ACS OMEGA

Article

Cite This: ACS Omega 2019, 4, 699−709

http://pubs.acs.org/journal/acsodf

Evidence for Hidden Involvement of N3-Protonated Guanine in RNA Structure and Function

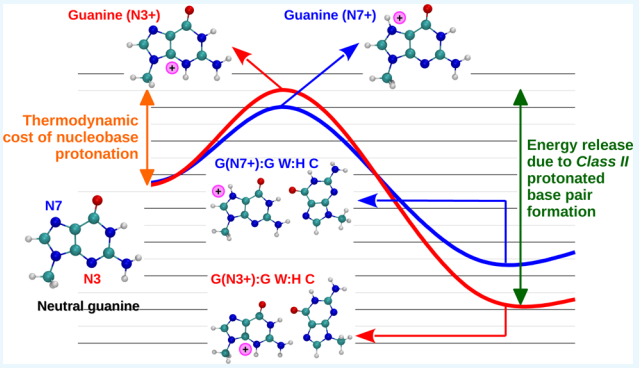
Antarip Halder,[†,§] Saurabh Vemuri,[†] Rohit Roy,[†] Jayanth Katuri,[†] Dhananjay Bhattacharyya,[‡] and Abhijit Mitra*[,†]

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

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

Supporting Information

ABSTRACT: Charged nucleobases have been found to occur in several known RNA molecules and are considered essential for their structure and function. The mechanism of their involvement is however not yet fully understood. Revelation of the role of N7-protonated guanine, in modulating the geometry and stability of noncanonical base pairs formed through its unprotonated edges [Watson−Crick (WC) and sugar], has triggered the need to evaluate the feasibility of similar roles of other protonated nucleobases [Halder et al., *Phys Chem Chem Phys*, **2015**, *17*, 26249]. In this context, N3 protonation of guanine makes an interesting case as its influence on the charge distribution of the WC edge is similar to that of N7 protonation, though its thermodynamic cost of protonation is significantly higher. In this work, we have carried out structural bioinformatics analyses and quantum mechanics-based calculations to show that N3 protonation of guanine may take place in a cellular environment, at least in the G:C W:W Trans and G:G W:H Cis base pairs. Our results provide a reasonable starting point for future investigations in order to address the larger mechanistic question.



INTRODUCTION

Noncoding RNAs participate in a wide variety of cellular processes,[1] including regulation of gene expression,[2] catalysis of biophysical processes,[3] providing scaffold to bring together multiple proteins into an active complex,[4,5] and so forth. To perform such diverse functions, RNA has to fold into functionally competent structures, whose complexities are often comparable to that of proteins.[6] In proteins, the complex folded structures are constructed using twenty physicochemically different amino acids,[7] whereas RNAs use only four nucleotides which are also physicochemically very similar.[8] So then, how does RNA display such structural complexity and functional diversity? A part of the answer possibly lies in biochemically mediated variations in nucleobase properties. Different sources of such modifications are post-transcriptional changes,[9] metal ion coordination,[10] ionization[11] and tautomerization[12] of nucleobases, and so forth. Among them, though ionization of nucleobases is thermodynamically unfavorable at the biological pH (~7.4),[a] involvement of charged (mainly protonated) nucleobases and their importance in modulating nucleic acid's structure[15−18] and function[13,19−21] have been highlighted in various contexts. In fact, charged nucleobases are expected to be associated with RNA functionalities just as charged amino acids have been found to be associated with different catalytic functionalities of proteins (e.g., lysine (LYS) and arginine (ARG) in oxyanion hole formation, histidine (HIS) in general acid−base catalysis, and so forth.).[7] It may however be noted that because of the absence of hydrogen atom coordinates in the crystal structures and ambiguity in detecting them within the solution NMR structures, the charged nucleobases remain "invisible" in the available RNA structure databases and their implications in the RNA structure and dynamics can be studied only on the basis of circumstantial evidence.[22,23]

It is unlikely that an RNA base would be deprotonated as it would give rise to a more negatively charged polyion. However, numerous investigations using both biophysical and theoretical methods have underlined the occurrence and importance of protonated nucleobases in RNA.[11,24] For example, the general acid role of the C75 residue in the cleavage mechanism of hepatitis delta virus ribozyme requires protonation at its N3 position,[25,26] N3 protonation of A2486 in 23S rRNA of *Haloarcula marismortui* (A2451 in *Escherichia coli*) is necessary for its general base role in the mechanism of peptide bond synthesis;[27] also, it has recently been established that c-GAMP binding riboswitches use an N1-protonated

Received: October 23, 2018
Accepted: December 25, 2018
Published: January 9, 2019

DOI: 10.1021/acsomega.8b02908
ACS Omega 2019, 4, 699−709

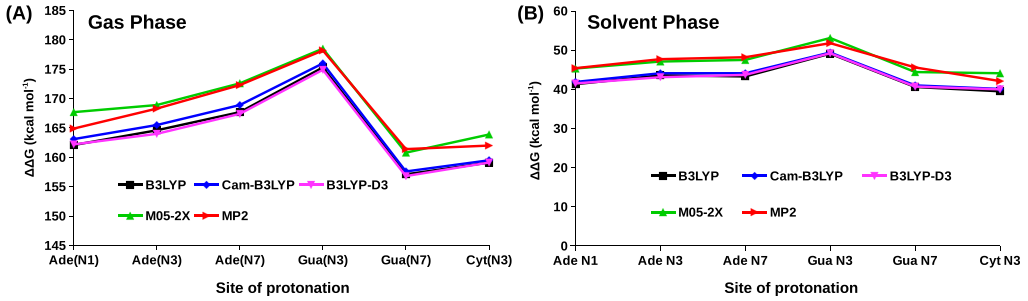


Figure 1. Thermodynamic cost of protonation for protonation at different nucleobase atoms. Calculations are performed at (A) gas phase and (B) conductor-like polarizable continuum model (CPCM) implicit solvent model.

adenine to bind to its near cognate ligand c-di-GMP.[21] Note that each nucleobase can be characterized by three interacting edges, e.g., Watson−Crick (WC), sugar, and Hoogsteen (C−H edge for pyrimidines).[28] Protonation, at a particular edge, changes its hydrogen bond donor−acceptor network, as it converts a potential hydrogen bond acceptor site to a hydrogen bond donor.[22] At the same time, protonation-induced charge redistribution modifies the hydrogen bonding potential of the hydrogen bond donor and acceptor sites of the other two edges.[29] Depending on whether the loaded proton is sequestered between two nucleobases or is free to exchange with the solvent, the protonated nucleobases have been categorized into Class I and Class II, respectively.[11] Class I protonated nucleobases, therefore, give rise to a new class of base pairs (protonated base pairs), where two bases are connected by at least one conventional interbase hydrogen bond and one additional proton-mediated hydrogen bond.[30,31] It is possible to identify the protonated base pairs in RNA crystal structures using in-silico algorithms, such as BPFIND.[32] When two mutually planar bases are found, where, in addition to one conventional interbase hydrogen bond, two potential hydrogen bond acceptor atoms from two different bases fall within the hydrogen bonding distance; BPFIND considers one of them as protonated and hence forms a proton-mediated hydrogen bond with the other. However, in Class II cases, the protonated edge is not involved in base-pairing, irrespective of whether it interacts with other bases or not. Thus, Class II protonation cannot be identified using similar approaches, and we needed to consider other indirect methods for the same.

Our approach for identifying possible instances of Class II nucleobase protonation mainly involved detailed characterization of geometrical features of relevant bases and base pairs, both in the crystal structures as well as after quantum chemical geometry optimization, with and without involving a protonation hypothesis.[23] The feasibility of existence of the putative base pairs, thus shortlisted, was also subsequently assessed through the computational evaluation of interaction energies. In the process, we could rationalize several base-pairing scenarios, among which the cases of G:C W:W Trans and G:G W:H Cis base pairs are remarkable. They occur frequently in RNA crystal structures. However, unlike other base pairs, in gas phase geometry optimization using different post Hartree−Fock (HF) methods, they deviate significantly from their respective crystal geometries.[33−35] Various attempts were made to identify and characterize the factors that stabilize these base pairs within the crystal environment.[36−39] It has been established that crystal geometry of the G:C W:W Trans pair (also known as the Levitt base pair) can be stabilized by positive charge buildup (in the form of metal ion coordination,[37,39] by post-transcriptional modification,[36] etc.) at the Hoogsteen edge of guanine or by its involvement in higher order interactions.[38] In addition to these factors, which are visible in the crystal structure, we have shown that nucleobase protonation at guanine's N7 position, which is invisible in the crystal structure, can also stabilize the crystal geometry of both G:C W:W Trans and G:G W:H Cis pairs.[23] These observations related to N7-protonated guanine underscore the importance of Class II nucleobase protonation in modulating the geometry and stability of RNA base pairs.

