PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Role of N7 protonation of guanine in determining structure, stability and function of RNA base pairs†

Antarip Halder,[a] Sohini Bhattacharya,[a] Ayan Datta,[b] Dhananjay Bhattacharyya[c] and Abhijit Mitra*[a]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

Roles of protonated nucleobases in stabilizing different structural motifs and in facilitating catalytic functions of RNA are well known. Among different polar sites of all the nucleobases, N7 of guanine has the highest protonation propensity under physiological pH. However, unlike with other easily protonable sites such as N1 and N3 of adenine or N3 of cytosine, N7 protonation of guanine does not lead to the stabilization of base pairs involving its protonated Hoogsteen edge. It also does not facilitate its participation in any acid-base catalysis process. To explore the possible roles of N7 protonated guanine, we have studied its base pairing potentials involving WatsonCrick and sugar edges, which undergo major charge redistribution on N7 protonation. We have carried out quantum chemical geometry optimization at M05-2X/6-311G+(2d,2p) level, followed by interaction energy calculation at MP2/aug-cc-pVDZ level, along with analysis of the context of occurrence for selected base pairs involving sugar edge or WatsonCrick edge of guanine within a non redundant set of 167 RNA crystal structures. Our results suggest that, four base pairs – G:C W:W Trans, G:rC W:S Cis, G:G W:H Cis and G:G S:H Trans may involve N7 protonated guanine. These base pairs deviate significantly from their respective experimental geometries on QM optimization, but they retain their experimental geometries if guanine N7 protonation is considered during optimization. Our study also reveals the role of guanine N7 protonation (i) in stabilizing important RNA structural motifs, (ii) in providing a framework for designing pH driven molecular motors and (iii) in providing alternative strategy to mimic the effect of post-transcriptional changes.

1 Introduction

The role of protonated nucleobases in shaping the folded three dimensional structures of nucleic acids, and in regulating the dynamics of functional RNAs, has been drawing substantial attention in recent literature.[1] Protonation of nucleobases is not thermodynamically favored in the physiological environment, since the pK_{a1} values of adenine, guanine and cytosine are usually 3-4 units away from neutrality.[2,3] Consequently, scientific activity in this area has focused on the exploration of the detailed mechanism of protonation induced activities of nucleobases and their functional roles.

Numerous examples of diverse activities of protonated nucleobases include the importance of **protonated cytosine** which is underlined by the role of (i) C(+)*G:C triples in formation of triple helical DNA which has putative regulatory roles,[4,5] (ii) intercalated C(+):C homodimers in the stabilization of DNA i-motifs[6] which are found in $(CCG)_n \cdot (CGG)_n$ trinucleotide repeats[7] whose expansion is correlated with the fragile X chromosome syndrome,[8] (iii) C8(+)*G12:C26 triple (PDB id 437D) in the construction of transcription regulatory pseudoknot which gets

[a]*Center for Computational Natural Sciences and Bioinformatics (CCNSB), International Institute of Information Technology (IIIT-H), Gachibowli, Hyderabad 500032, India; Email: abi_chem@iiit.ac.in*
[b]*Department of Spectroscopy, Indian Association for the Cultivation of Science, 2A&2B Raja S. C. Mullick Road, Jadavpur, Kolkata-700032, India; Email: spad@iacs.res.in*
[c]*Computational Science Division, Saha Institute of Nuclear Physics(SINP), 1/AF, Bidhannagar, Kolkata 700064, India; Email: dhananjay.bhattacharyya@saha.ac.in*

† Electronic Supplementary Information (ESI) available: [Computational details, NBO orbitals, context analysis of different base pairs, gas phase and solvent phase optimized geometries with inter-atomic distances and partial charges, details of interaction energy calculation, inter base pair geometry parameters, complete reference of Gaussian 09 package]. See DOI: 10.1039/b000000x/

destabilized when this triple is replaced by the isosteric U8:A12-U26 triple,[9] (iv) protonated C75 residue in activation of HDV ribozyme,[10] etc. **Protonated adenine** is known to form A(+):C base pair (isosteric with G:T/U wobble base pair) within a helical context and perform crucial structural and functional roles.[3,11–16] **Protonated guanine** residues have been reported to have pH driven switching potential with plausible applications in nucleic acid based nanotechnology.[17,18] Such a pH-driven DNA nanoswitch for responsive controlled release has recently been designed and demonstrated by Chen et al., where a cytosine rich single DNA strand has been used as the motor, which on changing the pH of the medium, switches to a folded quadruplex i-motif conformation stabilized by the formation of C(+):C base pairs.[19] It has also been demonstrated that the rates of RNA enzyme mediated catalysis are comparable to those of protein enzymes, when the functional groups of the bases involved are protonated.[10] Such protonated nucleobase mediated catalytic processes may take place in two different manners depending on whether the loaded proton is involved in any base pairing interactions (*Class I*) or not (*Class II*).[3,20] Paired protonated nucleobases of *Class I* participate in catalysis by participating in oxyanion hole formation (lysine-arginine type role). However, when the loaded proton is not involved in any base pairing type interaction (*Class II*), they can participate in general acid base catalysis (histidine like role).[3]

A major challenge, confronting a comprehensive study of protonated nucleobases, lies in their detection. This is because hydrogen atom coordinates are absent in X-ray crystal structures, and the characterization of exchangeable protons from NMR structures is rather ambiguous. For detecting protonated nucleobases belonging to *Class I*, we had earlier used BPFind software[21] to identify paired bases where, in addition to conventional inter base hydrogen bonds, there exist apparent hydrogen bond acceptors in close proximity – within hydrogen bonding distance. By considering these as putative protonated base pairs, we then evaluated the possibility of one of these apparent hydrogen bond acceptors as a potential donor because of 'invisible' protonation. Using this strategy, we had reported 18 different varieties of protonated base pairs from RNA crystal structures, and had carried out context analysis and quantum chemical investigations of their geometries and stabilities.[22] This strategy however restricts us to the detection of only *Class I* protonated bases. To the best of our knowledge, no comprehensive strategy has been proposed yet for detection of *Class II* protonated nucleobases, where the protonated edge is not involved in any base pairing interaction. Earlier we have shown that, in principle, protonation of a base could lead to (i) alteration of charge distribution and base pairing potentials of different edges and (ii) charge redistribution of the partner base.[18] Therefore, to investigate the implications of protonation at other than base pairing edges, *i.e.*, to detect some of the possible Class II protonation, we have followed the following steps — (i) short list a set of base pairs with protonable free edges, (ii) carry out quantum mechanical (QM) calculations to estimate the ground state electronic energy and charge distribution respectively with and without protonation of the free edge and (iii) infer regarding protonation and its effects by comparing the electronic properties of the systems in the presence and absence of the proton.

It is now widely accepted that, QM evaluation complements observations from thermodynamic experiments in understanding the nature of noncovalent stabilizing interactions such as hydrogen bonding and stacking in the context of biomolecules.[23–27] The experimentally measured thermodynamic stability in terms of free energy change with respect to some reference state, represents an overall effect of a large number of countervailing interactions. QM computations, on the other hand, helps us to accurately quantify the intrinsic stabilization due to some of these factors. Since, it is not possible to segregate these individual energy contributions from the overall free energy, the two approaches should be considered as addressing different aspects of the interactions and their values are not necessarily expected to be comparable.[28–31] In this study we have used QM methods to evaluate the impact of nucleobase protonation on the intrinsic stability (in terms of potential energy) of its base pairing interactions, even when the protonation site is not playing the role of a hydrogen bond donor, and have shown the distinct possibility of yet undetected protonation events within folded RNA structures.

In this context, there were three major considerations leading to the choice of N7 protonated guanine as the system for investigation. They are as follows,

1. Though among all the protonable sites of all the nucleobases, N7 position of guanine has the maximum protonation propensity in gas phase,[17,18] it is remarkable that, we found no references implying its occurrence in RNA structures. In fact, as we have earlier shown, both through qualitative theoretical considerations as well as through computational energy consideration, N7 protonated guanine cannot form any stable base pairing interaction using its N7 protonated hoogsteen edge.[17,18] It is also noteworthy that though Class II contexts have been implied, at least indirectly, for N3 protonated cytosine[32] and N1 protonated adenine[33] participating in general acid-base catalysis, the same cannot be said about N7 protonated guanine. That leaves us with the question of whether N7 protonated Guanine does occur in the Class II context in RNA structures.

2. The N7 of guanine, unless its hoogsteen edge is involved in base pairing, is exposed to the solvent and hence is easily susceptible to protonation. This is true even when the base pair lies within a double helical stretch, since the hoogsteen

edge remains solvent accessible through the major groove.[34]

3. Metal ion co-ordination and archaeosine modification at the N7 position of guanine have earlier been found to affect the base pairing potential of its WatsonCrick edge.[35–38] This supported our expectations regarding the possibility of Class II occurrence of N7 protonated guanine, where the protonation at N7 can induce similar effects on its electronic structure.

Available crystal structures of functional RNAs exhibit numerous examples of guanine base pairs, in addition to canonical WatsonCrick G:C base pairs, where the hoogsteen edge is not involved in base pairing interactions (a comprehensive list is available at RNABP COGEST database:[39] *http://bioinf.iiit.ac.in/RNABPCOGEST/*). Many of such base pairs are involved in important structural and functional contexts, such as, G:U W:W Cis base pair in group I self-splicing introns,[40,41] G:A S:H Trans base pair in Kink turn motif,[42] G:C W:W Trans base pair in preQ1 riboswitch,[43] etc. Therefore, in this study, we have selected those base pairs from RNA crystal structures, where guanine interacts with the partner base via its WatsonCrick or sugar edge and the hoogsteen edge remains free. The ground state optimized geometries of these base pairs have been calculated at M05-2X/6-311G+(2d,2p) level of theory for both models of these base pairs, respectively containing neutral and N7 protonated guanine. Comparing the ground state optimized geometries, MP2/aug-ccPVDZ level interaction energies and different electronic properties of these two models, we have shown that, consideration of guanine N7 protonation can explain the crystal geometry of G:C W:W Trans, G:rC W:S Cis, G:G W:H Cis and G:U S:H Trans base pairs,‡ which otherwise converge to alternate interacting geometries in ground state. Our study also reveals the role of guanine N7 protonation (i) in stabilizing important RNA structural motifs, (ii) in providing a framework for designing pH driven molecular motors and (iii) in providing alternative strategy to mimic the effect of post-transcriptional changes.

2 Methods

We have selected 167 crystal structures of RNA from the non redundant RNA crystal structure data set available at HD-RNAs database,[45] which (i) contains only structures with resolution better than 3.5 Å and (ii) excludes small synthetic RNA constructs by applying a lower cut off of 30 nucleotides for chain length. In these 167 structures, we have identified occurrences of all those base pairs, where guanine interacts via its WatsonCrick or sugar edge. To detect these base pairs with (i) two or more hydrogen bonds between the partner bases or (ii) only a single hydrogen bond between them, we have used the BPFind software[21] and our in-house tool INCAR,[46] respectively. The co-ordinates of a representative base pair of a particular geometry were extracted from the respective PDB files and were further edited for QM calculations by, (i) adding hydrogen atoms at appropriate positions and (ii) substituting the sugar moiety by a methyl group where the sugar atoms are not involved in hydrogen bonding interactions (as justified in the following section), using the molecule editing tool of GaussView5[47] package.

