

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/49774147>

Protonation of Base Pairs in RNA: Context Analysis and Quantum Chemical Investigations of Their Geometries and Stabilities

ARTICLE in THE JOURNAL OF PHYSICAL CHEMISTRY B · FEBRUARY 2011

Impact Factor: 3.3 · DOI: 10.1021/jp106848h · Source: PubMed

CITATIONS

14

READS

23

5 AUTHORS, INCLUDING:



Mohit Chawla

King Abdullah University of Science and Tec...

7 PUBLICATIONS 58 CITATIONS

[SEE PROFILE](#)



Purshotam Sharma

Panjab University

41 PUBLICATIONS 276 CITATIONS

[SEE PROFILE](#)



Dhananjay Bhattacharyya

Saha Institute of Nuclear Physics

116 PUBLICATIONS 1,105 CITATIONS

[SEE PROFILE](#)



Abhijit Mitra

International Institute of Information Techn...

68 PUBLICATIONS 349 CITATIONS

[SEE PROFILE](#)

Protonation of Base Pairs in RNA: Context Analysis and Quantum Chemical Investigations of Their Geometries and Stabilities

Mohit Chawla,^{†,‡} Purshotam Sharma,^{†,‡} Sukanya Halder,[§] Dhananjay Bhattacharyya,^{*,§} and Abhijit Mitra^{*,†}

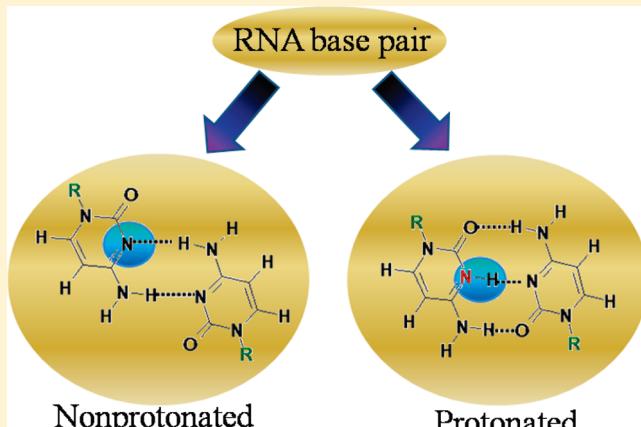
[†]Center for Computational Natural Sciences and Bioinformatics (CCNSB), International Institute of Information Technology (IIIT-H) Gachibowli, Hyderabad 500032, India

[‡]Lovely Professional University (LPU), Phagwara 144402, India

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

 Supporting Information

ABSTRACT: Base pairs involving protonated nucleobases play important roles in mediating global macromolecular conformational changes and in facilitation of catalysis in a variety of functional RNA molecules. Here we present our attempts at understanding the role of such base pairs by detecting possible protonated base pairs in the available RNA crystal structures using BPFind software, in their specific structural contexts, and by the characterization of their geometries, interaction energies, and stabilities using advanced quantum chemical computations. We report occurrences of 18 distinct protonated base pair combinations from a representative data set of RNA crystal structures and propose a theoretical model for one putative base pair combination. Optimization of base pair geometries was carried out at the B3LYP/cc-pVTZ level, and the BSSE corrected interaction energies were calculated at the MP2/aug-cc-pVDZ level of theory. The geometries for each of the base pairs were characterized in terms of H-bonding patterns observed, rmsd values observed on optimization, and base pair geometrical parameters. In addition, the intermolecular interaction in these complexes was also analyzed using Morokuma energy decomposition. The gas phase interaction energies of the base pairs range from -24 to -49 kcal/mol and reveal the dominance of Hartree–Fock component of interaction energy constituting 73% to 98% of the total interaction energy values. On the basis of our combined bioinformatics and quantum chemical analysis of different protonated base pairs, we suggest resolution of structural ambiguities and correlate their geometric and energetic features with their structural and functional roles. In addition, we also examine the suitability of specific base pairs as key elements in molecular switches and as nucleators for higher order structures such as base triplets and quartets.



1. INTRODUCTION

The availability of growing number of structures of functional RNA, accompanied by a spate of activities involving biophysical experiments^{1–3} as well as computational studies^{4–6} has led to the rapid unfolding of the variety of roles played by different non-covalent interactions governing RNA structure and dynamics. While tremendous progress has been made in our understanding of the mechanisms of, e.g., the self-splicing of ribozymes,^{7,8} peptidyl transfer reaction and associated ribosomal activity,⁹ ligand recognition by aptamers and riboswitch functioning,^{10,11} etc., these developments have also highlighted the gaps in our understanding of the details of the variety of molecular level interactions, particularly in the context of the role of protonation of bases, in several of these mechanisms.

1.1. Current Paradigms. Neutralization of the negatively charged backbone phosphate groups, by metal counterions, has earlier been shown as one of the primary events in RNA folding

*in vitro.*¹² Subsequent studies have also shown that the availability of specific divalent cations, in appropriate concentration, is often an important determinant for specific catalytic and folding processes in RNA.^{13,14} However, there is now a growing acceptance of the fact that, apart from phosphates and metal cations, the nucleobases themselves can acquire charges through protonation/deprotonation involving, often metal ion mediated, proton exchange with the aqueous environment.¹⁵ The accompanying paradigm shift has generated renewed interest toward understanding the roles of protonated nucleobases in defining structures, catalytic roles, and folding pathways of nucleic acids.^{16,17}

Compelling evidence for the formation of novel triple helical DNA structures, such as H-DNA and H*-DNA, with putative

Received: July 22, 2010

Revised: December 8, 2010

Published: January 21, 2011

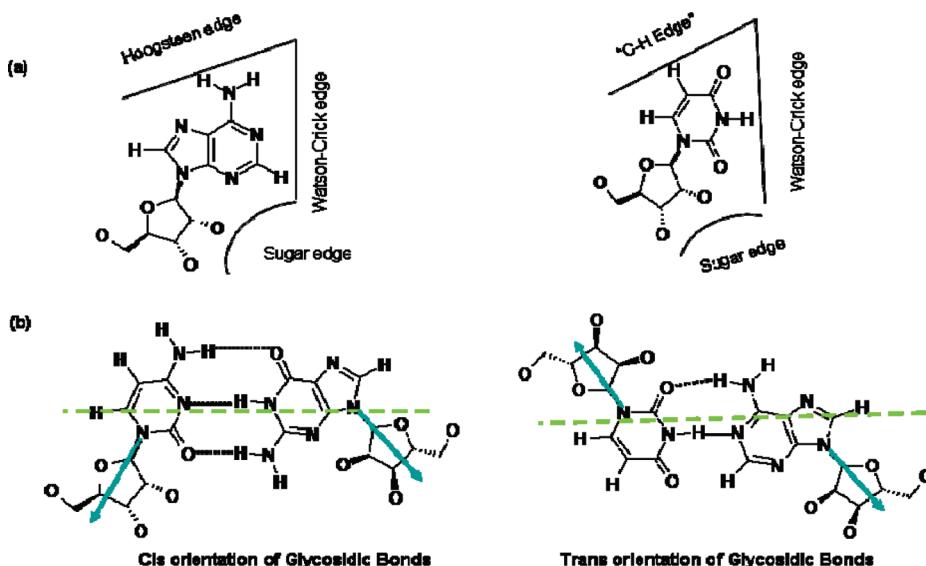


Figure 1. (a) Edge interaction in purine and pyrimidine bases taking adenine and uracil as examples. (b) Cis versus trans orientation of glycosidic bonds taking G:C W:W cis and A:U W:W trans, respectively, as examples.⁵²

regulatory roles,^{18–20} transcription regulatory pseudoknot formation involving a C8(+):G12-C26 triplet (PDB id 437D), which gets destabilized when this triplet is replaced by noncognate U8:A12-U26 triplet,²¹ formation of i-motifs, in telomeric regions²² involving tetrameric DNA structure formation consisting of 5'-d(TCCCCC) oligomer,²³ etc., underline the role of cytosine protonation. Similarly, adenine protonation has been shown to drive global conformational changes in the intramolecular stem-loop (ISL) involved in the U6 RNA of the spliceosome.^{24,25} Apart from these well studied specific examples of potentially pH dependent conformational switching, involving base pairs having protonated cytosine or adenine,^{22–25} we have recently discussed such potentials involving the protonation of guanine.²⁶

Based on their occurrence contexts and functional features, identified in known structures, protonated nucleobases have been divided into two classes, viz. class I, for which the loaded proton is sequestered in hydrogen bonding between paired bases, and class II sites, where the proton is not directly involved in base pairing.^{27,28} Protons bound to class I sites are not suitable for proton transfer processes, because of large activation energy required to open up the protonated base pair.²⁸ These sites are thus primarily involved in providing appropriate electrostatic environment, to facilitate catalytic pathways, for example, through oxyanion hole formation during peptidyl transfer reaction.²⁸ Class II sites, on the other hand, are known to participate in proton transfer processes (e.g., in HDV and hairpin ribozymes).^{29,30}

1.2. The Larger Questions. In general, nucleobases do not get protonated or deprotonated at physiological pH, since the pK_a values for imino nitrogens are ~ 3.5 and 4.2 for adenine and cytosine, respectively, and ~ 9.2 for guanine and uracil.^{31,32} In most related studies, the protonation of nucleobases have been explained either by analyzing the crystal and solution NMR structures at sufficiently low pH ranges or by the manipulation of the microenvironment, resulting in a shift in pK_a values of titratable groups.¹⁶ Such pK_a shifts may occur due to presence of cations and phosphates in the close proximity,²⁷ as well as because of electronic effects due to base pairing, stacking, or the presence of other hydrogen bonds near the site of protonation.^{15,16} All this agrees well with the general explanation that protonation, which may be thermodynamically unfavorable in the local context, is driven by global stability requirements.

Be that as it may, there remains a significant gap in our understanding of the rationale, in terms of both “why” and “how”, behind the protonation of nucleobases. As discussed above, significant understanding has evolved regarding the “why”. Further information leading to the correlation of the “presence” of protonated bases with RNA structure, dynamics, and functions is expected to eventually further enhance our understanding. The problem is that the detection of protonation from structures solved by X-ray crystallography is problematic as, due to its low scattering cross-section, hydrogen atom positions are generally not determined by this method. The NMR method is also not suitable due to the possibility of proton exchange with solution, leading to intermediate average structures.

Different groups have used different methods, including several spectroscopic and other experimental techniques,^{33–37} as well as MD simulation studies,^{4–6} to understand the protonation and tautomerism of nucleobases.^{38,39} Some of the prominent protonated base pair structures, present in a variety of contexts, have also been reported.^{16,40–50} However, mere correlation of protonation events with their corresponding environmental context cannot provide the complete understanding of the physicochemical driving forces leading to the protonation events. This not only requires a comprehensive mining of occurrences of protonated base pairs and associated higher order structures but also calls for a detailed study of relative energies and geometric stabilities of such base pairs in their specific functional contexts. As discussed later, in conjunction with MD simulations, these studies are also particularly important for sorting out ambiguous cases of protonated base pairs.

