open-angle, Propeller, Stagger, Shear and Stretch.[48] We have used the standalone NUPARM package[49] to measure the intra base pair parameters.

Estimating the ease of protonation

The ease of protonation had been characterized by computing the free energy change associated with the process of protonation. The standard Gibbs free energy of a system in gas phase in its standard state (ideal gas at 1 atm and 298 K) was obtained from, $\triangle G^0_{gas} = E_{0K} + ZPE + \triangle\triangle G_{0\rightarrow 298K}$, where, the total energy of the system at 0 K (E_{0K}) was calculated at its optimum geometries and the zero-point energy (ZPE) and Gibbs free energy change from 0 to 298K at 1atm ($\triangle\triangle G_{0\rightarrow 298K}$) were given by vibrational frequency analysis. Translational and rotational free energy contributions were also calculated within the ideal gas approximation. Therefore, the lower the value of $\triangle\triangle G_{(de/)protonation,gas}$ for (de/)protonation at a specific site, the higher is the (de/)protonation propensity of that particular site.¶ To understand the effect of bulk solvation on the (de/)protonation propensity of different sites $\triangle\triangle G_{(de/)protonation,sol}$ was calculated following the thermodynamic cycle described by Verdolino et al.[12] and used by other groups.[50] Details of the procedure is explained under section 2 of Supplementary Information.†

¶ $\triangle\triangle G_{(de/)protonation,gas} = \sum_{products} \triangle G^0_{gas} - \sum_{reactants} \triangle G^0_{gas}$

Table 1 Change in free energy ($\triangle\triangle G_{(de/)prot,gas}$) and enthalpy ($\triangle\triangle H_{(de/)prot,gas}$) in gas phase of the process of (de/)protonation (in kcal/mol) calculated following direct addition/removal of proton from the neutral nucleobase.

Base	Charged State	$\triangle\triangle G_{gas}$		$\triangle\triangle H_{gas}$	
		B3LYP	MP2	B3LYP	MP2
Adenine	$N1^+$	-221.6	-216.3	-229.3	-223.7
	$N3^+$	-219.1	-213.1	-227.7	-221.2
	$N6^+$	-198.7	-197.8	-205.9	-204.9
	$N7^+$	-215.6	-210.2	-223.4	-217.3
	$N6(1)^-$	349.5	343.8	356.8	351.4
	$N6(2)^-$	350.8	343.8	356.5	351.4
Cytosine	$O2^+$	-216.3	-212.6	-224.2	-220.4
	$N3^+$	-224.8	-219.4	-232.5	-227.0
	$N4^+$	-192.8	-192.7	-200.2	-200.0
	$N4(1)^-$	342.6	337.6	350.7	345.6
	$N4(2)^-$	347.4	342.5	355.9	350.5
Guanine	$N2^+$	-184.7	-185.1	-192.2	-192.4
	$N3^+$	-208.3	-203.6	-215.5	-210.5
	$O6^+$	-220.5	-213.8	-227.2	-220.9
	$N7^+$	-226.7	-220.8	-234.3	-228.0
	$N1^-$	332.8	327.2	340.3	334.8
	$N2(1)^-$	331.0	333.1	339.4	340.9
	$N2(2)^-$	332.9	328.6	338.8	336.2
Uracil	$O2^+$	-192.9	-190.5	-200.4	-197.9
	$O4^+$	-200.5	-196.2	-208.5	-204.2
	$N3^-$	339.9	335.4	347.6	343.0

RNA crystal structure database analysis

For the purpose of RNA crystal structure database search, we selected HD-RNAS database[51] which provides us with the complete dataset along with a non-redundant dataset of available crystal structures of RNA. The non-redundant dataset contains only those structures of RNA which has at least one, 30 nucleotide or longer, chain and has a resolution of 3.5Å or better. The complete dataset is also compiled by applying chain length and resolution cut-offs as filters. However, it has an over representation of those molecules which have been studied more extensively, and hence is prone to adding bias to statistical analyses of occurrences.

Results and Discussions

Modeling (de/)protonation: conventional approaches

A large diversity of approaches have been adopted in literature to model the process of (de/)protonation in the con-

Table 2 Change in free energy ($\triangle\triangle G_{(de/)prot,gas}$) and enthalpy ($\triangle\triangle H_{(de/)prot,gas}$) in gas phase of the process of (de/)protonation (in kcal/mol). Calculations are performed considering different proton donors (MH).

Base		MH = Water				MH = Formic Acid				MH = Acetic Acid			
		$\Delta\Delta G_{gas}$		$\Delta\Delta H_{gas}$		$\Delta\Delta G_{gas}$		$\Delta\Delta H_{gas}$		$\Delta\Delta G_{gas}$		$\Delta\Delta H_{gas}$	
		B3LYP	MP2	B3LYP	MP2	B3LYP	MP2	B3LYP	MP2	B3LYP	MP2	B3LYP	MP2
Adenine	$N1^+$	161.6	161.8	160.5	161.0	112.9	116.1	112.7	116.2	116.5	119.4	117.3	120.2
	$N3^+$	164.0	164.9	162.1	163.5	115.4	119.3	114.3	118.7	119.0	122.6	118.8	122.7
	$N6^+$	184.5	180.3	183.9	179.8	135.8	134.6	136.1	135.0	139.4	137.9	140.6	139.0
	$N7^+$	167.6	167.9	166.4	167.4	118.9	122.2	118.6	122.6	122.5	125.5	123.1	126.7
	$N6(1)^-$	192.2	188.5	192.6	189.3								
	$N6(2)^-$	193.5	188.6	192.4	189.3								
Cytosine	$O2^+$	166.9	165.5	165.5	164.3	118.2	119.8	117.8	119.5	121.8	123.1	122.3	123.6
	$N3^+$	158.4	158.7	157.3	157.7	109.7	113.1	109.5	112.9	113.3	116.3	114.0	116.9
	$N4^+$	190.4	185.4	189.6	184.7	141.7	139.8	141.8	139.9	145.3	143.0	146.3	143.9
	$N4(1)^-$	185.3	182.3	186.6	183.5								
	$N4(2)^-$	190.1	187.2	191.7	188.4								
Guanine	$N1^+$	189.3	190.9	190.6	192.1	140.7	145.3	142.8	147.3	144.2	148.5	147.3	151.3
	$N2^+$	198.4	193.0	197.6	192.3	149.8	147.4	149.8	147.5	153.4	150.6	154.4	151.5
	$N3^+$	174.9	174.5	174.3	174.2	126.3	128.8	126.5	129.4	129.8	132.1	131.0	133.5
	$O6^+$	162.7	164.2	162.5	163.8	114.1	118.6	114.8	119.0	117.6	121.8	119.3	123.0
	$N7^+$	156.5	157.3	155.4	156.7	107.9	111.6	107.7	111.9	111.4	114.9	112.2	115.9
	$N1^-$	175.42	171.9	176.16	172.7								
	$N2(1)^-$	173.7	177.9	175.3	178.8								
	$N2(2)^-$	175.5	173.3	174.7	174.1								
Uracil	$O2^+$	190.3	187.6	189.4	186.8	141.6	142.0	141.6	142.0	145.2	145.2	146.1	146.0
	$N3^+$	209.5	205.9	209.5	205.7	160.8	160.2	161.8	160.9	164.4	163.5	166.3	164.9
	$O4^+$	182.7	181.8	181.3	180.5	134.0	136.2	133.5	135.7	137.6	139.5	138.0	139.7
	$N3^-$	182.6	180.1	183.5	181.0								

text of calculations of gas phase basicity and proton affinity of nucleobases (Adenine[52–55], Guanine[12,54,56], Cytosine[57–59] and Uracil[60,61]) and their derivatives. These conventional approaches may be categorized primarily into two classes:

