On the Role of the cis Hoogsteen:Sugar-Edge Family of Base Pairs in Platforms and Triplets−Quantum Chemical Insights into RNA Structural Biology

Purshotam Sharma,[†] Judit E. Šponer,[‡] Jiří Šponer,[‡] Sitansh Sharma,[†] Dhananjay Bhattacharyya,*[,§] and Abhijit Mitra*[,†]

Center for Computational Natural Sciences and Bioinformatics (CCNSB), International Institute of Information Technology (IIIT-H), Gachibowli, Hyderabad 500032, India, Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic, and Biophysics Division, Saha Institute of Nuclear Physics (SINP), 1/AF, Bidhannagar, Kolkata 700064, India

Received: October 26, 2009; Revised Manuscript Received: February 4, 2010

Base pairs belonging to the cis Hoogsteen:sugar-edge (H:S) family play important structural roles in folded RNA molecules. Several of these are present in internal loops, where they are involved in interactions leading to planar dinucleotide platforms which stabilize higher order structures such as base triplets and quartets. We report results of analysis of 30 representative examples spanning 16 possible base pair combinations, with several of them showing multimodality of base pairing geometry. The geometries of 23 of these base pairs were modeled directly from coordinates extracted from RNA crystal structures. The other seven were predicted structures which were modeled on the basis of observed isosteric analogues. After appropriate satisfaction of residual valencies, these structures were relaxed using the B3LYP/6-31G(d,p) method and interaction energies were derived at the RIMP2/aug-cc-pVDZ level of theory. The geometries for each of the studied base pairs have been characterized in terms of the number and nature of H-bonds, rmsd values observed on optimization, base pair geometrical parameters, and sugar pucker analysis. In addition to its evaluation, the nature of intermolecular interaction in these complexes was also analyzed using Morokuma decomposition. The gas phase interaction energies range between −5.2 and −20.6 kcal/mol and, in contrast to the H:S trans base pairs, show enhanced relative importance of the electron correlation component, indicative of the greater role of dispersion energy in stabilization of these base pairs. The rich variety of hydrogen bonding pattern, involving the flexible sugar edge, appears to hold the key to several features of structural motifs, such as planarity and propensity to participate in triplets, observed in this family of base pairs. This work explores these aspects by integrating database analysis, and detailed base pairing geometry analysis at the atomistic level, with *ab initio* computation of interaction energies. The study, involving alternative classification of base pairs and triplets, provides insights into intrinsic properties of these base pairs and their possible structural and functional roles.

1. Introduction

Complex 3D structures of ncRNA (noncoding RNA), and their enzyme-like functions, have attracted wide coverage in recent literature. While RNAs share similar noncovalent interactions, such as base stacking, hydration, and metal ion coordination, with DNA molecules, unlike the regular dsDNA helices comprised of discrete strands with complementary Watson−Crick base pairs, RNA molecules fold onto themselves. The major difference, which stands out, is the preponderance of noncanonical base pairs, which play crucial roles in scripting the structural variety and functional dynamics of RNA molecules. Understanding the characteristics of noncanonical base pairs thus constitutes a prerequisite toward meeting the challenge of deciphering the rules governing RNA structures and functions.

In RNA structures, nucleobases can potentially pair through any of their three edges: the Watson−Crick edge (W), the Hoogsteen edge (H) (C−H edge in the case of pyrimidines), and the sugar edge (S) (Figure 1).[1] The six consequent base pair combinations of edge to edge interaction, each having both cis and trans as possible mutual orientations of glycosidic bonds, lead to 12 geometric families, which have further been grouped into subfamilies based on isostericity and covariation principles.[2] A simple enumeration, considering the four different base identities along with 5′−3′ pairing asymmetry, gives 168 theoretically possible base pair combinations (Leontis et al.)[1] spanning these 12 families. However, since some edge combinations do not result in plausible molecular interactions, the number of observed base pair types is far less than 168. This is so even after considering the additional possibilities which arise out of multimodality of base pairing geometries, involving substates within individual base pair combinations.

The actual number of "observed" base pair types of course depends on the base pairing criteria set by different groups. Reported approaches vary widely on issues such as the inclusion of C−H···O bonds and the minimum number of interbase hydrogen bonds.[1−5] Complications posed by the flexibility of sugar puckering and glycosidic torsional freedom have also led to different approaches toward the explicit consideration of sugar O2′ in base pairings.[6,7] Additional issues include the status of strong biologically significant S:S base pairs which lack base−base H-bonds,[8] assignment of unique base pair family,

* Corresponding authors. E-mail: dhananjay.bhattacharyya@saha.ac.in (D.B.); abi_chem@iiit.ac.in (A.M.).
[†] International Institute of Information Technology (IIIT-H).
[‡] Academy of Sciences of the Czech Republic.
[§] Saha Institute of Nuclear Physics (SINP).

10.1021/jp910226e
Published on Web 02/17/2010

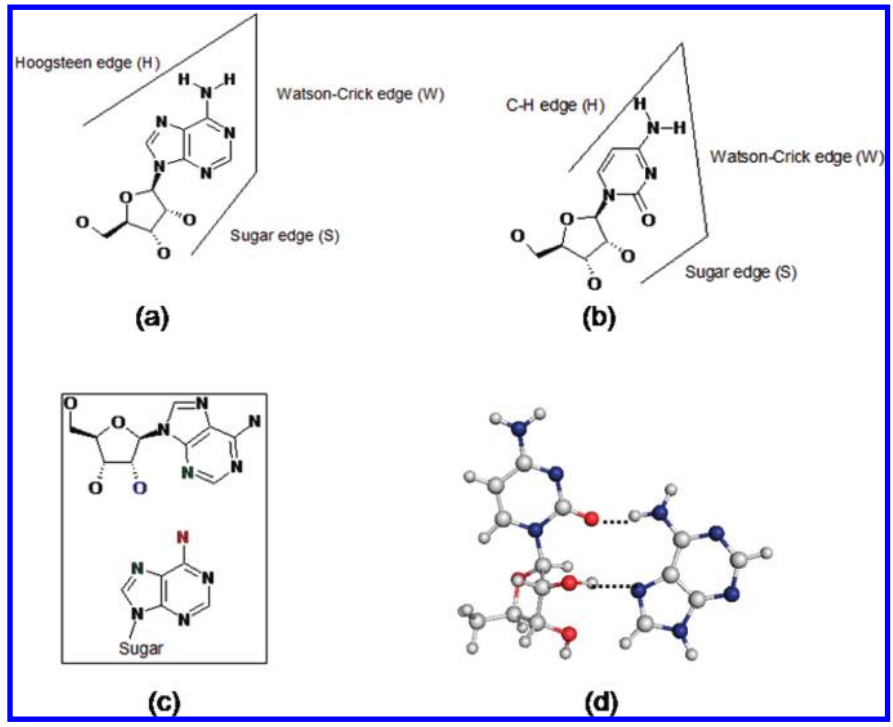


Figure 1. (a and b) Edge interaction in purine and pyrimidine bases taking (a) adenine and (b) cytosine as examples. (c) Illustration of H:S cis base pairing taking the A:rA base pair as an example. The donor atoms are colored as red, and the acceptor atoms are colored as green. The 2′-OH can act both as donor as well as acceptor, and is colored in blue. (d) Atomistic view of a variant of an A:rA H:S cis base pair showing the H-bonding pattern.

particularly to pairs exhibiting multimodality within the same edge interaction, consideration of bifurcated H-bonds and water/metal-ion-mediated H-bonds, etc.[9]

One approach toward addressing the multitude of such issues lies in the characterization of strengths and stabilities of noncanonical base pairs.[10−18] Techniques such as X-ray crystallography are able to determine static averaged RNA structures, while NMR methods can provide data on RNA dynamics. However, both of these methods are limited by their resolution, and by absence of interaction energy data. Experimental determination of the latter is not only difficult; it is not possible to experimentally estimate them at the local level, where individual base pairs exist in geometries which are away from their local minima.[15] Also, the C−H···O and C−H···N bonds require theoretical validation on a case to case basis. The structures and stabilities of noncanonical RNA base pairs can, thus, be accurately analyzed and compared only with the help of nonempirical quantum chemical calculations.[7,8,10,11,13−15] Computational studies have shown that, in several cases, the base pairing geometries, on isolated optimization, deviate greatly from those observed in the experimental structures.[10,11] In fact, there are cases where optimization leads even to changes in the edge interactions. While this variability is reflective of the inherent instability of the base pair involved, it has been shown earlier that it depends not only on the pairing strength but also on several other factors.[10,11] Given the structural and functional significance of base pairs involving their S-edge(s), as is highlighted by their involvement in several biological processes and interactions,[20−22] the factors linked with the involvement of the ribose O2′ are of particular importance.

In the absence of hydrogen atom coordinates in X-ray structures, the possible dual role of the ribose O2′, which can in principle act both as H-bond donor or acceptor, makes understanding of these factors problematic. This is highlighted in our recent report on the possibilities of amino acceptor interactions involving the amino group on the Hoogsteen/C−H edges of adenine and cytosine as acceptors and the O2′ of the corresponding sugar-edge base as donors.[14] Another related complication arises because of the geometries of the D−A network for the sugar edges, which, unlike the other edges, are variable and are highly dependent on δ (sugar pucker) and χ (glycosidic torsion) of the sugar-edge nucleoside. The problem is compounded by the possibility of rotation of the O2′−H bond about the C2′−O2′ axis, when O2′ acts as the donor. The issues involved here are similar to those related to the pyramidalization of amino groups of nucleobases.[19]

Until now, results of systematic quantum chemical studies have been reported for five of the six sugar-edge-containing families, viz., the cis and trans W:S, H:S, and S:S families of RNA base pairs. Here, we extend our investigations to the sixth family, the H:S cis, where the bases interact with each other through their respective H and S edges, with cis orientation of their glycosidic bonds. This family of base pairs is biochemically very different from other families and is poised to play distinct structural and functional roles. For example, this is the only family which shows a natural propensity for base pairing between successive residues ($i - (i + 1)$ type) in which case the backbone torsional constraints are expected to hinder the participation of O2′ in base pairing interactions. Accordingly, compared to base pairs belonging to the other five families, examples of H:S cis base pairs involving O2′-mediated H-bonds are less common.