1.3. Scope of This Study and Issues Involved. Experimental determination of the energies, for individual interactions involving base pair geometries that are away from their local minima in RNA crystal structures, is extremely difficult.⁵¹ Detailed quantum chemical studies, involving a variety of RNA base pairs, have addressed this issue to a large extent.^{26,42–50} In comparison, studies pertaining to quantum chemical evaluation involving protonated bases and their interactions are far rarer.^{26,52–60}

Out of the twelve RNA base pair geometric families described in the Leontis–Westhof (LW) classification scheme⁵² based on consideration of three interacting edges (Figure 1a) and two mutual orientations of glycosidic bonds (cis and trans) (Figure 1b), we have

Table 1. Occurrence Contexts, Geometries, Hydrogen Bonding Characteristics, and Flexibilities of Protonated Base Pairs

sys no.	protonated edge (site)	base pair geometry	PDB id; base pair id	environment	no. of H bonds ^a			rmsd (Å)
					N	BF	K	
1	adenine W-edge (N1)	A(+):C W:WC	402D; A10S:C112	stem	2 (1)	(1)		1.00
2		A(+):C W:WT	1NKP; A2371:C2403	stem	2(2)			2.24
3		A(+):G W:HC	1NJP; A1061:G2731	bulge	2 (2)			1.35
4		A(+):G W:HT	2TRA; A46:G22	triplet	2 (2)			0.81
5	adenine S-edge (N3)	rA(+):G S:HC	1FJG; A1001:G1002	DNP	1 (1)		2 (2)	1.73
6	cytosine W-edge (N3)	C(+):C W:WC	1MME; C170d:C30c	stem	2 (2)			1.05
7		C(+):C W:WT	1B23; C16:C59	tertiary int.	3 (3)			0.51
8		C(+):U W:WC	1J5A; C2554:U2490	stem	2(2)			0.56
9		C(+):U W:WT	modeled on 2D6F; U916:C959			2		
10		C(+):A W:WT	1NKP; C2403:A2371	stem	2(2)			2.73
11		C(+):A W:HT	1FFK; C505:A509	quartet	2 (2)			1.41
12		C(+):G W:WC	1FKA; C453:G459	stem	1	1 (1)		0.94
13		C(+):G W:HC	1GID; C260:G108	triplet	2 (2)			0.06
14		C(+):G W:HT	1DRZ; C141:G161	triplet	2 (2)			0.85
15		C(+):rC W:SC	1I94; C1364:C913	triplet	2 (2)			2.79
16		C(+):rU W:SC	1I94; C502:U499	triplet	2 (2)			1.85
17	guanine S-edge (N3)	G(+):C S:WC	1ET4; G326:C224	triplet	2 (2)			2.50
18		G(+):G S:HT	1DK1; G18:G51	triplet	2	(1)		1.91
19		G(+):G S:H T	1JZY; G1439:G1589	isolated	3 (3)			1.16

^a N indicates N—H---O/N type of hydrogen bond, BF indicates bifurcated hydrogen bonding, and K indicates C—H---O/N type of hydrogen bond. The values in parentheses are for Hopt geometries of base pairs. Since sys. no. 9 is a modeled geometry, rmsd is not calculated for this base pair.

used BPFind software⁶¹ to identify 18 distinct “possible” combinations (Table 1) of protonated base pairs from the analyzed crystal structure data set. In addition, we also report a putative base pair on the basis of modeling (Table 1, Figure 2). We report the analysis of their occurrence contexts, and calculate their intrinsic stabilities by using advanced quantum chemical methods.

As mentioned in section 1.2, a major hurdle toward a comprehensive and unambiguous identification of protonated base pairs is, barring few exceptions,⁶² the absence of hydrogen atom coordinates in crystal structures. This was addressed using a refined version of BPFind,⁶¹ which uses a “precursor atom” based algorithm to detect H-bonds without any explicit consideration of the position of hydrogen atoms. The strategy used was based on an initial short

listing of base pairs having at least one good hydrogen bond and also having a pair of closely situated acceptor atoms. The conspicuous closeness of two hydrogen bond acceptor atoms can in principle be explained in terms of one of these being protonated. These potential protonated base pairs were then further screened and validated using the usual BPFind protocols based on *E*-values and other considerations. It may be mentioned that the current version of BPFind also rules out instances of apparently protonated base pairs, which can be better understood as slightly distorted neutral base pairs and reports possibility of protonation only when it finds a pair of very good hydrogen bonds by protonation.⁶¹

It must, however, be noted that the “nearness of two acceptor atoms” approach does not always lead to unambiguous identification

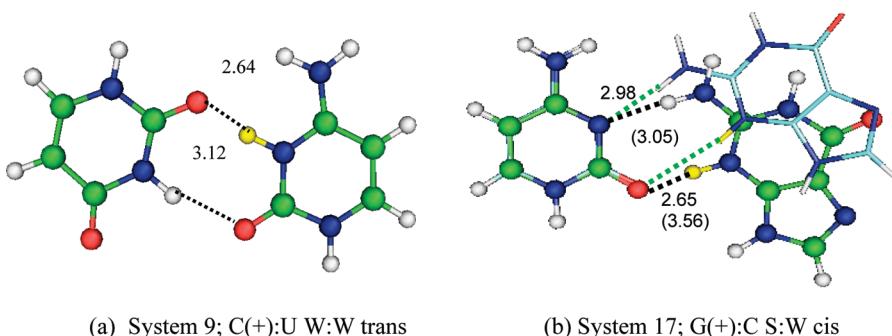


Figure 2. (a) Modeled optimized protonated base pairing geometries for system 9. (b) Superposition of system 17 done considering the protonation hypothesis.

of protonated base pairs. In fact, of the 19 different possibilities studied in this work, as many as 6 instances needed hypothesis driven manual intervention for addressing ambiguities that were resolved to different degrees using quantum chemical methods. We may classify these ambiguities into four broad categories.

1.3.1. Ambiguity of Protonation versus Water/Metal Ion Mediation. In some cases, closeness of two acceptor atoms can in principle be explained either in terms of the protonation hypothesis or in terms of water/ion mediated hydrogen bonding. As shown in the context of a possible G:C(+) S:W cis base pair (system 17), the latter possibility may be tested out, if explicit water molecules or ions are appropriately positioned in X-ray crystal structures.⁶⁰ In this case, not only was the water molecule in question, positioned out of the plane of the base pair, but also preliminary geometry optimizations of the protonated and water mediated models, respectively, showed far greater stability for the protonated model.

1.3.2. Ambiguity Due to Donor/Acceptor Duality of O2'. In four cases (system nos. 5, 15, 16, 19; Table 1), the 2'-OH of ribose sugar of one of the interacting nucleosides was found to interact with an acceptor atom of its partner base. However, in systems 15 and 16, the hydrogen bonding may also be modeled in terms of O2' as "the acceptor" while the interacting "acceptor" atom of the paired base is considered as a protonated "donor". Since the relevant pK_a values are also ambiguous,¹⁶ such cases may be addressed by computationally analyzing both the protonated and nonprotonated geometries (Figure 3).

1.3.3. Ambiguity Related to Multimodality of Edge Interaction. We have demonstrated in our earlier studies how the base pairs belonging to same geometric family⁵² can show multimodality in terms of pairing geometries and in terms of H-bonding patterns.⁵⁶ For example, similar G:G(+) H:S trans interaction shows multimodality in terms of hydrogen bonding pattern, in systems 18 and 19. In conjunction with the ambiguous possibilities of metal ions, water molecules and protonation, in available crystal structures, such multimodality introduces further concerns.

1.3.4. Ambiguity Related to Location of Protonation. During our database search, we could detect at least one example where the interactions were weak when the bases are considered neutral, but which had the potential of stronger interactions if one of the bases is considered protonated. Naturally, such arguments need to be used with caution, since, in the absence of solvent screening effects, protonated base pairs are expected to show unrealistically high interaction energies. It is, however, interesting to note that, depending on two alternative protonation positions, the structures converged to two different

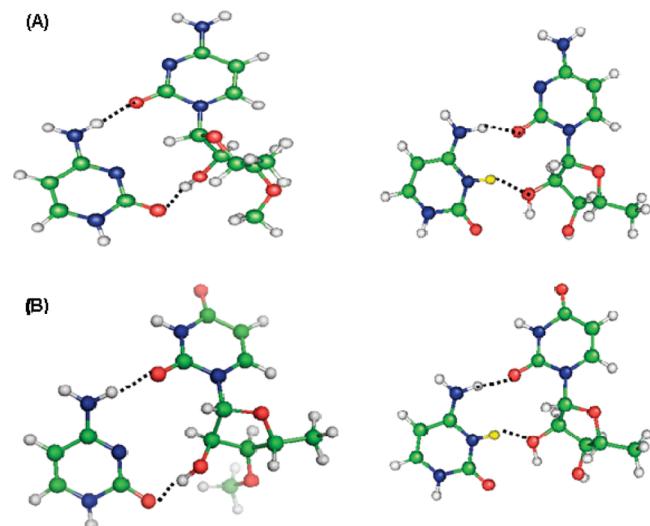


Figure 3. (A) Optimized geometries of nonprotonated C:rC W:S cis and protonated C(+):rC W:S cis base pairs. (B) Optimized geometries of nonprotonated C:rU W:S cis and protonated C(+):rU W:S cis base pairs. The protonation at the N3 atom of cytosine is colored yellow. Note that the O3' of ribose sugar is methylated in case of nonprotonated geometries.

geometries on optimization (systems 2 and 10, Table 1). Another type of ambiguity, under this category, relates to symmetrical pairs where the proton could in principle be on either of the two acceptor atoms. Such a situation is observed in system 7.

2. METHODS

2.1. Identification and Structural Analysis of Protonated Base Pairs. We used a data set (cf. Supporting Information) of selected high resolution representative RNA crystal structures from protein data bank (PDB) for our analysis. We have used the BPFind tool to detect protonated base pairs from RNA crystal structures, and to understand the structural context in which they occur. In the absence of hydrogen atoms in the PDB files, this tool detects the base pairs using an algorithm to detect the positioning of two base and edge specific hydrogen bonding donor and acceptor atoms. It further calculates pseudoangles involving suitable precursor atoms to ensure linearity of hydrogen bonds and also to guarantee planarity of the two pairing bases.

2.2. Modifications in BPFind. Compared to the earlier version,⁶¹ the present version of BPFind software has two major changes. First, it now takes into consideration the H-bond donor/acceptor dual

property of the 2'-OH group of the ribose moiety. This modification leads to reclassification of some base pairs, which have previously been noted for their ambiguous H-bonding patterns. For example, C: C W:S C and C:U W:S C base pairs are now searched according to their nonprotonated geometry. Second, the pseudoangle criterion has been made more stringent for detection of protonated base pairs to ascertain the linearity of H-bonds. The requisite value of pseudoangle to detect protonated H-bond geometry is increased to 150°, compared to the default value of 120° for nonprotonated base pairs. The recent version can be downloaded from www.saha.ac.in/biop/bioinformatics.html.

2.3. Model Building. The initial structural models of base pairs were built, from selected examples from Protein Data Bank (PDB) by extracting coordinates of the corresponding nucleotides from PDB files (see Table 1). Protonated bases A, C, and G were generated by adding single hydrogen atoms to the N1 or N3 positions of adenine depending on whether the Watson–Crick side or sugar edge is protonated, N3 position of cytosine and N3 of guanine; without disturbing the positions of the heavy atoms of the heterocyclic rings. In cases where the sugar O2' groups do not participate in base pairing, we replaced the ribose sugar attached to the bases with hydrogen atoms to reduce the computation cost for subsequent quantum chemical calculations. In other cases, we retained the sugar, with their respective 5'-OH groups replaced by methyl groups, in our models. Use of similar modeling approaches, for studying RNA base pairs, are well documented in the literature,^{26,56–58} and are justified in the present context. We also examined several base pairs for potential alternative hydrogen bonding patterns by optimizing geometries of models based on experimental structures of similar, though nonidentical, base pairs.

2.4. Computational Details. Gaussian03⁶³ and GAMESS-US⁶⁴ suites of quantum chemical programs were used for all quantum chemical calculations carried out in this work.