The selected crystal geometries of base pairs were relaxed on their respective potential energy surface using Gaussian09 package.† The optimization in gas phase was carried out using Truhlar's M05-2X functional[48] which was chosen based on its remarkable success in (a) describing dispersion interactions in a variety of molecular systems[49–51] and (b) predicting the structures and interaction energies between noncovalently bonded molecules of biological relevance,[52] including nucleobases.[53–59] The 6-311G+(2d,2p) basis set,[60] a split valance triple ζ augmented with two sets of d type polarization function for all non-hydrogen atoms and two sets of p type polarization function for hydrogen atoms, also including s-p diffused orbitals for non-hydrogen atoms, was used for ground state geometry optimization. Intrinsic stability (ΔE_{int}^{gas}) of the M05-2X/6-311G+(2d,2p) level ground state optimized geometries of isolated interacting base pairs were calculated using second order Møller-Plesset perturbation theory (MP2) with augmented correlation-consistent polarized valence only double ζ (aug-cc-pVDZ) basis set.[61] The basis set superposition error (BSSE) for the interaction energies was corrected via counterpoise method[62] using the same level of theory. The Hartree–Fock and correlation components of the MP2/aug-ccPVDZ level interaction energy were analyzed separately in order to quantify the contribution of electrostatic and dispersion energy, respectively, in stabilization of base pair. Further we have calculated the impact of a polar solvent environment (ε = 78.4) over these interaction energies (ΔE_{int}^{sol}) using a computationally efficient and robust conductor-like polarizable continuum model (CPCM),[63,64] which uses united atom topological model to define the atomic radii. This was found to be more appropriate for polar liquids, where the electrostatic potential goes to zero on the surface and is considered to be less sensitive to outlying charge error.[65] To obtain the solvent phase optimized geometries (at M05-2X/6-311G+(2d,2p) level of theory), we have included the solvent effect within the optimization subroutine using the same CPCM method. Interaction energy of the solvent phase optimized geometries was also calculated at MP2/aug-cc-pVDZ level of theory with BSSE corrections calculated at gas phase for the solvent phase optimized geometries and denoted as $\Delta E_{int}(Sol)$ instead of ΔE_{int}^{sol}. Details of the calculation steps are given in the section 'Computational details' (S1.2 of ESI†).

‡ Base pairs have been annotated using the nomenclature adopted by Koripella et al.[44] Along with that, if the sugar atoms of a residue are involved in the base pairing interaction, the base is prefixed with 'r', *e.g.*, G:rC W:S Cis

The wave function corresponding to the gas phase ground state optimized geometries at M05-2X/6-311G+(2d,2p) level was used to identify and quantitatively analyze the existing non covalent interactions within the base pairs, via atoms in molecules (AIM)[66] and natural bond orbital (NBO)[67,68] approach, respectively. NBO package[69] implemented in Gaussion 09 and the standalone AIMALL package (version 11.12.19)[70] were used respectively for NBO and AIM calculations, the relevent parameters for which are described elsewhere.[71,72] The bond paths detected by AIM analysis, associated with a positive slope of the electron density gradient ($\bigtriangledown^2\rho > 0$) at the bond critical point (where $\bigtriangledown\rho = 0$), were identified as non covalent interactions.[66] Among them, those which simultaneously satisfy the following criteria were identified as hydrogen bonds, — (i) the range of ρ values must be between 0.002 a.u. and 0.035 a.u. and (ii) the range of $\bigtriangledown^2\rho$ values must be within the range 0.024 a.u. and 0.139 a.u.[73–75] Interactions between filled Lewis and empty non-Lewis orbitals corresponding to such non covalent interaction detected by AIM analysis, were determined and energies of second order stabilization, $E^{(2)}$,[68,76] due to transfer of electron cloud from donor NBO(i) to acceptor NBO(j) were obtained using the equation,

$$E^{(2)} = q_i \left[\frac{(F(i,j))^2}{E_j - E_i} \right]$$

where q_i is donor orbital occupancy, E_i and E_j are energies of the orbitals i, j respectively, and F(i,j) is the off diagonal NBO Fock matrix element.

To study the effect of N7 protonation on the charge distribution and geometry of an individual guanine residue, we have compared the local geometry, dipole moment, local aromaticity and partial charges (from natural population analysis) of M05-2X/6-311G+(2d,2p) level optimized geometry of neutral and N7 protonated guanine with two different substitutions at N9 position: (a) ribose sugar and (b) methyl group. It was shown that, presence of the sugar moiety plays an important role in controlling the protonation dynamics of adenine.[77] Hence, we have modeled the guanine with sugar at N9 position. However, methyl substitution at N9 position of purines are known to reduce the computational effort without significantly affecting the properties of the real systems.[44,59,78]

Protonation at N7 position was reported to have a significant impact on the pyramidalization of the exocyclic amino group of guanine.[18] The extent of pyramidaliztion has been quantified by the improper dihedral angle (ψ) between the planes defined by N2-H22-H21 and H22-H21-C2, respectively, *i.e.*, $\psi = cos^{-1}[(\vec{r}_{N2-H22} \times \vec{r}_{N22-H21}) \cdot (\vec{r}_{H22-H21} \times \vec{r}_{H21-C2})]$, where $\vec{r}_{ij}$ is the vector connecting the atoms i and j.[79] To estimate the sugar puckering, we have calculated the pseudorotation angle (P) using the Altona and Sundaralingam method,[80]

$$\tan P = \frac{(\nu_2 + \nu_4) - (\nu_1 + \nu_3)}{2\nu_0(\sin(\pi/5) + \sin(2\pi/5))}$$

where, the ribose ring torsion angles are identified as: ν_0 about C2′-C3′, ν_1 about C3′-C4′, ν_2 about C4′-O4′, ν_3 about O4′-C1′ and ν_4 about C1′-C2′, respectively.

NICS(1)$_{zz}$,[81] which is the negative of the z-component of the magnetic field tensor calculated at 1Å above the geometrical center of an aromatic ring (in ppm unit), has been accepted as a good descriptor of the extent of delocalization of π electrons over the aromatic rings in contemporary literature.[82] We have calculated the values for the pyrimidine and imidazole rings of neutral and protonated guanine using the GIAO (Gauge Invariant Atomic Orbital)[83,84] method at the M05-2X/6-311G+(2d,2p) level. Dipole moment (μ) of the neutral species has been calculated following the standard protocol of Gaussian09 package.†

Details about the calculation procedure of the root mean square deviation (RMSD) between the crystal geometry and optimized geometries of the base pairs using VMD software[85] is given in section S1.1 in ESI†.

3 Results and Discussion

3.1 Effect of N7 protonation on charge distribution and on the geometry of guanine

Consistent with the earlier experimental[86,87] and theoretical[88–95] studies, involving guanosine carrying an electrophile at N7, we too have observed that protonation at N7 of guanine results in (i) elongation of glycosidic bond, and (ii) rotation of base around the glycosidic bond towards *anti* conformation (Table 1). Direct inclusion of solvent effects in the optimization routine, however, suppresses the extent of these two modifications (*e.g.*, for sugar substituted guanine $\Delta d1_{gas} \approx 0.03$Å, $\Delta\chi_{gas} \approx 21^\circ$ and $\Delta d1_{sol} \approx 0.02$Å, $\Delta\chi_{sol} \approx 15^\circ$). As expected, N7 protonation also shows a noticeable influence on the C2′-endo puckering of the sugar moiety (ΔP_{gas}=-2.2° and ΔP_{sol}=4.9°), as also reported earlier.[88] In addition, we have observed that, protonation at N7 promotes conjugation of lone pair of electrons of the exocyclic amino group with the ring π cloud. This can explain the observed enhancement of sp^2 character of the N2 nitrogen, characterized by, (i) shorter C2-N2 bond and significantly enhanced planarity of the amino group (significant reduction in improper dihedral), and (ii) the reduction in aromatic character of both the rings, characterized by the reduction in NICS(1)zz values (Table 1).

On the basis of natural population analysis (NPA), we have reported earlier[18] that, partial electronic charges on the polar sites of guanine get reduced on N7 protonation, and therefore, (i) the hydrogen bond donating potential of N1 and N2 gets improved while, (ii) hydrogen bond accepting potential of O6 and N3 is compromised. Though, both the factors, sugar substitution at N9

Table 1 Different geometric parameters are reported for gas phase and solvent phase (sol.) ground state optimized geometries of neutral (Neut.) and N7 protonated (Prot.) guanine with a substitution R (methyl/sugar) at N9 position. Bond lengths of the glycosidic bond, N9-C1′ (d1), and C2-N2 (d2) bond are reported in Å. The glycosidic torsion angle χ, pseudorotation angle (P)[34] and improper dihedral angle (ψ) at N2, (*i.e.*, N2-H22-H21-C2 angle) are reported in degree. Dipole moments (μ) of the neutral species have been reported (in debye unit). Partial charges at different base and sugar atoms, as obtained from natural population analysis, are reported in a.u. We have also reported NICS(1)$_{zz}$ values (in ppm unit) for both the pyrimidine (Pyr.) and imidazole (Imd.) rings of the gas phase optimized geometries.

R	State		Geometrical changes						NBO charges						NICS(1)$_{zz}$	
			N9-C1′	χ	P	C2-N2	ψ	μ	O6	N1	N2	N3	O2′	O3′	Pyr.	Imd.
methyl	gas	Neut.	1.45			1.37	28.6	7.2	-0.623	-0.639	-0.811	-0.606			-6.8	-25.7
		Prot.	1.46			1.33	0.2	–	-0.576	-0.624	-0.763	-0.599			-5.2	-22.7
sugar	gas	Neut.	1.44	-130.6	158.8	1.37	29.0	9.4	-0.625	-0.638	-0.811	-0.601	-0.735	-0.632	-6.5	-22.5
		Prot.	1.46	-151.6	156.6	1.34	1.8	–	-0.584	-0.624	-0.767	-0.589	-0.731	-0.633	-5.3	-20.4
methyl	sol.	Neut.	1.45			1.35	23.0	10.0	-0.707	-0.623	-0.799	-0.618				
		Prot.	1.46			1.33	3.4	–	-0.662	-0.612	-0.766	-0.609				
sugar	sol.	Neut.	1.44	-130.8	159.9	1.35	21.9	12.5	-0.703	-0.621	-0.795	-0.617	-0.759	-0.640		
		Prot.	1.46	-146.3	164.7	1.33	1.0	–	-0.660	-0.612	-0.764	-0.607	-0.753	-0.639		

and consideration of solvation, increase the charge separation (as indicated by higher dipole moment (μ) values), it is interesting to note that, extent of modification of the hydrogen bonding potentials of the base atoms is not affected by consideration of either solvent environment or of the N9 substituent. For example, the change in partial charges (Δq) over N1, due to protonation at the N7 position of guanine, is found to be similar for different combination of N9 substitution and environment: Δq_{gas}^{methyl} = -0.015, Δq_{gas}^{sugar} = -0.014, Δq_{sol}^{methyl} = -0.011 and Δq_{sol}^{sugar} = -0.009 in atomic units. It is pertinent to note that, while N7 protonation significantly affects the charge distribution of polar atoms of guanine, its effect is minimal on the polar atoms (O2′ and O3′) of the sugar moiety (Δq<0.01a.u.). Conversely, the charge distribution on the polar atoms of guanine, respectively, due to either N7 protonation or solvation is not significantly affected by the substitution of the sugar moiety at N9 by methyl group. These observations justify our strategy of using computationally lighter N9 methylated models wherever appropriate.