Leontis and Westhof have classified the base pairs of the H:S cis family into two isosteric subfamilies,[2] $\mathbf{I}_1$ corresponding to the type $i - (i + 1)$ where the paired bases are adjacent and $\mathbf{I}_2$ corresponding to the type $i - (i + 2)$ (the Structural Classification of RNA (SCOR) database[23,24] has apparently extended the definition of the I_2 subfamily by including $i - (i + 3)$ base pairs, with two bases separating the base pairs) where one base separates the paired bases. Of the 16 possible base pair

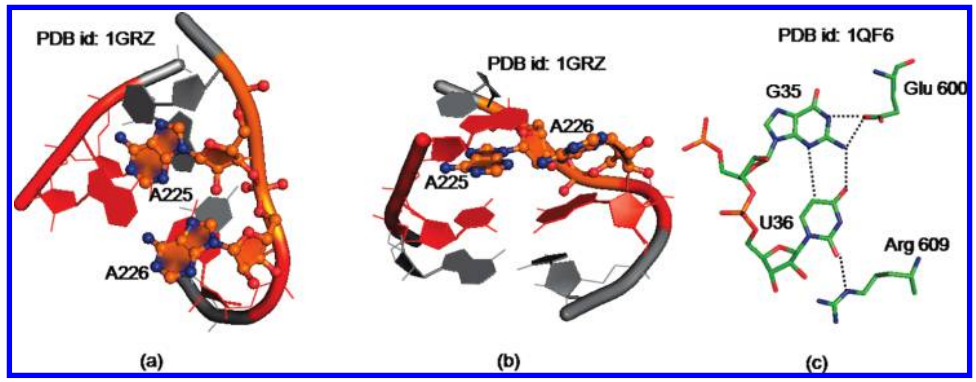


Figure 2. (a) Front view of the A225:A226 platform (adenosine platform) in the J6a/6ab tetraloop acceptor of the group I intron. (b) Side view of the platform to show the stacking bases (red). (c) U:G platform, in threonyl tRNA, observed in the crystal structure of its complex with its corresponding tRNA synthetase. Note that the Watson−Crick edges of the platform form H-bonds respectively with conserved Glu600 and Arg609 of the synthetase.

combinations, examples from the X-ray structure database have been reported for 14 combinations. Of these, nine combinations had examples only from I_1, whereas one had an example only from the I_2 isosteric subfamily. The remaining four combinations had examples from both the I_1 and I_2 isosteric subfamilies.[2,3] Apart from these 18 known H:S cis base pair examples, including isosteric variants of 4 combinations, our database search resulted in the characterization of 5 previously unreported base pair examples which were not of the $i - (i + 1/2/3)$ types and possessed a different H-bonding pattern compared to those observed in the conventional I_1/I_2 subfamilies. In this work, instead of attempting to modify the conventional functional definitions of the isosteric families, we have referred to the I_1 or I_2 isosteric subfamilies, respectively, as T1 and T2 base pair categories. This allowed us to place the five newly found base pair examples in a third category, referred to as T3. In addition to carrying out a systematic study of these $18 + 5 = 23$ observed base pair variants, we have used isostericity principles to predict 7 more base pair examples, spanning 6 base pair combinations. These include models of two base pair combinations (G:rC and G:rU) which do not have any H:S cis representative reported in crystal structures. The others have variant representatives either in T1 or T2 categories.

As expected, the base pairs of the I_1 subfamily constitute dinucleotide platforms (DNPs),[23−28] where coplanar and adjacent bases are involved in noncanonical base pairing. The most common examples of such interaction are the adenosine platform,[25] and the U:rG platform interactions, with the latter far more frequently occurring than the former.[27] A striking feature of these platforms is their remarkable planarity despite the involvement of only one H-bond, which in turn may even be a C−H···O/N type or bifurcated or, for that matter, water-mediated. Equally striking is the planarity of the base pairs of the I_2 subfamily which also implies a flipping out of the intervening base(s).

The contextual significance of some of these features can be understood in terms of the occurrence of these base pairs in internal loops and bulges. They are thus critically placed for mediating important tertiary interactions such as those involving adenosine platforms within tetraloop receptors in group I intron,[25] interaction of U:rG (r refers to the ribose attached to sugar edge base; see Methods section for details) platforms through their W-edges with conserved amino acids in threonyl tRNA synthtase complexed with tRNA[27,29] (Figure 2), and formation of higher order motifs utilizing the W-edge and/or H-edge of the 3′-end base or sugar-edge base of U:G platforms.[27] What is however lacking is a comprehensive physicochemical understanding of these structural features.

On the basis of the analysis of the SCOR database, it has been reported that DNPs frequently participate in base triple formation, while simple stand alone DNPs are quite uncommon.[23,24] With DNPs frequently occurring in internal loops, it is not surprising that the same report describes more than half of all internal loops as "loops with DNP in a triple". It is easy to understand the triple formation propensity of these platforms, where the W-edges of both of the bases and the H-edge of the 3′-end base are available for further pairing. While participation in triples can explain the planarity, and the surprising apparent stability, of these platforms, the question is, "Are these platforms stable enough to constitute the core structural element that drives triple formation?" We have addressed this issue in the current study and, as discussed later, have proposed an elegant alternative description which can facilitate our understanding of the structural and functional role of such triplets.

Since most of these base pairs, as observed in their respective crystal structure contexts characteristically show poor H-bonding, their planar geometry can mainly be ascribed to their respective environmental contexts. We have tried to assess the nature and influence of environmental factors affecting their respective experimental geometries by carrying out gas phase optimization and interaction energy computation on isolated models extracted from their crystal structures. The question asked was, "What if the environmental factors were absent?" Overall, we report the results of our detailed quantum chemical and structural analysis of 30 distinct base pairing patterns, spanning all 16 possible base pair combinations, of the H:S cis family.

2. Methods

(a) Crystal Structure Database Analysis. Given the geometric diversity in the D−A networks of the S-edge of nucleosides resulting in a high propensity for multimodality of their edge interactions, we have attempted to correlate the modalities, of the 23 different observed examples of cis H:S base pairs, with their occurrence contexts and interaction energies. Since different existing base pair detection strategies can identify different, though not necessarily distinct, subsets of the complete spectrum, we analyzed the examples identified in the SCOR[23,24] database and those identified by the tool[28] available at the NDB[31] site and compared them with the examples identifiable using BPFind,[4] including the five new base pair representatives which were detected (Table S1, Supporting Information) by analyzing a selected set of nonredundant RNA structures, solved by X-ray crystallography at 3.5 Å resolution or better (PDB id's listed in the Supporting Information).

TABLE 1: Source Crystal Structures and Corresponding Base Pair id's for H:S cis Base Pairs

base pair	PDB id; interaction	environment[a]	optimization	constrained parameter
		Adjacent Base Pairs (T1)		
A:rA	1GID; A226:A225	DNPT, MG	full	
A:rC	1JJ2; A1106:C1105	DNP	full	
A:rG	1JJ2; A1236:G1235	DNP	constrained	dihedral: N7(A)−C8(A)−N3(rG)−C2(rG) angle: N7(A)−C8(A)−C2(rG)
A:rU	1J5E; A1499:U1498	DNPT, mg	constrained	dihedral: N7(A)−C8(A)−N1(rU)−C2(rU) angle: C8(A)−N1(rU)−C2(rU)
C:rA	1JJ2; C2533:A2532	DNPT, MG	constrained	dihedral: C4(C)−C5(C)−O2′(rA)−C4(rA)
C:rC	1FJG; C748:C749	DNPT, MG	full	
C:rG	1J5E; C596:G595	DNPT, MG	constrained	dihedral: C5(C)−C4(C)−C4(rG)−N3(rG) angle: C5(C)−C4(C)−N3(rG) angle: C4(C)−C6(rG)−C8(rG) angle: C6(C)−C5(C)−N7(rG) angle: C2(C)−N3(C)−N2(rG)
C:rU	1DRZ; C156:U155	DNP	full	
G:rA	2J01; G2052:A2051	DNPT, MG	full	
G:rG	1JJ2; G2093:G2092	DNPT, MG	full	
U:rC	1ET4; U223-*C*222	DNPT, mg	constrained	dihedral: C6(U)−C5(U)−C2(rC)−N1(rC) angle: C5(U)−N1(rU)−C2(rU) angle: O2(U)−C1′(rU)−C5(rU)
U:rG	430D; U11:G10	DNPT	full	
U:rU	1JJ2; U831:U832	DNPT, wMG	constrained	dihedral: C6(U)−C5(U)−C2(rU)−N1(rU) angle: C5(U)−N1(rU)−C2(rU) angle: O2(U)−C1′(rU)−C5(rU)
		Near-Adjacent Base Pairs (T2)		
A:rA	1JJ2; A56:A54	mgT	full	
C:rC	1JJ2; C2787:C2785	B	constrained	dihedral: C5(C)−C4(U)−O2(rC)−N1(rC) angle: C4(C)−C2(rC)−N1(rC)
C:rU	1JJ2; C2475:U2473	mgT	full	
U:rA	1GRZ; U412:A410	MGT	constrained	dihedral: N3(U)−C4(U)−N3(rA)−C2(rA) angle: C4(U)−N3(rA)−C2(rA)
U:rG	1JJ2; U2527:G2525	MGT	full	
		Distant Base Pairs (T3)		
A:rC	1NWX; A1143:C2008	mgT	full	
A:rG	1NWX; A2246:G963	MGT	full	
A:rU	1JJ2; A923:U2107	mgT	full	
C:rC	1J5E; C899:C810	Q, mg	full	
U:rG	1FJG; U173:G198	mgT	full	

[a] DNP, simple dinucleotide platform lacking a base triplet; DNPT, dinucleotide platform in the triplet; B, base pair; T, base triplet; Q, quartet involving kissing interaction between hairpin loop and stem region; MG, major groove interaction with the W:W cis pair; mg, minor groove interaction with the W:W cis; wMG, water-mediated major groove interaction with both nucleotides of the W:W cis base pair; MGT, triplet involving major groove interaction with the W:W cis pair; mgT, triplet involving minor groove interaction with the W:W cis pair.

Occurrence contexts of these base pairs were analyzed, both in terms of their mutual separation in the sequence space as well as in terms of their participation in the formation of triplets and other higher order structures (Table 1).

(b) Choice of Base Pairs and Model Building. Base pair models were built using crystal structure coordinates of one example of each of these 23 types of base pairs. Examples of 16 types were picked up directly from the classic classification paper by Leontis et al.,[2] whereas, for the 2 other types, examples were taken from one of their more recent publications.[3] The remaining five base pairs, belonging to the T3 class, which could be detected using BPFind,[4] were chosen from the occurrences identified from the PDB data set. While building models for optimization, we replaced the ribose sugar of the H-edge base with hydrogen, while retaining it for the S-edge base. In the latter cases, the 5′-OH group of the ribose sugar was replaced by a hydrogen atom, in order to avoid formation of unrealistic H-bonds between the nucleobase atoms and the 5′-OH group of the ribose sugar. A trial calculation with one of the smallest systems (C:rU(T1)) indicated that the inclusion of 5′-OH in the model did not alter the interbase H-bonding of optimized geometry. The models, represented as X:rY, thus consist of one base (X) and one nucleoside (rY). Such models have been used earlier for studying RNA base pairing interactions involving ribose sugar, and are justified in the present context.[7,8,13]

In addition, the geometries of predicted structures were modeled, either on the basis of H-bonding potentials of the individual bases constituting the pairs or by using the geometries of known isosteric members. Apart from providing some clues toward understanding the nature of H-bonding in this family, these results help in examining the viabilities of the predicted examples.