2.4.1. Gas Phase Optimization. The base pairs were geometry optimized using two different procedures: “frozen” and “relaxed”, which we term as “Hopt” and “Fopt”, respectively. In Hopt, we optimized only the position of hydrogen atoms, keeping all the Cartesian coordinates of heavy atoms of the base pair as frozen, using the opt=modredundant option in the Gaussian03. During Fopt, we removed this option, and the geometries of the base pairs obtained from crystal structures were fully optimized, free from any geometrical constraints. The comparison of geometries and energies obtained using Hopt and Fopt procedures help us to analyze the contribution of interbase hydrogen bonding interactions to the overall stability of the base pair in the crystal environment. The details of this procedure are described in our earlier papers.^{26,56,60}

2.4.2. Optimization Methods. Geometry optimization was carried out using density functional theory (DFT). Becke's three parameter exchange⁶⁵ and Lee–Yang–Parr's correlation functional⁶⁶ (abbreviated as B3LYP) was used with large cc-pVTZ basis set of atomic orbitals (abbreviated as B3LYP/cc-pVTZ). The results for geometry optimizations at this level have been shown to correlate very well with the reference RIMP2/aug-cc-pVTZ structures and are sufficiently accurate to study base pairing interactions in RNA molecules.⁶⁷

2.4.3. Interaction Energies. For a base pair dimer A–B, the interaction energy $\Delta E_{(A-B)}$ can be given as

$$\Delta E_{(A-B)} = E_{(A-B)} - (E_{(A)} + E_{(B)})$$

where $E_{(A-B)}$ is the energy of the dimer A–B and $E_{(A)}$ and $E_{(B)}$ are the energies of the respective isolated monomers.

We calculated the interaction energies for the B3LYP/cc-pVTZ optimized geometries obtained using Hopt as well as Fopt methods, at the MP2 level, combined with large diffuse aug-cc-pVDZ basis set of atomic orbitals (abbreviated as MP2/aug-cc-pVDZ). This method is shown to give a fairly good estimation of interaction energies in case of RNA base pairs, with slight underestimation by 0.7–1.7 kcal/mol, compared to the reference values, calculated at the complete basis set limit (MP2/CBS).⁵⁸ These interaction energy values were corrected for basis set superposition error (BSSE) using the standard counterpoise method of Boys and Bernardi⁶⁸ and are represented as E_{int}^{MP2} in the subsequent text.

The BSSE corrected E_{int}^{MP2} energy has two components: the Hartree–Fock term (E_{int}^{HF}) and the correlation term (E_{int}^{corr})

$$E_{int}^{MP2} = E_{int}^{HF} + E_{int}^{corr}$$

E_{int}^{HF} mainly includes electrostatic, exchange repulsion, polarization, and charge transfer components. E_{int}^{corr} includes the attractive dispersion and electron correlation correction to the interaction energy.

The interaction energy values were also corrected for deformation energy, “ E_{def} ”, which is calculated as

$$E_{def(AB)} = (E_{(A)}^{AB} - E_{(A)}^o) + (E_{(B)}^{AB} - E_{(B)}^o)$$

Here $E_{(A)}^{AB}$ and $E_{(B)}^{AB}$ are the energies of the individual monomer base geometries, which are present in the optimized base pair AB. $E_{(A)}^o$ and $E_{(B)}^o$ are the energies of the isolated optimized monomers A and B.

The total interaction energy (E_{int}^{tot}) can thus be represented as

$$E_{int}^{tot} = E_{int}^{MP2} + E_{def}$$

For the “Fopt” geometries, “ E_{def} ” can be calculated as described above. However, as the deformation is not defined in the context of base pairs present in their respective RNA crystal structures, we did not calculate the E_{def} for Hopt geometries. “ E_{def} ” values are also not calculated for “Fopt” geometries of those base pairs where ribose sugar is retained during calculations, as in such cases, “ E_{def} ” may acquire large values due to the flexibility of the ribose sugar, which may bias the final interaction energy values. As E_{int}^{MP2} values are computed for both “Hopt” and “Fopt” base pairs, these values are used to estimate the comparative strength of different base pairs.

2.4.4. Energy Decomposition Analysis. In addition to its evaluation, the interaction energy of each of the base pairs was decomposed at the HF/6-31G(d,p) level, using the Kitaura–Morokuma (KM) decomposition scheme,⁶⁹ to understand the role of individual components of intermolecular interaction. It may be noted that although BSSE is partially corrected in KM analysis according to the scheme proposed by Sokalski et al.⁷⁰ and Cammi et al.,⁷¹ the interaction energy components obtained from this method are, to some extent, contaminated by BSSE. Noncontaminated values of interaction energy components can be obtained from other methods such as symmetry adapted perturbation theory⁷² or hybrid variation-perturbation decomposition methods.⁷³ However, considering the large number of base pair systems studied in the present work, the computationally less expensive KM method was used for interaction energy decomposition.

2.5. Assessment of Change in Geometries on Optimization. Two parameters were utilized to measure the relative

change in the geometry of each of the base pair in the crystal context from the gas phase optimized geometry.

- (1) rmsd: To evaluate the rmsd between Hopt and Fopt geometry of the same base pair, first we superposed geometries of one of the bases from both structures and calculated the rmsd for the other base. Then the superposition is done for the second base and rmsd is calculated for the first base. The final rmsd values are taken as the average of the two values. It may be noted that superposition is performed only for non-hydrogen atoms of the bases. The algorithms for these superposition programs have been described in our previous work.²⁶
- (2) Base pair parameters: To get a quantitative estimation of change in base pair geometry, on optimization of the crystal pair in the crystal context, we have calculated the rotational and translational base pair parameters (buckle, open, propeller, stagger, stretch, and shear) using the recently upgraded version of NUPARM software, which takes into account the edge specific base pair system for parameter calculation in RNA base pairs.⁷⁴

3. RESULTS AND DISCUSSION

Nineteen protonated base pair combinations have been investigated. Table 1 summarizes their occurrence contexts, geometries, and hydrogen bonding characteristics and compares their experimental “Hopt” geometries (see Methods), with those of their corresponding “Fopt” geometries located at the nearest local energy minima on potential energy surfaces of the isolated base pairs, in terms of the rmsd. In principle, the “Fopt” structures correspond to the ideal H-bonded geometries of protonated base pairs that would be obtained in the absence of perturbation by external effects, due to the presence of solvent molecules and metal ions as well as those due to the backbone topology related constraints present in the RNA structure. The flexibility or geometric variability, of these base pairs, as indicated by the rmsd values and structural superposition (Figure 4), have also been characterized in terms of changes observed in hydrogen bonding (Table S1 of Supporting Information) and base pair parameters (Table S2 of Supporting Information). Interaction energies ($E_{\text{int}}^{\text{MP2}}$) of protonated base pairs, in both “Hopt” and “Fopt” geometries, have also been computed (Table 2) and decomposed into components within the HF approximation (Table S3 of Supporting Information). Table 3 summarizes the propensity of protonated base pair systems to participate in higher order structures, their pH dependent switching potential, and highlights the resemblances of their donor–acceptor networks with those in otherwise unrelated neutral base pairs. Analysis of our computational results, augmented by appropriate database analysis, revealed several features related to the understanding of the role of protonated RNA base pairs.

3.1. Features Related to Interaction Energies. Though all the protonated base pairs reported here possess two or more H-bonds; they show large variation in their interaction energies and noticeable diversity in their respective binding patterns. Thus, the interaction energy is generally observed to be high for cases where a protonated nucleobase (A/G/C) interacts with a C or G because of the high dipole moment of the latter, which in turn results in strong ion–dipole interactions⁵⁵ (Table 1). Additional factors contribute to the increase in interaction energy in three cases, systems 17–19, which involve interaction of sugar edge of N3 protonated guanine with other bases. In these cases,

apart from the contributions due to increased dipole moment of protonated guanine, additional stability can arise due to N3 protonation-induced double bond character of C2–N2 bond of guanine. This increase in double bond character, as indicated by a decrease in C2–N2 bond length, compared to the respective bond length in monomeric guanine, can restrict the rotation of the N2-amino group of guanine, thus improving its interactions with the other base.

The gas phase interaction energies for the “Hopt” and “Fopt” geometries of base pairs range from −21 to −47 and −24 to −49 kcal/mol, respectively. These values correctly reflect the expected trend of “Fopt” geometries being more stable than their corresponding “Hopt” geometries.⁵⁶ The actual increase in interaction energies, as observed on optimization of individual base pairs, however, varies a lot. This is because of the diversity in the extent of variation in geometries and H-bonding patterns observed on full optimization of the crystal geometry. It may be pointed out that a large increase in interaction energy is not necessarily a consequence of the optimization induced increase in the “number” of H-bonds. Interaction energies may increase, irrespective of the “number”, so long as the optimization driven change in geometry is accompanied with formation of ‘better’ H-bonds. Thus, three base pair systems (systems 7, 12, 19) possess three H-bonds in their “Fopt” geometry, and their interaction energies range from −40 to −49 kcal/mol. For the rest of the base pairing geometries, possessing two H-bonds of N–H···O/N type, the interaction energy ($E_{\text{int}}^{\text{MP2}}$) values range from −25 to −49 kcal/mol in their Fopt geometries. The $E_{\text{int}}^{\text{MP2}}$ values of the “Fopt” base pairs involving 2'-OH interaction (systems 5, 15, 16, 19) range from −34 to −49 kcal/mol. E_{def} values of the studied protonated base pairs range from 3 to 9 kcal/mol and are significantly larger compared to those found in case of neutral RNA base pairs.^{56,59}

On comparing the relative contribution of HF ($E_{\text{int}}^{\text{HF}}$) and correlation ($E_{\text{int}}^{\text{corr}}$) components, respectively, it is found that for all the base pairs studied, the interaction energies, in both “Hopt” and “Fopt” geometries, are dominated by the $E_{\text{int}}^{\text{HF}}$ term, with the values ranging from 73% to 98% (Table 1). It may be noted that compared to nonprotonated pairs, the $E_{\text{int}}^{\text{HF}}/E_{\text{int}}^{\text{corr}}$ ratio is significantly larger in the case of protonated base pairs. This is because the positive charge enhances the electrostatic term, which is a major component of $E_{\text{int}}^{\text{HF}}$.

The interaction energy values from the Morokuma decomposition roughly correspond (with a difference of 1–4 kcal/mol) to the $E_{\text{int}}^{\text{HF}}$ component of $E_{\text{int}}^{\text{MP2}}$ energies. We found that in all the studied protonated base pairs, the magnitude of the major interaction energy components, viz. electrostatic (E^{Elec}), exchange repulsion (E^{EX}), polarization (E^{Pol}), and charge transfer (E^{CT}), increases when we go from “Hopt” geometries to “Fopt” geometries (Table S3). Also, the contribution of E^{Elec} to the interaction energy is larger than that of E^{EX} , both in crystal context (“Hopt”) and in gas phase optimized (“Fopt”) context, indicating the importance of electrostatic interactions in charged systems. In most of the cases, E^{Elec} is greater than that of E^{EX} by more than 17 kcal/mol in the gas phase optimized geometries. In general, E^{Pol} has larger magnitude than E^{CT} in “Hopt” as well as “Fopt” geometries.