3.2 Effect of N7 protonation of guanine on canonical and wobble base pairing interactions

As discussed above, N7 protonation redistributes the charges on the polar atoms of guanine, thereby reducing the hydrogen bond acceptor potential of O6 and enhancing the hydrogen bond donor potentials of N1 and N2. What would this mean to the canonical G:C and the wobble[96] G:U W:W Cis base pairs, where the N7 atoms of the guanines are accessible to water, and hence may easily be protonated? It may be noted that, though these base pairs are predominantly present in double helical regions, where access to the N7 of guanine may be somewhat constrained because of the narrowness of the major groove of the A-form RNA double helices, access for small water molecules is not an issue. This is particularly so, since in RNA the double helical regions typically occur only in short stretches.

Figure 1 shows the hydrogen bonding interactions (as identified by AIM analysis) for the M05-2X/6-311G+(2d,2p) level ground state optimized geometries of neutral (B,E) and N7 protonated (C,F) canonical G:C and wobble G:U base pairs. Each hydrogen bonding interaction is associated with a stabilization energy (E(2)) due to charge delocalization between the corresponding acceptor and donor NBO orbitals. Change in the E(2) value of a particular hydrogen bond on protonation, provides a quantitative estimate of the modification in the strength of that hydrogen bond. As expected for the G:C W:W Cis system, we have observed, (i) protonation induced weakening of the inter base hydrogen bonding where O6 of guanine acts as a hydrogen bond acceptor (ΔE(2) = -8.3kcal/mol) and (ii) protonation induced strengthening of the other two hydrogen bonds, where N1 and N2 of guanine acts as a hydrogen bond donor (ΔE(2) = 9.7 and 12.8kcal/mol, respectively). Corresponding changes in the energy gaps between the NBOs of the hydrogen bond donors and acceptors are shown in Figure S1, ESI†. The significant conclusion arising out of these computations is that, N7 protonation results in an improved base pairing interaction without the disruption of its inherent geometry (RMSD < 0.3Å). The small RMSD (0.26Å) is possibly largely due to the increase in planarity (Buckle = 0.20, Open angle = 1.94 and Propeller twist = 0.24), characteristic of improved base pairing, as can be seen in Table S2, ESI†.

Similar trends in the protonation induced modification of inter base pair hydrogen bond interactions are observed in G:U W:W Cis wobble base pair. However, though noticeable, the protonation induced gain in stability of the GU wobble pair is not as remarkable ($\Delta\Delta E_{int}^{gas}$ = 6.4 and $\Delta\Delta E_{int}^{sol}$ = 0.4 in kcal/mol) as in the GC case ($\Delta\Delta E_{int}^{gas}$ = 10.6 and $\Delta\Delta E_{int}^{sol}$ = 1.8 in kcal/mol). This is because protonation mainly impacts the HF component (dominated

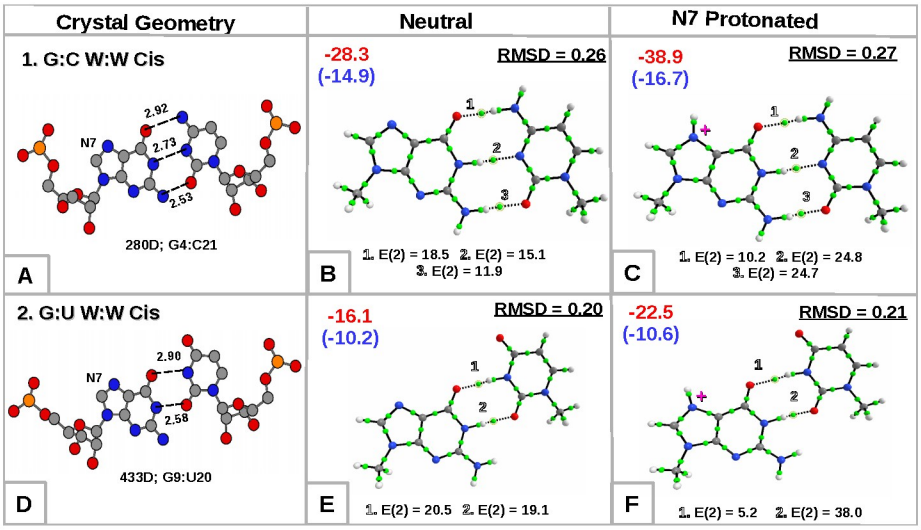

Fig. 1 Crystal geometry (A & D) and the ground state optimized geometries (in gas phase) of (i) neutral (B & E) and (ii) N7 protonated (C & F) base pairs are shown for (1) canonical G:C and (2) wobble G:U W:W Cis base pairs. Heavy atom distances (in Å) for possible hydrogen bonding interactions are provided for crystal geometries. The same for optimized geometries are shown in Figure S3 of ESI†. For the ground state optimized geometries, the BCPs are shown as green spheres and the bond paths corresponding to BCPs with $\nabla^2\rho > 0$ (noncovalent interactions) are shown as dotted lines. Stabilization energy (E(2)) corresponding to noncovalent interactions are reported in kcal/mol. Intrinsic interaction energies of the base pairs ΔE_{int}^{gas} (without solvent screening) and ΔE_{int}^{sol} (with solvent screening) are reported (in kcal/mol) in red and blue color respectively. We have also reported the root mean square deviation (RMSD) between the crystal geometry and the optimized geometries. These results remain consistent when solvation effects are considered within the optimization subroutine (Table S1, S2 and Figure S4, ESI†.)

by electrostatic interactions) of the MP2 level interaction energy, rather than the correlation component (Table S1, ESI†); and contribution of the HF component in the interaction energy of G:U W:W Cis pair is comparatively lower (%HF = 69.4) than that of the G:C W:W Cis pair (%HF = 78.0).¶ Be that as it may, our observations strongly suggest that, the interaction energies of both canonical G:C and wobble G:U base pairs increase on guanine N7 protonation. Such observations open up possibilities of designing thermostable DNA/RNA duplexes at low pH environment where guanine can exist in its N7 protonated state.

3.3 Base pairing geometries, that are unstable in the absence of guanine N7 protonation

During QM level geometry optimization, in the absence of the crystal environment, structures of noncanonical base pairs in RNA usually converge to their local minima, which are considerably close to their respective crystal structure geometries.[97] However, there are exceptions, where geometry optimization leads to significantly large deviation from that observed in their respective crystal structure contexts. In such cases the optimized geometry is often associated with a change in the interacting edge. It has been shown in several such cases that the crystal geometries, away from their intrinsic optimal geometries, owe their stabilities to tertiary interactions occurring in their respective crystal context.[98] We propose that, even in the absence of other tertiary interactions, protonation induced charge redistribution may have sufficient influence on the base pairing interactions, which are required to maintain the crystal geometry. In this study, we have identified four such systems (Figure 2) which have important structural and functional roles.

1. **G:C W:W Trans :** The reverse WatsonCrick Guanine-Cytosine base pair, commonly known as the Levitt base pair,‖ is a conserved base pair which mediates the interaction between the D-loop and the V-loop in tRNAs and is associated with vital functionalities.[101] G:C W:W Trans base pair is also found in (i) helix turn regions and kissing loop motifs of 23s rRNA[71] and (ii) in class II preQ1 riboswitch, where it plays a key role in the recognition of the metabolite preQ1.[43] Quite

¶ It is to be noted that, in the absence of a dielectric medium, gas phase interaction energies of the protonated systems are highly dominated by the electrostatic interaction which is further reflected in its %HF component (>80% for GC systems, Table S1, ESI†). Therefore, to compare the interaction energies of a neutral and a protonated system, we have implemented a solvent screening effect while calculating the interaction energy of the base pairs.

‖ In the context of tRNA, existence of a tertiary interaction between nucleotides G15 and C48 was first predicted by Levitt in 1969,[99] which was later identified as a G:C W:W Trans base pair.[100]

remarkably, the geometry of the Levitt base pair, on isolated optimization, converges to a significantly different geometry compared to its recurrently occurring crystal geometry.[71,102] Oliva et al. have shown earlier that, metal ion co-ordination or archaeosine modification at N7 position of guanine may stabilize the structure of the Levitt base pair in the context of tRNA.[35,36] Further, they have observed that, approximately two-third of the detected G:C W:W Trans base pairs in RNA crystal structures are also engaged in additional hydrogen bonding interactions, which include interaction with surrounding water molecules and phosphate groups, and formation of base triples and quadruples.[98] Our analysis of the non redundant set of 167 RNA crystal structures also reveal similar trends, where among the 118 occurrences of the Levitt base pair, it is found to be involved in extra hydrogen bonding interactions with neighboring bases (12 times), with phosphate groups (51 times) and with water molecules (30 times), along with metal ion coordination at the hoogsteen edge of guanine (13 times).