In a similar way, analyzing the influence of other protonated nucleobases, on subsequent base-pairing interactions involving their neutral edges, should help in developing a comprehensive understanding about the potential roles of Class II protonation. However, the large variety of possible base-pairing interactions corresponding to different types of Class II protonation (Table S1, Supporting Information), respectively, makes a comprehensive analysis a formidable proposition. Therefore, to begin with, this study focuses on the interesting case of N3 protonation of guanine, where, though biophysical experiments suggest that it is unfavorable,[40] its possible involvement in base-pairing interactions is supported by the analysis of RNA crystal structures[35,41] and quantum chemical studies.[22] It is significant to note here that though the protonation at both N7 and N3 of guanine leads to similar charge redistribution at its WC edge,[29] unlike in the case of N7 protonation, N3 protonation at guanine does not result in weakening of the corresponding glycosidic bond in RNA.[42] The question is whether this difference makes up for the energetic cost of protonation. We have therefore limited our investigations, using RNA crystal structure analysis and quantum mechanics (QM)-based calculations, to possible biophysical consequences of Class II guanine N3 protonation in the relatively high occurring model systems, G:C W:W Trans and G:G W:H Cis base pairs. The objective is to explore whether N3 protonation of guanine has similar impacts on their geometries and stabilities, as found in the case of N7 protonation.

RESULTS AND DISCUSSION

It may be reiterated here at the outset that this approach toward identifying possible Class II protonated base pairs is based on two considerations: one is the closeness of the optimized structure to the observed crystal structure and the other is the extent of base-pairing stabilization, as computed for model systems without protonation versus those with protonation at different possible sites. In principle, for the protonated model systems, one could consider the base-pairing stabilization as composed of two factors—the energy required for protonation and the subsequent stabilization due to base

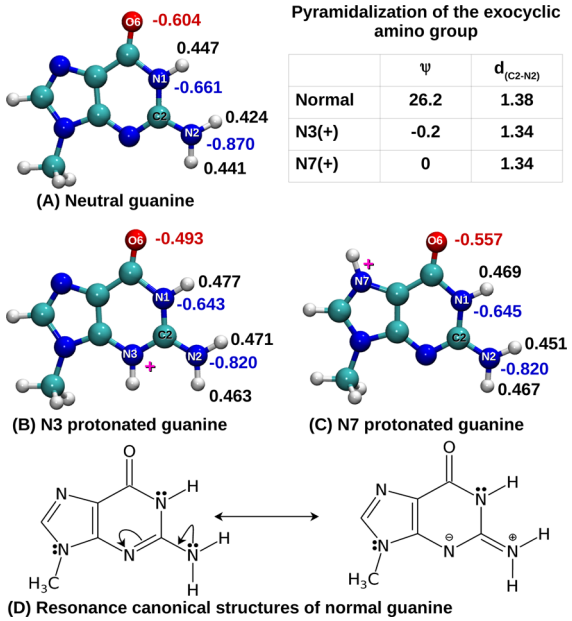
DOI: 10.1021/acsomega.8b02908
ACS Omega 2019, 4, 699−709

pair formation. This is expected to provide insights into "how" the energetically unfavorable protonation takes place and how, through consequent charge redistribution, the base pair formation is energetically compensated. In this context, the relative ease of protonation at N3 vis-a-vis that at N7 of guanine has also been examined.

Guanine N3 Protonation is Thermodynamically less Feasible than N7 Protonation. The thermodynamic cost of protonation ($\Delta\Delta G_{prot}$) at the N3 position of guanine has been reported to be significantly higher than that at the N7 position.[29] Because it was also reported that the $\Delta\Delta G_{prot}$ values are dependent on the level of theory and dielectric constant of the medium,[29] the $\Delta\Delta G_{prot}$ values calculated at the MP2 level were compared with the same calculated at four different density functional theory (DFT) functionals, for example, M05-2X, B3LYP, its long range-corrected (CAM-B3LYP) and its dispersion-corrected version (B3LYP-D3(BJ)). Our results show that out of all the possible sites of nucleobase protonation, the thermodynamic cost of protonation is the maximum for protonation at guanine's N3 position (Figure 1). This is generally valid for both the gas and solvent phases and for all the five levels of theory (Figure 1). However, the difference in the $\Delta\Delta G_{prot}$ values for N3 and N7 protonation of guanine is the least at the MP2 level; 16.8 kcal mol^{-1} in the gas phase and 6.2 kcal mol^{-1} in the solvent phase. In the gas phase, the difference is maximum at CAM-B3LYP (18.4 kcal mol^{-1}), and in the solvent phase the difference is maximum at M05-2X (8.7 kcal mol^{-1}). Essentially, based on this understanding that N7 protonation intuitively appeared to be more probable, we had focused on base pairs involving N7 protonation in our earlier work.[39] However, detailed studies with the two putatively N3-protonated base pairs tell a different story, as is discussed below.

Comparison of N3 and N7 Protonation-Induced Charge Redistribution at the WC Edge of Guanine. In general, protonation at any imino nitrogen reduces the electronic charge density over all the polar atoms in a nucleobase. As a consequence, the hydrogen bonding potential gets improved for the hydrogen bond donor sites. At the same time, the hydrogen bonding potential of the hydrogen bond acceptor sites get compromised.[29] In case of guanine (Figure 2), protonation at the Hoogsteen edge (at the N7 position) reduces the electronic charge density on all the three polar sites of the WC edge (O6, N1, and N2). As a result, hydrogen bonding potential of both N1 and N2 gets improved and that of O6 gets compromised. Previously, it was shown that the resulting stronger H-bonds further stabilize the crystal geometry of the G:C W:W Trans and G:G W:H Cis base pairs.[23] Interestingly, analysis of partial charges obtained from natural population analysis (NPA) suggests that protonation at the sugar edge of guanine (at N3 position) also induces similar changes, especially on the polar sites of the WC edge (Figure 2). In both the cases (N3 and N7 protonation), such charge redistributions are associated with enhanced delocalization of the lone pair of the exocyclic amino group into the ring π system (Figure 2D). The consequent changes in the hybridization state of the amino nitrogen ($sp^3 \rightarrow sp^2$) are characterized by (a) shorter C2–N2 bonds (d_{C2-N2}) and (b) depyramidalization of the exocyclic amino group. Hence, it is expected that guanine N3 protonation should be able to stabilize the G:C W:W Trans and G:G W:H Cis base pairs.

Note that for G:G W:H Cis pair, the N3 protonation may take place at any of the two guanines. However, protonation at



	ψ	$d_{(C2-N2)}$
Normal	26.2	1.38
N3(+)	-0.2	1.34
N7(+)	0	1.34

Figure 2. Partial charges obtained from NPA (at B3LYP/6-31G +(d,p) level) are reported as a fraction of elementary charge (e) for (A) neutral, (B) N3 protonated and (C) N7 protonated guanine. (D) Conjugation between the lone pair of electrons of the exocyclic amino group and the ring π electrons. Red, blue, and black colors are used for oxygen, nitrogen, and hydrogen atoms, respectively. Pyramidalization of the exocyclic amino group has been characterized by the improper dihedral angle (ψ) between C2, H_S, H_W, and N2, where, H_S and H_W represent the hydrogen atoms of the exocyclic amino group that belongs to the sugar edge and the WC edge, respectively. The improper dihedral angle (ψ) and the C2–N2 bond length (d_{C2-N2}) are reported in degrees and Å, respectively.

the N3 position of guanine reduces the hydrogen bonding potential of both the polar sites (N7 and O6) of its Hoogsteen edge (Figure 2B). Therefore, N3 protonation at the second guanine (whose Hoogsteen edge is involved in the W:H cis pairing) is not expected to stabilize the G:G W:H Cis pair. For this reason, in this article, we have considered N3 protonation at only the first guanine of the G:G W:H Cis pair (i.e., whose WC edge is involved in the W:H cis pairing).

Analysis of the Neighboring Residues: Indirect Support for Guanine N3 Protonation Hypothesis. A protonated N3 is likely to be closely associated with crystal water molecules. Accordingly, it has been analyzed and found that (Table 1) out of the 202 occurrences [116 in HDRNAS and 86 in nucleic acid database (NDB)] of G:C W:W Trans base pair and 345 occurrences (123 in HDRNAS and 222 in NDB) of G:G W:H Cis base pair, ~15% instances have at least one water molecule within the 3.5 Å distance of the N3 atom. It is to be noted that, 3.5 Å is a well-accepted cut-off value for the donor–acceptor distance of water-mediated hydrogen bonds in biological systems.[43] Abundance of the water molecules in the neighborhood of guanine N3 is indicative of the possibility of charge transfer between the nucleobase and the neighboring environment.