(c) Computational Details. ***Optimization.*** Base pair geometries were optimized employing density functional theory (DFT) using Gaussian 03.[32] The hybrid B3LYP functional[33,34] was used with the 6-31G(d,p) basis set. The geometries optimized at this level are shown to correlate well with the reference geometries obtained at the RIMP2/cc-pVTZ level, with a slight overestimation of H-bond distances on average by 0.01−0.05 Å.[35] Ten base pairs underwent substantial rearrangement of the interbase orientations on full optimization (Table S2, Supporting Information). To retain proper crystal-like H:S cis orientations of these base pairs, some of the angles and dihedrals were frozen during the course of optimization.

Interaction Energy. For each base pair A−B, the interaction energy $\Delta E_{(A-B)}$ can be given as

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

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

In the case of water-mediated base pairs, the energy of interaction between the three entities B(base), rB(nucleoside), and W(water) was calculated as

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

Interaction energies were calculated at the RIMP2/aug-cc-pVDZ level[36] using Turbomole.[37] The final interaction energy values were corrected for basis set superposition error (BSSE) using the Counterpoise method.[38]

Energy Decomposition. The BSSE corrected interaction energy obtained using the RIMP2 method (E_{RIMP2}) has two components: the Hartree−Fock term (E^{HF}) and correlation term (E^{corr})

$$E_{RIMP2} = E_{HF} + E_{corr}$$

E^{HF} mainly consists of electrostatic (E^{elec}), exchange repulsion (E^{EX}), polarization (E^{pol}), and charge transfer components (E^{CT}). E^{corr} includes the attractive dispersion component and a (usually repulsive) electron correlation correction to the electrostatic interaction energy. We carried out Kitaura−Morokuma (KM) decomposition[39] of interaction energy into E^{elec}, E^{EX}, E^{CT}, and higher order coupling (E^{HOC}) terms.

KM decomposition was carried out at the HF/6-31G(d,p) level, using Gamess,[40] separately for base−base and base−sugar interactions. For each base pair, the energy decomposition was carried out for the following two complexes: (i) **BB′**, where the nucleobase B is interacting via the H-edge with the second base B′ after sugar deletion, and (ii) **BS′**, where the nucleobase B is interacting via the H-edge with the sugar of the second base (S′) after base deletion. The structures of these two complexes were obtained as follows: The nucleoside interacting via the sugar edge was split up into sugar and base portions along the glycosidic bond. The dangling bonds were saturated with hydrogen atoms keeping N1/N9−H bond lengths as 1.0 Å and the C1′−H distance as 1.09 Å.

Weak H-Bonding Interactions. The influence of the C−H···O and C−H···N type of H-bonds on the overall stabilization energies of the base pairs was studied, following an earlier reported protocol,[11,14] by carrying out natural bond orbital[41] (NBO) analysis and atoms in molecules[42] (AIM) analysis of these weak bonds using Gaussian 03.[32]

(d) Deviation in Base Pair Geometry on Optimization. Quantitative estimation of variation in the geometries, observed on full optimization of the crystal geometries, was carried out using three different methods, viz., rmsd evaluation, sugar pucker analysis, and base pair parameter analysis. These are briefly described below:

rmsd. The rmsd's of the optimized geometries of base pairs with respect to the X-ray geometries were calculated by superimposing X-ray and computed positions of one of the nucleobases in the pair. rmsd was then derived for the rest of the structure or its parts (Figure S1, Supporting Information).

Base Pair Parameters. Conformational specificities of the base pairs were characterized in terms of intrabase parameters, viz., buckle, open angle, propeller, stagger, shear, and stretch, using the recently upgraded version of NUPARM[5] software, which uses a new edge specific axis system.

Sugar Puckering Parameters. The endocyclic torsion angles, puckering amplitude, and pseudorotation phase angles of the ribose sugar were calculated using standard definitions.[43]

TABLE 2: RIMP2/aug-cc-pVDZ Interaction Energies (kcal/mol) for the H:S cis Base Pairs

base pair	E_{HF}	E_{Corr}	E_{RIMP2}
Adjacent Base Pairs (T1)			
A:rA	−8.6	−8.3	−16.9
A:rC	−8.7	−6.9	−15.6
A:rG	−7.7	−4.4	−12.1
A:rU	−3.1	−5.3	−8.4
C:rA	−0.2	−5.0	−5.2
C:rC	−6.8	−4.9	−11.7
C:rG	−0.5	−5.8	−6.3
C:rU	−3.2	−4.7	−7.9
G:rA	−2.8	−3.6	−6.4
G:rG	−8.5	−6.1	−14.6
U:rC	−0.8	−4.3	−5.1
U:rG	−7.2	−4.6	−11.8
U:rU	−5.5	−6.0	−11.5
Near-Adjacent Base Pairs (T2)			
A:rA	−6.0	−7.6	−13.6
C:rC	−3.1	−4.6	−7.7
C:rU	−5.4	−2.5	−7.9
U:rA	−2.9	−3.4	−6.3
U:rG	−7.6	−2.4	−10.0
Distant Base Pairs (T3)			
A:rC	−8.1	−7.5	−15.6
A:rG	−5.6	−10.2	−15.8
A:rU	−9.3	−5.9	−15.2
C:rC	−7.2	−2.3	−9.5
U:rG	−8.4	−4.0	−12.4
Canonical Base Pairs			
A:U W:Wcis	−10.4	−4.9	−15.3
G:C W:Wcis	−25.5	−3.9	−29.4
G:U W:Wcis	−13.5	−5.3	−16.8

3. Results

Table 1 describes the 23 distinct examples of the H:S cis family, of which 13 are adjacent or DNPs (T1), 5 are near-adjacent (T2), and 5 are distant (T3) in the sequence space. The entries include details of the source, description of the local environment in the context of the folded RNAs in the crystal structures, and the constrained parameters (if any) used for optimizations. The BSSE-corrected interaction energies for optimized base pairs, along with the corresponding comparable values for the canonical base pairs, are reported in Table 2, while those for the predicted structures are reported in Table 3. Table S3 in the Supporting Information provides the KM energy decomposition for the studied interactions. Optimized geometries of all of the base pair structures are given in Figures 3 and 4.

Table S4 in the Supporting Information describes the relationships between the crystal and optimized geometries of base pairs, characterized by their RMSDs observed on optimization. A summary of the H-bonding D−A distances and corresponding geometrical parameters of interbase contacts is given in Tables S5 and S6 of the Supporting Information. Table S7 in the Supporting Information summarizes the NBO and AIM analysis of weak H-bonds.

As discussed in the Introduction, one of the important characteristics of the H:S cis base pair family is their propensity to form planar DNPs and their participation in higher order structures. The physicochemical rationalization of this phenom-

TABLE 3: Source Crystal Structures and Corresponding Base Pair id's on Which the Predicted Structures Were Modeled and Interaction Energies in Their Optimized Geometries

predicted structure (P)	modeled on	optimization	constrained parameter	E_{HF}	E_{Corr}	E_{RIMP2}
A:rU (P1)	1JJ2; A923:U2107	full		−8.8	−8.0	−16.8
A:rU (P2)	1JJ2; A923:U2107	full		−7.1	−7.3	−14.4
A:rG	1JJ2; A1236:G1235	full		−5.3	−11.6	−16.9
G:rA	1JJ2; G2093:G2092	constrained	dihedral: C2(rA)−N3(rA)−N7(G)−C5(G) angle: C2(rA)−N3(rA)−N7(G)	−1.7	−5.1	−6.8
G:rC	1JJ2; G2093:G2092	full		−4.3	−9.9	−14.2
G:rU	1JJ2; G2093:G2092	full		−5.4	−8.0	−13.4
U:rC	1JJ2; U831:U832	constrained	dihedral: N1(rC)−C2(rC)−C5(U)−C6(U) angle: N1(rC)−C2(rC)−C5(U) angle: C5(rC)−C1′(rC)−C5(U) angle: C5(rC)−C1′(rC)−C2(U)	−13.8	−6.8	−20.6

enon required an investigation of the flexibility in terms of sugar pucker and the variations of χ, both of which contribute toward a high degree of variability in the geometry of the D−A network of the sugar edge. Table S8 in the Supporting Information summarizes descriptions of this network, and provides interesting patterns and trends for the four different bases. For example, during optimization, there is a marked tendency to form intranucleoside base−sugar H-bonding within the sugar-edge nucleoside (section 4.4). Table S9 in the Supporting Information summarizes a description of the D−A networks of the H-edge.

Investigations into the remarkable planarity of these DNPs also require a relative evaluation of interaction energies of the base pairs in their intrinsically stable geometries, and of the nature and extent of physicochemical forces, potentially operative in their experimental contexts. In order to facilitate this, we have collated information regarding their participation in