2. **G:rC W:S Cis :** This base pair has also been reported to change its edges of interaction, on geometry optimization, and, to converge to the canonical G:C W:W Cis geometry.[103] Nevertheless, we have found that, it is a conserved base pair between C366 and G394 residues of 16S rRNA and occur at the end of a helical stem, where it remains stacked between a canonical GC pair and a flanking cytosine base (as shown in Figure S2, ESI†). Earlier, for the case of Levitt base pair, it has been shown that such stacking interactions are not sufficient for stabilizing the inherently unstable crystal geometry.[35] Further, analysis of RNA crystal structures, unlike that in the Levitt base pair case, rules out the possibility of participation of other higher order interactions (base triple, quadruple, etc.) in stabilizing the W:S Cis geometry. In fact, we have observed only one instance in the non-redundant dataset where the G:rC W:S Cis base pair is involved in extra hydrogen bonding interactions with a third base.**

3. **G:G W:H Cis :** It has been reported that, on both ab-initio as well as on DFT based optimization, the geometry of G:G W:H Cis base pair changes to an alternative G:G W:H Trans geometry.[104] Our context analysis reveals that, in 88 among the 123 occurrences of the G:G W:H Cis base pair in the non redundant set of 167 RNA crystal structures, it occurs within a higher order structure, such as, in a base triple or quadruple. In most of the cases, the G:G W:H Cis pair is found as a part of a GGC triple, where the second guanine of the G:G pair forms a G:C canonical pair via its Watson-Crick edge. For example, in G9*G12:C23 triple in tRNA-Leu which stabilizes the D-arm, G44*G10:C25 triple of tRNA-Trp which mediates the interaction of the D-arm with the variable loop, G251*G254:C272 triple which occurs frequently in 16S rRNA, etc. In other triples, the second guanine of the G:G pair is found to participate in a G:A W:W Cis type base pairing, e.g., in the G1182*G1160:A1176 motif that occurs frequently in 16S rRNA. In 15 cases, the second base of these GGC and GGA triples further interact with an adenine with its sugar edge to form a G:A S:W Trans or G:A S:S Trans type base pair, respectively. It is expected that, in these triples and quadruples, the neighboring interactions account for the stability of the intrinsic geometry of the G:G W:H Cis pair, that is absent in QM level geometry optimization protocols. Interestingly, we have identified 35 instances, where the G:G W:H Cis pair is found to be stabilized solely by two inter base hydrogen bonds, as observed in some conserved positions of 16S rRNAs, 23S rRNAs and eukaryotic TPP specific riboswitch in *Arabidopsis thaliana*. This appears to open up the possibility of existence of other stabilizing factors apart from the hydrogen bonding interactions with the neighboring bases. A comprehensive list of all such occurrences is provided in section S7 of ESI†.

4. **G:U S:H Trans :** G:U S:H trans base pair is stabilized by only one hydrogen bond between N2 of the Guanine and O4 of the Uracil. During unconstrained geometry optimization, its interaction geometry changes to S:H Cis form, where an additional hydrogen bond forms between C5 of Uracil and N3 of Guanine.[105] During database analysis we have identified 75 instances of single hydrogen bonded G:U S:H Trans base pair in the non-redundant dataset of 167 RNA crystal structures, out of which 73 instances are part of some higher order interaction.

We have observed that, ground state optimized geometries of these four base pairs diverge significantly from their respective crystal geometries (Figure 2) and are associated with high RMSD values. In the absence of the solvent effect within the optimization subroutine, we observe that the base pairs converge to an entirely different base pairing geometry characterized by alteration of the edge of interactions, for example, the G:G W:H Cis base pair (Figure 2.G) is converged to G:G W:W Trans base pairing geometry on gas phase optimization (Figure 2.H). Consequently, the gas phase optimized geometries are highly nonisosteric with the crystal geometry, as suggested by the remarkable variation in C1′-C1′ distances (Table S2, ESI†). Interestingly, except for in the case of G:U S:H Trans pair (Figure 2.J), optimization in the solvent phase does not lead to any major

** In crystal structure of FMN riboswitch (pdb id: 3F2Q) a base triple is observed where, Watson-Crick edge of G28 residue interacts with S edge of C13 residue with cis glycosidic bond orientation and the sugar edge of G28 interacts with hoogsteen edge of A14 to form a G:A S:H trans base pair.

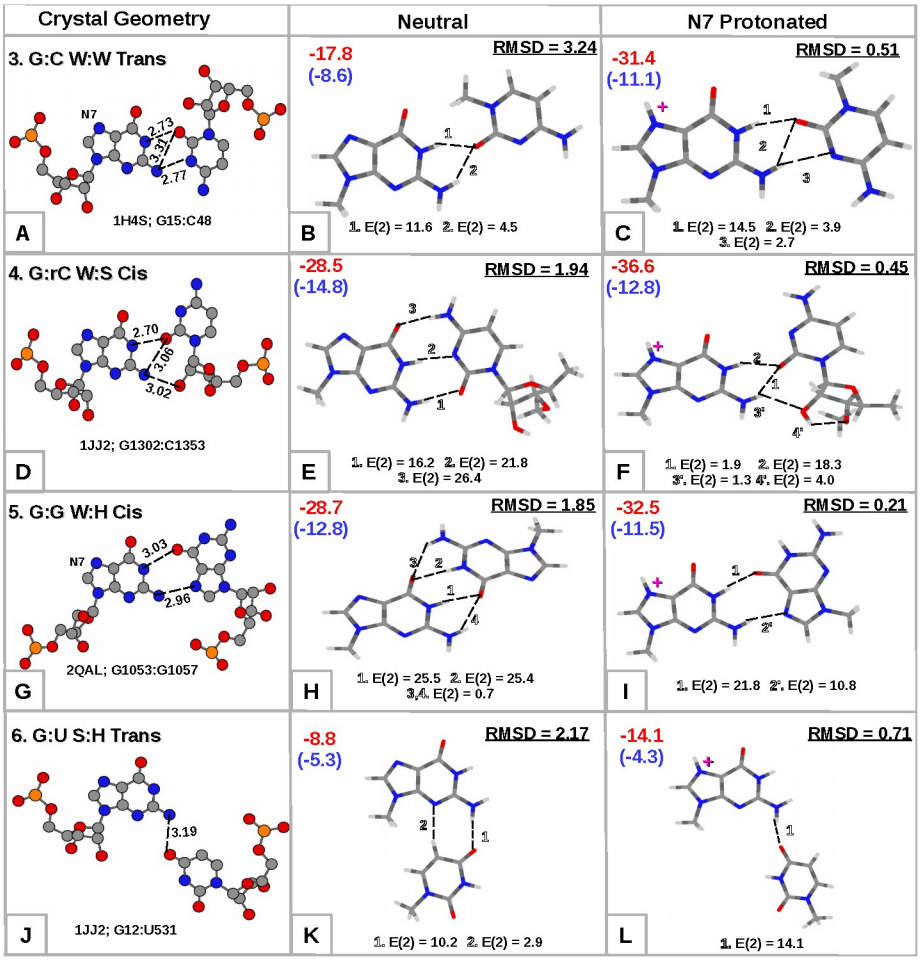

Fig. 2 Crystal geometry (left column), gas phase optimized geometry as neutral system (central column) and guanine N7 protonated system (right column) are shown for G:C W:W Trans (A,B & C), G:rC W:S Cis (D,E & F), G:G W:H Cis (G,H & I) and G:U S:H Trans (J,K & L) base pairs. Figures are annotated as in Figure 1.

change in the edges involved in the interactions. The solvent phase optimized geometries are however highly propeller twisted in all four cases (Table S2, ESI†). Such observations highlight the importance of the tertiary interactions in which these four base pairs are involved in the crystal environment.

In this context it is noteworthy that, even in the absence the crystal environment, consideration of only the guanine N7 protonation is sufficient to maintain the crystal geometry of all these base pairs (Figure 2). As we observe, on protonation at N7 of guanine, the optimized geometries remain isosteric with the corresponding crystal geometries (change in C1′-C1′ distances < 0.5Å) and exhibit remarkably small RMSD with the same. Such remarkable consequences of guanine N7 protonation can be explained in terms of the resultant charge redistribution leading to, (i) stronger hydrogen bond interaction via N1 and N2 of guanine and (ii) enhanced sp^2 character of N2 of guanine, which reduces the electrostatic repulsion with the neighboring electronegative atoms of the partner base.

Our observations highlight the possibility of occurrence of N7 protonated guanine within the crystal structures of RNA, which are involved in apparently unstable base pairing through their WatsonCrick and sugar edges. For example, among 1113 instances of detected G:C W:W Trans base pair in the 1487 available RNA structures (as considered by Chawla et al.), 375 are not involved in any other extra hydrogen bonding interactions.[98] Our results suggest that, for those isolated Levitt base pairs, guanine N7 protonation may be considered as a potential stabilizing factor. Interestingly, in G:rC W:S Cis base pair, the solvent effect alone, appears to be sufficient to maintain the crystal geometry (Figure S4 and Table S2, ESI†). This is consistent with the results of our context analysis which shows G:rC W:S Cis base pair, unlike in the other three cases, is not

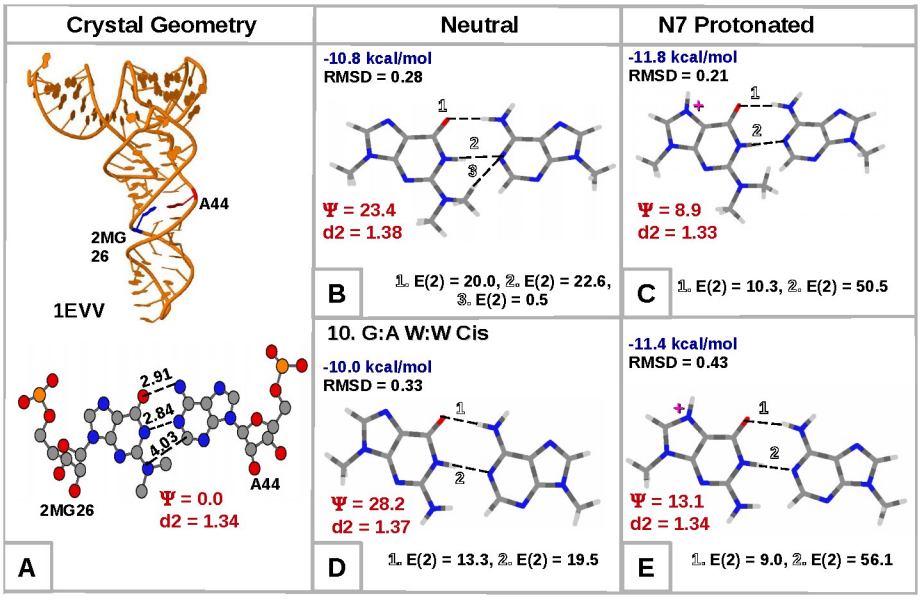

Fig. 3 In the left column (A) we have shown (i) the folded cartoon structure of tRNA with highlighted G26 (red) and A44 (blue) resides and (ii) the crystal geometry of N2 methylated G26 residue interacting with A44 residue in W:W Cis geometry. In the central and right columns, we have shown the ground state optimized geometries of the neutral and guanine N7 protonated G26:A44 pair with (B & C) and without (D & E) the methyl modification. Figures are annotated as in Figure 1.

involved in any other higher order interactions. These four base pairs also constitute a potential test bed for designing pH driven molecular motors or switches.

