Noncanonical Base Pairing in RNA: Topological and NBO Analysis of Hoogsteen Edge - Sugar Edge Interactions

Purshotam Sharma, Harjinder Singh, and Abhijit Mitra

Center for Computational Natural Sciences and Bioinformatics,
International Institute of Information Technology, Hyderabad 500032, India
laltu@iiit.ac.in

Abstract. Hoogsteen sugar pattern of RNA base pairing is studied using density functional theory. Hydrogen bonding patterns of these base pairs are characteristized using NBO analysis and AIM analysis. Correlation between strength of base pairing and the nature of donor-acceptor combinations is also carried out.

1   Introduction

Molecular recognition abilities, the constitution of active sites and the associated conformational dynamics, all of which are guided by noncovalent interactions in their respective three dimensional structures, are important aspects in understanding of the functioning of biological macromolecules [1]. The molecular biology community has greatly benefitted from the concomitant understanding of their sequence-structure-function space in terms of the physicochemical properties of constituent amino acids in protein enzymes. Recent surge in the interest in understanding of the functioning of ribozymes, the functional RNA molecules, has highlighted the need for developing fresh understanding about these aspects. RNA molecules, with a preponderance of negatively charged phosphate groups and a choice of univalent and divalent counter ions, present a different set of physicochemical environments. They need to function with the help of only four types of bases, apart from a few modified bases, instead of the twenty different amino acids. Though there exists abundant literature related to the understanding of molecular recognition and conformational dynamics of DNA molecules [2], the scenario in RNA molecules is far more complex than in DNA.

In DNA molecules we predominantly deal with double stranded helices with complementary Watson-Crick base pairs holding the two strands together. RNA molecules, on the other hand, are single strands folded on to themselves. The relative positioning of different regions in an RNA strand is thus not necessarily mediated by canonical Watson-Crick base pairs. More than 40% of base pairing interactions in RNA molecules are noncanonical [3]. The second important difference in the RNA molecules is the presence of the $2'$ OH group in the ribose moiety. This gives rise to extra possibilities for interstrand interactions in RNA as distinct from that in DNA. Each nucleobase has three distinct edges

the Watson-Crick, Hoogsteen and Sugar (W, H and S) edge. Leontis and West-hof have shown how, on the basis of the six possible edge interactions and with two possible glycosidic bond orientations (cis and trans) for each, pairing be-tween any two interacting bases can be classified into twelve geometric families [3]. With sixteen possible pair combinations arising out of an alphabet of four bases, this leads to a total of 168 distinct base pair types, only four out of which are canonical [4],[5]. This does not include the variety that may arise out of modifications, including protonation and tautomerism, of the interacting bases. Also not included are the possibilities of multimodality within each geometric family. Notably, this count considers for example A:G W:W cis and G:A W:W cis as distinct types. This is not unreasonable since studies within evolutionar-ily related sequences have, in many cases, shown a lack of covariation and has highlighted the distinct nature of such combinations. Several studies exist in-volving compilation and characterization of noncanonical base pairs in available RNA structures [4],[6]. These not only highlight the variety in number and na-ture of hydrogen bonds involved in base - base interactions, but also underline the need for evaluating associated interaction energies. Several groups are cur-rently involved in *ab-initio* quantum chemical studies of interaction energies of RNA molecules [5],[7,8,9],[12]. In our studies involving more than 45 represen-tative base pairs, most of which are mediated by two or more good N-H...N/O hydrogen bonds [7,8,9], we have argued in favor of the possibility that the vari-ety in physicochemical properties of several of these base pairs may define their respective unique roles in the context of structure and dynamics of functional RNA moelcules [7],[9]. While choosing representative examples and while build-ing models for computations, we have so far not considered hydrogen bonding interactions involving sugar O2′-H group. In this paper, we describe our results of *ab-initio* quantum chemical studies on some representative members of two structurally and biochemically important families of RNA base pairs, the cis and trans Hoogsteen:Sugar edge families, where the sugar O2′ interactions have been explicitly considered.

2   Computational Methods

Gaussian [10] suite of quantum chemical programs was used for all electronic and structural calculations carried out in this work. The initial structures of the base pairs were built on the basis of their respective crystal geometries. The ribose sugar of the base interacting through its sugar edge was retained in our computations. The 5′-OH group of the retained ribose was replaced by a hydrogen atom during model building. The ribose sugar of the base interacting through its Hoogsteen edge was not involved in base pairing interaction, and was therefore, replaced by hydrogen atom. Such models have been used recently in literature and are fully justified for studying base pairing interactions in RNA [5],[12].