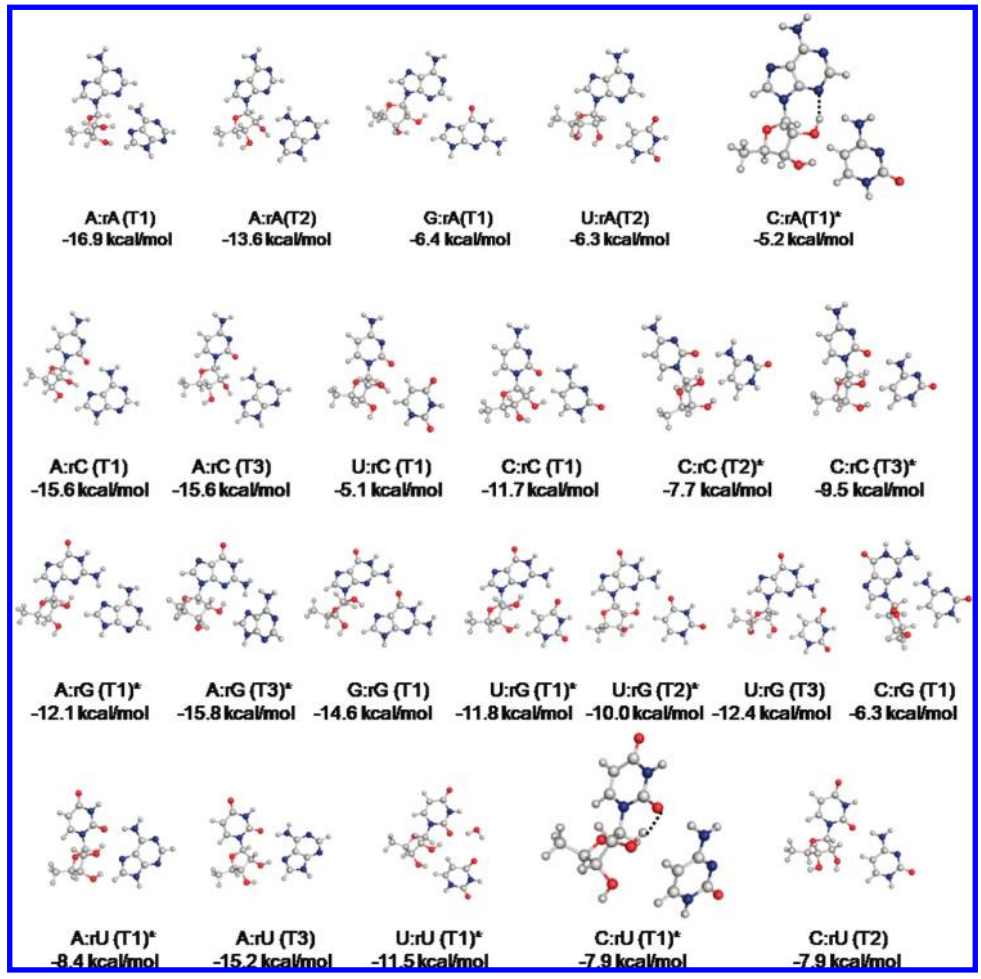


Figure 3. Optimized cis H:S base pair geometries calculated at the B3LYP/6-31G(d,p) level of theory. The layout is organized in four rows, corresponding to the four different bases as the sugar-edge nucleoside of the pair, respectively. Note the variation in the hydrogen bond donor/acceptor geometry, of the sugar edges, within each row. The layout also highlights the rich variety in how the H-edge bases orient themselves with respect to the donor/acceptor networks of the different sugar-edge bases. Each of the pairs marked with "*" possesses an intranucleoside (internal) hydrogen bond, of type O2′−H···N3 in cases where the sugar-edge nucleobase is a purine and O2′−H···O2 in cases where the sugar-edge nucleobase is a pyrimidine. Two pairs, C:rA(T1) and C:rU(T1) (representing purine and pyrimidine nucleoside on the sugar edge, respectively) are enlarged to show these internal hydrogen bonds. Interaction energies calculated at the RIMP2/aug-cc-pVDZ level are also displayed appropriately. Structural details for these geometries are given in Table S8 of the Supporting Information.

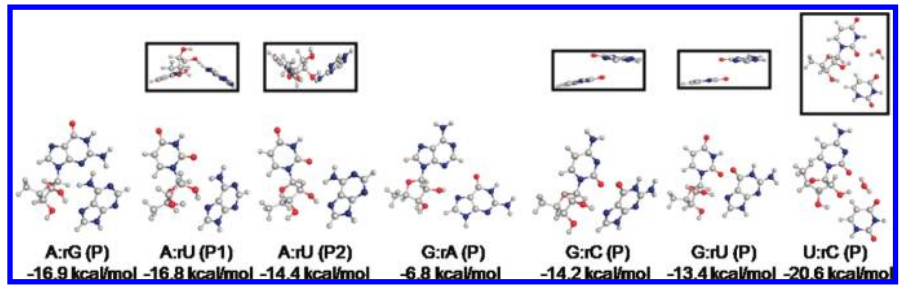


Figure 4. Optimized geometries of predicted base pair structures, at the B3LYP/6-31G(d,p) level of theory. The interaction energies calculated at the RIMP2/aug-cc-pVDZ level are also displayed corresponding to each other. Side views of A:rU(P1) and A:rU(P2) (in boxes) base pairs show the mutual difference in buckle for these structures, despite having a similar H-bonding scheme. Side views of G:rC(P) and G:rU(P) illustrate large stagger values for these base pairs. This is possibly the reason why both of these base pairs have not been found in known RNA structures. The box corresponding to U:rC(P) shows the water-mediated U:rU(T1) pair, on which U:rC(P) is modeled.

triplets, the extent of their planarity within and without the crystal context, and the requirement of constraints during optimization (Table S10, Supporting Information). Further, since H-bonding with neighboring nucleotides constitutes a major component of noncovalent interactions in the crystal, we have also looked into the optimization related changes in characteristics of H-bonds, operative between the paired bases, and also into the possible role of the third base in cases where the crystal context involves triplets. Table S11 in the Supporting Information describes these features under four categories consisting of entries characterizing 11 base pairs which are planar, 5 base pairs which are buckled, 4 which are propelled, and 3 which are both buckled as well as propelled in the crystal context.

4. Discussion

4.1. Crystal Structure Analysis: Issues and Insights. One of the major requirements for analyzing the characteristics of different base pairs in their RNA structure contexts is the availability of automated base pair detection algorithms, which can comprehensively detect all possible examples of each type of base pair with high sensitivity and specificity. Interestingly, out of the five T3 base pairs which were stabilized by at least two detectable interbase H-bonds could be detected by BPFind; only four could be detected using the tool available at NDB and none of them were mentioned in our primary source reference.[2,3] On the other hand, with the surprising exception of G:rA(T1), most of the examples taken from the T1 and T2 categories were found to possess either a single H-bond or only weak H-bonds in their respective crystal geometries, and thus failed to meet the BPFind criteria. BPFind could however detect G:rA(T1) by hypothesizing an extra hydrogen formed involving a potential protonated N3 of guanine. Overall, only 8 of the 10 base pairs from the T2 and T3 category are detected by the base pair detection tool available at NDB. Out of all 23 pairs, only two are detected by 3DNA[44] (Table S1, Supporting Information).

Problems related to the sensitivity of different base pair detection approaches also complicate the prospects of contextual characterization of these base pairs. For example, reports[24] based on the analysis of entries from the SCOR database[23] have suggested that DNPs exist mainly as a part of triplets, and that isolated DNPs are rare with only two examples: those of A:rA and A:rU. Our search however revealed that, apart from these, five more combinations can also exist as isolated DNPs, where three of the examples appear in our primary source reference (Table S12, Supporting Information).[2,3] Thus, it becomes difficult to draw conclusive inferences regarding the importance of "a third base" toward the stabilization and planarity of DNPs. It is particularly so since any given base pair type can, and some of them have been found to, occur as DNPs both in an isolated context as well as in a triplet context.

These observations indicate that there is a need to develop automated base pair finding algorithms, with enhanced sensitivity and specificity, for comprehensive base pair searches. The current study combines the results obtained using multiple (hybrid) approaches as the best alternative available.

4.2. General Observations from Quantum Chemical Analysis of Base Pair Structures and Interaction Energies. ***4.2.1. Interaction Energies Correlate with the Number and Nature of H-Bonds Involved.*** The total interaction energies for the studied base pairs range between −5.2 and −20.6 kcal/mol (Tables 2 and 3) and are comparable to the corresponding values for H:S trans base pairs recently analyzed by us (−1 to −17 kcal/mol). On the basis of their interaction energies, the cis H:S base pairs fall in three ranges which also correlate well with the number and nature of H-bonds involved (Table S13, Supporting Information):

(1) Seven base pairs possess sufficiently high interaction energies (in the range of −13 to −17 kcal/mol), comparable to those of the A:U and G:U base pairs (−15.3 and −16.8 kcal/mol) when calculated at the same level of theory. These include three T1, one T2, and three T3 base pairs. All of these base pairs have at least two strong (N−H···O/N type) H-bonds, which account for their high interaction energies.

(2) Seven base pairs possess interaction energies in the range of −9 to −13 kcal/mol. These include four T1, one T2, and two T3 base pairs. Except for two, all other base pairs possess at least one weak (C−H···O/N type) and one strong (N−H···O/N type) H-bond. The U:rU(T1) possess water-mediated H-bonds, whereas C:rC(T3), with a single strong H-bond, is the weakest pair in this category.

(3) The remaining nine base pairs possess interaction energies below −9 kcal/mol. These include six T1 and three T2 base pairs. Five of them possess a single strong H-bond, three possess one strong and one weak H-bond, and one (U:rA(T2)) possesses two weak H-bonds.

Of the predicted base pairs, water-mediated U:rC is the strongest with an interaction energy of −20.6 kcal/mol. Five structures belong to the category of strong base pairs, whereas G:rA is the weakest base pair. Note that, since the water-mediated U:rC base pair is a three component system, its interaction energy is not directly comparable with the other base pairs.

4.2.2. Correlation Component of Interaction Energy Varies with Type of H-Bonds. The difference between the RIMP2 interaction energies (E_{RIMP2}) and their respective HF component

is mainly due to the attractive dispersion energy, which is partly reduced by repulsive electron correlation correction to the electrostatic energy. This difference (correlation component; E_{corr}) is in the range −2.4 to −11.6 kcal/mol for the H:S cis base pairs, and constitutes anywhere between 24 and 96% of the E_{RIMP2}. Thus, the E_{corr} component for H:S cis base pairs is enhanced compared to that for the three standard W:W cis base pairs (G:C, A:U, and G:U; E_{corr} = 13−30% of the total interaction energy, Table 2). This observed increase in dispersion contribution may represent a considerable advantage for zipping the folded RNA molecules in polar environments. We have shown recently that E_{corr} in RNA base pairs derived by MP2/CBS extrapolation is by about 2−3 kcal/mol lower than the dispersion term directly derived by DFT-SAPT, with a fairly good correlation.[15]

Three base pairs, viz., C:rA(T1), C:rG(T1), and U:rC(T1), where the correlation component is particularly high (E_{corr} > 84% of E_{RIMP2}) (Table 2), were found to possess single base−base H-bonds with slightly loose D−A contacts (distance > 3.2 Å). As a result, the relative magnitude of the HF component decreases and the long-range dispersion forces become dominant. The other set of base pairs showing markedly higher dispersion components (E_{corr} > 60%), in keeping with our earlier finding regarding enhanced E_{corr} components of base pairs involving O2′ interactions,[8,13,14] were those with stronger B−S interactions compared to B−B interactions (Table 2 and Tables S5 and S6 of the Supporting Information). This enhanced role of the dispersion component in B−S contacts also explains the dissimilarity in the contribution of HF and the correlation component to the overall interaction in C:rC(T2) and C:rC(T3) base pairs, despite the presence of similar single base−base H-bonding interaction in both complexes. C:rC(T2) possesses a conformation where the cytosine, interacting through its H-edge, is near the ribose sugar of rC, whereas, in C:rC(T3), the geometry of the base pair does not allow such a close proximity (cf. Figure 3). It may thus be inferred that the E_{corr} contribution is generally enhanced for base pairs with weak H-bonding or where B−S interaction is dominant.