It is interesting to note that, among the two differently optimized geometries shown in Figure 2, ΔE_{int}^{sol} for system 4, 5 and 6 are higher for the neutral species than for the protonated species. This reflects a reversal in the trend observed in almost all other systems, where the protonated species show a higher value of ΔE_{int}^{sol}. Since the ΔE_{int}^{gas} values do not display any such discrepancies, we investigated this feature by trying to understand the variation in the effect of solvation due to the variation in ΔE_{HF}^{gas} and ΔE_{corr}^{gas} components for these systems. From Figure S5 of ESI†, for systems 4 and 6, it is easy to see that the ΔE_{HF}^{gas} drives the stability of the protonated species over neutral species to the extent that $\Delta\Delta E_{HF}^{gas} > \Delta\Delta E_{int}^{gas}$. In both these cases, the correlation components favor the neutral species. We therefore observe higher ΔE_{int}^{sol} for the neutral compared to the protonated species, where the solvent correction essentially reduces the ΔE_{HF}^{gas} contribution, while enhancing the correlation contribution. We see a similar situation in system 12 (section 3.7), where the ΔE_{int}^{sol} values of the protonated and neutral species are more or less equal. In the case of system 5, $\Delta\Delta E_{HF}^{gas}$ is less than $\Delta\Delta E_{int}^{gas}$. However, the latter itself is very small (3.8 kcal/mol). In this case, the relative reduction of HF component and enhancement of the correlation component respectively seem to be responsible for the reversal of trend. Here too a similar situation occurs in system 13 (section 3.7), where the ΔE_{int}^{sol} values of the protonated and neutral species are nearly the same. Detailed understanding of the reason behind such variation in the magnitude of different components is beyond the scope of this work. However, for the systems 4–6, the fact that the geometry of the neutral species is very different from the protonated species may have an important role to play in this context.

3.4 Blocking tertiary interaction potential of the exocyclic amino group of guanine in G:A W:W Cis base pair

The G:A W:W Cis base pair is one of the top ten most frequently occurring noncanonical base pairs in RNA (347 times in our non redundant dataset). It is highly nonisosteric with canonical base pairs, as characterized by larger C1′-C1′ distance (Table S2, ESI†). However, crystal structure and phylogenetic analysis reveals that G:A W:W Cis pairs occur frequently at the end of canonical helices, where they participate in the interfaces between the helices and other motifs such as, internal loop, bulge, etc.[106] It was also found that the geometry of the G:A W:W Cis base pair is inherently propeller twisted and that the sp^3 hybridized amino group of guanine participates in out of the plane hydrogen bonding interactions with the neighboring residues, by positioning itself out of the base pairing plane. G:A W:W Cis interaction has also been reported to be highly conserved in sequence, and a covariation with self isosteric A:G W:W Cis pair or other isosteric W:W Cis

Table 2 Change in the improper dihedral angle and C2-N2 bond length of the exocyclic amino group (N2) of guanine within a G:A base pair due to protonation at N7 of guanine. The amino group has two hydrogen atoms — *cis* (parallel to glycosidic bond) and *trans* (parallel to the carbonyl oxygen at O6). Contribution of s–character and p–character towards the total hybridization state of the N2 nitrogen as detected by NBO analysis corresponding to the two (*cis* and *trans*) N-H bonds are given. For G:A W:W Cis pair, the analysis was performed over N-C bonds.

				cis hydrogen			*trans* hydrogen		
Base pair		ψ	C2-N2	%s	%p	total	%s	%p	total
G:A W:W Trans	Neut.	33.2	1.38	26.2	73.5	sp 2.8	25.5	74.2	sp 2.9
	Prot.	0.4	1.33	30.0	69.8	sp 2.3	31.8	68.1	sp 2.1
G:A W:H Cis	Neut.	29.8	1.37	26.8	72.9	sp 2.7	26.7	73.0	sp 2.7
	Prot.	14.0	1.34	29.3	70.5	sp 2.4	29.9	69.9	sp 2.3
G:A W:H Trans	Neut.	32.0	1.38	25.9	73.9	sp 2.8	26.4	73.3	sp 2.8
	Prot.	1.6	1.33	32.3	67.5	sp 2.1	29.1	70.8	sp 2.4
G:A W:W Cis	Neut.	28.2	1.37	27.0	72.7	sp 2.7	27.2	72.6	sp 2.7
	Prot.	13.1	1.34	30.8	69.2	sp 2.2	30.5	69.4	sp 2.3
G:G W:W Trans	Neut.	18.6	1.35	28.6	71.2	sp 2.5	29.5	70.3	sp 2.4
	Prot.	0.3	1.33	30.1	69.8	sp 2.3	31.2	68.6	sp 2.2

pairs (that lack the NH_2 group at C2 position) takes place only when the NH_2 group of guanine is not involved in other hydrogen bond interactions (nonparticipating G:A W:W Cis base pairs).[106] Thus, where covariation is observed, it is possible that the amino group does not maintain its inherent pyramidal shape. Of course, due to the absence of the hydrogen atom coordinates, the extent of pyramidalization of the amino group can not be detected within the crystal structure. Our QM level ground state geometry optimizations reveal that (Figure 3 D), the exocyclic amino group of neutral G:A W:W Cis pair, as expected, exhibits a nonplanar geometry accompanied with a longer C2-N2 bond (d2 = 1.37Å) and a higher degree of pyramidalization (ψ = 28.2°). However, protonation at the guanine N7 position leads to an increase in sp^2 hybridization of the exocyclic amino group (Table 2) as characterized by a shorter C2-N2 bond (d2 = 1.34Å) with a partially planar (ψ = 13.1°) disposition of the amino group (Figure 3 E). So, in principle, it is possible that there are several nonparticipating G:A W:W Cis pairs where the guanine is N7 protonated and is associated with a planar orientation of the amino group. Further support can be obtained by looking at the crystal structures of G:A W:W Cis pair, where guanine's N2 position is methylated. It is to be noted that, the methyl substitution allows us to detect the extent of pyramidalization of the amino nitrogen within the crystal structure.

One interesting example is found in tRNA, where the G(26):A(44) W:W Cis base pair is not involved in any kind of neighboring or tertiary contacts via the NH_2 group of guanine and therefore shows covariation with A(26):G(44) and other W:W Cis pairs like A:A, A(+):C, G:U, G:C, etc.[106] At the same time, in many higher organisms (eukaryote), G26 residue undergoes a post transcriptional modification (methylation) at N2 position. In the non redundant dataset, there are three examples (PDB Id: 1EHZ, 1EVV and 1YFG) where both the hydrogens at N2 of G26 are substituted by methyl (Figure 3.A). The modified G:A W:W Cis pair (Figure 3.B) is only 0.8 kcal/mol more stable than the normal base pair (Figure 3.D). Further, we have observed that, in crystal environment, such double methylation at N2 induces the amino group to obtain a sp^2 geometry characterized by shorter C2-N2 distance (d2=1.34Å) and planarity of the amino group (ψ = 0°). For example, d2=1.36Å, 1.34Å and 1.33Å and ψ = 1.83°, 0° and 0° for double methylated G26 residue found in 1EHZ, 1EVV and 1YFG PDB files, respectively. In the absence of the crystal environment (Figure 3.B), the optimized geometry of the modified 2M2G:A W:W Cis pair maintains the sp^3 geometry of the amino group with accompanying pyramidalization (ψ = 23.4°) and longer C2-N2 bond (d2 = 1.38Å). However, invoking of protonation at N7 position of guanine, results in a sp^2 hybridized geometry of the exocyclic nitrogen (d2 = 1.33Å and ψ = 8.9°) and the optimized geometry is significantly close to the crystal geometry (RMSD = 0.21Å). Therefore, our results suggest that, in a G:A W:W Cis pair, N2 methylation in conjunction with N7 protonation of guanine can play significant roles in restricting the pyramidalization of the exocyclic amino group of guanine. In the absence of post-transcriptional changes like methylation, especially in prokaryotes, N7 protonation of guanine alone can exhibit similar effects.

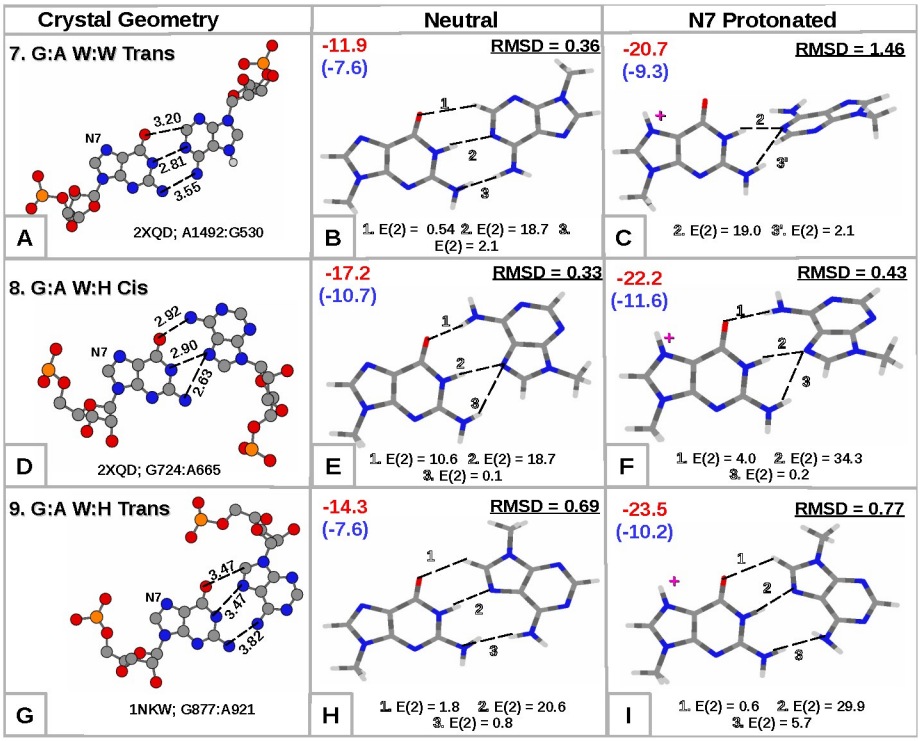

Fig. 4 We have shown the crystal structure (left column) and gas phase optimized geometry of the (i) neutral (central column) and (ii) guanine N7 protonated (right column) G:A base pairs, which undergo a geometric switching due to protonation at N7 of guanine. Figures are annotated as in Figure 1.