Within the crystal environment, if positive charge build up is frequently observed at a particular edge of a base pair, it is expected to have an influence on the geometry and stability of the base pair. This is well-documented for the case of the Levitt base pair, where positive charge build up is frequently observed in different forms (post-transcriptional modification,

DOI: 10.1021/acsomega.8b02908
ACS Omega 2019, 4, 699–709

Table 1. Number of Occurrences of G:C W:W Trans and G:G W:H Cis Base Pairs in NDB and HDRNAS Dataset. Number of Occurrences of These Two Base Pairs, Where They Have Atleast One Water Molecule, Metal Cation, or Amino Acid Residue within a Cutoff Distance (3.5, 6 and 10 Å) from the N3 Atom of Guanine

	total occurrences	distance cutoff (Å)	water molecule	metal cation	amino acid residue
G:C W:W Trans					
HDRNAS	116	3.5	6	0	0
		6	16	6	4
		10	33	30	32
NDB	86	3.5	5	1	0
		6	11	7	5
		10	30	26	20
G:G W:H Cis					
HDRNAS	123	3.5	5	1	0
		6	14	24	18
		10	21	52	41
NDB	222	3.5	67	0	0
		6	120	36	10
		10	152	117	29

metal ion coordination, etc.) at the Hoogsteen edge of the guanine.[36,39] This positively charged environment further stabilizes the crystal geometry of the Levitt base pair.[36,37,39] Hence, the RNA crystal structures have been analyzed to find out similar signs of detectable positively charged environment (such as that provided by charged amino acid residues or metal cations) near guanine's sugar edge in G:C W:W Trans and G:G W:H Cis base pairs. Contrary to expectations, Table 1 shows that metal ions or amino acid residues are rarely observed within the 3.5 Å cut-off distance. It is known that metal ions occurring within 3.5−6 Å of an imino nitrogen of a nucleobase may interact with it through the water molecules of its first coordination shell.[44] A search in this range also revealed only some more instances (13%) with at least one metal cation, and even fewer instances (7%) with amino acid residues, near the N3 atom. However, further increasing the cut-off distance to 10 Å reveals an interesting picture. As shown in Table 1, ∼41% instances have at least one metal cation and ∼22% instances have at least one amino acid residue within the 10 Å distance of the N3 atom. Figure S1A,B (in the Supporting Information) further show that among the amino acid residues, the ones with positively charged side chains, for example, LYS, ARG, and HIS are the most frequent in both the nonredundant datasets. Similarly, among the metal cations, Mg^{2+} and K^{+} are the most frequent ones (Figure S1C,D). Occurrences of water molecules also increase significantly on increasing the cut-off distance, for example, ∼29% (at 6 Å) and ∼43% (at 10 Å). In this context, it is important to note that a number of biophysical studies have highlighted the importance of such long-range electrostatic interactions in biological molecules.[45,46] For example, Fersht and co-workers have shown that the pK_a of the active site residue His-64 is affected by the change of only one surface charge which is 14−15 Å away from the active site of the serine protease subtilisin.[47,48] Other studies on metal ion binding to proteins have revealed that long-range electrostatic interactions can modulate the protein-Ca^{2+} binding affinity.[49] It is evident from the above discussion that similar to guanine's Hoogsteen edge in the Levitt base pair, positive charge build up is also frequently observed at guanine's sugar edge in these two base

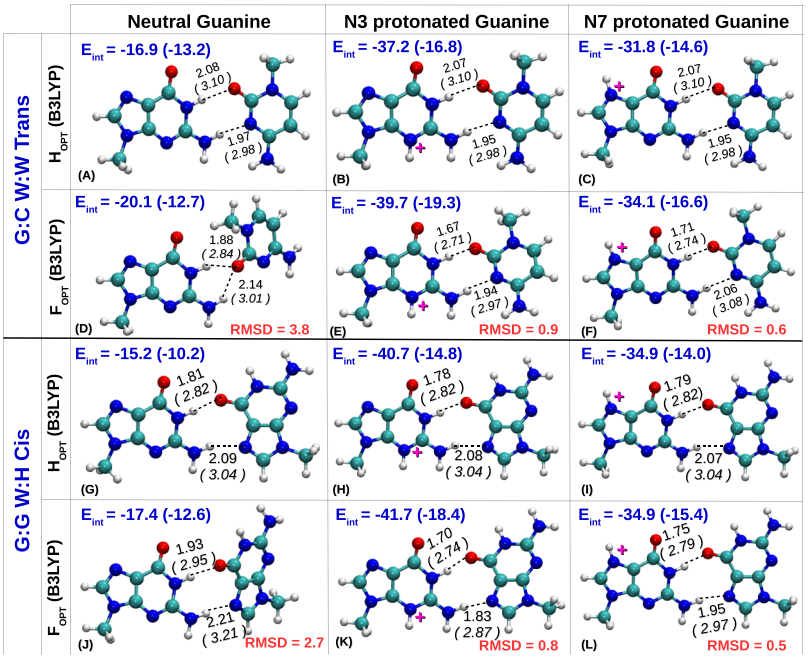


Figure 3. H_{OPT} and F_{OPT} geometries of G:C W:W Trans (A−F) and G:G W:H Cis (G−L) base pairs obtained at B3LYP/6-31+G(d,p). The optimizations have been performed with three different protonation states of guanine: neutral, N3-protonated, and N7-protonated. Interbase hydrogen bonds are shown in broken lines. For each hydrogen bond, the distance between hydrogen and acceptor atom (in normal font) and the distance between the donor and acceptor atom (in italic font, within parentheses) are reported in Å. Interaction energies (E_{int}) of the base pairs (calculated at MP2/aug-cc-pVDZ) have been reported in kcal mol^{-1}. The values in parenthesis represent the solvent-phase interaction energies. RMSD values between the F_{OPT} and H_{OPT} geometries are reported in Å.

DOI: 10.1021/acsomega.8b02908
ACS Omega 2019, 4, 699−709

Table 2. Optimized Geometries of the Base Pairs are Characterized by Three Rotational (Buckle, Open Angle, and Propeller Twist) and Three Translational (Stagger, Shear, and Stretch) Parameters

base pair		buckle	open	propel	stagger	shear	stretch
G:C W:W Trans							
H_{OPT} (B3LYP)	neutral	−2.96	2.72	−0.29	0.24	−2.49	3.09
	Gua-N3(+)	−2.94	2.71	−0.29	0.24	−2.49	3.09
	Gua-N7(+)	−2.95	2.70	−0.31	0.24	−2.49	3.09
F_{OPT} (B3LYP)	neutral	29.90	71.95	41.50	−0.03	−3.53	3.15
	Gua-N3(+)	6.18	13.02	−21.81	−0.03	−2.65	2.85
	Gua-N7(+)	1.97	12.42	−4.17	0.02	−2.79	2.88
F_{OPT} (B3LYP-D3(BJ))	Gua-N3(+)	−4.85	11.38	28.22	0.04	−2.49	2.80
	Gua-N7(+)	−5.77	9.76	19.67	−0.05	−2.61	2.82
F_{OPT} (M05-2X)	Gua-N3(+)	0.06	64.18	0.00	0.00	−4.01	2.88
	Gua-N7(+)	−2.27	19.72	−1.98	−0.02	−3.48	2.84
F_{OPT} (MP2)	Gua-N3(+)	−8.21	11.73	35.07	−0.03	−2.43	2.80
	Gua-N7(+)	−9.78	8.84	36.64	−0.18	−2.43	2.80
G:G W:H Cis							
H_{OPT} (B3LYP)	neutral	−1.01	−6.19	−6.88	0.30	3.22	2.94
	Gua-N3(+)	−1.02	−6.20	−6.87	0.29	3.22	2.94
	Gua-N7(+)	−1.03	−6.20	−6.88	0.29	3.22	2.94
F_{OPT} (B3LYP)	neutral	−40.63	−14.96	39.25	−0.16	2.68	3.03
	Gua-N3(+)	3.54	−5.93	−19.72	0.03	2.84	2.84
	Gua-N7(+)	0.51	−3.80	−0.54	0.01	2.97	2.89
F_{OPT} (B3LYP-D3(BJ))	Gua-N3(+)	7.41	−7.51	−28.51	0.06	2.70	2.78
	Gua-N7(+)	0.60	−3.75	−0.48	0.00	2.91	2.85
F_{OPT} (M05-2X)	Gua-N3(+)	5.83	−8.57	−28.71	0.07	2.77	2.83
	Gua-N7(+)	0.46	−4.58	−0.08	0.01	3.07	2.92
F_{OPT} (MP2)	Gua-N3(+)	12.72	−11.00	−40.40	0	2.48	2.78
	Gua-N7(+)	6.20	−6.81	−39.64	−0.04	2.60	2.78

pairs. Hence, possibility of positive charge buildup in the form of "invisible" protonation cannot be ruled out.