1. Direct addition of a proton (H^+) to a neutral species BH.

$$BH + H^+ \longrightarrow BH_2^+ \quad (1)$$

2. Considering a neutral molecule MH that acts as a proton donor which protonates the neutral species BH to BH_2^+ and itself gets deprotonated to M^-.‖

$$BH + MH \longrightarrow BH_2^+ + M^- \quad (2)$$

The former approach considers, (i) the stabilization of a free proton by its association with a base in the process of protonation and (ii) creation of two free charges from a neutral species in the process of deprotonation. As expected, the first process is associated with a high negative value of free energy change ($\triangle\triangle G_{prot,gas}$) and a high positive value for the second process [Table 1]. Although, the negative values of $\triangle\triangle S_{prot,gas}$ associated with the protonation process increase $\triangle\triangle G_{prot,gas}$ values, the high negative values of $\triangle\triangle H_{prot,gas}$ because of charge delocalization in the product (BH_2^+) is the major factor leading to the 'nonintuitive' large negative values of free energy change [Table 1]. The essential issue with this approach is that it does not consider the free energy of formation of a proton from any proton donor.** This is taken care of in the the latter approach which reflects the thermodynamic barrier characterized by high positive value for $\triangle\triangle G_{(de/)prot,gas}$ associated with $\triangle\triangle S_{prot,gas} < 0$ and, hence, is physicochemically more relevant, though highly dependent on the choice of the proton donor/acceptor (MH) [Table 2]. Calculations with proton donors of different acidic strength (water, formic acid and

‖ In the first approach, direct removal of proton from a neutral species will constitute the deprotonation process, $BH \longrightarrow B^- + H^+$. In the second approach, deprotonation process may be represented as, $BH + MH \longrightarrow B^- + MH_2^+$.

** We have considered, $\triangle H^0_{gas}$ (H^+) = 2.5 RT = 1.48 kcal/mol and $\triangle G^0_{gas}$ (H^+) = 2.5 RT - $T\triangle S^0_{gas}$ = 1.48 - 7.76 = -6.28 kcal/mol where, gas phase (at 298 K and 1 atm pressure) entropy of the proton has been calculated using the Sackur-Tetrode equation.[62,63]

acetic acid) show that considering a stronger acid as proton donor, significantly reduces the magnitude of $\triangle\triangle G_{prot,gas}$. Therefore, although there is a limited scope of comprehensive analysis of the absolute values of $\triangle\triangle G_{(de/)prot}$ for different polar sites of nucleobases, potency of these conventional approaches in studying the relative order of the (de/)protonation propensities of different sites is well demonstrated in earlier works.[12,50,52–55,57,64,65] With that confidence, we have further evaluated the relative trend of $\triangle\triangle G_{prot}$ in solvent phase considering, physicochemically more relevant, water molecule as a proton donor, i.e.,

$$Base + H_2O \longrightarrow Protonated\ Base + OH^- \quad (3)$$

Site specific protonation propensity

The relative orders of ease of protonation of different polar sites as obtained from our QM calculations in solvent phase [Table 3] are well correlated with the experimental reports[66–69] which suggest that protonation is more feasible at imino nitrogens compared to carbonyl oxygens and primary amino nitrogens remain unprotonated even at a very low pH.[70] The order of site specific protonation propensities of imino nitrogen sites on the basis of (i) $\Delta\Delta E_{prot,sol}$ values (change in total electronic energy) at both the level of theory (B3LYP and MP2) and (ii) $\Delta\Delta G_{prot,sol}$ values at B3LYP level [Table 3] is as follows:

- Cytosine N3 > Adenine N1 > Guanine N7 > Adenine N7 > Adenine N3 > Guanine N3

Values of $\Delta\Delta G_{prot,sol}$ at MP2 level, however, suggest that N1 of Adenine is the most preferable site for protonation ($\Delta\Delta G_{prot,sol}$ = 35.7 kcal/mol). Despite that, it is interesting to note that, N7 of Guanine and Adenine (polar sites at the Hoogsteen edges of purines), specially N7 of Guanine††, are thermodynamically very preferable sites for protonation. But, neither any instance of N7 protonated Guanine, nor that of Adenine, has been detected in reported structures of RNA and consequently, the possibility of N7 protonation of purines has not been seriously considered earlier.[24,71] Moreover, earlier computational studies by Jissy *et al.*,[72,73] in the context of pH driven molecular switching action of nucleobases, have suggested that N7 protonated Guanine forms weak nonplanar base pairs. The computationally predicted ease of Guanine protonation at N7 however appears to be validated by the observation of DNA and RNA structures with Mg^{2+} coordinated at N7 of Guanine.[74–77] Nevertheless, in the context of RNA, (i) occurrence of N3 protonated Cytosine in its base pairs (formation of I-motif, triple helical DNA, *etc.*), (ii) presence of N1/N3 protonated Adenine in biologically significant regions (active site of hairpin ribozyme,[78] intramolecular stem-loop of U6 RNA of the spliceosome,[79,80] *etc.*) are well known and we have earlier detected instances of N3 protonated Guanine forming base triples in novel RNA structures.[23] We have addressed these apparently conflicting observations by exploring the RNA crystal structure database to detect possible protonated base pairs with N7 protonated Guanine and Adenine and have evaluated their corresponding intrinsic stability.

Table 3 Change in total electronic energy ($\triangle\triangle E_{prot,sol}$) and free energy ($\triangle\triangle G_{prot,sol}$) in solvent phase of the process of protonation (in kcal/mol) following Equation 3.

Base	Protonation Site	$\triangle\triangle E_{prot,sol}$ B3LYP	$\triangle\triangle E_{prot,sol}$ MP2	$\triangle\triangle G_{prot,sol}$ B3LYP	$\triangle\triangle G_{prot,sol}$ MP2
Adenine	N1	33.1	35.4	34.0	35.7
	N3	37.0	40.2	38.9	41.2
	N6	58.8	52.0	59.8	52.7
	N7	36.9	39.8	37.7	40.0
Cytosine	O2*	47.6	-	47.5	-
	N3	31.7	33.6	32.5	37.0
	N4	56.1	53.4	56.9	55.2
Guanine	N2	57.9	54.6	58.4	53.7
	N3	41.3	43.3	40.8	42.2
	O6	44.7	46.6	42.8	45.0
	N7	34.6	37.5	34.1	36.8
Uracil	O2	54.1	53.5	53.2	52.5
	O4	48.7	49.3	48.5	48.9

* O2 protonated Cytosine converges to a different minima on geometry optimization at MP2 level and therefore $\triangle\triangle E_{prot,sol}$ and $\triangle\triangle G_{prot,sol}$ for the same is not reported for MP2 level.

The curious case of N7 protonated Guanine

In silico search of the complete RNA crystal structure database using BPFind software reveals that there are only three instances with a possibility of protonated base pair formation involving N7 protonation of Guanine [Fig. 3]. Interestingly, these three base pairs were studied earlier by Chawla *et al.*[23] considering Guanine as the neutral partner and the proton was assigned to the second base. Hence we have reoptimized those systems considering protonation of either base (Guanine or its partner base) following model (A) – where Guanine is considered N7-protonated and model (B) – where the partner base is protonated [Fig. 3]. Comparison of the results (optimized geometry and intrinsic stability) obtained in model (A) and model (B) highlights that:

†† Gas phase free energy change values ($\Delta\Delta G_{prot,gas}$) in Table 1 and Table 2 suggest that N7 site of Guanine has the highest protonation propensity. $\Delta\Delta G_{prot,sol}$ values at MP2 level in Table 3 also suggest that N7 of Guanine (36.8 kcal/mol) is more preferable site for protonation than N3 of Cytosine (37.0 kcal/mol)