3.5 Switching action based on protonation driven depyramidalization of exocyclic amino group

We have discussed how protonation at N7 position of guanine enhances the sp^2 character in the exocyclic amino group (N2) in guanine which, in turn, forces the amino hydrogen atoms to be coplanar with the pyrimidine ring of guanine. Such a protonation induced change in local geometry may result in a protonation mediated switching mechanism involving certain G:A base pairs as shown in Figure 4. In all the three cases, protonation causes larger deviation in gas phase optimized geometry with respect to crystal geometry with only minor perturbation of the C1′-C1′ distance (Table S2, ESI†). In the first two systems, the most significant change in geometry is noticed in terms of larger propeller twist, and, in the G:A W:H Trans base pair, protonation causes a highly buckled geometry. Table 2 explains how, on protonation at N7 position of guanine, the sp^3 hybridized state of the nitrogen atom of the exocyclic amino group transforms into a sp^2 hybridized state which is characterized by the shortening of C2-N2 bond and depyramidalization of amino group. It is to be noted that, in G:A W:W Trans (Figure 4.A) and G:A W:H Trans (Figure 4.G), the pyramidalized amino group of the neutral base pair acts as a hydrogen bond acceptor (labeled as 3 in Figure 4.B and 4.H). Role of such amino-acceptor interactions in stabilizing noncanonical base pairs[105,107] and stacking interactions between dinucleotide steps[108] have been highlighted earlier. Forced depyramidalization of the amino group due to N7 protonation, therefore, interrupts such amino-acceptor type interactions in G:A W:W Trans and G:A W:H Trans, causing a steric hindrance and electrostatic repulsion with the amino group of adenine (N6). In G:A W:H Cis, similar electrostatic repulsion takes place between the hydrogen at C8 position of adenine and hydrogens of the guanine amino group (Figure 4.F). As a result, protonation drives the systems to new geometries which are non planar, and yet, are more stable on the potential energy surface than their respective neutral geometries. The protonation driven buckling of G:A W:H Trans base pair, however, facilitates formation of a stronger hydrogen bonding interaction ($\Delta E(2)$ = 4.9kcal/mol) between N2 of guanine and N6 of adenine (labeled as 3 in Figure 4.I). The gain in stability due to protonation driven switching of geometry is therefore maximum for G:A W:H Trans (2.6kcal/mol). Such protonation driven geometric switching does not interrupt the backbone orientation, but induces some important functional groups of the participating bases (N6 of adenine in G:A W:W Trans and G:A W:H Trans, N2 of guanine in G:A W:H Cis) to flip

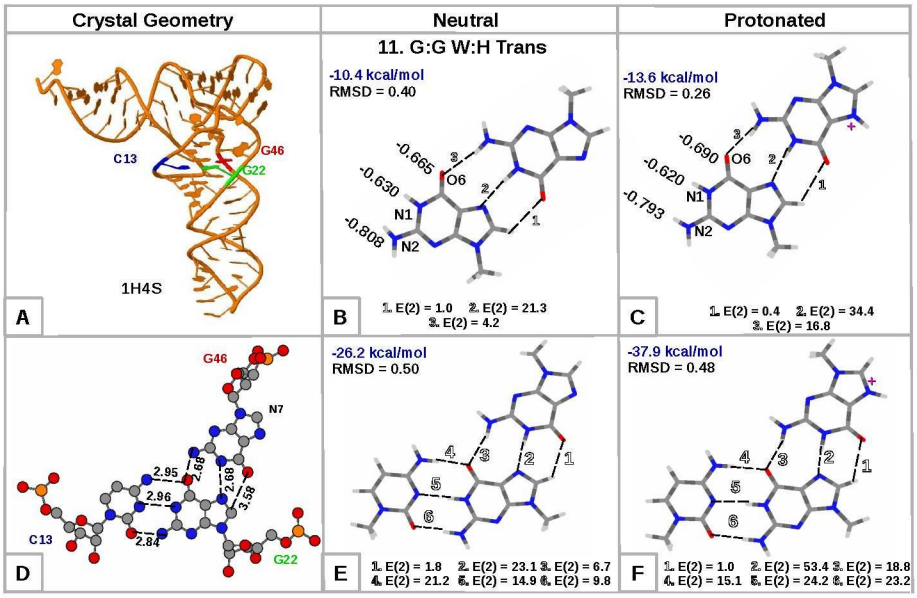

Fig. 5 (Top row) G46*G22:C13 base triple in folded structure of tRNA is highlighted (A). Optimized geometries of neutral (B) and protonated (C) G46:G22 W:H Trans pair along with the NBO charges at the polar sites of the free edges are given in atomic units. (Bottom row) Crystal geometry of the G46*G22:C13 base triple is shown (D) along with the analysis of the effect of N7 protonation on its geometry and stability (E & F). Figures are annotated as in Figure 1.

out from the base pairing plane, and thus possibly participate in out of the plane interactions. We have detected 3 instances of G:A W:H Cis base pair among its 22 occurrences in our non-redundant set, where atleast one or more values of the rotational parameters (Buckle, Open and Propeller Twist) are more than two standard deviation away from the average value and therefore resemble the N7 protonation induced highly nonplanar geometry. The base pairs G68(X):A126(X) of 3DIO, G169(A):A124(A) of 1U9S and G2447(A):A2451(A) of 3UYE are some such examples. As reported in the RNABP COGEST database,[39] the other two base pairs (G:A W:W Trans and G:A W:H Trans) do not occur frequently in the non-redundant dataset. Hence, it is possible that protonation at N7 of guanine acts as a trigger for such a switch in their geometries. However, one should not undermine the fact that, the geometries of the base pairs in the X-ray structures may be affected also by other factors, such as the overall structural context, water bridges, ions, packing, etc.

It may be noted that, since the depyramidalization is accompanied by charge separation between the N2 and the aromatic ring, one would expect that the base pair would show similar non planarity when the optimization is done in a high dielectric medium (Figure S4, ESI†).

3.6 Modification in hydrogen bonding potential of the free edges of the partner base in base pairs involving N7 protonated guanine

N7 protonation of guanine also has a significant influence on the charge distribution of the partner base which pairs with the WatsonCrick edge of the protonated guanine. Protonation at N7 position of the first base (guanine) naturally causes a depletion of electronic partial charges over the free (not involved in base pairing) polar sites of the second base (Figure S3, ESI†). This modifies the hydrogen bonding potential of the free edges of the second base in similar fashion in all but one out of 14 base pairs studied. It gets enhanced for hydrogen bond donors and reduced for hydrogen bond acceptors. The lone base pair, G:G W:H Trans, which shows a distinct exception in the trend, has the hydrogen bonding potential of all the three polar sites of the free WatsonCrick edge of the second guanine undergoing a significant improvement ($\Delta q^{(O6)}$ = +0.025, $\Delta q^{(N1)}$ = -0.010, $\Delta q^{(N2)}$ = -0.015, Figure 5.B and 5.C). Interestingly, G:G W:H Trans base pair is a conserved interaction between G46 and G22 residues of tRNA, where G46 residue is a part of the loop, joining the anticodon domain and T-domain in tRNA. The G22 residue further interacts with C13 residue, via its WatsonCrick edge, and constitutes a conserved canonical closing base pair of the loop of the D-domain (Figure 5.A) of tRNA. We have observed that (Figure 5.E and 5.F), stability of this conserved G46*G22:C13 triple is highly dependent on the protonation state of the N7 position of G46 ($\Delta\Delta E_{int}^{sol}$ = -

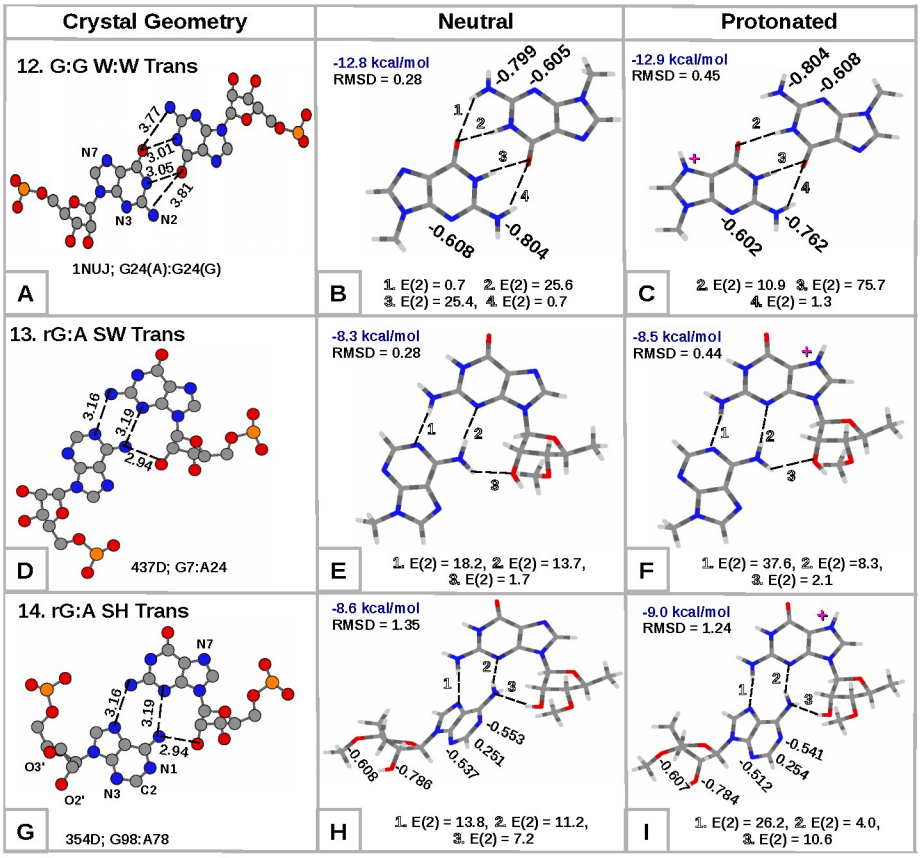

Fig. 6 Crystal geometry and optimized geometry of neutral and protonated G:G W:W Trans (A,B & C), rG:A S:W Trans (D,E & F) and rG:A S:H Trans (G,H & I) are shown. For G:G W:W Trans base pair, the NBO charges at the polar sites of the free edges are also given in atomic units. Figures are annotated as in Figure 1.

11.7kcal/mol). Such a remarkable improvement is a consequence of two factors, (i) protonation induced charge redistribution in G46, which improves the strength of the hydrogen bonds in the G46:G22 pair (labeled as 2 and 3 in Figure 5.E and 5.F) and (ii) charge redistribution in G22 due to pairing with N7 protonated G46, which improves the strength of the hydrogen bonds in the G22:C13 pair (labeled as 5 and 6 in Figure 5.E and 5.F).[††] Therefore, by tuning the protonation state of N7 position of guanine, stability of higher order interactions may be modulated.

[††] It is to be noted that, protonation at G46 lowers the E(2) value corresponding to the fourth hydrogen bond [(C13)N4-H···O6(G22)] by 6.1kcal/mol, but the damage is compensated by the remarkable rise of E(2) values corresponding to the hydrogen bonds which involve the exocyclic amino group of guanine as donor (3 and 6 in Figure 5).

3.7 Base pairing interactions that are unaffected by protonation at N7 position of guanine

Earlier we have discussed that, protonation at N7 position does not have a major influence on the charge distribution of the sugar edge of guanine. Therefore, base pairs involving sugar edge of guanine are not expected to experience any large change on N7 protonation. Figure 6 explains the effect of N7 protonation on two such base pairs, rG:A S:W Trans (Figure 6.D) and rG:A S:H Trans (Figure 6.G). Among all the base pairs involving sugar edge of guanine, the rG:A S:H Trans base pair has the maximum occurrence (1299 times in non redundant dataset). It occurs in the noncanonical stem (NC-stem) in kink-turn motif which is a recurrent asymmetric RNA internal loop that occurs in different type of RNAs and causes a sharp bend in the RNA backbone.[109–113] This base pair further participates in A-minor interaction with the residues of the canonical stem via its sugar edge.[114,115] Our analysis shows that, protonation at N7 position of guanine does not have any significant impact on the stability ($\Delta\Delta E_{int}^{sol}$ = -0.4 kcal/mol) of the rG:A S:H Trans pair (Figure 6.H and 6.I). Despite

protonation, the charge distribution at the sugar edge of adenine (as realized by the NBO charges at the polar sites) also remains unchanged and therefore is not expected to influence the A-minor interaction.