It may be noted that the “high” interaction energy values, obtained from gas phase calculations, may be misleading, unless they are understood in the proper context. In the actual nucleic acids environment, the ion–molecular dipole interactions in protonated base pairs are screened by solvent. Also the free

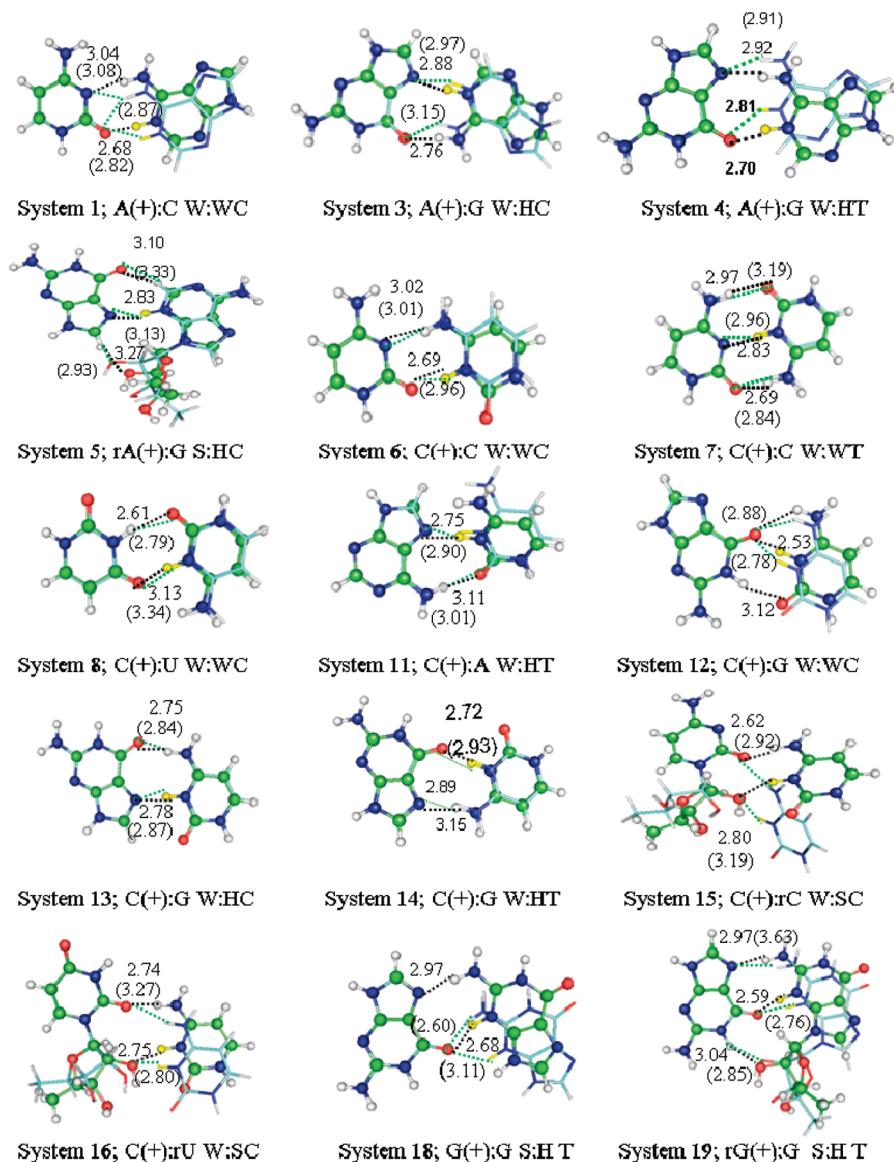


Figure 4. Superposition of gas phase optimized (ball and stick) and crystal geometries (sticks) of protonated base pair geometries. The H-bonds present in the crystal contexts of these base pairs are shown in green, whereas those observed on gas phase optimization are shown in black. In both the gas phase optimized and crystal geometries, the protonated hydrogen atom is colored yellow. The donor–acceptor distances in parentheses correspond to that of crystal geometries, and those outside parentheses correspond to the optimized geometry. Note that all the distances are in angstroms.

energy cost of changing the local pK_a is quite high. In the physiological context, therefore, protonation of bases occurs only when there are no other viable options to get the structure desired for the specific structure/function in context of RNA. However, the gas phase interaction energies reflect the intrinsic stability of the ideal geometries of protonated base pairs, as would be obtained in the absence of any other perturbing forces.

3.2. Features Related to Structures and Geometries. With the exception of the systems 1, 12, and 18, where relaxed optimization results in significant changes in the H-bonding pattern, most protonated base pairs have reasonably stable experimental (or crystal) geometries. This can be concluded from the absence of significant differences in H-bond pattern corresponding to “Hopt” and “Fopt” geometries (Table S1, Supporting Information). Values of base pair parameters, as given in Table S2 (Supporting Information) also indicate similar hydrogen bonding schemes in the “Hopt” and “Fopt” structures.

In the case of system 1 ($A(+):C(W:W\text{ cis})$), the O2-mediated bifurcated H-bonding pattern present in the crystal context is lost on optimization, resulting in formation of two separate H-bonds, N6–H(A)…N3(C) and N1(+)-H(A)–O2(C). For the $C(+):G(W:W\text{ cis})$ base pair (system 12), the crystal geometry has a high buckle and propeller, with O6 of guanine forming a bifurcated hydrogen bond with the protonated N3(+) and amino N4 of cytosine. On optimization, the base pair becomes coplanar with a slight increase in opening angle, resulting in formation of an extra H-bond (Table S1, Supporting Information). Similarly, the highly nonplanar system 18 ($G18(+):G51\text{ (S:H trans)}$) presents a bifurcated H-bond pattern in which O6(G51) interacts with N3(+)(G18) and N2(G18) atoms. Optimization involves a large relative movement of the bases (rmsd of 1.99 Å), accompanied by huge reduction in buckle and propeller and an increase in the open angle, resulting in the convergence of the bifurcated H-bond to a

Table 2. MP2/aug-cc-pVDZ Interaction Energies for Protonated Base Pair Optimized at the B3LYP/cc-pVTZ Level of Theory^a

system no.	base pair	edge interaction	$E_{\text{int}}^{\text{tot}}$ (kcal/mol)	$E_{\text{int}}^{\text{MP2}}$ (kcal/mol)	$E_{\text{int}}^{\text{HF}}$ (kcal/mol)	$E_{\text{int}}^{\text{corr}}$ (kcal/mol)	E_{def}^{\pm} (kcal/mol)
1	A(+):C	W:WC	-35.37	-39.84 (-38.51)	-37.03 (-36.25)	-2.81 -2.26	4.47
2	A(+):C	W:WT	-33.26	-37.86 (-22.12)	-34.08 (-22.30)	-3.78 (0.18)	4.60
3	A(+):G	W:HC	-35.15	-40.90 (-35.06)	-36.01 (-31.55)	-4.89 (-3.51)	5.75
4	A(+):G	W:HT	-35.99	-41.79 (-38.50)	-38.01 (-33.95)	-3.78 (-4.55)	5.80
5	rA(+):G	S:HC		-34.79 (-24.45)	-29.83 (-20.23)	-4.96 (-4.22)	
6	C(+):C	W:WC	-34.57	-38.51 (-37.71)	-35.36 (-35.86)	-3.15 -1.85	3.94
7	C(+):C	W:WT	-42.45	-47.88 (-46.43)	-43.33 (-43.62)	-4.55 (-2.81)	5.43
8	C(+):U	W:WC	-24.58	-29.42 (-27.40)	-26.88 (-25.70)	-2.54 (-1.70)	4.84
9	C(+):U	W:WT	-23.46	-26.68	-23.46	-3.22	3.22
10	C(+):A	W:WT	-23.08	-30.75 (-13.48)	-23.42 (-12.34)	-7.33 (1.14)	7.97
11	C(+):A	W:HT	-19.32	-25.45 (-21.54)	-19.10 (-15.82)	-6.35 (-5.72)	6.13
12	C(+):G	W:WC	-30.97	-39.97 (-30.88)	-38.10 (-29.07)	-1.87 (-1.81)	9.00
13	C(+):G	W:HC	-39.02	-44.95 (-42.66)	-40.39 (-37.24)	-4.56 (-5.42)	5.93
14	C(+):G	W:HT	-35.42	-40.87 (-39.36)	-36.53 (-37.34)	-4.34 (-2.02)	5.45
15	C(+):rC	+W:SC		-48.83 (-33.96)	-45.29 (-30.50)	-3.54 (-3.46)	
16	C(+):rU	W:ST		-34.12 (-26.57)	-30.02 (-23.71)	-4.10 (-3.86)	
17	G(+):C	S:WC	-38.83	-43.78 (-33.10)	-41.82 (-32.54)	-1.96 (-0.56)	4.95
18	G(+):G	S:HT	-40.48	-47.02 (-36.83)	-44.11 (-35.69)	-2.91 (-1.14)	6.54
19	G(+):rG	S:H T		-48.97 (-37.79)	-43.11 (-33.88)	-5.86 (-3.91)	

^a “ E_{def} ” and $E_{\text{int}}^{\text{tot}}$ values are not calculated for “Fopt” geometries of those base pairs where ribose sugar is retained during calculations, as “ E_{def} ” may acquire large values due to flexibility of the ribose sugar, which may bias the final interaction energy values. ^a The values in parentheses correspond to that of H-optimized crystal geometries. Note that these values are not available for sys. no. 9, since it is a modeled geometry.

Table 3. Protonated Base Pair Systems with Donor–Acceptor Network Analogous with Neutral Base Pairs (Column 4), the Potential of Switching from Protonated to Nonprotonated Geometries (Column 5), and Their Propensity to Nucleate in to Higher Order Structures (Column 6)

system no.	base pair	edge interaction	D/A network analogous with	switching potential ^a	higher order structure ^a
1	A(+):C	W:W cis		—	+
2	A(+):C	W:W trans		+	—
3	A(+):G	W:H cis		—	—
4	A(+):G	W:H trans		—	+
5	rA(+):G	S:H cis		+	—
6	C(+):C	W:W cis	G:C W:W trans	—	—
7	C(+):C	W:W trans	G:C W:W cis	+	—
8	C(+):U	W:W cis	G:U W:W trans	+	—
9	C(+):U	W:W trans	G:U W:W cis	—	—
10	C(+):A	W:W trans	G:A W:W cis	+	—
11	C(+):A	W:H trans	G:A W:H cis	+	+
12	C(+):G	W:W cis	G:G W:W trans	+	+
13	C(+):G	W:H cis	G:G W:H trans	—	+
14	C(+):G	W:H trans	G:G W:H cis	—	+
15	C(+):rC	W:S cis		—	—
16	C(+):rU	W:S trans		—	—
17	G(+):C	S:W cis		—	—
18	G(+):G	S:H trans		—	—
19	G(+):G	S:H trans		+	+

^a “+” indicates that the switching potential or higher order structure is present for a particular base pair, whereas “—” indicates the absence of such propensities.

single 0.43 Å shorter, and hence far stronger, N3(+)(G18)–H···O6(G51) H-bond.

In all the studied cases, on optimization, a decrease in H-bond D–A distances is observed for those hydrogen bonds that involve interaction of a protonated atom. As is indicated by the

relatively small changes in base pair parameters (Table S2), and their respective rmsd values compared to crystal geometries (Table 1), this is accompanied mostly with minor changes in relative orientation of two nucleobases. This observation supports the protonation hypothesis in most cases and may be used for the

purpose of disambiguation related to protonation site (e.g., system 2 vs system 10).

3.3. Addressing Ambiguities. In the Introduction, we listed several issues that lead to ambiguities in identifying and characterizing protonated base pairs. We have addressed some of these instances in our study.

3.3.1. Ambiguity Due to Donor/Acceptor Duality of O2'. Molecular interactions within the base pairs in systems 15 ($C(+)$:
rC (W:S cis,) and 16 ($C(+)$:rU W:S cis) can also be considered without assuming protonation. As explained in the previous section, and as encountered earlier in our investigations related to amino acceptor interactions,⁵⁸ the interacting ribose O2' in these cases can, in principle, act both as H-bond donor or acceptor.