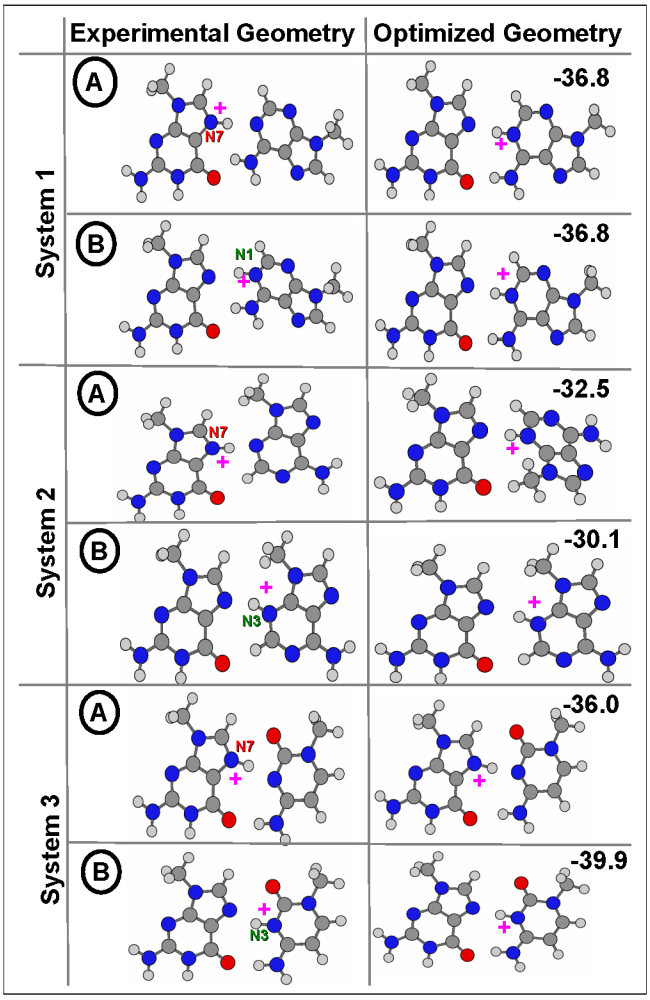

Fig. 3 Experimental geometries of detected base pairs with possibility of protonation at N7 of Guanine have been modeled with the proton placed at (A) the N7 site of Guanine and (B) the corresponding nitrogen atom of the second base. The optimized geometries at B3LYP/6-31G++(2,d2p) level are shown with their interaction energies (in kcal/mol unit) calculated at MP2/aug-ccpVDZ level. PDB Ids and Base pair Ids of these systems are given in Section 1 of Supplementary Information.†

1. Only in System 3, geometry optimization of model (A) results in a planar base pair with N7 protonated Guanine. For the other two cases, in the B3LYP/6-31++G(2d,2p) geometry of model (A), the proton is found to be attached with N1 or N3 of Adenine and Guanine gets converted to the neutral partner.

2. In System 3, although model (A) produces a planar [Table 4] and stable base pair, base pairing interaction between neutral Guanine and N3 protonated Cytosine (model (B)) is equally planar and even stronger.

3. On geometry optimization, both model (A) and model (B), converge to the same minima in System 1, producing a planar [Table 4] and stable G:A(+) H:W Cis base pair. Whereas, in System 2, they converge to two different minima. Unlike System 1, the B3LYP/6-31++G(2d,2p) geometry of model (A) is not planar, rather characterized by large propeller twist and buckle [Table 4] and that of model (B) is planar but weaker than that obtained from model (A).

4. Middle range dispersion interactions might have a significant contribution in these systems. We have, therefore, calculated the dispersion corrections (ΔE_{disp}) for the B3LYP/6-31G++(2d,2p) optimized geometries. In System 1 and 3, for both the models, we obtain equal amount of dispersion correction: ΔE_{disp} = -4.0 kcal/mol for System 1 and ΔE_{disp} = -5.1 kcal/mol for System 3. However, for System 2, in comparison to model (B), dispersion interactions stabilize model (A) by ~1 kcal/mol (ΔE_{disp} = -4.7 kcal/mol for model (A) and -3.6 kcal/mol for model (B)).

The above observations raise the question: why does the N7 protonated Guanine get converted to a neutral base in the optimized geometries of model (A) for System 1 and System 2? This may be understood by considering two factors,

1. The protonated base pairs are stabilized by charge dipole interactions.[23,83] Calculations show that dipole moments of Guanine and Cytosine are approximately three times higher than that of Adenine [Table 5]. Neutral Guanine, because of its high dipole moment, is naturally preferred in the minimum energy structure of the base pairs of System 1 and 2, even though in the initial geometry Adenine was considered as the neutral base (model (A)). On the other hand, difference of dipole moments of Cytosine and Guanine being considerably small, in System 3, difference of charge dipole interaction does not provide sufficient driving force to cause a proton transfer from Guanine N7 to Cytosine N3 on geometry optimization of model (A).

2. Analysis of protonation induced charge redistribution demonstrates that, protonation at ring atoms results in the withdrawal of electrons from the ring [Table 6], which, in turn, positively influences the hydrogen bond donor potential and negatively influences the hydrogen bond acceptor potential of the hydrogen bond donor and acceptor sites, respectively. This effect is substantiated by the analysis of protonated base pairing induced shift in vibrational bond stretching frequency of N-H bonds which participate in hydrogen bond formation between the two interacting bases [Section 4 in supplementary information†]. Basis of such charge redistribution may

Table 4 NUPARM calculated intra base pair parameters and the dihedral angle between the planes of the two bases in their B3LYP/6-31++G(2d,2p) geometry.

System	Buckle	Open	Twist	Stagger	Shear	Stretch
1(A)	-0.11	-1.90	-0.07	-0.01	0.41	2.83
1(B)	0.11	-1.91	0.06	0.01	0.41	2.83
2(A)	30.29	-34.91	110.96	-0.77	1.17	2.44
2(B)	0.17	-4.31	3.27	0.02	0.76	2.77
3(A)	-0.51	7.22	-0.07	0.00	0.46	2.63
3(B)	-0.15	-1.00	-0.08	0.00	0.26	2.75
4(A)	6.28	0.24	47.24	0.34	0.32	2.68
4(B)	-17.28	-8.90	38.43	0.60	-0.26	2.80
A+:G H:H Cis	-0.35	-0.47	0.69	0.01	-0.27	2.81
A+:G H:H Trans	-0.38	-5.98	2.06	0.01	2.94	2.81

be understood by analyzing the charged state of the most stable resonance canonical structure (Valance Bond structure) of the neutral and protonated bases [Fig. 4]. The reverse effect is observed in the case of deprotonation. N7 protonation of Guanine, therefore, reduces the electronic charge at the O6 site, making it a poor hydrogen bond acceptor. Hence, although N7 protonation confirms one strong hydrogen bond involving the N7 site as hydrogen bond donor, the second hydrogen bond between the two interacting bases becomes considerably weak since that involves the O6 site as hydrogen bond acceptor.

Therefore, it may be concluded that, although it is comparatively easier to protonate the N7 site of Guanine, after getting protonated at N7 site, Guanine can not form stable and planar base pairs. The point to be noted is that, our method for inferring protonation of bases in PDB structures can only consider instances where the protonated edge is involved in base pairing. The possibility of N7 protonated Guanine participating in base pairs involving other edges, for example, the Watson-Crick edge of Guanine can not be ruled out. This is particularly notable since there has been a lot of discussions on the role of base pairs containing Guanine bases where Mg ion is co-ordinated at N7 [74–77] or where there is an archaeosine modification at N7. [84] The significance of our observations regarding the ease of protonation at N7 of Guanine has been elaborated in the conclusion section.

Table 5 Dipole moment (μ) of nucleobases (in Debye) calculated in B3LYP/6-31G+(2d,2p) level and MP2/aug-cc-pVDZ level for gas phase and solvent phase.