The other base pair, rG:A S:W Trans, has the second highest occurrence among the guanine sugar edge interactions (566 times in non redundant data set) and is found to occur frequently in the hairpin regions, *e.g.*, in motif id HL_4QCN_22, HL_88311.7, etc. in FR3D database of RNA loops (rna.bgsu.edu/rna3dhub/loops). As shown in Figure 6.E and 6.F, among the three inter base hydrogen bonds, protonation at N7 position of guanine (i) strengthens (ΔE(2) = 19..4kcal/mol) the first hydrogen bond (labeled as 1), (ii) weakens (ΔE(2) = -5.4kcal/mol) the second one (labeled as 2) and (iii) does not affect the third hydrogen bond between the O2′ of guanine and N6 of adenine (ΔE(2) = 0.4kcal/mol). The N7 protonation of guanine therefore does not influence the overall stability of the rG:A S:W Trans base pair.

The other base pair, G:G W:W Trans, also does not show any noticeable N7 protonation induced changes, in terms of, (a) interaction energy ($\Delta\Delta E_{int}^{sol}$ = -0.1kcal/mol), (b) intra base pair translational and rotational geometry and change in C1′-C1′ distance (Table S2, ESI†), (c) hybridization state of exocyclic amino group of guanine (Table 2) and (d) charge redistribution of the free sugar edges of guanine (Figure 6.A, 6.B and 6.C). However it is interesting to note that, G:G W:W Trans pair is a conserved interaction between the residues G1025 and G1139 in 23S rRNA where the G1139 residue further interacts with the sugar edge of A1143 residue in *trans* orientation. Protonation also does not have any significant impact on the stability of this particular GGA triple (ΔE_{int}^{sol}(neutral) = -18.9kcal/mol and ΔE_{int}^{sol}(protonated) = -19.5kcal/mol). However, there is an interesting consequence of the fact that neutral as well as the protonated variant of the G:G W:W Trans base pair have equally high interaction energy. We have seen earlier that neutral G:G W:H Cis base pair (system 5) converges to the W:W Trans geometry on ground state optimization, and show similar interaction energy (ΔE_{int}^{sol}(neutral) = -12.8kcal/mol). This change of geometry 'away from experimental geometry' is prevented by N7 protonation of the first guanine. It implies that, the W:H Cis geometry would not possibly have occurred in crystal structures, unless the guanine was N7 protonated. Interestingly, protonation of guanine in the new G:G W:W Trans geometry does not allow the reverting back to the W:H Cis geometry, since, the ΔE_{int}^{sol} of protonated W:W Trans is -12.9kcal/mol, which is 1.4kcal/mol higher than the protonated W:H Cis geometry (Figure 2.I and Figure 6.C).

Such versatile roles of nucleobase protonation is not limited to guanine N7 protonation only. Earlier we have reported guanine N3 protonation mediated geometric switches between G:G H:S Trans and G:G W:H Trans geometries.[44,71] In the absence of N3 protonation, the G:G H:S Trans base pair converges to the G:G W:H Trans geometry, which again on protonation at N3 position does revert back to G(N3+):G H:S Trans geometry.

4 Conclusions

The occurrence of *Class II* protonated nucleobases, where the protonated edge does not participate in base pairing, is difficult to detect directly in experimental X-ray crystal structures of nucleic acids. Their occurrence may however be indirectly inferred by following up on the experimental evidences of protonation induced changes in electronic structure, and consequent changes in the geometry and interactive properties of the base, as predicted on the basis of quantum chemical calculation.

To establish the proof of concept, we have systematically investigated neutral and N7 protonated guanine, and experimentally observed guanine containing base pairs and triples, using detailed bioinformtics methods and QM calculations. Our studies reveal the distinct possibilities of the involvement of the, hitherto unreported, N7 protonated guanine in base pairs and triples in RNA structures. Apart from canonical G:C W:W Cis and wobble G:U W:W Cis, we report four other functionally important guanine containing base pairs which show possibilities for protonation at the N7 site of their respective guanine moieties. With hydrogen atoms being 'invisible' in X-ray crystal structures, and, in the absence of other methods of their detection, all these base pairs had been earlier reported as neutral base pairs. For the canonical and wobble pairs, we have shown that N7 protonation at guanine leads to base pairs having higher or equivalent energy of interaction, while retaining their respective crystal geometries, on geometry optimization. The other four base pairs, G:C W:W Trans, G:rC W:S Cis, G:G W:H Cis and G:U S:H Trans (Figure 2) deviated greatly from their crystal geometries when geometry optimized as neutral base pairs. However, when geometry optimized as protonated base pairs, with each having a guanine moiety protonated at N7, the crystal geometries were retained. Notably, for G:C W:W Trans, similar divergence from crystal geometry during optimization have earlier been explained by invoking cation (Mg^{2+}) ion coordination, or positively charged archaeosine modification, at N7 of guanine. The difference in our case is that unlike metal ions or substituents at N7, proton at N7 can not be detected using X-ray crystallography.

Following up on the possible role of N7, on other guanine containing base pairs, we could identify three different G:A pairs (Figure 4) where optimization of the protonated models resulted in stronger, yet distorted, base pairs without any major changes in C1′-C1′ distances. Relating the N7 protonation to depyramidalization of guanine's primary amino group (N2), which in turn leads to a reduction in the amino acceptor potential of the amino group, we have proposed the possible involvement of these base pairs in pH dependent switches. Using the same argument, we have also proposed the possibility of guanine N7 protonation in

apparently neutral G:A W:W Cis base pairs, where the amino group of guanine is sp^2 hybridized and is not involved in tertiary and neighbouring interactions, *e.g.*, G26:A44 pair in tRNA (Figure 3). Our QM analyses further suggest that, due to its influence on the charge distribution of the free edges of the partner base in a base pair, guanine N7 protonation can be a crucial factor in modulation of the stability of higher order interactions in functional RNAs. Such an example is well demonstrated by the G46*G22:C13 triple in tRNAs (Figure 5), where protonation at N7 position of G46 increases the interaction energy of the triple by $\sim$10 kcal/mol.

Revelation of possibility of undetected protonation, as we show here for N7 protonated guanine, opens up the possibility of the discovery of undetected protonation in other base pairing systems also. This, in turn, is expected to have significant applications in the emerging field of RNA nanotechnology.[116] Unraveling of such versatile roles of nucleobase protonation in RNA, also supports the hypothesis that, evolutionary origin of RNA may have taken place in an acidic environment.[117,118]

Acknowledgement

S.B. and A.H. acknowledge UGC and CSIR, respectively, for JRF support. A.M., S.B. and A.H. thank DBT, Government of India project BT/PR-14715/PBD/16/903/2010 for partial funding and financial support. A.M. and D.B. thank DBT, Government of India project BT/PR-11429/BID/07/271/2008 for supporting computational infrastructure. A.D. thanks DST, CSIR, and INSA for partial funding.

References

1 J. L. Wilcox, A. K. Ahluwalia and P. C. Bevilacqua, *Acc Chem Res*, 2011, **44**, 1270–1279.
2 V. Verdolino, R. Cammi, B. H. Munk and H. B. Schlegel, *J Phys Chem B*, 2008, **112**, 16860–16873.
3 P. C. Bevilacqua, T. S. Brown, S.-i. Nakano and R. Yajima, *Biopolymers*, 2004, **73**, 90–109.
4 S. M. Mirkin and M. D. Frank-Kamenetski, *Ann Rev Biophys and Biomol Struct*, 1994, **23**, 541 – 576.
5 R. Zain and J.-S. Sun, *Cell and Mol Life Sci*, 2003, **60**, 862 – 870.
6 K. Gehring, J.-L. Leroy and M. Gueron, *Nature*, 1996, **363**, 561 – 565.
7 P. Fojtík and M. Vorlícková, *Nucleic Acids Res*, 2001, **29**, 4684–4690.
8 P. Jin and S. T. Warren, *Human Mol Gen*, 2000, **9**, 901–908.
9 P. L. Nixon and D. P. Giedroc, *J Mol Biol*, 2000, **296**, 659 – 671.
10 S.-i. Nakano, D. M. Chadalavada and P. C. Bevilacqua, *Science*, 2000, **287**, 1493–1497.
11 Z. Cai and I. Tinoco, *Biochemistry*, 1996, **35**, 6026–6036.
12 P. Legault and A. Pardi, *J Am Chem Soc*, 1994, **116**, 8390–8391.
13 T. M. Iverson, C. Luna-Chavez, I. Schröder, G. Cecchini and D. C. Rees, *Curr. Opin. Struct. Biol.*, 2000, **10**, 448–455.
14 N. A. Siegfried, B. O'Hare and P. C. Bevilacqua, *Biochemistry*, 2010, **49**, 3225–3236.
15 P. C. Durant and D. R. Davis, *J Mol Biol*, 1999, **285**, 115 – 131.
16 G. Chen, S. D. Kennedy and D. H. Turner, *Biochemistry*, 2009, **48**, 5738–5752.
17 A. K. Jissy and A. Datta, *J Phys Chem B*, 2010, **114**, 15311–15318.
18 A. Halder, S. Halder, D. Bhattacharyya and A. Mitra, *Phys Chem Chem Phys*, 2014, **16**, 18383–18396.
19 L. Chen, J. Di, C. Cao, Y. Zhao, Y. Ma, J. Luo, Y. Wen, W. Song, Y. Song and L. Jiang, *Chem Commun*, 2011, **47** 2850–2852.
20 P. C. Bevilacqua, T. S. Brown, D. Chadalavada, J. Lecomte, E. Moody and N. S. I., *Biochem Soc Trans*, 2005, **33**, 466–470.
21 J. Das, S. Mukherjee, A. Mitra and D. Bhattacharyya, *J Biomol Struct Dyn*, 2006, **24**, 91–202.
22 M. Chawla, P. Sharma, S. Halder, D. Bhattacharyya and A. Mitra, *J Phys Chem B*, 2011, **115**, 1469–1484.
23 J. Šponer, J. E. Šponer, A. I. Petrov and N. B. Leontis, *J Phys Chem B*, 2010, **114**, 15723–15741.
24 J. Šponer, J. E. Šponer and N. B. Leontis, *RNA 3D Structure Analysis and Prediction*, Springer, 2012, pp. 239–279.
25 J. Šponer, J. E. Šponer, A. Mládek, P. Banáš, P. Jurečka and M. Otyepka, *Methods*, 2013, **64**, 3–11.
26 K. Kino, K. Sugasawa, T. Mizuno, T. Bando, H. Sugiyama, M. Akita, H. Miyazawa and F. Hanaoka, *ChemBioChem*, 2009, **10**, 2613–2616.
27 M. Morikawa, K. Kino, T. Oyoshi, M. Suzuki, T. Kobayashi and H. Miyazawa, *Bioorganic & Medicinal Chemistry Letters*, 2015.
28 J. Šponer, C. A. Morgado and D. Svozil, *J Phys Chem B*, 2012, **116**, 8331–8332.
29 D. Svozil, P. Hobza and J. Šponer, *J Phys Chem B*, 2009, **114**, 1191–1203.
30 M. Y. Anzahaee, J. K. Watts, N. R. Alla, A. W. Nicholson and M. J. Damha, *J Am Chem Soc*, 2010, **133**, 728–731.
31 N. A. Siegfried, R. Kierzek and P. C. Bevilacqua, *J Am Chem Soc*, 2010, **132**, 5342–5344.
32 B. Gong, J.-H. Chen, E. Chase, D. M. Chadalavada, R. Yajima, B. L. Golden, P. C. Bevilacqua and P. R. Carey, *J Am Chem Soc*, 2007, **129**, 13335–13342.