Resolution of ambiguity was attempted by optimizing Hopt models representing the two alternative possibilities corresponding to both systems 15 and 16, respectively. In the first case, O2' was considered as a donor (without N3 protonation) implying a nonprotonated geometry, whereas in the second case, O2' was considered as an acceptor (with N3 protonation). Obviously, due to an increase in electrostatic interaction because of protonation, the protonated base pair does not undergo any change in H-bonding pattern on optimization, whereas a change in H-bonding pattern is observed in the nonprotonated geometry in both systems 15 and 16. However, the ribose sugar undergoes greater deviation from crystal geometry in the case of protonated geometries, compared to their respective nonprotonated counterparts. Given the large flexibilities observed within the sugar moieties, further corroborative evidence is needed before drawing firm conclusions.

3.3.2. Ambiguity Pertaining to Location of Protonation. System 2 ($A(+)$:C W:W trans) and system 10 ($C(+)$:A W:W Trans) are two alternative models for the A:C W:W trans geometry (A2371:C2403 in 1NKW: Table 1) where, in principle, either of the two closely situated acceptor atoms, namely N1 of adenine or N3 of cytosine, could be considered as a protonated hydrogen bond donor. Disambiguation was attempted by carrying out the geometry optimization and interaction energy calculations for the two models.

On optimization, both systems showed high propeller twisting, possibly to avoid the repulsion between the hydrogen atoms of C2(A) and N4(C), while retaining the other hydrogen bonds in the respective starting models. In fact, apart from the reversal of hydrogen bond donor/acceptor roles of N1(A) and N3(C), with system 2 having a N1(+)(A)—N3(C) H-bond and system 10 having a N3(+)(C)—N1(A) H-bond, there were no major differences between the two final geometries. They could, however, be distinguished in terms of their respective rmsd values compared to the experimental structure, their N1(A)—N3(C) distances, and their interaction energies. Our results come out clearly in favor of system 2, which shows lower rmsd (Table 1, Figure 5), has a N1(A)—N3(C) distance smaller by 0.27 Å (Table S1, Supporting Information), and is stabler by more than 10 kcal/mol (Table 2), compared to system 10. Further studies involving comparison of energies required to protonate N3(C) vs N1(A) and pK_a of these two bases, respectively, may shed some more light on the relative importance of the two geometries.

The extra stabilization of system 2 can be understood in terms of three factors:

- (I) Increased protonation-induced acidity of the N6—H of adenine, and hence better hydrogen bond donor efficiency in the N6(A)—O2(C) H-bond,⁷¹ in system 2.
- (II) High exchange repulsion component, higher than the corresponding electrostatic component, in system 10 (Table S3, Supporting Information).

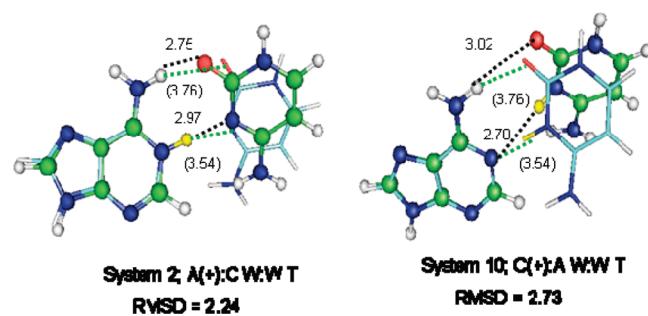


Figure 5. Superposition of gas phase optimized (ball and stick) and crystal geometries (sticks) of $A(+)$:C W:W trans and $A:C(+)$ W:W trans protonated base pair geometries.

(III) Stronger ion—dipole interactions⁷² in the $A(+)$:C W:W trans model in system 2 compared to the $C(+)$:A W:W trans model in system 10. Not only does cytosine have a much higher dipole moment than adenine, whereas N1 protonation increases the dipole moment of adenine, but also N3 protonation of cytosine actually decreases the dipole moment of cytosine.

In contrast to the above example, one cannot distinguish between two possible alternative protonated models corresponding to system 7, having a C(+):C W:W trans, since it exists in a hemiprotonated symmetrical Y:Y form, where either of the two interacting cytosines must be protonated at N3 to allow their close approach. In such cases, because of the symmetry involved, optimization of both possible models will lead to identical geometries as well as interaction energies.

3.3.3. Ambiguity Related to Multimodality of Edge Interaction. BPFind⁶¹ is able to detect examples of G:G H:S trans base pair that can best be modeled as an N3 protonated G:G(+):H:S trans base pair (system 19), where both the Hopt and Fopt geometries are stabilized by three hydrogen bonds including the one involving the ribose O2' with N1 of guanine. On the other hand, the example of a “nonprotonated” G:G H:S trans base pair, which was reported earlier by Leontis et al.,⁵² has been shown to converge to a G:G H:W trans geometry,²⁶ on fully relaxed optimization. Interestingly, the latter example may also be modeled as in system 18, which considers an N3 protonated guanine interacting with its sugar edge involved in a bifurcated hydrogen bonding pattern. On relaxed optimization, the bifurcated hydrogen bonding pattern changes to a stable geometry with a two hydrogen bond pattern quite different from the three hydrogen bond pattern observed in system 19. This raises the question as to whether system 18 is neutral or protonated.

The fact that the G:G H:S trans geometry is maintainable, on full optimization, only in the event of protonation at N3 of guanine, argues in favor of N3 protonation of one of the guanines in system 18. This implies that the optimization of the two distinct types of crystal geometries, corresponding to systems 18 and 19, in turn leads to two different G:G(+):H:S trans geometries. The existence of two distinct minimum energy structures corresponding to the two systems suggests that these geometries constitute examples of multimodal interactions involving protonation. It also suggests that there may be scope of inter conversion between the two structures.

The edge interaction pattern observed in system 19 deserves special mention, particularly since this base pair seems to play a crucial role in a four-way RNA junction motif⁷³ and that, because of its strength, it may be forcing the formation of stems, involving

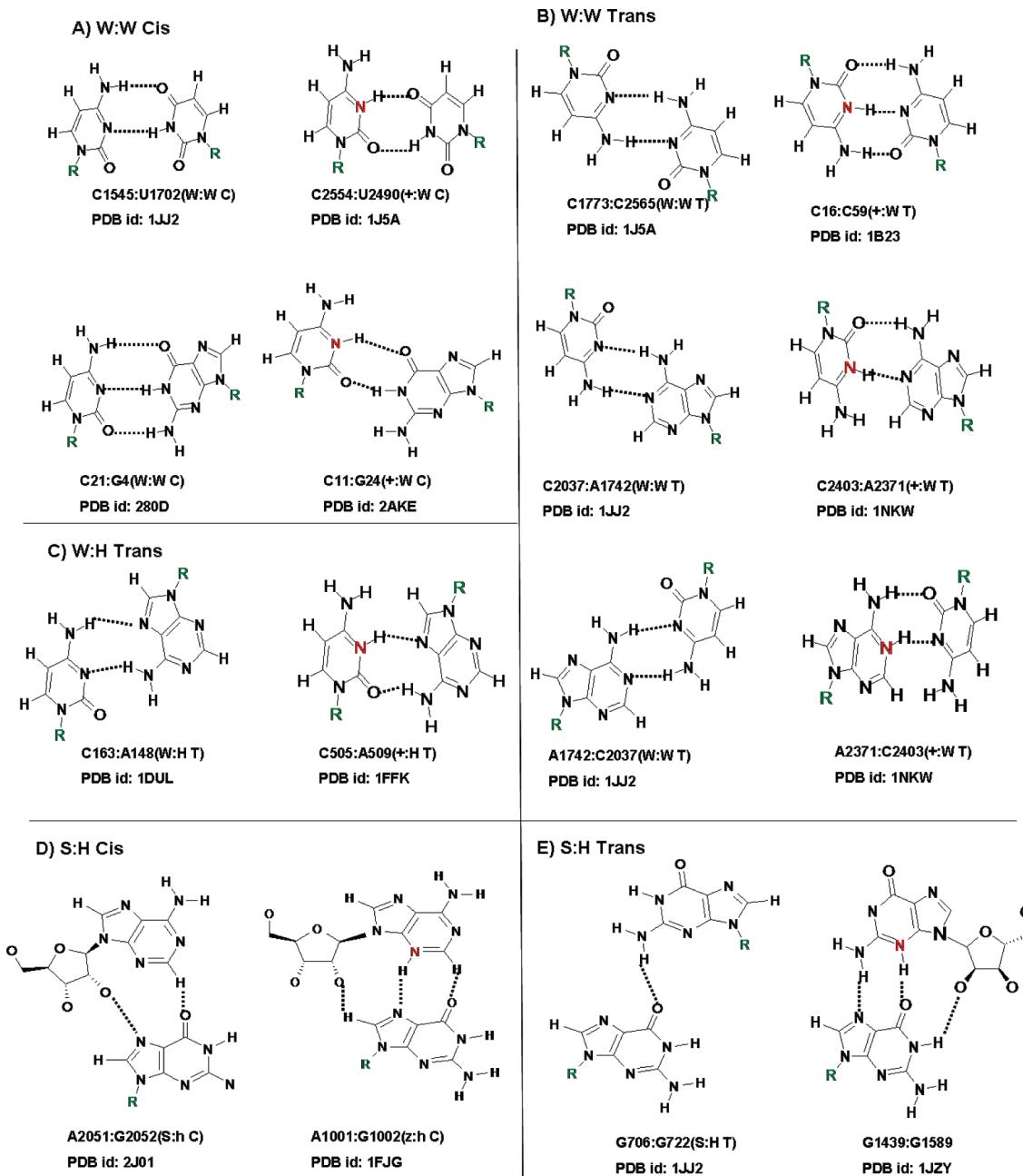


Figure 6. Possible base pair switches that can exist both in protonated and in nonprotonated geometries.

highly distorted base pairs, on either side of the junction. Note that system 19 represents a unique case where the guanine interacting from its sugar edge spans both the Hoogsteen edge and Watson–Crick edge of the partner guanine. Such an interaction, where two edges of the same base interact with the partner base, is not defined in the LW nomenclature.⁵²

3.4. In Search of the Rationale: Significance and the Driving Force for Protonation of Nucleobases. Having carried out a comprehensive search of the RNA crystal structure database, to compile a list of protonated base pairs along with their contexts, and having characterized their geometric and energetic features, we attempt here to address the question of the “why” and “how” of protonation in the context of RNA. In the process we propose a few hypotheses regarding why nature may have invoked the protonation of bases in functional RNA

molecules. Apart from the significance of protonated base pairs in the context of providing catalytically significant electrostatic environment, we have looked into several other possible reasons. Protonated base pairs may be involved in pH dependent conformational switches and may also play important roles in nucleating higher order structures. In addition, we have also looked into the interesting possibilities arising out of protonation induced changes in the hydrogen bonding functionality of bases.