Base	Gas Phase		Solvent Phase	
	B3LYP	MP2	B3LYP	MP2
Adenine	2.67	2.92	3.56	3.82
Cytosine	6.23	6.86	9.85	10.54
Guanine	7.14	7.60	10.59	10.81
Uracil	5.03	5.52	6.81	7.41

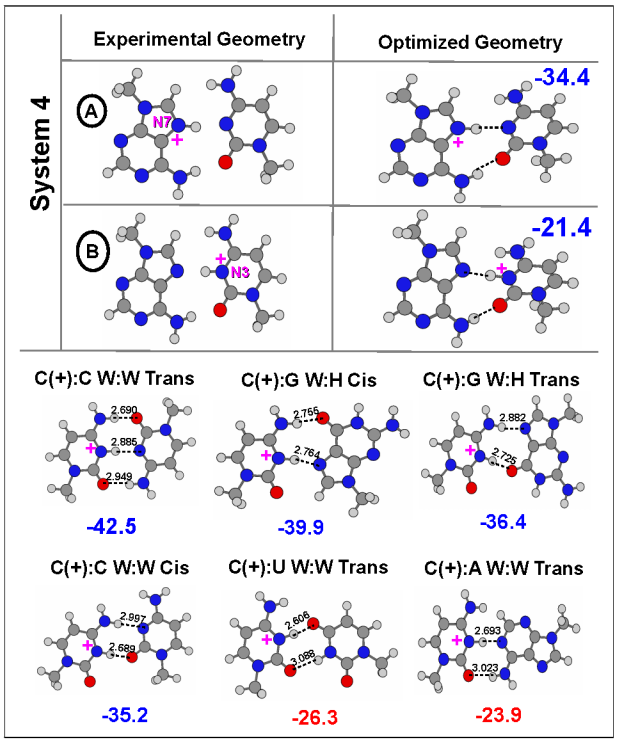

Fig. 6 Experimental geometry of *in-silico* detected base pairing interaction between Hoogsteen edge of Adenine and Watson-Crick edge of Cytosine (System 4) has been modeled for geometry optimization by (A) adding proton at N7 of Adenine and (B) adding proton at N3 of Cytosine. B3LYP/6-31++G(2d,2p) geometry of these models with MP2/aug-cc-pVDZ level interaction energy (in kcal/mol) of System 4 is also reported. The same for different detected base pairs with N3 protonated Cytosine are also reported. Hydrogen bond donor-acceptor distances are also given in Å unit. PDB Ids and Base pair Ids of these systems are given in Section 1 of Supplementary Information.†

pH driven conformational switching based on N7 protonation of Adenine

Unlike Guanine, Adenine contains NH_2 group at C6 position which acts as a hydrogen bond donor in the base pair-

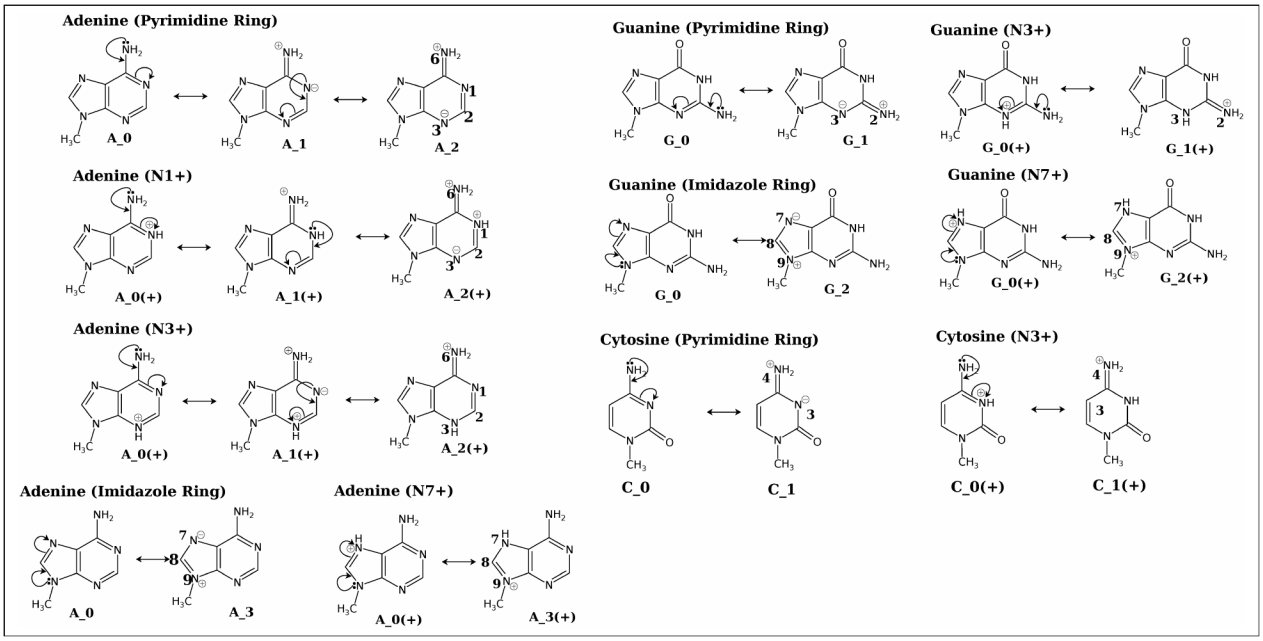

Fig. 4 Resonance canonical (Valance Bond) structures of modeled neutral and protonated RNA bases are shown.

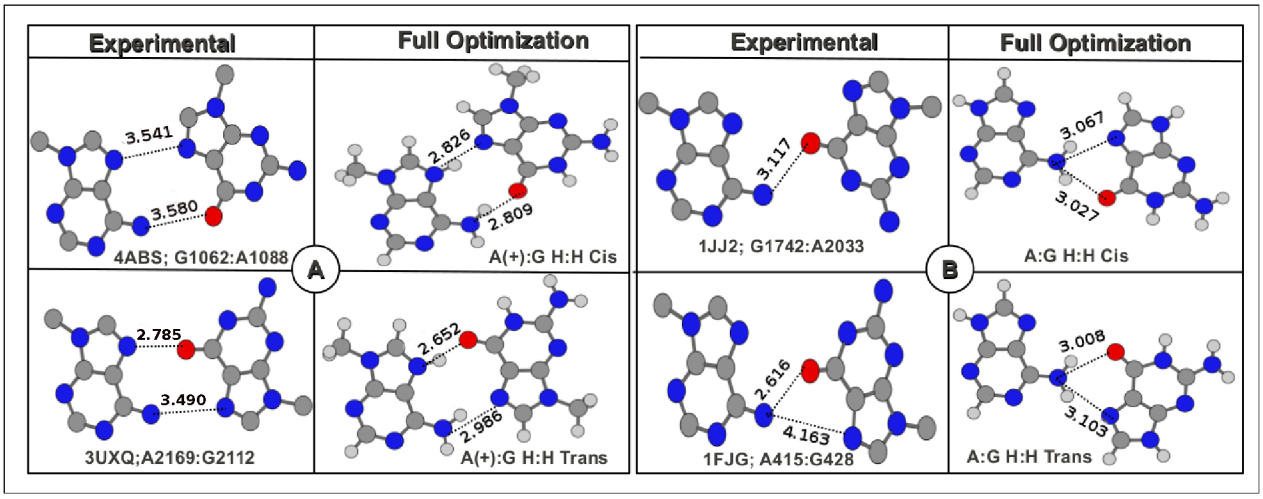

Fig. 5 The figure highlights the possibility of protonation driven multimodality within A:G H:H Cis and Trans base pair geometries respectively. (A) Examples of experimental and gas phase optimized structures of A(+):G H:H Cis and A(+):G H:H Trans as reported in this work. (B) Examples of experimental and gas phase optimized structures of A:G H:H Cis and A:G H:H Trans as reported earlier.[85] Hydrogen bond donor-acceptor distances are also given in Å unit.

ing interactions involving its Hoogsteen edge. We have already seen that charge redistribution due to protonation at ring nitrogen atoms results in improved hydrogen bond donor strength. Thus, with the expectation of observing stable base pairing interactions involving Hoogsteen edge of N7 protonated Adenine, we have searched the complete RNA crystal structure database and could detect two new classes of protonated base pairs: A(+):G H:H Cis and A(+):G H:H Trans [Fig. 5(A)]. It may be noted here that Leontis and Westhof had earlier reported examples of neutral A:G H:H Cis and A:G H:H Trans as weakly interacting base pairs involving single hydrogen bonds.[71] It has also been shown earlier that geometry optimization of these examples lead to significant changes in hydrogen bonding pattern, with optimized structures show-