Analysis of the Optimized Geometries: Further Support for Guanine N3 Protonation Hypothesis. *Optimized Geometries with N3- and N7-Protonated Guanine.* The geometries of G:C W:W Trans and G:G W:H Cis base pairs, optimized using various quantum chemical methods, show that in the absence of protonation of the guanine, their F_{OPT} geometries are remarkably different from their respective H_{OPT} geometries (Figure 3). Root mean square deviation (rmsd) between H_{OPT} and F_{OPT} geometries are 3.8 and 2.7 Å, respectively, for G:C W:W Trans (Figure 3A) and G:G W:H Cis (Figure 3G) base pairs. Interestingly, both the F_{OPT} geometries obtained with N3-protonated (Figure 3E,K) and N7-protonated guanine (Figure 3F,L) are significantly close to the respective H_{OPT} geometries (rmsd < 1 Å). Therefore, as with N7 protonation, N3 protonation of guanine may also be considered as a stabilizing factor for these two noncanonical base pairs.

Compensation for the Cost of Protonation. It is to be noted that the F_{OPT} geometries with N3-protonated guanine (Figure 3E,K) have noticeably shorter interbase hydrogen bonding distances and also are significantly more buckled and propeller-twisted (Table 2) compared to N7 protonated ones (Figure 3F,L). It was shown that such changes in planarity of base-pairing interaction (characterized by Buckle, Propeller Twist, and Stagger values), on geometry optimization, are mediated by the interplay and balance of different components of the interaction energies, of the exchange repulsion (EX), and electrostatic attraction (ES) components.[33] Thus, different components of the interaction energies of the H_{OPT} and F_{OPT} geometries of the N3- and N7-protonated systems, as shown in Figure 4, suggest that the comparatively more buckled and twisted geometries of N3-protonated systems are associated with comparatively larger increase in their EX component. This is of course also accompanied by a concomitant compensating increase in the attractive electrostatic component because of the shortening of the hydrogen bonding distances. Further, the importance of the charge−dipole interaction between the two pairing bases in determining the interaction energy has been discussed in the context of Class I protonated base pairs.[22] Here, the positive charge center is located closer to the base-pairing edge (i.e., the WC edge) for protonation at the N3 position and hence is expected to influence the charge−dipole interaction between the two pairing bases to a greater extent. Consequently, the N3 protonated systems are also associated with the comparatively larger increase in the polarization (PL), charge transfer (CT), and ES components (Figure 4).

Distribution of the frontier orbitals corresponding to the optimized geometries of the N3- and N7-protonated base pairs are shown in Figure S2 of the Supporting Information. Note that these two base pairs have two interbase hydrogen bonds (H-bond) between them. Formation of these interbase hydrogen bonds are such that both the H-bond acceptors belong to the neutral base and both the H-bond donors belong to the protonated guanine. Interestingly, in the N3-protonated G:G W:H Cis pair, the LUMO is distributed over the H-bond donors and the HOMO is distributed over the H-bond acceptors. Therefore, a CT between HOMO and LUMO may stabilize the H-bonds along the base pair. A similar complementary orientation of the frontier orbitals is missing in N7-protonated pair.

These factors may be collectively resulting in a higher interaction energy which, in turn, may compensate for the

DOI: 10.1021/acsomega.8b02908
ACS Omega 2019, 4, 699−709

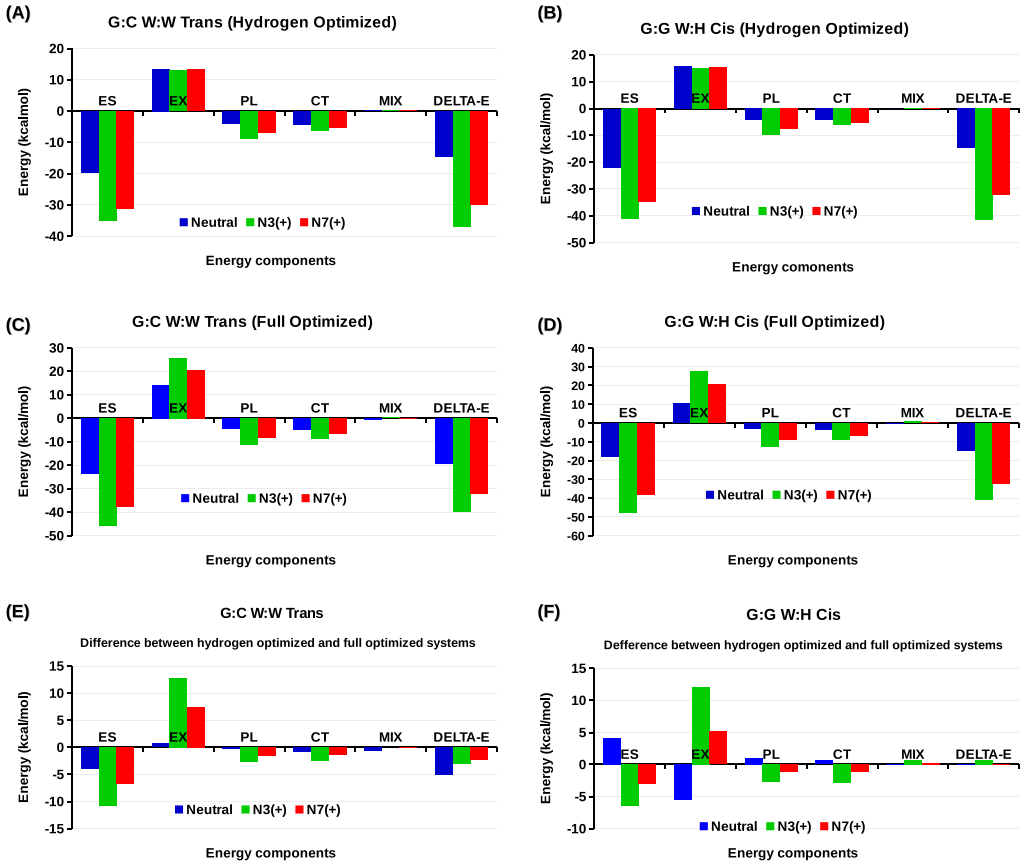


Figure 4. Morokuma–Kitaura energy decomposition of the (A,B) H_{OPT} and (C,D) F_{OPT} geometries of G:C W:W Trans and G:G W:H Cis base pairs with neutral, N3-protonated and N7-protonated guanine. (E,F) Difference between different energy components of the H_{OPT} and (C,D) F_{OPT} systems are also shown. Abbreviations used here for different components of interaction energy are as follows, ES = electrostatic, EX = exchange repulsion, PL = polarization, CT = charge transfer, and MIX = higher-order coupling.

higher thermodynamic cost of protonation at the N3 position. For example, in the gas phase, interaction energy of the N3-protonated systems are higher than that of the N7-protonated systems by 5.6 and 6.8 kcal mol^{-1} for the G:C W:W Trans and G:G W:H Cis pairs (Figure 3), respectively. Considering the effect of the solvent environment, the differences reduce to 2.7 and 3 kcal mol^{-1}, respectively (Figure 3). To understand the reason for such large reduction, the contribution of the Hartree–Fock (% HF) and the correlation (% corr) components of the gas phase interaction energies of these two base pairs have been calculated (Table S2 in the Supporting Information). It has been found that the interaction energies of both N3- and N7-protonated base pairs are dominated by their respective HF components (% HF > 75%), though the contribution of the correlation component is comparatively higher for N7-protonated systems.

Consequences of guanine N3 protonation are not limited only to the formation of stronger base-pairing interactions. Rather, analysis of the charge redistribution at the free edges of the two base pairs (Figure 5) reveals that both N3 and N7 protonation lead to improved H-bonding potential of these free edges. Charge redistribution at a particular site is quantified as the difference between the partial charges (Δq) of the paired base and isolated base. A negative Δq indicates better H-bonding potential for H-bond acceptors. Again, a positive Δq indicates better H-bonding potential for H-bond donors. Therefore, guanine N3 protonation in the G:C W:W Trans pair results in improved H-bonding potential of all the

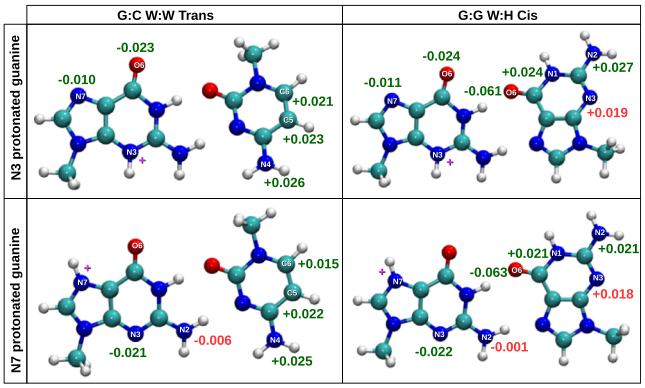


Figure 5. Change in partial charge distribution (Δq) at the H-bond donor–acceptor sites of the free edges due to base-pairing. Green text is used for sites with improved H-bonding potential and red text is used for sites with reduced H-bonding potential. All values are reported as a fraction of elementary charge.