3.4.1. Several Protonated Base Pairs May Play the Role of pH Dependent Switches. The significance of protonation of base pairs in some cases can be explained in terms of pH dependent conformational switch. We could find eight pairs of instances (Figure 6) of base pairs interacting with similar edges, but with different hydrogen bonding patterns, where one set of instances consists of neutral base pairs while members of the other set

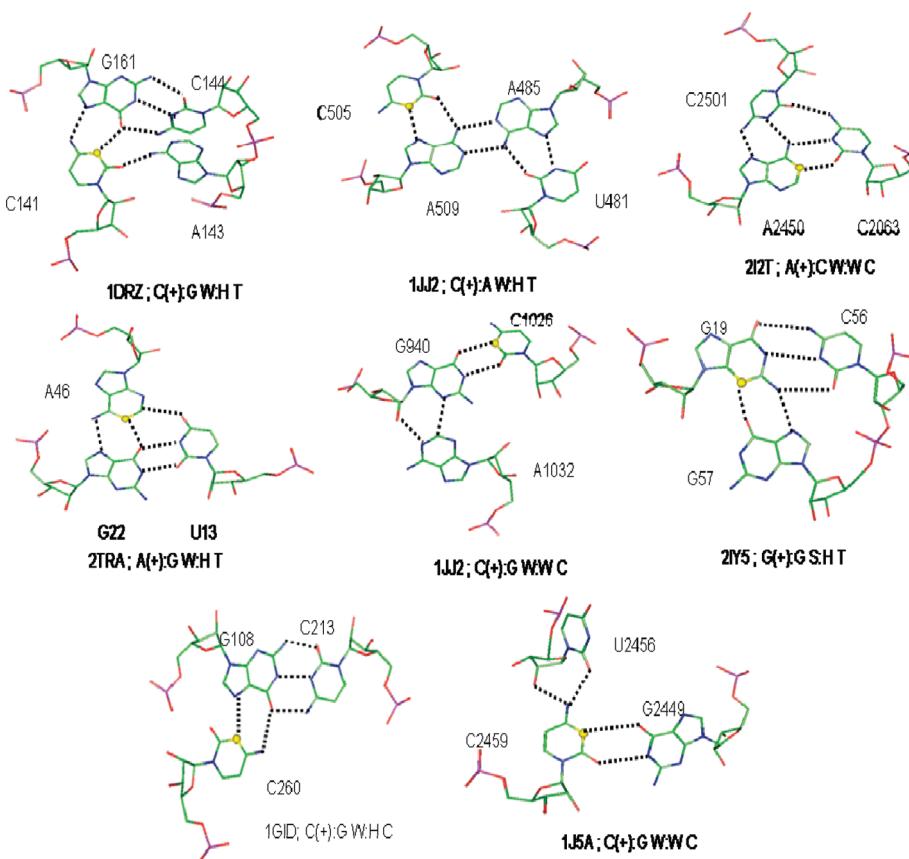


Figure 7. Some higher order recurrent motifs containing one protonated base. PDB ids and the nucleating protonated base pairs are indicated in parentheses.

involves protonation. In addition to comparing the geometries of protonated and their corresponding nonprotonated forms in terms of differences in their respective hydrogen bonding patterns, we have also carried out quantitative comparison of two of their base pair parameters, viz. shear and opening, to estimate the relative shift during a possible protonation/deprotonation induced switching event.

For example, in systems 8 and 12, the shear and opening values undergo large increases on protonation. In contrast, as a result of protonation, four of the systems viz. systems 2, 7, 10, and 19, showed increases in respective values of opening, with simultaneous decreases in corresponding values of shear. In addition to these, we have system 5, where the value of opening does not change much, but the nonprotonated geometry possesses a larger shear value. Note that the nonprotonated crystal geometry has two hydrogen bonds, compared to three in the protonated case, which explains the role of protonation in the sharp reduction in opening accompanying a slight increase in shear value. In all these cases, there is a large variation in the value of shear in the optimized geometries of the protonated and their corresponding nonprotonated models. This suggests that in the larger context of RNA structures, large pH dependent movements along the shear axes of these base pair combinations can be instrumental in initiating conformational switching.

3.4.2. Protonated Bases May Be Important for the Stabilization of Tertiary and Higher Order Interactions. Our studies show that several protonated base pairs are involved in the formation of higher order structures such as triplets and quartets. It is possible that protonation is crucial for the formation of such higher order structures. Figure 7 shows some such important

structures. Some such recurrent motifs involving protonated base pairs are discussed below:

3.4.2.1. C+GCA Motif. This is a structurally conserved motif,^{75,76} found in BWYV (beet western yellows virus) and PEMV (pea enation mosaic virus) pseudoknots,¹⁶ which can be analyzed in terms of a C(+):G W:H trans base pair as the core element. Here the Watson–Crick edge of the guanine participates in a stable canonical G:C W:W Cis base pairing with a distant cytosine, while the fourth base, an adenine, interacts with its exocyclic N6 with the O2 of the protonated cytosine. We have observed a similar interaction in 23S rRNA of *H. marismortui*, where a water molecule replaces the adenine nucleobase to form a water mediated triplet similar to the C+GCA motif. This quartet is also formed by C41 of genomic hepatitis delta virus ribozyme, where the effect of canonical C41 and possible compensatory ion binding, and the role of protonation–deprotonation in mRNA functional conformational switch of BYWV, has been studied using MD simulations.⁴⁷⁷

3.4.2.2. C+AAU Motif. This is a structurally conserved interaction observed in the 23s rRNA of both an Archaea (*H. marismortui*) and a bacteria (*E. coli*). In *H. marismortui*, the C505(+):A509 W:H trans base pair is a part of a lone pair–triloop (LPTL) motif where A509 also base pairs with A485 of A485:U481 H:W trans, which, in turn, is a part of different LPTL motif. As discussed earlier in the context of W:W trans base pairs,⁵⁶ nucleation of this quartet brings together tertiary regions of RNA.

3.4.2.3. CA(+):C Motif. This involves a conserved “spanning” type of interaction also observed in the 23S rRNA of both *H. marismortui* and *E. coli* with identical hydrogen bonding patterns. Instances of this motif can be analyzed from two perspectives, both

of which provide some ideas and insights into its role in the larger macromolecular context.

One way is to consider A(+):C W:W cis as the core base pair with another cytosine involved in a “spanning” interaction from the major groove side. The core A(+):C is a stable base pair, rightly labeled as the A:C counterpart of the G:U wobble base pair, and can be considered as a pair that can occur in helical regions in the same manner as the G:U wobble pair does. Accordingly, out of 19 nonredundant instances of this base pair, we find at least 9 instances where it occurs in the regular stem region (at terminal/middle), 7 instances where it involves interactions joining different junction loops and stem regions, 2 instance where it is present in internal loop, and 1 instance where it is a part of the C-loop motif.

Another way of looking at this triplet motif is to consider A:C H:W trans as the core and third base cytosine pairing with its sugar edge interacting with the Watson–Crick edge of protonated adenine. It is interesting to note that the A:C H:W trans interaction is an important recurrent component of C-loop motifs. This triple, with the potential of providing the third cytosine with a switching option of binding or not binding depending on the protonation of Adenine at N1, may be of functional importance.

3.4.2.4. A(+)GU Motif. This is another type of “spanning” base interaction observed in tertiary contacts of tRNA molecules (detailed context in the Supporting Information). Here, the Watson–Crick edge of an N1 protonated adenine interacts with three acceptor atoms spanning both the nucleobases, of a G:U W:W Cis wobble base pair, by docking in from the major groove of the pair. Please note that the protonation of the adenine at N1 not only converts it into a donor atom but also enhances the donor characteristics of the neighboring C2 and N6 atoms.

3.4.2.5. AGC(+) Motif. This type of interaction is observed in the 23s rRNA of *H. marismortui*. It may be considered similar to the type 1 A-minor motif interaction, except for that here, instead of a canonical G:C W:W cis base pair, we have a protonated G:C(+) W:W cis base pair interacting with the sugar edge of the adenine.

3.4.2.6. GG(+)C Motif. In this motif, observed in tRNA^{Phe}, an N3 protonated guanine can be seen to be simultaneously interacting with two consecutive bases, a 5' cytosine and a neighboring 3' guanine, with its Watson–Crick and sugar edge, respectively. This gives rise to an interesting situation, where the protonated guanine forms a stable canonical G:C W:W cis base pair with the cytosine, and a G(+):G S:H trans base pair with the other guanine. In the process, it holds the consecutive cytosine and guanine bases in the same plane, in a manner similar to a CrG S:H cis dinucleotide platform. The difference is that the consecutive bases do not have any direct hydrogen bond between them.

3.4.2.7. C(+)GC Motif. Here a protonated cytosine forms a C(+):G W:H cis base pair from the major groove of a canonical G:C W:W Cis base pair. Observed in the group I Intron, such C(+):GC triplets are well-known in the literature and have been studied intensively in the past in conjunction with nucleic acid triple helices.⁷⁸ In fact, the propensity of a protonated polycytidyllic strand to bind with G:C duplex sequences, through such C(+):GC triplet formation, has been exploited in antisense research using the TFO (major groove triplex forming oligonucleotide) approach.⁷⁹

3.4.2.8. UC(+)G Motif. This motif is observed in 23s rRNA of the large ribosomal sub unit of the extremophile *D. radiodurans*, where the N4 of the protonated cytosine of a C(+):G W:W cis

Table 4. Interactions Involving Covariation in 22:46^a Positions Involved in Tertiary Interaction Joining D-Loop and Variable Loop in tRNA Molecules

biomolecule; organism	PDB id	interacting nucleobases
tRNA ^{Glu} ; <i>E. coli</i>	2DER	A(+):46:G22:U13
tRNA ^{Glu} ; <i>Thermus thermophilus</i>	1N77	A(+):546:G522:U513
tRNA ^{Asp} ; <i>Saccharomyces cerevisiae</i>	2TRA	A(+):46:G22:PSU13
tRNA ^{ile} ; <i>Staphylococcus aureaus</i>	1FFY	G46:G22:C13
tRNA ^{met} ; <i>E. coli</i>	2J00	G46:G22:C13
tRNA ^{phe} ; <i>S. cerevisiae</i>	1EHZ	7MG46:G22:C13
tRNA ^{phe} ; <i>Th. thermophilus</i>	2IY5	G46:G22:C13
tRNA ^{pro} ; <i>Th. thermophilus</i>	1H4S	G46:G22:C13
tRNA ^{thr} ; <i>Bacillus subtilis</i>	2FK6	G46:G22:C13
tRNA ^{thr} ; <i>E. coli</i>	1QF6	7MG46:G22:C13
tRNA ^{tyr} ; <i>Methanococcus jannaschii</i>	1J1U	G523:G547:CS13
tRNA ^{val} ; <i>Th. thermophilus</i>	1GAX	G921:G945:CS13

^a The actual numbers may vary because of variation in chain numbering and in the sequences considered.

base pair is involved in bifurcated hydrogen bonding with the sugar edge of a uracil.

3.4.3. Protonation Confers Alternative Hydrogen Bonding Functionalities to Bases. While different constituent bases play distinct, often unique, roles in defining the structure and dynamics of RNA, the rich variety and complexity observed in functional RNA molecules cannot be understood without understanding the roles of base pairs and their higher order structures.^{80–82} A major pointer is the phenomenon of covariation observed among structurally and functionally important base pairs. While this can, to a very large extent, be understood in terms of isostericity and related principles, we also need to consider the variations in the stabilities of the base pairs and of the hydrogen bonding functionalities of both individual bases and base pairs.^{26,56,59,60}

In this context, we present some interesting observations related to hydrogen bonding functionalities of two protonated bases: N1-protonated adenine and N3-protonated cytosine.