Table 6 NBO charges at different hydrogen bond acceptor and donor sites of the four RNA bases are given for different charged state of the base. On protonation at ring nitrogen atoms, the proton acts as a charge withdrawing group causing a depletion of electronic charges over the ring atoms. The opposite result is obtained in the case of deprotonation. The same trend is followed in the analysis of Mulliken partial charges and ESP charges as shown in Table S2 and Table S3 of supplementary information.†

Base		H-bond Acceptor			H-bond donor	
		N1	N3	N7	N6	
Adenine	Neutral	-0.55	-0.526	-0.485	-0.823	
	$N1^+$	-	-0.441	-0.45	-0.775	
	$N3^+$	-0.489	-	-0.438	-0.742	
	$N7^+$	-0.499	-0.492	-	-0.798	
		N3	O6	N7	N1	N2
Guanine	Neutral	-0.575	-0.6	-0.439	-0.655	-0.852
	$N3^+$	-	-0.49	-0.395	-0.636	-0.803
	$N7^+$	-0.565	-0.552	-	-0.64	-0.803
	$N1^-$	-0.66	-0.686	-0.464	-	-0.882
		O2	N3		N4	
Cytosine	Neutral	-0.631	-0.592		-0.826	
	$N3^+$	-0.54	-		-0.766	
		O2	O4		N3	
Uracil	Neutral	-0.627	-0.598		-0.672	
	$N3^-$	-0.729	-0.708		-	

ing large RMSD from their respective experimental structures [Fig. 5(B)].[85] In contrast, the examples detected in this study, when optimized as protonated base pairs, yielded highly stabilized structures (E_{int}^{gas} = -35.27 kcal/mol for Cis and -36.95 kcal/mol for Trans) while retaining their hydrogen bonding patterns as well as their geometries [Table 4]. These observations strongly indicate the possibility of protonation induced multimodality in A:G H:H Cis and Trans geometries; thereby suggesting the possible role of these base pairs in pH driven conformational switching processes.

Variation of stability of base pairs involving N3 protonated Cytosine

We have identified another instance of possible protonated base pairing interaction involving Hoogsteen edge of Adenine and Watson-Crick edge of Cytosine (System 4 in Fig. 6). The same system was studied earlier by Chawla *et al.*[23] considering Adenine as neutral partner and Cytosine as the protonated base with protonation at N3 site (model (B) in Fig. 6). Interestingly, optimized geometry of System 4 with N7 protonated Adenine as initial geometry (model (A) in Fig. 6) turns out to be 13 kcal/mol stabler and significantly planar [Table 4] than that of model (B). Analysis of intrinsic interaction energies of all the protonated base pairs involving N3 protonation of Cytosine [Fig. 6] suggests that, although N3 site of Cytosine is one of the most thermodynamically favorable site of protonation, N3 protonated Cytosine produces significantly weak base pairs along with highly stable ones. The reason may be explained by observing the effect of N3 protonation on the other hydrogen bond donor (N4) and acceptor (O2) sites of the Watson-Crick edge of Cytosine. Since due to N3 protonation (i) N4 site acts as a stronger hydrogen bond donor and (ii) O2 site acts as a weaker hydrogen bond acceptor, significantly lower interaction energy is observed for the base pairs where there are only two hydrogen bonds and the second hydrogen bond is formed via the O2 site (e.g., C(+):U W:W Trans and C(+):A W:W Trans).

How do protonated base pairs get stabilized?

On the basis of the above discussed examples we may conclude that, it is not the ease of protonation but the protonation induced charge redistribution which dominates the stabilizing forces (base pairing) and hence the occurrence propensity of a nucleobase protonated at a specific site. Therefore, it is also expected that for a protonated base pair, apart from its intrinsic stability, the charge redistribution within the individual bases

due to the base pairing interaction will also play a major role in determining their occurrence propensity.

To evaluate the extent and nature of base pairing induced charge redistribution in the individual partner bases, we have calculated the difference of NBO charges (Δq) between different sites of isolated bases and protonated base pairs involving them. It is interesting to note that, canonical base pairing (AT/U and GC) does not make any significant influence on the charge distribution of the partner bases. Whereas, as shown in Fig. 7, protonated base pairing significantly redistributes the electronic charges in a way that promotes the formation of base triples via the free edges of the protonated base pair. Fig. 7 describes six instances of base triples found in RNA crystal structures involving six different geometries of protonated base pairs. The hydrogen bond donor and acceptor sites of the free edges of the base pairs have been modified positively (with a few exceptions, such as, N7 of Adenine in example 2 and O6 of Guanine in example 4) by the effect of base pairing interaction to facilitate base pairing interaction with the third base.

In protonated base pairs, the positive co-operative effect of base pairing is not limited to opening up new avenues for strong hydrogen bonding interactions only. Changes of NICS(1) values of the aromatic rings of the parter bases on base pairing, as reported in Table S5†, also suggest that compared to canonical base pairing, protonated base pairing enhances the aromatic character of the corresponding rings except the one which contains the site of protonation. Such enhancement of aromatic character may further influence the stacking interactions which constitute an interesting area of ongoing research.[86,87]

Lack of correlation between stability of protonated base pairs and their occurrence frequency in RNA crystal structures

A comprehensive search and analysis of a non-redundant set of 156 RNA crystal structures, as obtained from HD-RNAS database,[51] has revealed another important issue regarding occurrence propensity of protonated base pairs. Among several protonated base pairs found in the non-redundant set, we have found A(+):C W:W Cis base pair to be the most abundant (45 instances), followed by C(+):C W:W Cis (13 instances). The occurrence frequencies of all other examples of protonated base pairs are however much lower. To investigate this variation in occurrence frequencies, we have increased our search space and performed a rigorous search over the complete crystal structure database to identify six base pairs having high to moderate occurrence frequencies. Interestingly, optimized geometries of all the six base pairs have shown similar planar geometries [Table S4†] and high stabilities [Table S5†]. Apart from (i) A(+):C W:W Cis (604 instances) and (ii) C(+):C W:W Cis (172 instances), this list contains four more pairs: (iii) A(+):G W:H Cis (79 instances), (iv) A(+):G S:H Cis (48 instances), (v) C(+):G W:H Trans (91 instances) and (vi) C(+):U W:W Cis (77 instances). Clearly the occurrence of protonated base pairs in the non-redundant data set, do not correlate with their extents of planarity and stabilities, nor with their occurrence frequencies in the complete crystal structure data base.

Our hypothesis is that, because of evolutionary pressure, the occurrence of protonated base pairs is dependent on their specific structural and functional roles respectively. Base pairs which can participate within double helical regions should therefore have greater occurrence frequencies whereas those needed only in special motifs will have lower occurrence in a non-redundant dataset. Analysis of the context of occurrence of these base pairs show that A(+):C W:W Cis is more abundant within double helical stretches: (a) in the complete database – 322 out of 604 instance and (b) in the non-redundant data set – 19 out of 45 instances. The C(+):C W:W Cis base pair also has a lesser, albeit noticeable, occurrence within helical stretches: (a) in the complete database – 20 out of 173 instances and (b) in the non-redundant set – none. Other base pairs on the other hand have little or no occurrence within double helical stretches.

Why does A(+):C W:W Cis base pair occur frequently within double helical stretches?

Two factors, apart from stability, appear to be relevant in the context of occurrence potential of base pairs in double helical stacks: (a) isostericity with canonical base pairs and (b) stacking potential with the canonical base pairs. Base pairs which do not possess these two characteristics may, and do, occur in double helical regions, but this may be because of some specific functional requirement. The C(+):C W:W Cis base pair, for example, in spite of having a noticeable occurrence in double helical regions in the complete database have no such occurrence in the non-redundant database. Another base pair G(+):G H:S Trans has low frequency of occurrence, but is found to be conserved within a double helical stretch in a set of similar crystal structures of 23S rRNA of *Thermus thermophilus* (between 1589G:1439G) and *Escherichia coli* (between 1723G:1737G).[51] While these base pairs are not isosteric with canonical base pairs, their occurrence pattern is indicative of some specific functional role.