H-bond donor–acceptor sites at the Hoogsteen edge of guanine and the C–H edge of cytosine. Similarly, guanine N3 protonation in the G:G W:H Cis pair results in improved H-bonding potential of all the H-bond donor–acceptor sites at the Hoogsteen edge of the first guanine and the WC edge of the second guanine. Such improvements in the H-bonding potentials of the free edges are expected to result in stronger higher order interactions, such as base triples. Those stronger higher order interactions will further compensate for the

DOI: 10.1021/acsomega.8b02908
ACS Omega 2019, 4, 699–709

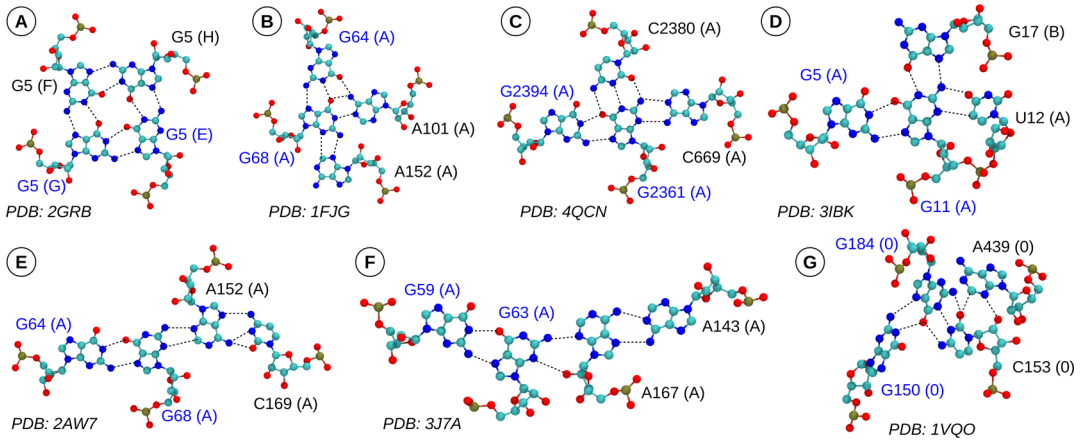


Figure 6. Examples of base quartets that involve a G:G W:H Cis pair. The two guanine bases that form the W:H cis pair are labeled in blue. Interbase hydrogen bonds are shown in broken lines. Out of the seven base quartets shown, one belongs to cyclic topology (A), three belong to star topology (B–D), and three belong to linear topology (E–G). Details of the quartet topologies are described in http://quarna.iiit.ac.in/.

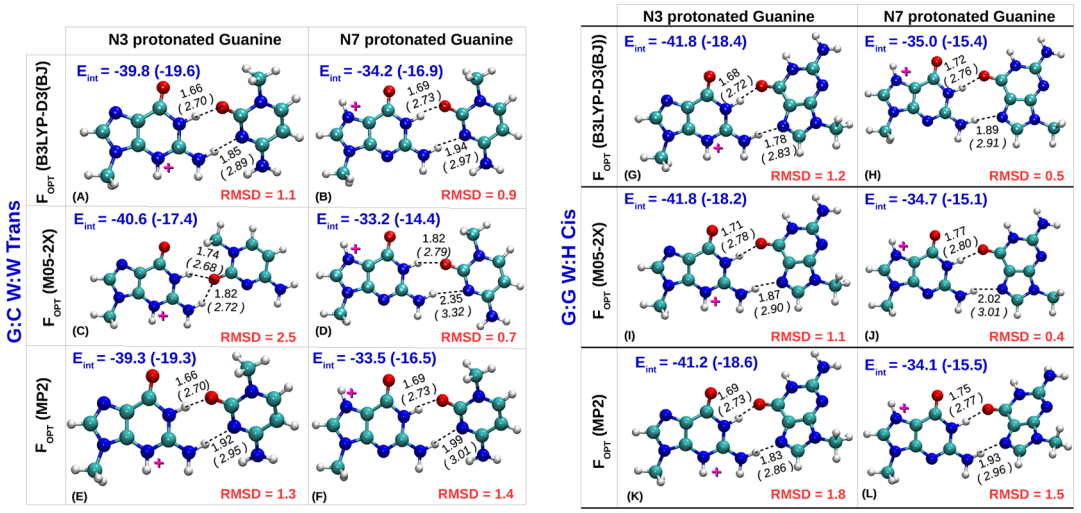


Figure 7. Optimized geometries of the G:C W:W Trans (A–F) and G:G W:H Cis (G–L) pairs with N3- and N7-protonated guanine obtained by the MP2 theory and DFT with M05-2X and B3LYP-D3(BJ) functionals. Other annotations are same as shown in Figure 3.

thermodynamic cost of protonation. Note that improvement in the H-bonding potential of the free edges because of the N7 protonation is somewhat limited. Rather, it has a negative impact on the H-bonding potential of its sugar edge.

Involvement of G:C W:W Ttrans pairs in base triples and quartets have already been studied comprehensively by Chawla and co-workers.[38] Interestingly, G:G W:H Cis pairs also occur frequently in isolated base triples and quartets. We have found that in HDRNAS dataset, 88 out of 123 occurrences of G:G W:H Cis pairs (i.e., 71.5%) interact with a third base to form a base triple. In the NDB data set, the ratio is 54.5%, that is, 121 out of 222 occurrences. Many of these base triples (22 in HDRNAS and 30 in NDB) are also part of a base quartet. These base quartets are of 7 types as illustrated in Figure 6.

Effect of Dispersion Interactions. In order to account for the dispersion interactions, F_{OPT} geometry optimizations have been performed using the MP2 theory and with dispersion corrected DFT functionals, B3LYP-D3(BJ) and M05-2X. Note that M05-2X is highly parameterized to account for middle-range dispersion effects, whereas the DFT-D3 method in conjunction with the Becke–Johnson (BJ) damping function accounts for both long-range and middle-range correlation effects.[50] Figure 7 illustrates the optimized geometries of G:C W:W Trans and G:G W:H Cis base pairs with N3- and N7-protonated guanine. It shows that though there are noticeable variations depending on whether dispersion correction is used or how it is implemented, in general, the results obtained using the B3LYP-D3(BJ) functional and MP2 theory are consistent with the results obtained using the B3LYP functional, that is, (a) both N3 and N7 protonation of guanine can stabilize the crystal geometry of G:C W:W Trans and G:G W:H Cis pairs, which are otherwise unstable under gas phase geometry optimization, (b) N3-protonated systems are more buckled and propeller twisted than N7-protonated systems and also have shorter interbase hydrogen bonds, and (c) interaction energy of N3-protonated systems are higher than that of N7-protonated systems. It may however be kept in mind that parameterized approach toward incorporating dispersion corrections may involve error compensations arising out of different factors, and one may observe discrepancies, in the general trends listed above, for specific cases.

This may be the reason why in the case of computations on G:C W:W Trans pair, using M05-2X functional, the results obtained show that the F_{OPT} geometry with N3-protonated guanine (Figure 7C) is similar to the geometry obtained with neutral guanine (Figure 3D) and is significantly deviated from

DOI: 10.1021/acsomega.8b02908
ACS Omega 2019, 4, 699–709

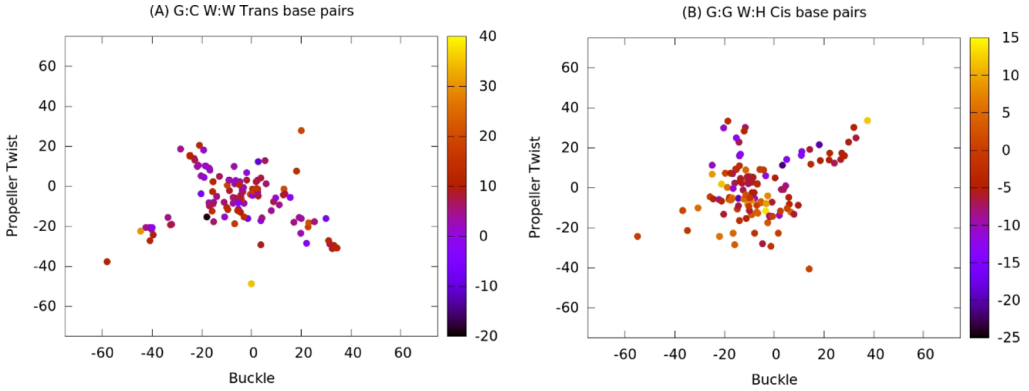


Figure 8. Distribution of the (A) G:C W:W Trans and (B) G:G W:H Cis base pairs observed in HDRNAS dataset in the plane defined by Buckle and Propeller Twist parameters. Data points are colored according to the corresponding values of the open angle.

its crystal geometry (rmsd = 2.5 Å). Another interesting observation in this context is that, for B3LYP-optimized geometries, the buckle and propeller twist observed in the F_{OPT} geometries of N3-protonated systems are comparatively higher than those for N7-protonated systems (Table 2). Such differences between the N3- and N7-protonated systems are not observed at other levels of theory. It may be mentioned in this context that graphene sheets, on geometry optimization at the B3LYP level, retain their planarity but develop a curvature at the DFT-D levels of the theory.[51] This demands for a closer look at the variation of the rotational (buckle, open angle, and propeller twist) and translational (stagger, shear, and stretch) parameters of the optimized base pairs because of the level of theory used for optimization. As illustrated in Figures S3 and S4 of the Supporting Information, the three parameters that define the planarity of a base-pairing interaction (i.e., buckle, propeller twist and stagger) are more sensitive to the level of theory used for optimization, at least for G:C W:W Trans and G:G W:H Cis base pairs. Also, in comparison to the M05-2X functional, the trends shown by the B3LYP-D3(BJ) functional are closer to that shown by the MP2 level of theory.