4.2.3. Characterization of Different Types of H-Bonds and Their Stabilization Potentials Using Energy Decomposition Analysis. In order to understand the relative balance of different interaction energy components toward the stabilization of H-bonded RNA systems, we carried out KM energy decomposition analysis of the HF term of the interaction energy (Table S3, Supporting Information). In the case of noncovalent interactions, the HF component of the total interaction energy has been shown to be largely determined by the attractive E_{elec} term and the repulsive E_{EX} term, both of which generally have larger values compared to other components. In the case of H-bonding, the E_{elec} as well as E_{EX} are generally found to increase with decreasing D−A distance. Thus, the optimum geometry is defined by the net balance of the two energy terms: one attractive and the other repulsive. It may be mentioned that, since most of the energy decomposition terms exhibit exponential behavior in the short range, they are sharply affected by even small variations in interatomic contacts. Given that H-bonds can become compressed or extended on optimization of the overall structure, interpretation of energy decomposition results, for base pairs with multiple H-bonds, must be done with caution. In order to avoid this complication, and to understand the energy balance of B−S and B−B interactions in the H:S cis base pairs, energy decomposition was carried out separately for B−S and B−B complexes, modeled as described in the Methods section. The sum of B−S and B−B components of interaction energy from the KM decomposition roughly corresponds to the HF component of interaction energy evaluated from MP2/aug-cc-pVDZ calculations.

We find that, in 15 base pairs, B−B interaction is greater than the B−S interaction (Table S5, Supporting Information). In four of these cases, B−S interaction is repulsive, and the overall HF interaction is only due to the B−B component. This indirectly indicates that B−S interactions mainly benefit from the E_{corr} component of interaction energy. On the other hand, in 13 base pairs, the B−S contribution exceeds the B−B contribution, mainly due to poor B−B H-bonding. In both of the water-mediated base pairs, the B−W−B interaction exceeds the direct B−S interaction.

For the B−B interactions, in most cases, E_{elec} is greater than E_{EX}. In three of these, E_{elec} and E_{EX} are almost equal in magnitude. In two cases, B−B interaction is repulsive due to large E_{EX}. For the B−S interactions, E_{elec} and E_{EX} are almost equal in magnitude in four cases, whereas E_{EX} exceeds the E_{elec} component in eight cases. The contribution of E_{pol} and E_{CT} leads to overall attraction in these complexes. In general, E_{pol} is always larger than E_{CT} in all of these complexes.

4.2.4. Very Few C−H···X Interactions Play an Active Role in Base Pair Stabilization. Out of the total 30 optimized geometries of H:S cis base pairs, 13 geometries indicated the existence of weak C−H···X interactions. We investigated the significance of these interactions using NBO and AIM analysis. The electronic density, ρ, and the Laplacian of the electronic density, $\nabla^2\rho$, at the bond critical points, are comparable with the literature data for A:U W:W cis, U:U W:H trans (commonly known as UU "Calcutta" pair), and UU7 (involving N1 as H-bond donor) base pairs analyzed in ref 45. In that work, Hobza et al. have used ρ and $\nabla^2\rho$ values to show that the contact between C2(A) and O2(U) in the A:U W:W cis base pair is too weak to be considered as a viable C−H-mediated H-bond and to distinguish it from the C5−H···O4 and C5−H5···O2 H-bonds in the U:U W:H trans and UU7 pairs, respectively, which are relatively stronger and which participate more actively in the stabilization of the base pairs. Our studies, on the basis of comparison of ρ and $\nabla^2\rho$ values (Table S7, Supporting Information), reveal that, apart from in the A:rG(T1) and U:rG(T2) base pairs, C−H-mediated interactions seem unlikely to have an active stabilizing role in the rest of the H:S cis base pairs.

4.2.5. Isostericity Principle and Geometry Optimization Can Be Used to Predict Viable Base Pairs. Wherever possible, apart from optimizations starting from the experimental geometries, we also modeled the structures with alternative H-bonding patterns for the base pairs and tried to assess their viabilities by optimizing the model geometries. We hypothesized seven such structures and modeled the geometries based either on known crystal structures of similar base pair combinations or for the same pair with different H-bonding patterns (Table 3).

Using this approach, we could propose a plausible explanation for why 2 out of the 16 possible combinations, the G:rC and G:rU combinations, have not been detected experimentally. We found that their initial geometries, with a single H-bond between O2′(rC or rU) and N7(G), as modeled based on the known G:rG geometry, had high repulsion between O6 of G and O2 of pyrimidine and yielded structures with very high stagger on optimization (Table S14 in the Supporting Information, Figure 4).

Our studies also suggest that the remaining five hypothetical pairs are viable predictions, though they have not yet been detected in RNA X-ray structures. In the case of A:rU, the two

observed crystal geometries, one each corresponding to T1 and T3, respectively, gave rise to different H-bonding patterns on optimization (Figure 3). In addition, based on our analysis of ArU edge interaction, we could detect two other potential geometries which involved alternative O2′-mediated bifurcated H-bonding patterns. The gas phase optimization of these geometries enabled us to predict two new mutually different structures which are buckled in opposite orientations (Figure 4). A similar bifurcated H-bonding pattern is also predicted for the A:rG base pair. Similarly, geometries of two more base pairs, GrA and UrC, were predicted from constrained optimizations.

4.2.6. Diversity in Base Pairing Geometries and H-Bonding Patterns. There is great diversity in the number and nature of H-bonds in the H:S cis base pair family. Of the 30 optimized geometries, one possesses two B−B H-bonds, 15 possess at least one B−B and one B−S H-bond, 6 possess only one B−B H-bond, 4 possess bifurcated (three-center) B−S H-bonds, and 2 possess only one B−S H-bond. (Tables S5 and S6, Supporting Information).

The crystal structure of U:rU(T1) shows mediation of a water molecule, bridging the O2 atom of one base with the O4 atom of the interacting base, which is retained in the optimized geometry. On optimization of the similar predicted structure of the U:rC(P) base pair, the water molecule gets involved in an additional H-bond with the sugar O2′. The cooperative network of H-bonds, thus formed, appears to lend overall stabilization of base pairs involving water-mediated H-bonds (Figure 4).

Optimization of A:rG(T3) gives rise to an additional amino donor−amino acceptor type N2−H(rG)···N6(A) H-bond and involves extensive pyramidalization of the amino group. A similar H-bond was reported by Šponer et al. in the case of W:W trans geometries of G:C base pairs occurring in parallel stranded DNA structures.[46] Note that such interactions cannot be analyzed using empirical force field based calculations where amino group pyramidalization is neglected. This also emphasizes the importance of quantum chemical methods for the study of such base pair complexes.

In general, most of the H:S cis base pairs show a high rmsd, because of geometrical changes resulting from changes in H-bonding patterns on optimization. The B−B and S−S RMSDs contribute almost equally to the overall rmsd values (Table S4, Supporting Information). The H-bonding pattern, in some ways, determines the pucker of the sugar ring, and thus the relative positioning of the D/A atoms present in the S-edges of the nucleosides. The ribose in A-form RNA molecules, in keeping with the attendant steric constraints, generally possesses a C3′-endo pucker. While the T3 base pairs possess C3′-endo pucker in their crystal geometries, except for C:rC (T1) and A:rA, C:rU, and U:rA (all T2), all other T1 and T2 structures possess the less common C2′-endo pucker. This is possibly related to the backbone related constraints arising out of the adjacent or near adjacent positioning of bases belonging to the T1 and T2 classes. Interestingly, while the C3′-endo pucker in three of the T3 cases is retained even after optimization, optimization of ArG (T3) and CrC (T3) leads to a change of pucker to C2′-endo. As explained later (section 4.4), this is accompanied by the formation of internal H-bonds.

Some base pair combinations possess examples belonging to more than one class. The geometries of the different classes for any particular combination can be characterized both in terms of differences in H-bonding patterns and in terms of the base pairing parameters (Table S14, Supporting Information). For example, C:rC(T1) has −2.31 Å shear, whereas C:rC(T2) and C:rC(T3) have large positive shear, indicating that C:rC(T2) and C:rC(T3) have a similar H-bonding scheme. The stretch values are higher for base pairs possessing either a single H-bond or having one C−H-mediated H-bond in addition to one N−H···O/N H-bond. In general, most of the base pairs show good hydrogen bonding with small open angles.

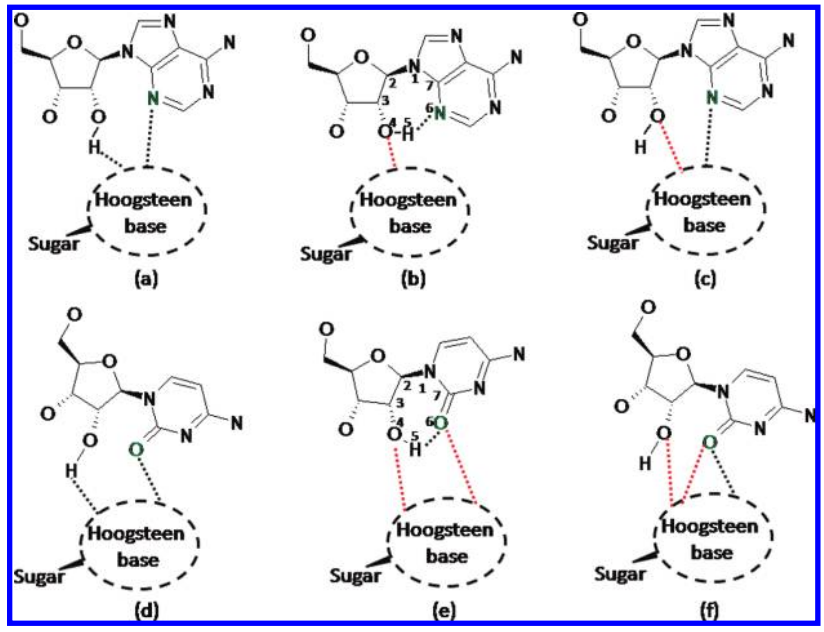
4.3. Planarity and Triplet Formation Propensities of DNPs. ***4.3.1. Planarity Is Not an Intrinsic Property for Most DNPs.*** Of the 13 T1 cases (DNPs), 2 base pairs have been found to occur only as isolated platforms in crystal structures, for 6 base pairs, all examples are found to occur as a platform in triplets, and the remaining 5 base pairs are found to occur both as isolated platforms as well as in triplets (Table S12, Supporting Information). Of these, three examples are nonplanar in their respective crystal geometries (Table S10, Supporting Information). Though the remaining 10 DNPs are planar in the crystal context, only three of these retain their planarity on optimization. In fact, for two of these three pairs, the retention of planarity may possibly be due to constraints applied during optimization (Table S10, Supporting Information). Special mention needs to be made regarding the most frequently occurring platforms, A:rA and U:rG. Both of these are planar in the crystal occurrences, and are historically connected with the nomenclature "platform family" for H:S cis base pairs.[28] That both of these examples end up losing their planarity on optimization strengthens the hypothesis that physicochemical forces operative locally in their respective crystal context are instrumental to their planarity in the crystal.