A(+):C W:W Cis base pair on the other hand is isosteric with the wobble base pair G:U W:W Cis and, like the latter, should be capable of occurring in double helical stretches in its own rights. The next question relates to the stacking potential of A(+):C W:W Cis. To have a qualitative understanding of the stacking potentials of A(+):C W:WC base pair with flanking AT/GC canonical base pairs in helical environment, we have looked into the spatial distribution of the Highest Occu-

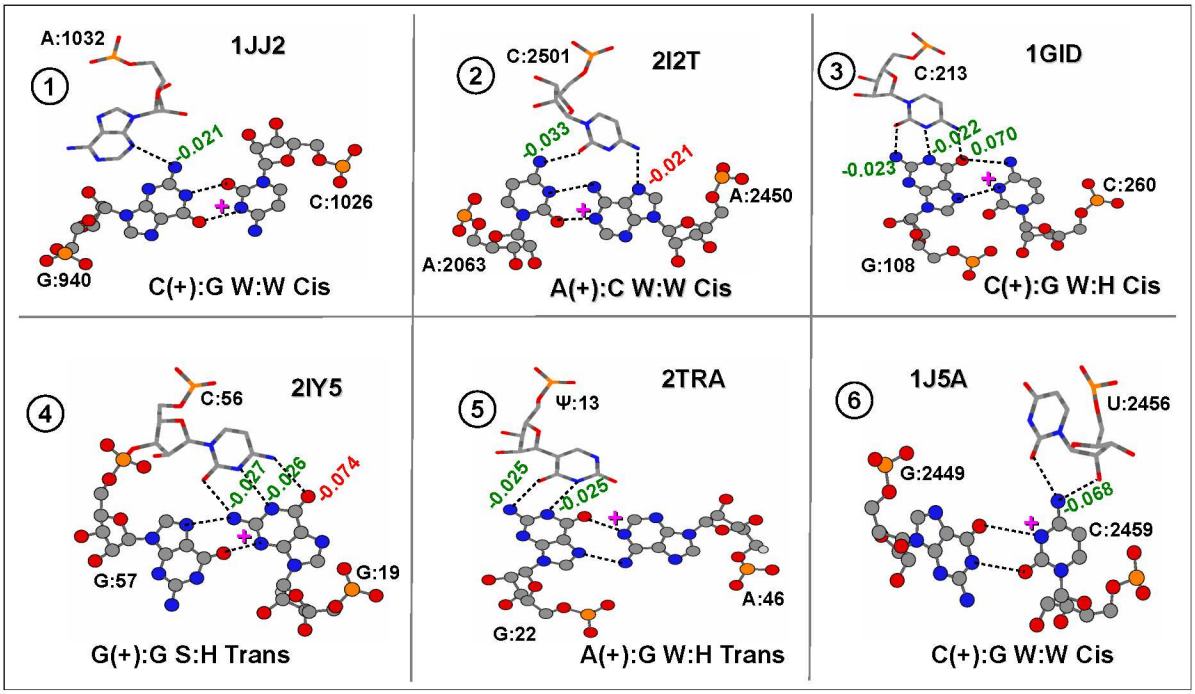

Fig. 7 Examples of base triples involving protonated base pairs (shown in ball and stick model) of different geometry are demonstrated. Change in NBO charges (Δq) from the isolated (unpaired) neutral state to the paired state at different sites of the protonated base pair (which interact with the third neutral base) are given. Negative value of Δq at a hydrogen bond donor site and positive value of Δq at a hydrogen bond acceptor site indicate that their hydrogen bonding potential improves due to the protonated base pair formation.

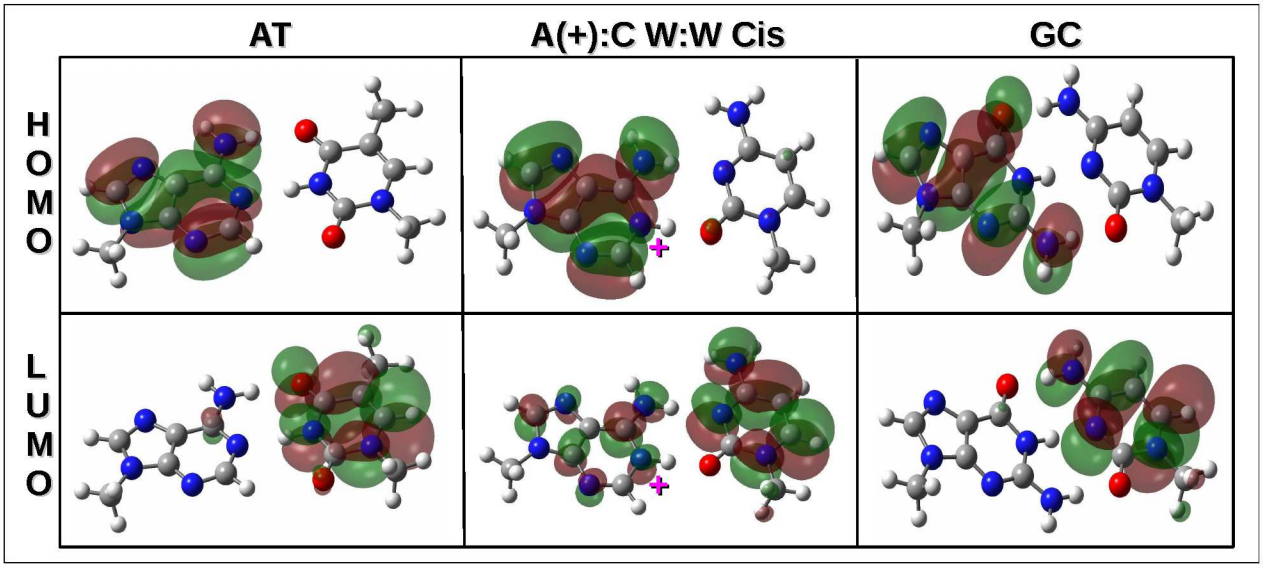

Fig. 8 HOMO and LUMO of Natural Bond Orbitals of A(+):C W:W Cis protonated base pair and AT and GC canonical base pairs.

pied and the Lowest Unoccupied NBOs of A(+):C W:W Cis base pair and have compared them with that of the canonical base pairs [Figure 8]. HOMO and LUMO of all these pairs are found to be symmetric p-orbitals. HOMOs of A(+):C W:W Cis, AT and GC are localized only over the purine base (A, A(+) or C). But, though the LUMOs of GC and AT are restricted only over the pyrimidines, that for A(+):C W:W Cis is distributed over both pyrimidine (C) and purine (A(+))

bases. This enhances the stacking potential of A(+):C W:W Cis flanked by AT and GC by increasing the spatial proximity of the LUMO of A(+):C W:W Cis with the HOMOs of GC and AT, an effect absent in other protonated base pairs [Fig. S1]†. In the context of DNA duplexes, similar complementary spatial distribution of frontier orbitals also has been observed between the canonical base pairs and synthetic base pair involving natural Adenine and 6-ethynylpyridone, a thermostabilizing Thymine analogue.[88]

Conclusions

Significance of protonated nucleobases in enabling the RNA molecule to perform catalytic functions like protein has been well established.[11] The thermodynamic barrier associated with the protonation of the nucleobase, makes it important to develop a molecular level understanding of the factors that stabilize the pK_a shifted bases and base pairs involving them.

We have benchmarked the conventional formalisms for modeling the process of protonation and concluded that considering water as a proton donor might provide a physicochemically relevant picture of the relative order of protonation propensity of different sites of the nucleobases. We performed QM calculations using both DFT based (B3LYP) and wave function based (MP2) formalisms. With the exception of O2 protonated Cytosine, all neutral and corresponding charged nucleobases, on ground state optimization, converge respectively to equivalent geometries at both the levels of theory. Although both the methods, by and large, result in similar trend in energetics data, it is interesting to note that there are two minor differences in the trends: (1) in solvent phase [Table 3], MP2 level calculations suggest that N1 site of Adenine is easier to protonate than N3 site of Cytosine and (2) for imino nitrogen protonation, B3LYP calculation results in underestimation of free energy change ($\Delta\Delta G_{prot}$) compared to that obtained from MP2 level calculations. For example, in Table 3, at B3LYP level $\Delta\Delta G_{prot,sol}$ for N1 protonation of Adenine is 1.7 kcal/mol less than the corresponding MP2 level calculated value (35.7 kcal/mol). But for deprotonation from secondary amino nitrogens and protonation at primary amino nitrogens, the trend reverses.