3.4.3.1. Apparent Covariation of A(+):G W:H Trans and G:G W:H Trans Geometry at 22:46 Positions in tRNA Molecules. A(+):G and G:G base pairs have respectively similar glycosidic C1'-C1' distances in their W:H trans geometries and are found to covary at 22:46 position involved in tertiary interaction joining D-loop and variable loop in tRNA molecule (Table 4). They are, however, not quite isosteric⁵² and do not exhibit a pairwise covariation in the strict sense. As discussed above, the A(+):G W:H trans base pair constitutes a recurrent A(+):GU motif, which is never found to be replaced by a corresponding GGU motif. This is because the Watson–Crick edge of N1-protonated A(+):46 has a unique donor–donor–donor arrangement, which allows it to conveniently dock with the acceptor–acceptor–acceptor network, spanning the major groove edge of the G22:U13 wobble pair. In contrast, replacement of A(+):46 by G46, which has a donor–donor–acceptor network on its Watson–Crick edge, will rule out the spanning interaction with U13 (Figure 8). In fact, in most tRNAs we have a geometrically different GGC motif, instead of the A(+):GU motif. Thus in those tRNA molecules, whenever the A(+) is replaced by neutral G, we have the G:U wobble base pair replaced by a canonical G:C base pair. Since these triplets mediate the interaction between the D-loop and V-loop, they are likely to be significant in the context of recent studies on the role of tRNA flexibility on the speed and fidelity of translation.^{83–85} This highlights the need for looking

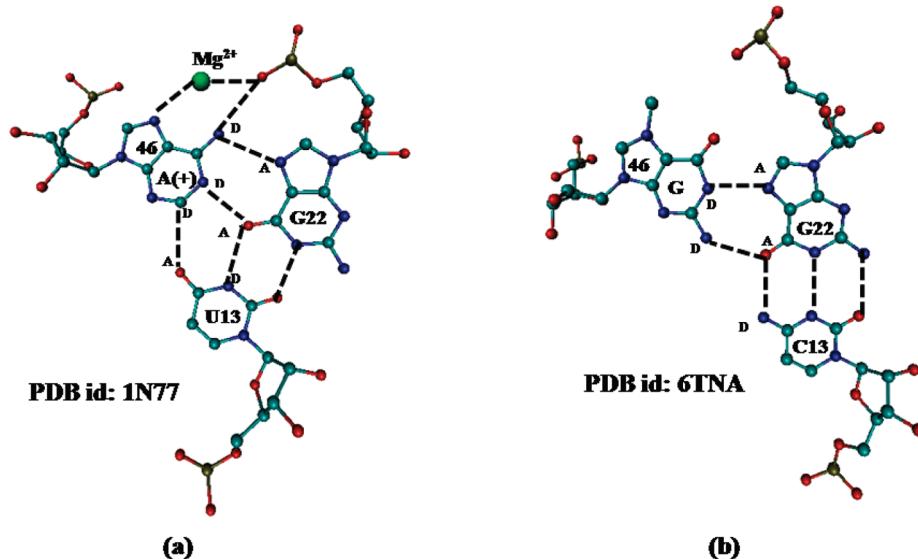


Figure 8. Replacement of A(+)46 by G46 rules out the spanning interaction of this nucleotide with pyrimidine at position 13. The hydrogen bonding sites are labeled as “D” for donor atoms, and “A” for acceptor atoms.

Table 5. Examples of Base Pairs Involving Neutral Guanine, Having Analogous Hydrogen Bonding Pattern as in Base Pairs Containing N3-Protonated Cytosine

neutral guanine containing base pair	corresponding base pair containing N3-protonated cytosine
G:C(W:W trans)	C(+):C(W:W cis)
G:G(W:W trans)	C(+):G(W:W cis)
G:U(W:W trans)	C(+):U(W:W cis)
G:A(W:W cis)	C(+):A(W:W trans)
G:C(W:W cis)	C(+):C(W:W trans)
G:U(W:W cis)	C(+):U (W:W trans)
G:G(W:H trans)	C(+):G(W:H cis)
G:A(W:H cis)	C(+):A(W:H trans)
G:G(W:H cis)	C(+):G(W:H trans)

beyond pairwise isostericity studies and that of devising strategies for rationalization of the roles of higher order structures.

3.4.3.2. Watson–Crick Edge of N3 Protonated Cytosine Is Very Similar to That of “Reverse” Guanine. A striking example of a protonated base edge having hydrogen bonding functionality analogous to that of a different neutral base edge is the protonation of cytosine at N3, which provides a donor–acceptor network similar to guanine in reverse fashion. Thus, base pairs involving interaction of W-edge of guanine with other base in a cis fashion, possess same donor–acceptor interactions as in case of W-edge of protonated cytosine interacting with other base in a trans fashion. Examples of such base pairs involving protonated cytosine and analogous neutral base pairs involving guanine in reverse orientation are given in Table 5. Though this does not imply isostericity, the commonality in hydrogen bonding functionality has demonstrated design implications.

A major challenge in implementing the TFO approach in antisense research, as mentioned in the context of C(+):GC motif, is the requirement of acidic condition for the protonation of cytosine.^{86,87} In an attempt to eliminate the necessity for protonation, major efforts have been directed toward the synthesis of nonnatural nucleosides that display the hydrogen bonding functionality of

protonated cytosine. One such effort that appears to hold a lot of promise involves replacement of C(+) with non-natural N7-substituted guanine (N7G) as its mimic in synthetic antisense constructs, which can participate in pH-independent triplex formation with cDNA sequences.^{87,88}

3.4.3.3. Hydrogen Bonding Interaction in C(+):C W:W Cis Base Pair Is Analogous with That of the G:C W:W Trans Base Pair: They, However, Have Different Stability Characteristics. It is known that the G:C W:W trans geometry is stabilized, in the crystal context, either by archeosine modification or by Mg²⁺ binding to N7 position of guanine.^{81,89} Energy minimization of the normal G:C W:W trans base pair, in the absence of Mg²⁺, changes this geometry to a bifurcated pattern, indicating inherent instability of this base pair.⁸⁹ Since hydrogen bonding interaction in the C(+):C W:W cis base pair is analogous with that of the G:C W:W trans base pair, it may be expected that it would show a similar change in geometry on optimization. In fact, our earlier calculations, at the HF level of theory, did indeed show a similar behavior.²⁶ However, our current investigations at a higher level of theory, with the inclusion of correlation, shows that the C(+):C W:W cis base pair retains its hydrogen bonding pattern on optimization.

To understand the possible effects of protonation on the C(+):C (W:W cis) geometry, atomic charges were calculated for both N3-protonated cytosine and nonprotonated cytosine, using natural bond orbital analysis. Comparison of atomic charges revealed that on N3 protonation, the negative charge is depleted from amino nitrogen and carbonyl oxygen. This depletion of charge may result in a decrease in amino–amino and carbonyl–carbonyl repulsions between protonated cytosine and the interacting cytosine in the C(+):C (W:W cis) geometry. In addition, the increase in positive charge on amino hydrogen involved in interbase H-bond formation, results in strengthening this H-bond (Figure S1, Supporting Information). Thus, it appears that protonation of cytosine has a stabilizing effect on the C(+):C W:W cis base pair, similar to the role of Mg²⁺ ion binding to or archeosine modification of guanine, in the stabilization of G:C W:W trans geometry. Further corroborative

experimental and theoretical evidence may be required to substantiate this observation.

4. CONCLUSIONS

We carried out detailed structural and energetic analysis of 19 protonated base pairs belonging to different families in RNA structures. Barring one, which is a predicted base pair, all the others have been found to occur in RNA X-ray structure database and, hence, may be playing important roles in maintaining the structure and dynamics of functional RNA molecules. Advanced quantum chemical techniques were used to analyze the relative order of magnitude of interaction energies of the base pairs studied. The gas phase interaction energies were computed at the MP2/aug-cc-pVDZ//B3LYP/cc-pVTZ level. In addition, Morokuma decomposition of the HF component showed that the interaction energies, being dominated by the electrostatic component, are significantly higher than those of neutral base pairing geometries. Within different protonated base pairs, higher interaction energies were observed for cases where a protonated nucleobase (A/G/C) interacts with C/G because of the high dipole moment of the latter. This could be explained in terms of stronger ion–dipole interactions. Apart from determining the occurrence frequencies of base pairs detected earlier, database analysis using BPFind software led to the identification of four hitherto undetected protonated base pairs which possess adenine and guanine molecules protonated at N3 positions.

Because of the absence of protons in experimental X-ray crystal structures, and the consequent lack of direct evidence for protonation from these structures, identification and characterization of protonated base pairs depend upon hypothesis driven analysis at two different levels: (1) Are these base pairs really protonated, and if so, where are the protons attached? (2) Do these protonations play any significant structural or functional role in the biological context? We have carried out quantitative computational determination of interaction energies, to evaluate these hypotheses.

For example, six out of the 19 possibilities, studied in this work, needed manual intervention for addressing ambiguities referred to in point 1 listed above. These were resolved to different degrees using structural and ab initio quantum chemical methods. During our database search we could find one instance of multimodal interaction corresponding to G:G S:H trans protonated base pairing pattern, where the H-bonding pattern was different from one that was proposed by Leontis et al.⁵² We also observed one unique case, where guanine interacting from its sugar edge spans both the Hoogsteen edge and Watson–Crick edge of the partner guanine. Such a combination, where the two edges of one nucleobase interact simultaneously with the single edge of the partner base was not defined in LW nomenclature. Also, on the basis of both geometries and interaction energies, we concluded that protonation pattern in A(+):C W:W trans is more favorable compared to C(+):A W:W trans base pairing geometry.

We have formulated a few hypotheses regarding why nature may have invoked the protonation of bases in functional RNA molecules. Thus, we propose that, apart from providing the electrostatic environment, protonated base pairs may be involved in pH dependent conformational switching and in nucleating into higher order interactions involving triplet and quartet formation. We have also evaluated the possible significance of examples where the protonated edge of a particular base mimics the hydrogen bond donor–acceptor network of an edge of a

different neutral base. One striking example in this context is the Watson–Crick edge of N3 protonated cytosine, which can mimic the same edge of neutral reverse guanine. We report our studies on the C(+):C W:W cis base pair as a positively charged model for the unstable G:C W:W trans (Levitt) base pair. Geometry optimization and analysis of natural charge distribution in C(+):C W:W cis geometry provides insights into the possible role of Mg²⁺ ion binding or archeosine modification, either of which is required for stabilization of G:C W:W Trans geometry. Many of the protonated base pairs are found as part of base triple or quadruple, where protonation of one of the bases may improve stabilities of the other associated base pairs. Moreover, deprotonation due to change in pH might have larger destabilizing effect on the other base pairs.

Our studies involving a comprehensive identification and characterization of protonated base pairs would help in addressing the larger question of characterizing the driving forces leading to the pK_a shifts needed for the protonation of nucleobases, which otherwise would not get protonated at physiological pH.

■ ASSOCIATED CONTENT

S Supporting Information. Tables of geometrical and base pair parameters, energies, context analysis of protonated base pairs, pdb files, and atomic coordinates. Structures showing natural charges. Gaussian input formats. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: D.B., dhananjay.bhattacharyya@saha.ac.in; A.M., abi_chem@iiit.ac.in

■ ACKNOWLEDGMENT

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]. P.S. thank CSIR, New Delhi, for a senior research fellowship.