33 M. Guo, R. C. Spitale, R. Volpini, J. Krucinska, G. Cristalli, P. R. Carey and J. E. Wedekind, *J Am Chem Soc*, 2009, **131**, 12908–12909.

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

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

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

37 R. K. Sigel and B. Lippert, *Chem Commun*, 1999, 2167–2168.

38 J. V. Burda, J. Šponer, J. Leszczynski and P. Hobza, *J Phys Chem B*, 1997, **101**, 9670–9677.

39 S. Bhattacharya, S. Mittal, S. Panigrahi, P. Sharma, S. P. Preethi, R. Paul, S. Halder, A. Halder, D. Bhattacharyya and A. Mitra, *Database*, 2015, **2015**, 1–9.

40 J. A. Doudna, B. P. Cormack and J. W. Szostak, *Proc Natl Acad Sci USA*, 1989, **86**, 7402–7406.

41 M. Hur and R. B. Waring, *Nucleic Acids Res*, 1995, **23**, 4466–4470.

42 S. A. McPhee, L. Huang and D. M. J. Lilley, *Nat Commun*, 2014, **5**, 5127.

43 J. A. Liberman, M. Salim, J. Krucinska and J. E. Wedekind, *Nat Chem Biol*, 2013, **9**, 353–355.

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

45 S. S. Ray, S. Halder, S. Kaypee and D. Bhattacharyya, *Front Genet*, 2012, **3**, 1–10.

46 R. S. Bangaru and A. Mitra, *MS by research thesis*, International Institute of Information Technology, Hyderabad, 2010.

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

48 Y. Zhao, N. E. Schultz and D. G. Truhlar, *J Chem Theory Comput*, 2006, **2**, 364–382.

49 D. Jose and A. Datta, *Cryst Growth Des*, 2011, **11**, 3137–3140.

50 Z. Mou, K. Uchida, T. Kubo and M. Kertesz, *J Am Chem Soc*, 2014, **136**, 18009–18022.

51 H. Ando, B. P. Fingerhut, K. E. Dorfman, J. D. Biggs and S. Mukamel, *J Am Chem Soc*, 2014, **136**, 14801–14810.

52 Y. Zhao and D. G. Truhlar, *J Chem Theory Comput*, 2007, **3**, 289–300.

53 F. Santoro, V. Barone and R. Improta, *J Comput Chem*, 2008, **29**, 957–964.

54 S. G. Stepanian, M. V. Karachevtsev, A. Y. Glamazda, V. A. Karachevtsev and L. Adamowicz, *J Phys Chem A*, 2009, **113**, 3621–3629.

55 W. Sun, Y. Bu and Y. Wang, *J Phys Chem C*, 2011, **115**, 3220–3228.

56 A. Jissy, U. Ashik and A. Datta, *J Phys Chem C*, 2011, **115**, 12530–12546.

57 M. Dargiewicz, M. Biczysko, R. Improta and V. Barone, *Phys Chem Chem Phys*, 2012, **14**, 8981–8989.

58 J. Wang, J. Gu and J. Leszczynski, *J Comput Chem*, 2012, **33**, 1587–1593.

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

60 R. Ditchfield, W. J. Hehre and J. A. Pople, *J Chem Phys*, 1971, **54**, 724–728.

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

62 S. F. Boys and F. Bernardi, *Mol Phys*, 1970, **19**, 553–566.

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

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

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

66 R. F. W. Bader, *Chem Rev*, 1991, **91**, 893–928.

67 J. P. Foster and F. Weinhold, *J Am Chem Soc*, 1980, **102**, 7211–7218.

68 A. E. Reed, L. A. Curtiss and F. Weinhold, *Chem Rev*, 1988, **88**, 899–926.

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

70 T. G. Todd A. Keith, *AIMAll (Version 11.12.19)*, 2012, Overland Park KS, USA.

71 P. Sharma, A. Mitra, S. Sharma, H. Singh and D. Bhattacharyya, *J Biomol Struct Dyn*, 2008, **25**, 709–732.

72 A. K. Jissy, S. Konar and A. Datta, *Chem Phys Chem*, 2013, **14**, 1219–1226.

73 U. Koch and P. L. A. Popelier, *J Phys Chem*, 1995, **99**, 9747–9754.

74 P. L. A. Popelier, *J Phys Chem A*, 1998, **102**, 1873–1878.

75 F. Sánchez-Viesca, F. Cortés, R. Gómez and M. Berros, *American Journal of Chemistry*, 2012, **2**, 343–346.

76 F. Weinhold and C. R. Landis, *Chem Educ Res Pract*, 2001, **2**, 91–104.

77 M. K. Sarangi, D. Bhattacharyya and S. Basu, *Chem Phys Chem*, 2012, **13**, 525–534.

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

79 D. Bhattacharyya, K. Sen and S. Mukherjee, *Indian J Chem*, 2006, **45A**, 58–67.

80 C. Altona and M. Sundaralingam, *J Am Chem Soc*, 1972, **94**, 8205–8212.

81 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.
82 J. I. Wu, J. E. Jackson and P. v. R. Schleyer, *J Am Chem Soc*, 2014, **136**, 13526–13529.
83 K. Wolinski, J. F. Hinton and P. Pulay, *J Am Chem Soc*, 1990, **112**, 8251–8260.
84 J. R. Cheeseman, G. W. Trucks, T. A. Keith and M. J. Frisch, *J Chem Phys*, 1996, **104**, 5497–5509.
85 W. Humphrey, A. Dalke and K. Schulten, *J Mol Graph*, 1996, **14**, 33 – 38.
86 J. T. Stivers and Y. L. Jiang, *Chem Rev*, 2003, **103**, 2729–2760.
87 P. J. Berti and J. A. B. McCann, *Chem Rev*, 2006, **106**, 506–555.
88 R. Ríos-Font, J. Bertrán, L. Rodríguez-Santiago and M. Sodupe, *J Phys Chem B*, 2006, **110**, 5767–5772.
89 R. Rios-Font, L. Rodríguez-Santiago, J. Bertran and M. Sodupe, *J Phys Chem B*, 2007, **111**, 6071–6077.
90 Y. Zheng, Y. Xue and G.-s. Yan, *J Theor Comput Chem*, 2009, **08**, 1253–1264.
91 E. J. Shim, J. L. Przybylski and S. D. Wetmore, *J Phys Chem B*, 2010, **114**, 2319–2326.
92 M. S. Ahmadi and A. Fattahi, *J Theor Comput Chem*, 2010, **09**, 585–609.
93 R. Rios-Font, J. Bertran, M. Sodupe and L. Rodriguez-Santiago, *Theor Chem Acc*, 2011, **128**, 619–626.
94 M. Ahmadi and A. Fattahi, *Scientia Iranica*, 2011, **18**, 1343 – 1352.
95 S. A. P. Lenz, J. L. Kellie and S. D. Wetmore, *J Phys Chem B*, 2012, **116**, 14275–14284.
96 G. Varani and W. H. McClain, *EMBO Rep*, 2000, **1**, 18–23.
97 J. Šponer, J. E. Šponer and N. B. Leontis, *RNA 3D Structure Analysis and Prediction*, Springer Berlin Heidelberg, 2012, vol. 27, pp. 239–279.
98 M. Chawla, S. Abdel-Azeim, R. Oliva and L. Cavallo, *Nucleic Acids Res*, 2014, **42**, 714–726.
99 M. Levitt, *Nature*, 1969, **224**, 759–763.
100 S. Kim, F. Suddath, G. Quigley, A. McPherson, J. Sussman, A. Wang, N. Seeman and A. Rich, *Science*, 1974, **185**, 435 – 440.
101 T. Christian, R. S. Lipman, C. Evilia and Y.-M. Hou, *J Mol Biol*, 2000, **303**, 503 – 514.
102 R. Oliva, L. Cavallo and A. Tramontano, *Nucleic Acids Res*, 2006, **34**, 865–879.
103 J. E. Šponer, N. Špačková, P. Kulhánek, J. Leszczynski and J. Šponer, *J Phys Chem A*, 2005, **109**, 2292–2301.
104 S. Panigrahi, R. Pal and D. Bhattacharyya, *J Biomol Struc Dyn*, 2011, **29**, 541–556.
105 A. Mládek, P. Sharma, A. Mitra, D. Bhattacharyya, J. Šponer and J. E. Šponer, *J Phys Chem B*, 2009, **113**, 1743–1755.
106 J. Šponer, A. Mokdad, J. E. Šponer, N. Špačková, J. Leszczynski and N. B. Leontis, *J Mol Biol*, 2003, **330**, 967 – 978.
107 P. Sharma, J. E. Šponer, J. Šponer, S. Sharma, D. Bhattacharyya and A. Mitra, *J Phys Chem B*, 2010, **114**, 3307–3320.
108 S. Mukherjee, S. Kailasam, M. Bansal and D. Bhattacharyya, *Biopolymers*, 2014, **101**, 107–120.
109 I. Vidovic, S. Nottrott, K. Hartmuth, R. Lührmann and R. Ficner, *Molecular Cell*, 2000, **6**, 1331–1342.
110 K. T. Schroeder, S. A. McPhee, J. Ouellet and D. M. Lilley, *RNA*, 2010, **16**, 1463–1468.
111 T. Moore, Y. Zhang, M. O. Fenley and H. Li, *Structure*, 2004, **12**, 807 – 818.
112 J. A. Chao and J. R. Williamson, *Structure*, 2004, **12**, 1165 – 1176.
113 A. Lescoute, N. B. Leontis, C. Massire and E. Westhof, *Nucleic Acids Res*, 2005, **33**, 2395–2409.
114 K. Réblová, J. E. Šponer, N. Špačková, I. Beššeová and J. Šponer, *J Phys Chem B*, 2011, **115**, 13897–13910.
115 F. Razga, N. Špačková, K. Réblová, J. Koča, N. B. Leontis and J. Šponer, *J Biomol Struct Dyn*, 2004, **22**, 183–193.
116 P. Guo, *Nat Nano*, 2010, **5**, 833–842.
117 V. K. Jayasena and L. Gold, *Proc Natl Acad Sci USA*, 1997, **94**, 10612–10617.
118 H. Bernhardt and W. Tate, *Biology Direct*, 2012, **7**, 4.