Our QM calculations, with excellent correlation with experimental reports, indicate relatively better protonation propensity of N7 site of purines, specially Guanine. Although we have detected two new geometries of protonated base pairs (A(+):G H:H Cis and Trans) with putative role as pH driven conformational switches and involving N7 protonated Adenine, N7 protonated Guanine, however, limited by the protonation induced charge redistribution, can not form strong planar base pairs. For the same reason, a large diversity in the geometry and stability is observed for the base pairs involving N3 protonated Cytosine, another site of high protonation propensity. Like protonated bases, occurrence propensities of base pairs involving them are also influenced by the base pairing induced charge redistribution. We have found that, unlike canonical base pairing, protonated base pairing causes a significant charge redistribution at the free edges, which, in turn, facilitates the formation of base triples.

A statistical analysis of the occurrence frequencies of protonated base pairs in a non-redundant RNA crystal structure dataset, however, suggest that, the occurrence of a protonated base pair is finally decided by its context of occurrence and the functional requirement. Therefore, we have observed that A(+):C W:W Cis pair, which is capable of occurring inside a double helical stretch due to its isostericity with canonical base pairs and stacking potential with canonical base pairs, has a surprisingly high occurrence frequency, whereas, other protonated base pairs with similar stability does not occur at all in the non-redundant dataset.

Our current approach towards detection of possible protonated base pairs, limits our search to those where the protonated edge is involved in base pairing. But there is sufficient evidence suggestive of modifications in hydrogen bonding interactions through the non-protonated edges of protonated bases. For example, metal ion co-ordination at N7 of Guanine has been found to modify the hydrogen bonding potential of its WC edge and therefore the stability and geometry of the non-canonical base pairs formed through the WC edge shows significant variation.[77,84] We have proved that similar modifications can be achieved simply by adding a proton at the N7 site of Guanine. Hence, interactions through the non-protonated edges of protonated bases need to be investigated further.

Acknowledgements

A.M. and D.B. thank DBT, India, for research grants [grant numbers BT/PR-11429/BID/07/272/2008 and BT/PR5451/BID/07/111/2004]. AM and AH thank DBT project BT/PR-14715/PBD/16/903/2010 for financial support.

References

1 B. C. Stark, R. Kole, E. J. Bowman and S. Altman, *Proc. Natl. Acad. Sci. U.S.A.*, 1978, **75**, 3717–3721.
2 B. L. Bass and T. R. Cech, *Nature*, 1984, **308**, 820–826.
3 H. Tabara, M. Sarkissian, W. G. Kelly, J. Fleenor, A. Grishok, L. Timmons, A. Fire and C. C. Mello, *Cell*, 1999, **99**, 123 – 132.
4 A. P. Carter, W. M. Clemons, D. E. Brodersen, Morgan, B. T. Wimberly and V. Ramakrishnan, *Nature*, 2000, **407**, 340–348.
5 E. Birney, J. A. Stamatoyannopoulos, A. Dutta, R. Guigó, T. R. Gingeras, E. H. Margulies, Z. Weng, M. Snyder, E. T. Dermitzakis, J. A. Stamatoyannopoulos and et al., *Nature*, 2007, **447**, 799–816.
6 K. Hirota, T. Miyoshi, K. Kugou, C. S. Hoffman, T. Shibata and K. Ohta, *Nature*, 2008, **456**, 130–134.
7 M. Gimpel, H. Preis, E. Barth, L. Gramzow and S. Brantl, *Nucleic Acids Res.*, 2012, **40**, 11659–11672.

8 *Ribozymes and RNA Catalysis*, ed. D. M. J. Lilley and F. Eckstein, The Royal Society of Chemistry, 2007, pp. P001–318.

9 D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*, W. H. Freeman and Co., 5th edn., 2009.

10 P. C. Bevilacqua, T. S. Brown, S.-i. Nakano and R. Yajima, *Biopolymers*, 2004, **73**, 90–109.

11 J. L. Wilcox, A. K. Ahluwalia and P. C. Bevilacqua, *Acc. Chem. Res.*, 2011, **44**, 1270–1279.

12 V. Verdolino, R. Cammi, B. H. Munk and H. B. Schlegel, *J. Phys. Chem. B*, 2008, **112**, 16860–16873.

13 P. B. Rupert and Ferre, *Nature*, 2001, **410**, 780–786.

14 P. B. Rupert, A. P. Massey, S. T. Sigurdsson and A. R. Ferr-D'Amar, *Science*, 2002, **298**, 1421–1424.

15 P. Banáš, N. G. Walter, J. Šponer and M. Otyepka, *J. Phys. Chem. B*, 2010, **114**, 8701–8712.

16 V. Mlýnský, P. Banáš, D. Hollas, K. Reblova, N. G. Walter, J. Šponer and M. Otyepka, *J. Phys. Chem. B*, 2010, **114**, 6642–6652.

17 S. M. Mirkin and M. D. Frank-Kamenetski, *Ann. Rev. of Biophys. and Biomol. Struct.*, 1994, **23**, 541 – 576.

18 R. Zain and J.-S. Sun, *Cell. and Mol. Life Sci.*, 2003, **60**, 862 – 870.

19 K. Gehring, J.-L. Leroy and M. Gueron, *Nature*, 1996, **363**, 561 – 565.

20 I. Berger, C. Kang, A. Fredian, R. Ratliff, R. Moyzis and A. Rich, *Nat. Struct. Mol. Biol.*, 1995, **2**, 416 – 425.

21 A. Kuttan and B. L. Bass, *Proc. Natl. Acad. Sci. U.S.A.*, 2012, **109**, E3295E3304.

22 S.-i. Nakano, D. M. Chadalavada and P. C. Bevilacqua, *Science*, 2000, **287**, 1493–1497.

23 M. Chawla, P. Sharma, S. Halder, D. Bhattacharyya and A. Mitra, *J. Phys. Chem. B*, 2011, **115**, 1469–1484.

24 J. Das, S. Mukherjee, A. Mitra and D. Bhattacharyya, *J. Biomol. Struct. Dyn.*, 2006, **24**, 91–202.

25 D. Bhattacharyya, S. C. Koripella, A. Mitra, V. B. Rajendran and B. Sinha, *J. Biosci.*, 2007, **32**, 809–825.

26 M. Rooman and R. Wintjens, *J. Biomol. Struct. Dyn.*, 2014, **32**, 532–545.

27 H. Fernando, G. A. Papadantonakis, N. S. Kim and P. R. LeBreton, *Proc. Natl. Acad. Sci. U.S.A.*, 1998, **95**, 5550–5555.

28 P. Slavíček, B. Winter, M. Faubel, S. E. Bradforth and P. Jungwirth, *J. Am. Chem. Soc.*, 2009, **131**, 6460–6467.

29 R. Dennington, T. Keith and J. Millam, *GaussView Version 5*, Semichem Inc. Shawnee Mission KS 2009.

30 M. J. Frisch, G. W. Trucks, 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, *Gaussian 03, Revision E.01*, Gaussian, Inc., Wallingford, CT, 2004.

31 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, 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, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09 Revision C.01*, Gaussian Inc. Wallingford CT 2009.

32 B. Yang and M. T. Rodgers, *J. Am. Chem. Soc.*, 2014, **136**, 282–290.

33 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

34 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.

35 A. Mládek, P. Sharma, A. Mitra, D. Bhattacharyya, J. Šponer and J. E. Šponer, *J. Phys. Chem. B*, 2009, **113**, 1743–1755.