4.3.2. Participation in Triplet Formation Is Neither Necessary nor Sufficient for Ensuring Planarity of DNPs. As listed above, at least two examples of DNPs, U:rC and A:rC, participate in triplets and yet are nonplanar in the crystal. Conversely, there is only one instance, C:rU, from among 10 DNPs which are planar in the crystal context, which does not participate in triplet formation. For eight of the remaining nine base pairs, the planarity in crystal geometries appears to be due to the presence of a third base, particularly since on optimization in the absence of third base, these base pairs become nonplanar (Table S10, Supporting Information). The case of G:rA, which is planar in the crystal, participates in a triplet, and also retains the planarity on optimization, is intriguing, and has been discussed later in section 4.4.2. Thus, though in general the third bases in respective dinucleotide platform triplets appear to be responsible for planarity in a majority of cases, our studies suggest a major role of "other" environmental factors.

It is interesting to note that, in both A:rA and U:rG, the most frequent DNPs, there exists only one strong H-bond in between the planar paired bases in the crystal, and both of them lose their planarity while concomitantly acquiring a second stabilizing H-bond on optimization of the isolated pairs. Also noteworthy is the fact that both of these base pairs show high stability involving other geometric modes as derived by optimization of their T2 and T3 crystal occurrences. This may hold the key to the high frequency of occurrence, and their high triplet formation potentials.

4.4. Geometric Characteristics of Sugar Edge and Multimodality in H:S cis Family. ***4.4.1. O2′ of Sugar Has a Tendency to Form an Intranucleoside H-Bond with the S-Edge of Its Attached Base.*** Observations as reported in Table S8 and S15 of the Supporting Information indicate that, during the process of optimization, the S-edge nucleoside has a propensity to form seven-membered intranucleoside H-bonding (referred to as internal H-bonding), of O2′−H···O2 type in pyrimidines and O2′−H···N3 in purines (Scheme 1). Such H-bonds have been observed in 11 optimized geometries (7 T1, 2 T2, and 2

SCHEME 1: Observed Scenarios of Hydrogen Bonding in Optimized Geometries of H:S cis Base Pairs[a]



[a] Parts a−c involve purine as a sugar-edge base, whereas parts d−f involve pyrimidine as the sugar-edge base. The 2′-OH acts as a donor for the acceptor present on the Hoogsteen edge of the interacting base in parts a and d, as a donor for the intramolecular hydrogen bond, with N3 acting as an acceptor in part b and O2 acting as an acceptor in part e, and only as an acceptor in parts c and f. Bonds indicated by red lines are not necessarily present in all cases.

T3 geometries; Table S15, Supporting Information), where the corresponding donor−acceptor distances are found to be less than 2.9 Å. All of them also have C2′-endo pucker and have χ values of around 180 ± 10°. In all cases other than that of G:rA (T1), the internal H-bonding is clearly absent in the crystal geometry, and the seven-membered ring formation during optimization is accompanied by a change either in the χ value or of the sugar pucker of the sugar-edge nucleoside. In all of the T1 and T2 cases, where the sugar pucker is C2′-endo in both crystal as well as optimized geometries, the χ values need to undergo a large deviation, from the crystal geometry range of −115 ± 15°, in order to form the internal H-bonds. On the other hand, the crystal geometries for the two T3 cases follow the general trend of having a C3′-endo sugar pucker and a χ value of around 180°. The optimization induced formation of an internal H-bond, in these cases, is accompanied mainly by a marked change in sugar pucker from C3′-endo to C2′-endo. The only exception, G:rA (T1), is also the only case observed, where the internal H-bond seems to be present even in the crystal structure, and is retained on optimization. Note that, in order to form the internal H-bonds, both the sugar pucker, as well as the χ value, needs to come within the prescribed range. Such a combination is rarely observed in RNA crystal structures.

Of the 12 cases where purine acts as an S-edge base, O2′−H···N3 internal H-bonds are observed in the six cases (Table S15, Supporting Information). Out of these, in five cases, neither N3 as an acceptor nor O2′ as a donor are found to participate in interbase H-bonding. They are thus free for O2′−H···N3 H-bond formation. In the case of C:rA(T1), N4−H(C)···N3 interbase interaction is very weak (~3.5 Å), and consequently, an O2′−H···N3 internal H-bond is formed. In five of the six cases where an internal H-bond is not formed, either N3 or O2′ or both are blocked through interbase interaction. An internal H-bond is also not formed in the sixth case, i.e., C:rG(T3), where the C5−H···O4′ interaction restricts the χ value and, consequently, the geometry of the ribose sugar.

Of the 11 cases where pyrimidine acts as an S-edge base, 5 cases involve the formation of an O2′−H···O2 internal H-bond (Table S15, Supporting Information). Unlike the base pairs where purine acts as an S-edge base, in these cases, due to the presence of two lone pairs, the exocyclic O2 has the tendency to form bifurcated H-bonds with the two donor atoms on either side of the C2−O2 axis. The involvement of two different lone pairs of O2 in bifurcated bonding implies that it can participate in internal H-bonding even in the presence of interbase H-bonds.

Since internal H-bond formation within the S-edge base may potentially constrain its interactions with the H-edge base, we have investigated the possible correlation between the formation of such internal H-bonds and the geometric changes observed on optimization of base pairs (Table S16, Supporting Information).

4.4.2. Formation of Internal H-Bond Correlates Inversely with Planarity of Optimized Base Pairs. Of the 11 base pair geometries where an internal H-bond is observed, 5 T1 pairs, 1 T2 pair, and 2 T3 pairs are nonplanar in their optimized geometries. This nonplanarity is possibly related to the involvement of internal H-bonding. Only three base pairs appear to be able to attain planarity in the optimized geometry, despite internal H-bond formation. Out of these, in the case of C:rA(T1), the planarity may be due to constraints applied for optimization. It is interesting to note that U:rG(T2) is the only non-T1 base pair, which acquires an internal H-bond while retaining its planarity on optimization (Table S16, Supporting Information).

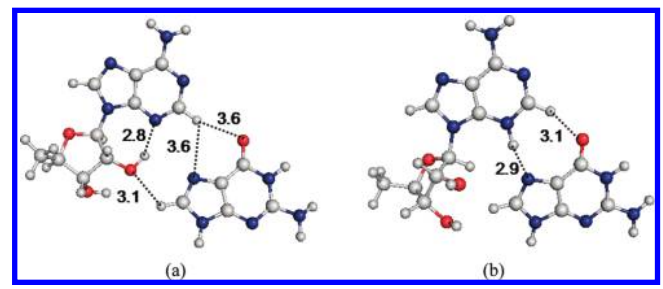


Figure 5. Two possible binding patterns of the G:rA base pair: (a) nonprotonated structure involving a network of bifurcated hydrogen bonds; (b) N3-protonated geometry.

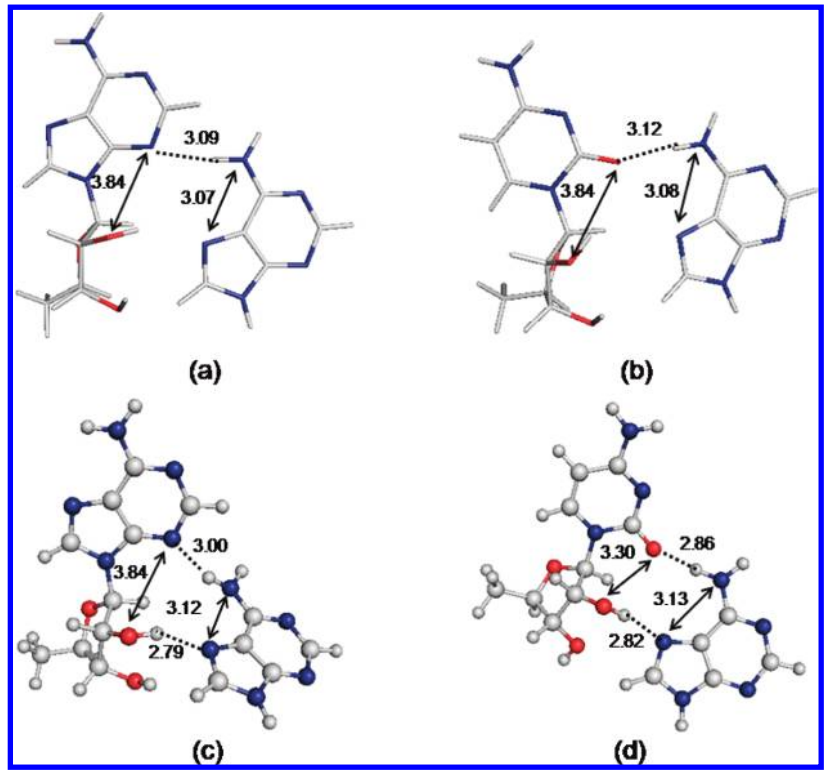


Figure 6. (a and b) Crystal geometries of A:rA(T1) and A:rC(T1). (c and d) Respective optimized geometries. Both of the base pairs are planar in their crystal geometries and acquire an additional O2′−H···N7 interaction on optimization. However, A:rA(T1) becomes nonplanar on optimization, whereas A:rC(T1) retains its planarity. All distances are in angstroms.

It is true that one cannot rule out the "absence of better hydrogen bonding potential for alternate nonplanar geometries" as a possible reason behind the occurrence of weakly stabilized planar geometries. However, the planarity of G:rA both in the crystal (i.e., in the triplet context) as well as in the full optimized geometry is surprising and it is difficult to explain this merely in terms of a single C2−H(A)···O6(rG) H-bond. It may be noted here that CH···O/N H-bonds are generally longer than usual polar H-bonds, and it is possible that a network of several such H-bonds with a length of about 3 Å may give significant stability. Thus, one way to understand this planarity is to consider a network of weak bifurcated H-bonds, as shown in Figure 5. The results of NBO and AIM analysis also support the existence of these H-bonding interactions (Table S17, Supporting Information).

An alternative interesting explanation arises from the fact that BPFind identified the occurrences of G:rA by considering protonation of N3 of A, and considered the existence of a N3(+)−H···N7 H-bond between rA and G. The inclusion of this extra bond, in addition to the C2−H···O6 H-bond, can explain the planarity of the G:rA base pair in its crystal as well as in optimized geometries (Figure 5).

The behaviors of A:rC(T1) and A:rA(T1) are interesting. In their respective crystal contexts, both of these pairs are planar, and are held with only one H-bond involving N6 of the H-edge of A as a donor and O2 and N3 as acceptors. On optimization, both of them retain these H-bonds and acquire a second hydrogen bond involving O2′−H of the S-edge base and the N7 of A. However, in the case of ArC, the optimized geometry remains planar, whereas ArA undergoes a severe change in buckle and propeller (Figure 6). This difference in behavior seems to be related to the variation in the O2′−O2 or O2′−N3 distances, for cytosine and adenine, respectively, as a function of χ. For optimized ArC to be planar, the N6−N7 distance of A needs to match the O2′−O2 distance of rC. For this, the χ for rC changes substantially and causes a reduction of the O2′−O2 distance from 3.84 to 3.31 Å on optimization. However, to ensure such a match with rA geometry, χ has to attain a very small value, and the planar positioning of the A becomes sterically hindered, resulting in large buckle and propeller.