Overall, the trends obtained from QM calculations are also consistent with the trends observed in instances of G:C W:W Trans and G:G W:H Cis base pairs present in the nonredundant RNA crystal structure datasets. As shown in Figure 8, a number of instances of both the base pairs are associated with similar high buckled and propeller-twisted geometries in the HDRNAS dataset.

CONCLUSIONS

Following up on our earlier work demonstrating the possibility of base pairs involving N7-protonated guanine,[23] we have investigated the possibility of other protonation sites. As test cases, we have carried out quantum chemical studies on the effect of N3 protonation of guanine on the geometry and stability of the G:C W:W Trans and G:G W:H Cis base pairs, which are otherwise unstable when subjected to gas-phase geometry optimization. Consistent trends revealed by our studies at different levels of theory indicate that, while both N7 and N3 protonation of guanine have very similar effects on the geometries and stabilities of these base pairs, the N3-protonated systems have greater stability associated with them.

Detailed energy decomposition studies have also been carried out to rationalize the higher stability of the N3-protonated systems. It may be concluded that both in terms of energetics as well as in terms of consistency with observed nonplanar geometries, N3 protonation appears to be a possible alternative stabilizing factor in the two cases studied. This is despite the higher energy cost involved in N3 protonation compared to N7 protonation.

This study opens up an important area of investigation in understanding the structure, folding, and function of RNA molecules in terms of hitherto undetected protonation of bases.

METHODS

Nomenclature of Base-Pairing Interactions. As per Leontis and Westhof (LW) nomenclature scheme, base-pairing interaction between edge X of base M and edge Y of base N in cis orientation is represented as MN cXY.[28] However, the nomenclature used in this manuscript is slightly different. The same base pair is represented as M:N X:Y Cis. This sort of representation is also intuitive, and at the same time, because of the presence of the text delimiter (:), it is more convenient for parsing of metadata files, especially when modified nucleobases (represented by more than one character, e.g. 7MG, 2MA, etc) are involved. If the interbase hydrogen bonds in a particular base pair involve any of the sugar atoms (O2′, O3′, etc) a prefix "r" is added with the base, e.g., G:rC W:S Cis.

Analysis of RNA Crystal Structure Datasets. For this study, two different nonredundant sets of RNA crystal structures have been analyzed, one is provided by the NDB[52] maintained by the BGSU server and the other is provided by the HDRNAS[53] database. As discussed earlier,[39,54] on the basis of resolution (cutoff = 3.5 Å), chain length, *R*-factor, and so forth, further 838 and 167 structures have been shortlisted from NDB and HDRNAS dataset, respectively (see the Supporting Information). Occurrences of G:C W:W Trans and G:G W:H Cis base pairs within these sets of RNA crystal structures were identified using the BPFIND[32] software. Details of the BPFIND algorithm are discussed in the Supporting Information. Geometries of the detected base pairs have been characterized by 3 rotational (buckle, open, and propeller twist) and 3 translational (stagger, shear, and stretch) parameters, as calculated by NUPARM[55,56] package. Large values of buckle, propeller twist, and stagger parameters indicate the nonplanarity of the two bases. Large open angle, shear, and stretch indicate deformation in the interbase hydrogen bonding. BPFIND estimates the deviation of a base pair from its ideal planar geometry by the geometry-based hybrid parameter called "*E*-value".[32] On the basis of the E-value, the least distorted examples of G:C W:W Trans

DOI: 10.1021/acsomega.8b02908
ACS Omega 2019, 4, 699−709

(515G:548C; 1N78 chain C) and G:G W:H Cis (251G:254G; 3KNH chain A) were selected for QM calculations.

Geometry Optimization. The base-pairing systems were modeled, from the extracted crystal coordinates, by adding hydrogen atoms at appropriate positions and by replacing the sugar and phosphate backbone by a methyl group. As explained by Gatti and co-workers in their recent article,[57] both experimental[58] and computational[59] studies support that replacing sugar-phosphate backbone of nucleotides by a methyl group does not influence the physicochemical properties of base−base interactions. Further, our earlier calculations on purine−purine reverse WC base pairs[60] reinforce the argument that in the context of base pairs, such type of modeling reduces the computational cost without compromising with any important chemical information. The optimized geometries of the modeled systems (base pairs and individual bases) were obtained in two steps. First, all the non-hydrogen atoms were kept fixed at their crystal coordinates and only the hydrogen atoms were allowed to change their positions. Thus, these hydrogen-optimized (or H_{OPT}) geometries represent the crystal geometry with hydrogens placed at their respective minimum energy positions. The H_{OPT} geometries were further optimized, by removing all constraints, to obtain the corresponding full optimized or F_{OPT} geometries, representing the minimum energy structure of the base pairs in the absence of the crystal environment. Hessian calculations performed over the F_{OPT} geometries confirm that they are not associated with any imaginary frequencies. All the QM calculations have been performed using Gaussian 09 software.[61] RMSD between the H_{OPT} and F_{OPT} geometries have been calculated using the VMD software.[62] The geometry optimizations (H_{OPT} and F_{OPT}) have been performed using the Becke, three-parameter (parameterized by Lee−Yang−Parr) hybrid generalized gradient approximation (GGA) functional (B3LYP[63,64]) and a double zeta basis set [6-31G+(d,p)]. B3LYP, is arguably the most well-tested[65−67] DFT functional for studying the ground state electronic properties of nucleobases and similar systems. However, B3LYP does not account for the dispersion interactions. To understand the role of dispersion interactions in these base-pairing systems, the F_{OPT} geometry was also obtained using the Møller−Plesset perturbation theory[68] with second order perturbation correction (MP2) and dispersion-corrected DFT with B3LYP-D3(BJ) and M05-2X functionals. The way of considering the dispersion interactions are fundamentally different in B3LYP-D3(BJ) and M05-2X. In B3LYP-D3(BJ), the dispersion correction has been empirically added by Grimme's method (third order) with Becke−Johnson damping,[69,70] whereas in M05-2X,[71] the dispersion correction has been incorporated during the parameterization itself. For systematic comparison, we have also used the CAM-B3LYP functional,[72] which is the long-range corrected version of B3LYP.

Interaction Energy Calculation. Interaction energies of the optimized systems have been calculated as $E^{int} = E$(base pair) − ΣE(individual bases), using the MP2 theory and aug-cc-pVDZ basis set. The CPCM[74,75] implicit solvent model was implemented to estimate the impact of a polar solvent environment (ϵ = 78.4) over these interaction energies. Kitaura−Morokuma decomposition analysis[76] was performed using the GAMESS-US software,[77] to study the partitioning of the two-body intermolecular interaction energies into electrostatic, PL, CT, and higher-order coupling terms, within the HF approximation. Though we have attempted to calculate the basis set superposition error (BSSE) for the optimized base pairs using the counterpoise[73] method, the MP2/aug-cc-pVDZ level wavefunctions do not converge for some of the cases. Hence, the interaction energies are reported without the BSSE correction. However, it may be noted that earlier studies[23] have reported that for both the neutral and protonated G:C W:W Trans and G:G W:H Cis base pairs optimized at the M05-2X/6-311G+(2d,2p) level, the BSSE energy component is within 3−4 kcal mol^{-1} at the MP2/aug-cc-pVDZ level of theory.

As suggested earlier,[29] the thermodynamic cost of protonation at a particular site of a nucleobase has been determined by the change in free energy ($\Delta\Delta G_{prot}$) of the following process: neutral base + H_2O → protonated base + OH^-. NPA[78,79] was performed using the NBO package[80] incorporated in the Gaussian 09 software to study the charge distribution of the individual bases.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.8b02908.