36 J. Šponer, P. Jurečka and P. Hobza, *J. Am. Chem. Soc.*, 2004, **126**, 10142–10151.

37 T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007–23.

38 M. Cossi, N. Rega, G. Scalmani and V. Barone, *J. Comput. Chem.*, 2003, **24**, 669–681.

39 V. Barone and M. Cossi, *J. Phys. Chem. A*, 1998, **102**, 1995–2001.

40 A. Klamt and G. Schuurmann, *J. Chem. Soc., Perkin Trans. 2*, 1993, 799–805.

41 E. D. Glendening, A. E. Reed, J. E. Carpenter and F. Weinhold, *NBO Version 3.1*, Gaussian, Inc., Wallingford, CT, 2004.

42 M. K. Cyraski, M. Gilski, M. Jasklski and T. M. Krygowski, *J. Org. Chem.*, 2003, **68**, 8607–8613.

43 P. Cysewski, *J. Mol. Struct.*, 2005, **714**, 29 – 34.

44 P. v. R. Schleyer, C. Maerker, A. Dransfeld, H. Jiao and N. J. R. v. E. Hommes, *J. Am. Chem. Soc.*, 1996, **118**, 6317–6318.

45 K. Wolinski, J. F. Hinton and P. Pulay, *J. Am. Chem. Soc.*, 1990, **112**, 8251–8260.

46 J. R. Cheeseman, G. W. Trucks, T. A. Keith and M. J. Frisch, *J. Chem. Phys.*, 1996, **104**, 5497–5509.

47 H. Fallah-Bagher-Shaidaei, C. S. Wannere, C. Corminboeuf, R. Puchta and P. v. R. Schleyer, *Org. Lett.*, 2006, **8**, 863–866.

48 W. K. Olson, M. Bansal, S. K. Burley, R. E. Dickerson, M. Gerstein, S. C. Harvey, U. Heinemann, X.-J. Lu, S. Neidle, Z. Shakked, H. Sklenar, M. Suzuki, C.-S. Tung, E. Westhof, C. Wolberger and H. M. Berman, *J. Mol. Biol.*, 2001, **313**, 229 – 237.

49 S. Mukherjee, M. Bansal and D. Bhattacharyya, *J. Comput. Aid. Mol. Dsgn.*, 2006, **20**, 629–645.

50 D. M. Close, *J. Phy. Chem. A*, 2013, **117**, 473–480.

51 S. S. Ray, S. Halder, S. Kaypee and D. Bhattacharyya, *Front. Genet.*, 2012, **3**,.

52 J. E. Šponer, J. Leszczynski, F. Glahé, B. Lippert and J. Šponer, *Inorg. Chem.*, 2001, **40**, 3269–3278.

53 D. T. Major, A. Laxer and B. Fischer, *J. Org. Chem.*, 2002, **67**, 790–802.

54 A. Zhachkina, M. Liu, X. Sun, F. S. Amegayibor and J. K. Lee, *J. Org. Chem.*, 2009, **74**, 7429–7440.

55 D. Touboul, G. Bouchoux and R. Zenobi, *J. Phys. Chem. B*, 2008, **112**, 11716–11725.

56 B. H. Munk, C. J. Burrows and H. B. Schlegel, *J. Am. Chem. Soc.*, 2008, **130**, 5245–5256.

57 A. K. Chandra, D. Michalska, R. Wysokisky and T. Zeegers-Huyskens, *J. Phys. Chem. A*, 2004, **108**, 9593–9600.

58 M. J. G. Moa, M. Mandado and R. A. Mosquera, *Chem. Phys. Lett.*, 2006, **428**, 255 – 261.

59 M. Liu, T. Li, S. F. Amegayibor, D. S. Cardoso, Y. Fu and J. K. Lee, *J. Org. Chem.*, 2008, **73**, 9283–9291.

60 M. T. Nguyen, A. K. Chandra and T. Zeegers-Huyskens, *J. Chem. Soc., Faraday Trans.*, 1998, **94**, 1277–1280.

61 J. A. Frey, A. Mller, M. Losada and S. Leutwyler, *J. Phys. Chem. B*, 2007, **111**, 3534–3542.

62 G. J. Tawa, I. A. Topol, S. K. Burt, R. A. Caldwell and A. A. Rashin, *J. Chem. Phys.*, 1998, **109**, 4852–4863.

63 S.-G. Hwang and D.-S. Chung, *Bull. Kor. Chem. Soc.*, 2005, **26**, 589 – 593.

64 Y. Podolyan, L. Gorb and J. Leszczynski, *The Journal of Physical Chemistry A*, 2000, **104**, 7346–7352.

65 B. H. Allehyani, S. A. Elroby, S. G. Aziz and R. H. Hilal, *J. Biomol. Struct. Dyn.*, 2014, **0**, 1–14.

66 R. L. Benoit and M. Frchette, *Canad. J. Chem.*, 1985, **63**, 3053–3056.

67 J. Clauwaert and J. Stockx, *Z Naturforsch B.*, 1968, **23**, 25–30.

68 *Handbook of Biochemistry and Molecular Biology Vol.1 Nucleic Acids*, ed. G. D. Fasman, Chem. Rubber Co., Cleveland Ohio, 1975, pp. 76 – 206.

69 *Handbook of Biochemistry, Selected Data for Molecular Biology*, ed. H. A. Sober, R. A. Harte and E. K. Sober, Chem. Rubber Co., Cleveland Ohio, 1970, pp. G3 – G98.

70 M. Egli and W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, 1984, pp. P001–556.

71 N. B. Leontis, J. Stombaugh and E. Westhof, *Nucleic. Acid. Res.*, 2002, **30**, 3497–3531.

72 A. K. Jissy and A. Datta, *J. Phys. Chem. B*, 2010, **114**, 15311–15318.

73 A. K. Jissy and A. Datta, *J. Phys. Chem. Lett.*, 2014, **5**, 154–166.

74 A. K. Katz, J. P. Glusker, S. A. Beebe and C. W. Bock, *J. Am. Chem. Soc.*, 1996, **118**, 5752–5763.

75 D. Bandyopadhyay and D. Bhattacharyya, *J. Biomol. Struct. Dyn.*, 2003, **21**, 447–458.

76 S. Mukherjee and D. Bhattacharyya, *J. Biomol. Struct. Dyn.*, 2013, **31**, 896–912.

77 R. Oliva and L. Cavallo, *J. Phys. Chem. B*, 2009, **113**, 15670–15678.

78 P. C. Bevilacqua, *Biochemistry*, 2003, **42**, 2259–2265.

79 V. Venditti, L. Clos II, N. Niccolai and S. E. Butcher, *J. Mol. Biol.*, 2009, **391**, 894 – 905.

80 N. J. Reiter, H. Blad, F. Abildgaard and S. E. Butcher, *Biochemistry*, 2004, **43**, 13739–13747.

81 N. Marom, A. Tkatchenko, M. Rossi, V. V. Gobre, O. Hod, M. Scheffler and L. Kronik, *J. Chem. Theory Comput.*, 2011, **7**, 3944–3951.

82 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *The Journal of Chemical Physics*, 2010, **132**, –.

83 J. E. Šponer, J. Leszczynski, V. Sychrovský and J. Šponer, *J. Phys. Chem. B*, 2005, **109**, 18680–18689.

84 R. Oliva, L. Cavallo and A. Tramontano, *RNA*, 2007, **13**, 1427–1436.

85 P. Sharma, J. E. Šponer, J. Šponer, S. Sharma, D. Bhattacharyya and A. Mitra, *J. Phys. Chem. B*, 2010, **114**, 3307–3320.

86 J. W. G. Bloom and S. E. Wheeler, *Angew. Chem. Int. Ed.*, 2011, **50**, 7847–7849.

87 A. C. Tsipis and A. V. Stalikas, *Inorg. Chem.*, 2013, **52**, 1047–1060.

88 A. Halder, A. Datta, D. Bhattacharyya and A. Mitra, *J. Phys. Chem. B*, 2014, **118**, 6586–6596.