4.5. Analysis of Triplets Containing H:S cis Base Pairs: Looking beyond DNPs. As mentioned in Table 1, 18 of the H:S cis base pairs analyzed involved triplet formation in their respective crystal geometries. Out of these, 10 base pairs involved the interaction of the S-edge base with one of the bases of the W:W cis base pair from their major groove side (Figure S2, Supporting Information). Among these, triplets involving three H:S cis pairs possess the interaction of G:U wobble or U:U pairs with the S-edge base, respectively. The rest of the base pairs involve the interaction of the S-edge base with either A:U or G:C canonical base pairs.

Seven triplets and one quartet involve the interaction of the H-edge base with the minor groove of the W:W cis pair (Figure S3, Supporting Information). Only one triplet (containing the U:rG(T1) base pair) involved the interaction of the noncanonical A:U H:W trans base pair with the third base (Figure 7d).

4.5.1. N2 of Guanine Enhances Its Potential for Participating as the Sugar-Edge Base in DNP Triplets. The major groove edge of the U:U and G:U wobble base pairs has been found to have an ideal geometry of D−A network for molecular recognition by basic amino acids such as lysine.[27] In the same manner, we find that the sugar edge of guanine, because of its N2-amino group, can easily replace such amino acids to dock with these base pairs to give rise to different types, including T1 type, of G:rG and U:rG instances, with support from the third base involving the N2-amino group (Figure 7).

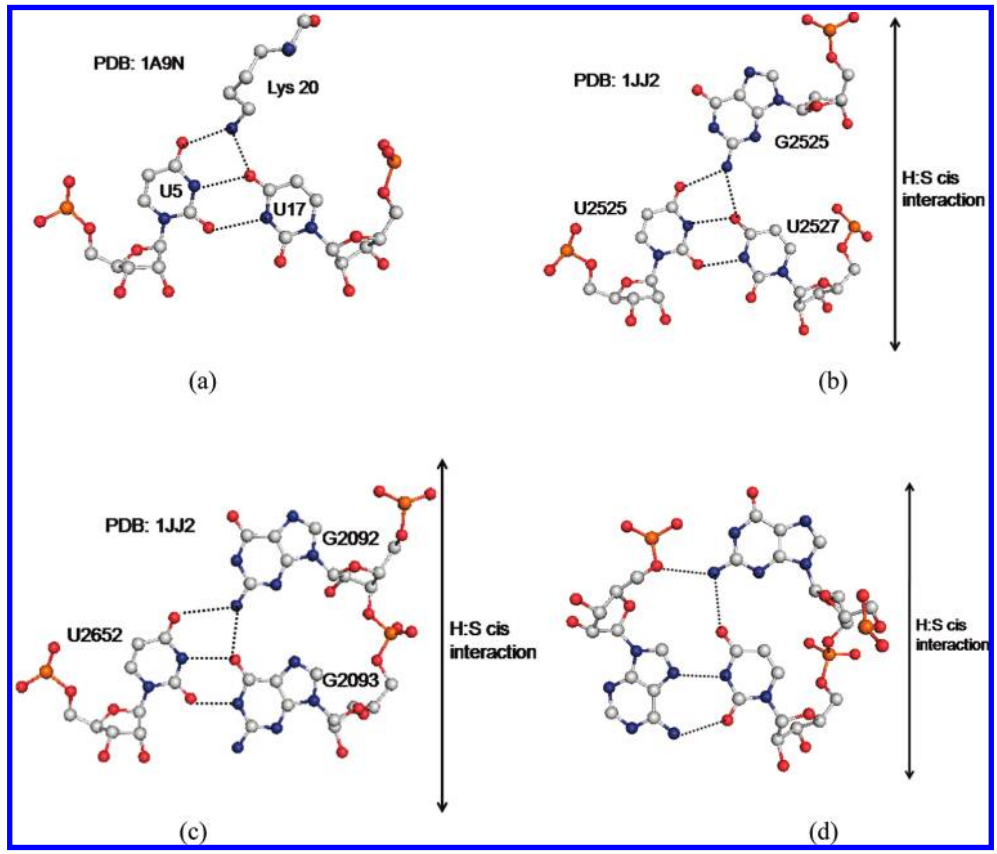


Figure 7. The N2-amino group of guanine can mimic side chain amino groups involved in protein−nucleic acid interactions: (a) the amino group of lysine interacts with the U:U W:W cis base pair; (b) the N2-amino of guanine interacts with the U:U W:C cis base pair in a similar fashion; (c) N2 of guanine also interacts with the G:U wobble base pair in a similar fashion due to similarity in geometries of the U:U and G:U base pairs; (d) a common motif involving the U:G H:S cis base pair. Note that the N2-amino group of guanine forms bifurcated hydrogen bonds with one of the phosphate oxygens and the O4 group of guanine.

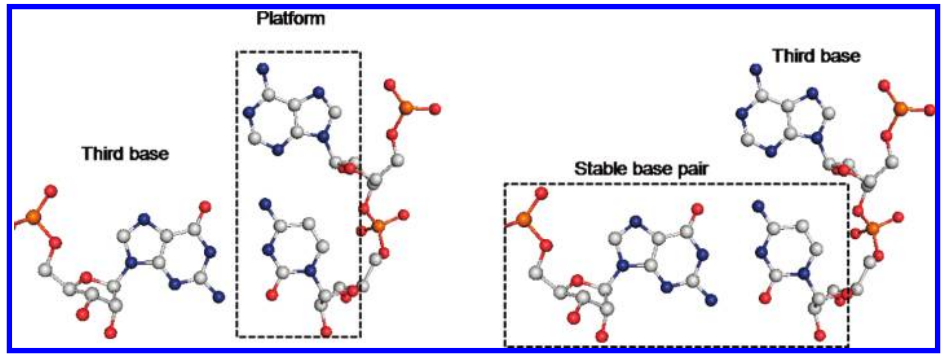


Figure 8. (a) Traditional view of triplet containing platforms considering the platform to be a stable entity with which the third base interacts. (b) Alternative view involving interaction of the stable base pair with the third base to form platform interaction.

Another interesting highly recurrent motif has the N2 of guanine engaged in triplet formation with the A:U H:W trans base pair (Figure 7d), involving a unique donor−acceptor network, where rG, apart from its dinucleotide interaction with O4 of U, also interacts with one of the oxygens of the 3′-phosphate of adenine. This donor−acceptor network pattern is itself also very highly occurring and contributes toward the high occurrence frequency of U:rG platforms in triplets. Some examples of contextual occurrences of such a motif are described in Table S17 of the Supporting Information.

4.5.2. H:S cis Containing Triplets Can Be Alternatively Described in Terms of a Stable Core Base Pair, with a Third Base Docking from Major/Minor Groove. As has been described in the case of the A:rA and U:rG platform,[25,27] the traditional way of looking at triplets involving DNPs implies the platform to be a stable entity, where the third base comes and interacts to form a triplet. Such platforms are shown to be positioned in such a way that the H-edge base of the H:S cis pair is poised to interact through its W-side with the third base (Figure 8).[27] However, our analysis suggests that, in several cases, the DNP "*needs*" a third base for its stability. Therefore, considering the DNP as the "*stable*" core, to which the third base docks, may not be conducive for understanding the mechanism of such triplet formation.

An alternative way of describing these triplets is in terms of a stable base pair, canonical or otherwise, with an appropriately positioned third base docking through H-bonding interactions, involving its S-edge or its H-edge, from the major or minor groove side, respectively (Figure 8). Searching our database, we found seven examples of T1, two of T2, and one of T3 types of base pairs which clearly involve the interaction of the S-edge of the third base with the major groove of stable base

pairs. Similarly, our search for triplet motifs, with the H-edge of the third base interacting with the minor groove of a stable base pair, yielded two examples each of T1 and T2 types and three of T3 type. Apart from suggesting a proactive role of T1 type DNPs toward stabilization of bulges, the elegance of such an alternative characterization lies in its striking similarity with the description of A-minor motifs.[30] In operative terms, such a description opens the way for evaluating the potential of such triplets to participate in switching mechanisms, as has been widely implicated for A-minor motifs.[30] Though such an analogy with A-minor motifs is particularly apparent for the type T3, it may also be significant for type T2 base pairs. It is possible that, depending on the environment of the intervening unpaired base, T2 base pairs may be involved in switching between alternative conformations of bulge loops. The importance of interactions available to the intervening base can be understood from our observations related to the stabilization of their flipped out conformations in their crystal contexts. (Figure S4, Supporting Information).

5. Conclusions

We report the first detailed theoretical investigation of H:S cis base pairs, several of which play fundamental roles in DNPs present in internal loops in RNA structures, with or without the involvement of sugar-base H-bonding through the 2′-OH group of ribose. In order to explore the structure and stability of these base pairs, we have integrated database analysis with advanced quantum chemical analysis of base pairing geometries and their molecular level interactions.

As a result of variability in the geometry of the H-bonding D/A network of the S-edge, the H:S cis family displays great diversity in terms of distinct base pairing patterns observed in crystal structures. Our studies include 18 examples which have been reported earlier, 5 additional examples which were newly observed during our analysis using BPFind,[4] 5 predicted examples which have not yet been observed, and 2 which do not lead to viable geometries. Database analysis involving different base pair finding algorithms highlight the need to enhance the sensitivity and specificity of currently available options. In order to accommodate the five new base pairs, all of which are strongly paired and involve two good hydrogen bonds, we have supplemented the isostericity criterion for classifying the base pairs of this family.[2] We propose a new classification scheme involving three classes of base pairs: T1 for $i - (i + 1)$, T2 for $i - (i + 2/3)$, and T3 for $i - (i + n, n \gg 3)$. While retaining the I_1 and I_2 subfamilies as distinct groups T1 and T2, respectively, this supplementary classification scheme includes all five new instances of base pairs within the T3 class and thus justifies its appropriateness for classifying members of the cis H:S family.

Interaction energies were calculated at the RIMP2/aug-cc-pVDZ level, for the optimized geometries, and were further analyzed using KM decomposition. The interaction strengths correlate well with the number and nature of the hydrogen bonds involved and span a considerably wide range (−5.2 to −20.6 kcal/mol). While all of the H:S cis base pairs are less stable than the G:C canonical base pair (−29.4 kcal/mol), as many as 7 out of 23 H:S cis base pairs observed are equivalent to or are stronger than the A:U canonical base pair (−15.3 kcal/mol). The interaction energies of base pairs in this family are characterized by large correlation energy contributions. In many cases, the correlation component exceeds even the HF term, making the dispersion energy the leading stabilization force. By implication, they might be more hydrophobic compared to standard base pairs as well as to their H:S trans counterparts.