Gua-N3-protonation-SuppInfo: details of some computational methods, possible sites of Class II protonation in different RNA base pairs, neighboring residues of G:C W:W Trans and G:G W:H Cis base pairs, HF and correlation components of MP2 interaction energy, and frontier orbitals. Coordinates: Cartesian coordinates of the optimized geometries (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: abi_chem@iiit.ac.in (A.M.).

ORCID

Antarip Halder: 0000-0002-3921-5207
Rohit Roy: 0000-0001-8306-121X
Dhananjay Bhattacharyya: 0000-0001-9227-6490
Abhijit Mitra: 0000-0002-6876-4838

Present Address

§Solid State and Structural Chemistry Unit (SSCU), Indian Institute of Science (IISc), Bangalore 560012, Karnataka, India.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

A.H. acknowledges CSIR for SRF support. A.M. 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.

■ ABBREVIATIONS

RNA, ribonucleic acid; NMR, nuclear magnetic resonance; DFT, density functional theory; HF, Hartree−Fock; HOMO, highest occupied molecular orbital; LUMO, lowest unoccupied molecular orbital

DOI: 10.1021/acsomega.8b02908
ACS Omega 2019, 4, 699−709

ADDITIONAL NOTE

[a]Adenine (pK_{a1} = ∼4.1), cytosine (pK_{a1} = ∼4.4) and guanine (pK_{a1} = ∼3.2, pKa2 = ∼9.2), uracil (pK_{a2} = ∼9.2).[13,14]

REFERENCES

(1) Mattick, J. S.; Makunin, I. V. Non-coding RNA. *Hum. Mol. Genet.* **2006**, *15*, R17.

(2) McCaffrey, A. P.; Meuse, L.; Pham, T.-T. T.; Conklin, D. S.; Hannon, G. J.; Kay, M. A. Gene expression: RNA interference in adult mice. *Nature* **2002**, *418*, 38−39.

(3) Doudna, J. A.; Lorsch, J. R. Ribozyme catalysis: not different, just worse. *Nat. Struct. Mol. Biol.* **2005**, *12*, 395−402.

(4) Yik, J. H. N.; Chen, R.; Nishimura, R.; Jennings, J. L.; Link, A. J.; Zhou, Q. Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. *Mol. Cell* **2003**, *12*, 971−982.

(5) Zappulla, D. C.; Cech, T. R. From The Cover: Yeast telomerase RNA: A flexible scaffold for protein subunits. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10024−10029.

(6) Rother, K.; Rother, M.; Boniecki, M.; Puton, T.; Bujnicki, J. M. RNA and protein 3D structure modeling: similarities and differences. *J. Mol. Model.* **2011**, *17*, 2325−2336.

(7) Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*, 5th ed.; W. H. Freeman and Co., 2009.

(8) Egli, M.; Saenger, W. *Principles of Nucleic Acid Structure; Springer Advanced Texts in Chemistry*; Springer-Verlag, 1984; pp P001−P556.

(9) Helm, M. Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res.* **2006**, *34*, 721−733.

(10) Woodson, S. A. Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr. Opin. Chem. Biol.* **2005**, *9*, 104−109.

(11) Wilcox, J. L.; Ahluwalia, A. K.; Bevilacqua, P. C. Charged nucleobases and their potential for RNA catalysis. *Acc. Chem. Res.* **2011**, *44*, 1270−1279.

(12) Singh, V.; Fedeles, B. I.; Essigmann, J. M. Role of tautomerism in RNA biochemistry. *RNA* **2014**, *21*, 1−13.

(13) Bevilacqua, P. C.; Brown, T. S.; Nakano, S.-i.; Yajima, R. Catalytic roles for proton transfer and protonation in ribozymes. *Biopolymers* **2003**, *73*, 90−109.

(14) Verdolino, V.; Cammi, R.; Munk, B. H.; Schlegel, H. B. Calculation of pK_a values of nucleobases and the guanine oxidation products guanidinohydantoin and spiroiminodihydantoin using density functional theory and a polarizable continuum model. *J. Phys. Chem. B* **2008**, *112*, 16860−16873.

(15) Mirkin, S. M.; Frank-Kamenetskii, M. D. H-DNA and related structures. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 541−576.

(16) Zain, R.; Sun, J.-S. Do natural DNA triple-helical structures occur and function in vivo? *Cell. Mol. Life Sci.* **2003**, *60*, 862−870.

(17) Gehring, K.; Leroy, J.-L.; Guéron, M. A tetrameric DNA structure with protonated cytosine. cytosine base pairs. *Nature* **1993**, *363*, 561−565.

(18) Berger, I.; Kang, C.; Fredian, A.; Ratliff, R.; Moyzis, R.; Rich, A. Extension of the four-stranded intercalated cytosine motif by adenineadenine base pairing in the crystal structure of d(CCCAAT). *Nat. Struct. Mol. Biol.* **1995**, *2*, 416−425.

(19) Kuttan, A.; Bass, B. L. Mechanistic insights into editing-site specificity of ADARs. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, E3295−E3304.

(20) Nakano, S.; Chadalavada, D. M.; Bevilacqua, P. C. General Acid-Base Catalysis in the Mechanism of a Hepatitis Delta Virus Ribozyme. *Science* **2000**, *287*, 1493−1497.

(21) Keller, H.; Weickhmann, A. K.; Bock, T.; Wöhnert, J. Adenine protonation enables cyclic-di-GMP binding to cyclic-GAMP sensing riboswitches. *RNA* **2018**, *24*, 1390−1402.

(22) Chawla, M.; Sharma, P.; Halder, S.; Bhattacharyya, D.; Mitra, A. Protonation of base pairs in RNA: context analysis and quantum chemical investigations of their geometries and stabilities. *J. Phys. Chem. B* **2011**, *115*, 1469−1484.

(23) Halder, A.; Bhattacharya, S.; Datta, A.; Bhattacharyya, D.; Mitra, A. The role of N7 protonation of guanine in determining the structure, stability and function of RNA base pairs. *Phys. Chem. Chem. Phys.* **2015**, *17*, 26249−26263.

(24) Goh, G. B.; Knight, J. L.; Brooks, C. L. pH-Dependent Dynamics of Complex RNA Macromolecules. *J. Chem. Theory Comput.* **2013**, *9*, 935−943.

(25) Veeraraghavan, N.; Ganguly, A.; Chen, J.-H.; Bevilacqua, P. C.; Hammes-Schiffer, S.; Golden, B. L. Metal Binding Motif in the Active Site of the HDV Ribozyme Binds Divalent and Monovalent Ions. *Biochemistry* **2011**, *50*, 2672−2682.

(26) Mlnsk, V.; Walter, N. G.; Šponer, J.; Otyepka, M.; Banáš, P. The role of an active site Mg^{2+} in HDV ribozyme self-cleavage: insights from QM/MM calculations. *Phys. Chem. Chem. Phys.* **2015**, *17*, 670−679.

(27) Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A. The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science* **2000**, *289*, 920−930.

(28) Leontis, N. B.; Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **2001**, *7*, 499−512.

(29) Halder, A.; Halder, S.; Bhattacharyya, D.; Mitra, A. Feasibility of occurrence of different types of protonated base pairs in RNA: a quantum chemical study. *Phys. Chem. Chem. Phys.* **2014**, *16*, 18383−18396.

(30) Sowers, L. C.; Fazakerley, G. V.; Eritja, R.; Kaplan, B. E.; Goodman, M. F. Base pairing and mutagenesis: observation of a protonated base pair between 2-aminopurine and cytosine in an oligonucleotide by proton NMR. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 5434−5438.

(31) Sowers, L. C.; Shaw, B. R.; Veigl, M. L.; Sedwick, W. D. DNA base modification: Ionized base pairs and mutagenesis. *Mutat. Res.* **1987**, *177*, 201−218.

(32) Das, J.; Mukherjee, S.; Mitra, A.; Bhattacharyya, D. Non-Canonical Base Pairs and Higher Order Structures in Nucleic Acids: Crystal Structure Database Analysis. *J. Biomol. Struct. Dyn.* **2006**, *24*, 149−161.

(33) Sharma, P.; Mitra, A.; Sharma, S.; Singh, H.; Bhattacharyya, D. Quantum Chemical Studies of Structures and Binding in Non-canonical RNA Base pairs: The Trans Watson-Crick:Watson-Crick Family. *J. Biomol. Struct. Dyn.* **2008**, *25*, 709−732.

(34) Panigrahi, S.; Pal, R.; Bhattacharyya, D. Structure and Energy of Non-Canonical Basepairs: Comparison of Various Computational Chemistry Methods with Crystallographic Ensembles. *J. Biomol. Struct. Dyn.* **2011**, *29*, 541−556.

(35) Bhattacharya, S.; Mittal, S.; Panigrahi, S.; Sharma, P.; Preethi, S. P.; Paul, R.; Halder, S.; Halder, A.; Bhattacharyya, D.; Mitra, A. RNABP COGEST: a resource for investigating functional RNAs. *Database* **2015**, *2015*, bav011.