■ REFERENCES

- (1) Ellington, A. D.; Chen, X.; Robertson, M.; Syrett, A. *Int. J. Biochem. Cell Biol.* **2009**, *41*, 254–265.
- (2) Boese, B. J.; Corbino, K.; Breaker, R. R. *Nucleosides Nucleotides Nucleic Acids* **2008**, *27*, 949–966.
- (3) Chen, X.; Li, N.; Ellington, A. D. *Chem. Biodivers.* **2007**, *4*, 633–655.
- (4) Csaszar, K.; Špačková, N.; Stefl, R.; Šponer, J.; Leontis, N. B. *J. Mol. Biol.* **2001**, *313*, 1073–1091.
- (5) Mlynšky, V.; Banas, P.; Hollas, D.; Reblova, K.; Walter, N. G.; Šponer, J.; Otyepka, M. V. *J. Phys. Chem. B* **2010**, *114*, 6642–6652.
- (6) Banas, P.; Walter, N. G.; Šponer, J.; Otyepka, M. V. *J. Phys. Chem. B* **2010**, *114*, 8701–8712.
- (7) Cate, J. H.; Gooding, A.; Podell, E.; Zhou, K.; Golden, B.; Kundrot, C.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1678–1685.
- (8) Pyle, A. M. *Crit. Rev. Biochem. Mol. Biol.* **2010**, *45*, 215–232.
- (9) Steitz, T. A. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 242–253.
- (10) Mulhbacher, J.; Lafontaine, D. A. *Nucleic Acids Res.* **2007**, *35*, 5568–5580.
- (11) Sharma, M.; Bulusu, G.; Mitra, A. *RNA* **2009**, *15*, 1673–1692.
- (12) Misra, V. K.; Shiman, R.; Draper, D. E. *Biopolymers* **2003**, *69*, 118–136.
- (13) Cech, T. R.; Bass, B. L. *Annu. Rev. Biochem.* **1985**, *55*, 599–629.

- (14) Dahm, S. C.; Uhlenbeck, O. C. *Biochemistry* **1991**, *30*, 9464–9469.
- (15) Lippert, B. *Chem. Biodivers.* **2008**, *5*, 1455–1474.
- (16) Tang, C. L.; Alexov, E.; Pyle, A. M.; Honig, B. *J. Mol. Biol.* **2007**, *366*, 1475–1496.
- (17) Rupert, P. B.; Ferre-D'Amare, A. R. *Nature* **2001**, *410*, 780–786.
- (18) Mirkin, S. M.; Frank-kamenetskii, M. D. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 541–576.
- (19) Zain, R.; Sun, J. S. *Cell. Mol. Life Sci.* **2003**, *60*, 862–870.
- (20) Lu, Q.; Teare, J. M.; Granok, H.; Swede, M. J.; Xu, J.; Elgin, S. C. R. *Nucleic Acids Res.* **2003**, *31*, 2483–2494.
- (21) Nixon, P. L.; Giedroc, D. P. *J. Mol. Biol.* **2000**, *296*, 659–671.
- (22) Chen, L.; Cai, L.; Zhang, X.; Rich, A. *Biochemistry* **1994**, *33*, 13540–13546.
- (23) Gehring, K.; Leroy, J. L.; Gueron, M. *Nature* **1993**, *363*, 561–565.
- (24) Venditti, V.; Clos, L.; Niccolai, N.; Butcher, S. E. *J. Mol. Biol.* **2009**, *391*, 894–905.
- (25) Reiter, N. J.; Blad, H.; Abildgaard, F.; Butcher, S. E. *Biochemistry* **2004**, *43*, 13739–13747.
- (26) Bhattacharyya, D.; Koripella, S. C.; Mitra, A.; Rajendran, V. B.; Sinha, B. *J. Biosci.* **2007**, *32*, 809–825.
- (27) Bevilacqua, P. C.; Brown, T. S.; Nakano, S.; Yajima, R. *Biopolymers* **2004**, *73*, 90–109.
- (28) Bevilacqua, P. C.; Brown, T. S.; Chadalavada, L.; Lecomte, J.; Moody, E.; Nakano, S.-I. *Biochem. Soc. Trans.* **2005**, *33*, 466–470.
- (29) Perrotta, A. T.; Shih, I.; Been, M. D. *Science* **1999**, *286*, 123–126.
- (30) Nakano, S.; Chadalavada, D. M.; Bevilacqua, P. C. *Science* **2000**, *287*, 1493–1497.
- (31) Lilley, D. M. J. *RNA* **2004**, *10*, 151–158.
- (32) Bink, H. J.; Hellendoorn, K.; Meulen, J. V.; Pleij, C. W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 13465–13470.
- (33) Wu, R.; McMahon, T. J. *Am. Chem. Soc.* **2006**, *129*, 569–580.
- (34) Wang, H.; Zhang, J. D.; Schaefer, H. F. *Chem. Phys. Chem.* **2010**, *11*, 622–629.
- (35) Zhang, J. D.; Schaefer, H. F. *J. Chem. Theory Comput.* **2007**, *3*, 115–126.
- (36) Šponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers* **2001**, *61*, 3–36.
- (37) Liguori, A.; Napoli, A.; Sindona, G. *J. Mass Spectrom.* **2000**, *35*, 139–144.
- (38) Touboul, D.; Bouchoux, G.; Zenobi, R. *J. Phys. Chem. B* **2008**, *112*, 11716–11725.
- (39) Cai, Z.; Tinoco, I., Jr. *Biochemistry* **1996**, *35*, 6026–6036.
- (40) Legault, P.; Pardi, A. *J. Am. Chem. Soc.* **1994**, *116*, 8390–8391.
- (41) Walberer, B. J.; Cheng, A. C.; Frankel, A. D. *J. Mol. Biol.* **2003**, *327*, 767–780.
- (42) Knitt, D. S.; Narlikar, G. J.; Herschlag, D. *Biochemistry* **1994**, *33*, 13864–13879.
- (43) Kubarenko, A. V.; Sergiev, P. V.; Bogdanov, A. A.; Brimacombe, R.; Dontsova, O. A. *Nucleic Acids Res.* **2001**, *29*, 5067–5070.
- (44) Chen, G.; Kennedy, S. D.; Turner, D. H. *Biochemistry* **2009**, *48*, 5738–5752.
- (45) Durant, P. C.; Bajji, A. C.; Sundaram, M.; Kumar, R. K.; Davis, D. R. *Biochemistry* **2005**, *44*, 8078–8089.
- (46) Zimmerman, G. R.; Shields, T. P.; Jenison, R. D.; Wick, C. L.; Pardi, A. *Biochemistry* **1998**, *37*, 9186–9192.
- (47) Siegfried, N. A.; Hare, B. O.; Bevilacqua, P. C. *Biochemistry* **2010**, *49*, 3225–3236.
- (48) Moody, E. M.; Lecomte, J. T. J.; Bevilacqua, P. C. *RNA* **2005**, *11*, 157–172.
- (49) Izatt, R. M.; Christensen, J. J.; Rytting, J. H. *Chem. Rev.* **1971**, *71*, 439–481.
- (50) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
- (51) Šponer, J.; Hobza, P. *Collect. Czech. Chem. Commun.* **2003**, *68*, 2231–2282.
- (52) Leontis, N. B.; Stombaugh, J.; Westhof, E. *Nucleic Acids Res.* **2002**, *30*, 3497–3531.
- (53) Šponer, J. E.; Špačková, N.; Leszczynski, J.; Šponer, J. *J. Phys. Chem. B* **2005**, *109*, 11399–11410.
- (54) Šponer, J. E.; Špačková, N.; Kulhánek, P.; Leszczynski, J.; Šponer, J. *J. Phys. Chem. B* **2005**, *109*, 2292–2301.
- (55) Šponer, J. E.; Leszczynski, J.; Sychrovský, V.; Šponer, J. *J. Phys. Chem. B* **2005**, *109*, 18680–18689.
- (56) Sharma, P.; Mitra, A.; Sharma, S.; Singh, H.; Bhattacharyya, D. *J. Biomol. Struct. Dyn.* **2008**, *25*, 709–732.
- (57) Roy, A.; Panigrahi, S.; Bhattacharyya, M.; Bhattacharyya, D. *J. Phys. Chem. B* **2008**, *112*, 3786–3796.
- (58) Mládek, A.; Sharma, P.; Mitra, A.; Bhattacharyya, D.; Šponer, J.; Šponer, J. E. *J. Phys. Chem. B* **2009**, *113*, 1743–1755.
- (59) Sharma, P.; Chawla, M.; Sharma, S.; Mitra, A. *RNA* **2010**, *16*, 942–957.
- (60) Sharma, P.; Šponer, J. E.; Šponer, J.; Sharma, S.; Bhattacharyya, D.; Mitra, A. *J. Phys. Chem. B* **2010**, *114*, 3307–3320.
- (61) Das, J.; Mukherjee, S.; Mitra, A.; Bhattacharyya, D. *J. Biomol. Struct. Dyn.* **2006**, *24*, 149–161.
- (62) Petrova, T.; Podjarny, A. *Rep. Prog. Phys.* **2004**, *67*, 1565–1605.
- (63) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Ciosowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
- (64) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordian, M. S.; Jensen, J.; Koseky, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
- (65) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (66) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1993**, *37*, 785–789.
- (67) Šponer, J.; Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142–10151.
- (68) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.
- (69) Kitaura, K.; Morokuma, K. *Int. J. Quantum Chem.* **1976**, *10*, 325–340.
- (70) Sokalski, W. A.; Roszak, S.; Hariharan, P. C.; Kaufman, J. *J. Int. J. Quantum Chem.* **1983**, *23*, 847–854.
- (71) Cammi, R.; Bonnacorsi, R.; Tomasi, J. *Theor. Chim. Acta* **1985**, *68*, 271–283.
- (72) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887–1930.
- (73) Sokalski, W. A.; Roszak, M. *J. Mol. Struct.: THEOCHEM* **1991**, *234*, 387–400.
- (74) Mukherjee, S.; Bansal, M.; Bhattacharyya, D. *J. Comput. Aided Mol. Des.* **2006**, *20*, 629–645.
- (75) Nixon, P. L.; Cornish, P. V.; Suram, S. V.; Geidroc, O. P. *Biochemistry* **2002**, *41*, 10665–10674.
- (76) Nixon, P. L.; Rangan, A.; Kim, Y.-G.; Rich, A.; Hoffman, D. W.; Hennig, M.; Giedroc, D. P. *J. Mol. Biol.* **2002**, *322*, 621–633.
- (77) Krasovska, M. V.; Sefcikova, J.; Špačková, N.; Šponer, J.; Walter, N. G. *J. Mol. Biol.* **2005**, *351*, 731–748.
- (78) Cheng, Y. K.; Pettit, B. M. *Prog. Mol. Biophys.* **1992**, *58*, 225–257.
- (79) Dagle, J. M.; Weeks, D. L. *Differentiation* **2001**, *69*, 75–82.

- (80) Sühnel, J. *Biopolymers* **2002**, *61*, 32–51.
- (81) Oliva, R.; Tramontano, A.; Cavallo, L. *RNA* **2007**, *13*, 1427–1436.
- (82) Stombaugh, J.; Zirbel, C. L.; Westhof, E.; Leontis, N. B. *Nucleic Acids Res.* **2009**, *37*, 2294–2312.
- (83) Daviter, T.; Murphy, F. V., 4th; Ramakrishnan, V. *Science* **2005**, *308*, 1123–1124.
- (84) Sanbonmatsu, K. Y. *Biochimie* **2006**, *88*, 1075–1089.
- (85) Schmeing, T. M.; Voorhees, R. M.; Kelley, A. C.; Gao, Y. G.; Murphy, F. V., 4th; Weir, J. R.; Ramakrishnan, V. *Science* **2009**, *326*, 688–694.
- (86) Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*, 1360–1363.
- (87) Brunar, H.; Dervan, P. B. *Nucleic Acids Res.* **1996**, *24*, 1987–1991.
- (88) Kumar, V. A.; D'Costa, M.; Ganesh, K. N. *J. Org. Chem.* **2003**, *68*, 4439–4445.
- (89) Oliva, R.; Cavallo, L.; Tramontano, A. *Nucleic Acids Res.* **2006**, *34*, 865–879.