In all 30 cases, geometry optimization resulted in variations, which were quantified using rmsd and base pair parameter calculations and were explained in terms of their respective environmental contexts and attendant intermolecular as well as intramolecular (intranucleoside or internal) H-bonding patterns. Thus, backbone conformational restrictions present in T1 type H:S cis base pairs provide the rationale for the propensity of their S-edge bases to attain a C2′-endo pucker, as opposed to the default C3′-endo sugar pucker. In addition, in nearly all cases, in the optimized geometries, the C2′-endo pucker is stabilized by internal H-bonds, the formation of which correlates well with the pattern of H-bonding interaction between the two bases, and with the planarity of the platform. The importance of the C2′-endo pucker with respect to such internal H-bond formation is highlighted by the two cases where the sugar pucker changes from C3′-endo to C2′-endo, and by the fact that, in all such cases, the O2′−O2 distances for pyrimidines and O2′−N3 distances of purines have values within a very narrow range.

Barring notable exceptions characterized by the involvement of H-bonding networks, most of the cis H:S base pairs are weakly H-bonded and are intrinsically nonplanar. The rationale for the remarkable planarity of DNPs, as observed in experimental structures, thus lies in their crystal contexts. Given their high propensity to participate in base triples, the interactions provided by the third base could be the most important structural context which ensures the planarity. However, identification of several isolated DNPs, which are nevertheless planar in their experimental contexts, strongly suggests possible roles for "other" environmental factors.

As the planarity of most DNPs, including the most highly occurring U:rG and A:rA platforms, strongly correlated with their participation in triplets, the rationale behind their propensity to such participation needs an explanation. We have addressed this issue by providing an alternative description of the triplets in terms of a third base docking to a base pair, from either the major or minor groove. Thus, for example, our analysis provides a possible explanation for the high occurrence of U:rG platform, and their participation in highly occurring triplet motifs, in conjunction with A:U H:W trans base pairs. Such a description not only provides an understanding of triplets involving DNPs belonging to the T1 class, it also helps us understand the triplet participation propensity and concomitant planarity for T2 and T3 classes of H:S cis base pairs. In addition, this description allows us to propose possible functional significance of these triplets. Thus, base pairs of T1 type may be involved in stabilizing bulge loops, T2 type in switching options for bulge conformations and T3 type in larger conformational switches. Further investigations are needed to evaluate these hypotheses.

Acknowledgment. P.S. and S.S. thank CSIR, New Delhi, India, for SRFs. A.M. and D.B. thank DBT, India, for research grants. A.M. thanks CDAC, Bangalore, India, for computational support. J.S. and J.E.S. were supported by the Ministry of Education of the Czech Republic (Grant Nos. AVOZ50040507, AVOZ50040702, LC06030, and MSM0021622413), by the Grant Agency of the Academy of Sciences of the Czech Republic (Grant No. IAA400040802), and the Grant Agency of the Czech Republic (Grant No. 203/09/1476).

Supporting Information Available: Additional information, tables, figures, and Cartesian coordinates of optimized geom-

etries of base pairs. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Leontis, N. B.; Westhof, E. *RNA* **2001**, *7*, 499–512.

(2) Leontis, N. B.; Stombaugh, J.; Westhof, E. *Nucleic Acids Res.* **2002**, *30*, 3497–3531.

(3) Stombaugh, J.; Zirbel, C. L.; Westhof, E.; Leontis, N. B. *Nucleic Acids Res.* **2009**, *37*, 2294–2312.

(4) Das, J.; Mukherjee, S.; Mitra, A.; Bhattacharyya, D. *J. Biomol. Struct. Dyn.* **2006**, *24*, 149–162.

(5) Mukherjee, S.; Bansal, M.; Bhattacharyya, D. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 629–646.

(6) Nagaswamy, U.; Voss, N.; Zhang, Z.; Fox, G. E. *Nucleic Acids Res.* **2000**, *28*, 375–376.

(7) Roy, A.; Panigrahi, S.; Bhattachharyya, D.; Bhattacharyya, M. *J. Phys. Chem. B* **2008**, *112*, 3786–3796.

(8) Šponer, J. E.; Leszczynski, J.; Sychrovsky, V.; Šponer, J. *J. Phys. Chem. B* **2005**, *109*, 18680–18689.

(9) Lemieux, S.; Major, F. *Nucleic Acids Res.* **2002**, *30*, 4250–4263.

(10) (a) Bhattacharyya, D.; Koripella, S. C.; Mitra, A.; Rajendran, V. B.; Sinha, B. *J. Biosci.* **2007**, *32*, 809–825. (b) Sharma, P.; Mitra, A.; Sharma, S.; Singh, H. *J. Chem. Sci.* **2007**, *119*, 525–531.

(11) Sharma, P.; Mitra, A.; Sharma, S.; Singh, H.; Bhattachharyya, D. *J. Biomol. Struct. Dyn.* **2008**, *25*, 709–732.

(12) Gesteland, R. F.; Cech, T. R.; Atkins, J. F. *The RNA World*, 3rd ed.; Cold Spring Harbor Laboratory Press: New York, 2006.

(13) (a) Šponer, J.; Hobza, P. *Collect. Czech. Chem. Commun.* **2003**, *68*, 2231–2282. (b) Šponer, J. E.; Špačkova, N.; Kulhánek, P.; Leszczynski, J.; Šponer, J. *J. Phys. Chem. A* **2005**, *109*, 2292–2301. (c) Šponer, J. E.; Špačkova, N.; Leszczynski, J.; Šponer, J. *J. Phys. Chem. B* **2005**, *109*, 11399–11410.

(14) Mladek, A.; Sharma, P.; Mitra, A.; Bhattachharyya, D.; Šponer, J.; Šponer, J. E. *J. Phys. Chem. B* **2009**, *113*, 1743–1755.

(15) Šponer, J.; Zgarbová, M.; Jurečka, P.; Riley, K. E.; Šponer, J. E.; Hobza, P. *J. Chem. Theory Comput.* **2009**, *5*, 1166–1179.

(16) Sharma, P.; Singh, H.; Mitra, A. *Lect. Notes Comput. Sci.* **2008**, *5102*, 379–386.

(17) Ramakrishnan, V.; Moore, P. B. *Curr. Opin. Struct. Biol.* **2001**, *11*, 144–154.

(18) (a) Oliva, R.; Tramontano, A.; Cavallo, L. *RNA* **2007**, *13*, 1427–1436. (b) Oliva, R.; Cavallo, L.; Tramontano, A. *Nucleic Acids Res.* **2006**, *34*, 865–879.

(19) Šponer, J.; Mokdad, A.; Šponer, J. E.; Špačková, N.; Leszczynski, J.; Leontis, N. B. *J. Mol. Bol.* **2003**, *330*, 967–978.

(20) Pyle, A. M.; Cech, T. R. *Nature* **1991**, *350*, 628–631.

(21) (a) Das, G. K.; Bhattacharyya, D.; Burma, D. P. *J. Theor. Biol.* **1999**, *200*, 193–205. (b) Rodnina, M. V.; Beringer, M.; Wintermeyer, W. *Q. Rev. Biophys.* **2006**, *39*, 203–225.

(22) Hou, Y.; Zhang, X.; Holland, J. A.; Davis, D. R. *Nucleic Acids Res.* **2001**, *29*, 976–985.

(23) Tamura, M.; Hendrix, D. K.; Klosterman, P. S.; Schimmelman, Brenner, S. E.; Holbrook, S. R. *Nucleic Acids Res.* **2004**, *32*, D182–D184.

(24) Klosterman, P. S.; Hendrix, D. K.; Tamura, M.; Holbrook, S. R.; Brenner, S. E. *Nucleic Acids Res.* **2004**, *32*, 2342–2352.

(25) (a) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; A. A.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1696–1699. (b) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Szewczak, A. A.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1678–1685.

(26) (a) Burkhardt, C.; Zacharias, M. *Nucleic Acids Res.* **2001**, *29*, 3910–3918. (b) Davis, J. H.; Foster, T. R.; Tonelli, M.; Butcher, S. E. *RNA* **2007**, *13*, 76–86.

(27) Westhof, E.; Fritsch, V. *Structure* **2000**, *8*, R55–R65.

(28) Yang, H.; Jossinet, F.; Leontis, N. B.; Chen, L.; Westbrook, J. V.; Berman, H.; Westhof, E. *Nucleic Acids Res.* **2003**, *31*, 3450–3460.

(29) Sankaranarayanan, R.; Dock-Bregeon, A. C.; Romby, P.; Caillet, J.; Springer, M.; Rees, B.; Ehresmann, C.; Ehresmann, B.; Moras, D. *Cell* **1999**, *97*, 371–381.

(30) (a) Noller, H. F. *Science* **2005**, *309*, 1508–1514. (b) Nissen, P.; Ippolito, J. A.; Ban, N.; Moore, B.; Steitz, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4899–4903. (c) Rázga, F.; Koča, J.; Leontis, N. B. *Biophys. J.* **2005**, *88*, 3466–3485. (d) Šponer, J. E.; Réblová, K.; Mokdad, A.; Sychrovský, V.; Leszczynski, J.; Šponer, J. *J. Phys. Chem. B* **2007**, *111*, 9153–9164.

(31) Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S.-H.; Srinivasan, A. R.; Schneider, B. *Biophys. J.* **1992**, *63*, 751–759.

(32) 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.; Bakken, V.; 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.; Cioslowski, 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.

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

(34) (a) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789. (b) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, *157*, 200–206.

(35) Šponer, J.; Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142–10151.

(36) (a) Weigend, F.; Haser, M. *Theor. Chem. Acc.* **1997**, *97*, 331–340. (b) Weigend, F.; Haser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143–152.

(37) (a) Eichkorn, K.; Treutler, O.; Oehm, H.; Haeser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *242*, 652–660. (b) Weigend, F.; Haeser, M. *Theor. Chem. Acc.* **1997**, *97*, 331–340. (c) Weigend, F.; Haeser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143–152.

(38) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.

(39) Kitaura, K.; Morokuma, K. *Int. J. Qunatum Chem* **1976**, *10*, 325–331.

(40) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordan, 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.

(41) (a) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735–746. (b) Reed, A. E.; Curtsiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899–926. (c) Reed, A. E.; Schleyer, P. v. R. *J. Am. Chem. Soc.* **1987**, *109*, 7362–7373.

(42) Bader, R. F. W. *Atoms in molecules. A Quantum Theory*; Clarendon Press: Oxford, U.K., 1990.

(43) Saenger, W. *Principles of Nucleic Acid Structure*; Springer Verlag: New York, 1984.

(44) (a) Lu, X.-J.; Olson, W. K. *Nucleic Acids. Res* **2003**, *31*, 5108–5121. (b) Lu, X.-J.; Olson, W. K. *Nat. Protoc.* **2008**, *3*, 1213–1227.

(45) Hobza, P.; Šponer, J.; Cubero, E.; Orozco, M.; Luque, F. J. *J. Phys. Chem. B* **2000**, *104*, 6286–6292.

(46) Šponer, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1994**, *12*, 671–680.

JP910226E