(36) Oliva, R.; Tramontano, A.; Cavallo, L. Mg^{2+} binding and archaeosine modification stabilize the G15 C48 Levitt base pair in tRNAs. *RNA* **2007**, *13*, 1427−1436.

(37) Oliva, R.; Cavallo, L. Frequency and Effect of the Binding of Mg^{2+}, Mn^{2+}, and Co^{2+} Ions on the Guanine Base in Watson−Crick and Reverse Watson−Crick Base Pairs. *J. Phys. Chem. B* **2009**, *113*, 15670−15678.

(38) Chawla, M.; Abdel-Azeim, S.; Oliva, R.; Cavallo, L. Higher order structural effects stabilizing the reverse Watson-Crick Guanine-Cytosine base pair in functional RNAs. *Nucleic Acids Res.* **2013**, *42*, 714−726.

(39) Halder, A.; Roy, R.; Bhattacharyya, D.; Mitra, A. How Does Mg 2+ Modulate the RNA Folding Mechanism: A Case Study of the G:C W:W Trans Basepair. *Biophys. J.* **2017**, *113*, 277−289.

(40) Wu, R. R.; Yang, B.; Berden, G.; Oomens, J.; Rodgers, M. T. Gas-phase conformations and energetics of protonated 2′-deoxyguanosine and guanosine: IRMPD action spectroscopy and theoretical studies. *J. Phys. Chem. B* **2014**, *118*, 14774−14784.

(41) Sharma, P.; Šponer, J. E.; Šponer, J.; Sharma, S.; Bhattacharyya, D.; Mitra, A. On the Role of the cis Hoogsteen: Sugar-Edge Family of

DOI: 10.1021/acsomega.8b02908
ACS Omega 2019, 4, 699−709

Base Pairs in Platforms and Triplets — Quantum Chemical Insights into RNA Structural Biology. *J. Phys. Chem. B* **2010**, *114*, 3307−3320.
(42) Lenz, S. A. P.; Kohout, J. D.; Wetmore, S. D. Hydrolytic Glycosidic Bond Cleavage in RNA Nucleosides: Effects of the 2′-Hydroxy Group and Acid-Base Catalysis. *J. Phys. Chem. B* **2016**, *120*, 12795−12806.
(43) Jeffrey, G. A.; Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press: New York, 1997; Vol. *32*.
(44) Bowman, J. C.; Lenz, T. K.; Hud, N. V.; Williams, L. D. Cations in charge: magnesium ions in RNA folding and catalysis. *Curr. Opin. Struct. Biol.* **2012**, *22*, 262−272.
(45) Harvey, S. C. Treatment of electrostatic effects in macromolecular modeling. *Proteins: Struct., Funct., Genet.* **1989**, *5*, 78−92.
(46) van Gunsteren, W. F.; Weiner, P. K.; Wilkinson, A. J. *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*; Springer Science & Business Media, 2013; Vol. *3*.
(47) Thomas, P. G.; Russell, A. J.; Fersht, A. R. Tailoring the pH dependence of enzyme catalysis using protein engineering. *Nature* **1985**, *318*, 375−376.
(48) Russell, A. J.; Fersht, A. R. Rational modification of enzyme catalysis by engineering surface charge. *Nature* **1987**, *328*, 496−500.
(49) Pantoliano, M. W.; Whitlow, M.; Wood, J. F.; Rollence, M. L.; Finzel, B. C.; Gilliland, G. L.; Poulos, T. L.; Bryan, P. N. The engineering of binding affinity at metal ion binding sites for the stabilization of proteins: subtilisin as a test case. *Biochemistry* **1988**, *27*, 8311−8317.
(50) Goerigk, L.; Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670−6688.
(51) Panigrahi, S.; Bhattacharya, A.; Banerjee, S.; Bhattacharyya, D. Interaction of nucleobases with wrinkled graphene surface: dispersion corrected DFT and AFM studies. *J. Phys. Chem. C* **2012**, *116*, 4374−4379.
(52) Leontis, N. B.; Zirbel, C. L. *Nonredundant 3D Structure Datasets for RNA Knowledge Extraction and Benchmarking*; Springer, 2012; Vol. *27*, pp 281−298.
(53) Ray, S. S.; Halder, S.; Kaypee, S.; Bhattacharyya, D. HD-RNAS: An automated hierarchical database of RNA structures. *Front. Genet.* **2012**, *3*, 1−10.
(54) Halder, A.; Roy, R.; Bhattacharyya, D.; Mitra, A. Consequences of Mg^{2+} binding on the geometry and stability of RNA base pairs. *Phys. Chem. Chem. Phys.* **2018**, *20*, 21934−21948.
(55) Olson, W. K.; Bansal, M.; Burley, S. K.; Dickerson, R. E.; Gerstein, M.; Harvey, S. C.; Heinemann, U.; Lu, X.-J.; Neidle, S.; Shakked, Z.; et al. A standard reference frame for the description of nucleic acid base-pair geometry 1 1Edited by P. E. Wright 2 2This is a document of the Nomenclature Committee of IUBMB (NC-IUBMB)/IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN), whose members are R. Cammack (chairman), A. Bairoch, H.M. Berman, S. Boyce, C.R. Cantor, K. Elliott, D. Horton, M. Kanehisa, A. Kotyk, G.P. Moss, N. Sharon and K.F. Tipton. *J. Mol. Biol.* **2001**, *313*, 229−237.
(56) Mukherjee, S.; Bansal, M.; Bhattacharyya, D. Conformational specificity of non-canonical base pairs and higher order structures in nucleic acids: crystal structure database analysis. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 629−645.
(57) Gatti, C.; Macetti, G.; Boyd, R. J.; Matta, C. F. An electron density source-function study of DNA base pairs in their neutral and ionized ground states. *J. Comput. Chem.* **2018**, *39*, 1112−1128.
(58) Candeias, L. P.; Steenken, S. Ionization of purine nucleosides and nucleotides and their components by 193-nm laser photolysis in aqueous solution: model studies for oxidative damage of DNA. *J. Am. Chem. Soc.* **1992**, *114*, 699−704.
(59) Gu, J.; Xie, Y.; Schaefer, H. F. Structural and Energetic Characterization of a DNA Nucleoside Pair and Its Anion: Deoxyriboadenosine (dA) − Deoxyribothymidine (dT). *J. Phys. Chem. B* **2005**, *109*, 13067−13075.
(60) Mittal, A.; Halder, A.; Bhattacharya, S.; Bhattacharyya, D.; Mitra, A. Reverse Watson-Crick purine-purine base pairs : the Sharp-turn motif and other structural consequences in functional RNAs. bioRxiv, **2017**.
(61) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian*, 09 Revision C.01; Gaussian Inc.: Wallingford CT, 2009.
(62) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(63) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(64) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **1994**, *98*, 11623−11627.
(65) Dutta, B. J.; Bhattacharyya, P. K. Reactivity and aromaticity of nucleobases are sensitive toward external electric field. *J. Phys. Chem. B* **2014**, *118*, 9573−9582.
(66) Jena, N. R.; Bansal, M.; Mishra, P. C. Conformational stabilities of iminoallantoin and its base pairs in DNA: implications for mutagenicity. *Phys. Chem. Chem. Phys.* **2016**, *18*, 12774−12783.
(67) Negi, I.; Kathuria, P.; Sharma, P.; Wetmore, S. D. How do hydrophobic nucleobases differ from natural DNA nucleobases? Comparison of structural features and duplex properties from QM calculations and MD simulations. *Phys. Chem. Chem. Phys.* **2017**, *19*, 16365−16374.
(68) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618−622.
(69) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(70) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(71) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Design of Density Functionals by Combining the Method of Constraint Satisfaction with Parametrization for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2006**, *2*, 364−382.
(72) Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange−correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51−57.
(73) Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998**, *102*, 1995−2001.
(74) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669−681.
(75) Kitaura, K.; Morokuma, K. A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation. *Int. J. Quantum Chem.* **1976**, *10*, 325−340.
(76) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; et al. General atomic and molecular electronic structure system. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(77) Boys, S. F.; Bernardi, F. The calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors. *Mol. Phys.* **1970**, *19*, 553−566.
(78) Foster, J. P.; Weinhold, F. Natural hybrid orbitals. *J. Am. Chem. Soc.* **1980**, *102*, 7211−7218.
(79) Reed, A. E.; Curtiss, L. A.; Weinhold, F. Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint. *Chem. Rev.* **1988**, *88*, 899−926.
(80) Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F. *NBO*, version 3.1. 2004; Gaussian, Inc.: Wallingford, CT, 2004.

DOI: 10.1021/acsomega.8b02908
ACS Omega 2019, 4, 699−709