   We carried out unconstrained geometry optimizations with all parameters relaxed, at the DFT level. The Becke's three parameter exchange [13] and

Lee-Yang-Parr's correlation functional [14] (known as B3LYP) was used with the 6-31G** basis set for geometry optimizations. The B3LYP/6-31G** optimized base pair geometries are shown to agree very well with the geometries obtained from RIMP2/cc-pVTZ (resolution of identity MØller-Plesset second order perturbation method/correlation consistent-polarization valence triple zeta basis set) optimizations [15]. This level of theory is well established in the literature for studying nucleic acid base pairing properties [5],[12]. Natural Bond Orbital [16] (NBO) and Atoms in molecules [17] (AIM) analysis have been performed in order to understand the nature of interbase binding and to characterize the hydrogen bonding interactions in terms of spatial distribution of electron densities.
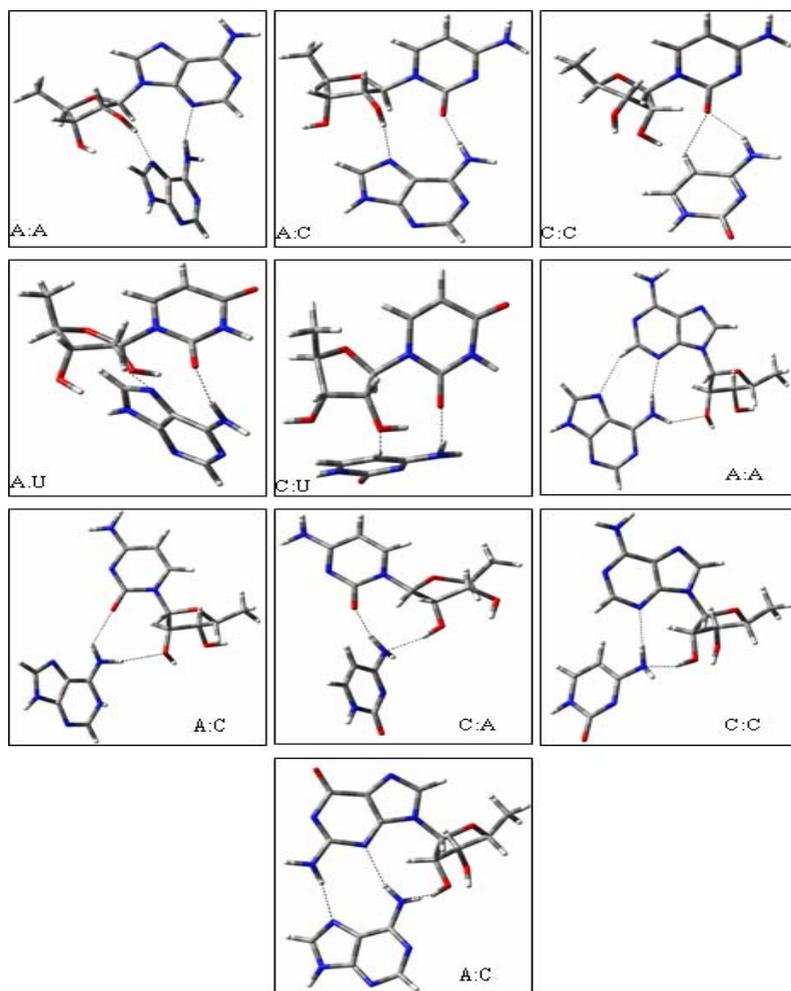

Fig. 1. Optimized geomerties of cis (first five) and trans (last five) H:S base pairs

3   Results and Discussion

We carried out geometry optimizations for ten selected base pair geometries, five of which belong to each of the cis and trans H:S base pair families. We have obtained minimum energy structures, which correspond to at least local minima on the intrinsic potential energy surfaces of these isolated molecular systems. All these studied structures are intrinsically stable, and do not deviate much from their respective crystal geometries. The main geometrical parameters for these structures are summarized in Table 1. All optimized structures possess at least two hydrogen bonds, and therefore, can be considered as stable base pairs. In order to characterize the hydrogen bonding in these base pairs, AIM and NBO analyses have been carried out. An analysis of results of our computations on base pairs of H:S family revealed several features, which could be helpful in getting insights into physicochemical basis of RNA structures, and developing a correlation between structural aspects and functional characteristics.

The 2′-OH group of the ribose sugar seems to have an equal tendency of behaving as a hydrogen bond donor as well as acceptor towards the base of the other nucleotide as revealed by small hydrogen bonding distances and relatively linear bond angles (Table 1). The additional hydrogen bonds formed by 2′-OH group of the sugar of a base with its partner base seem to be strong enough to provide intrinsic stability to these base pairs. Except A:U H:S cis and C:C H:S cis geometries, all other base pairs are stabilized by at least one hydrogen bond involving O2′ interactions. In case of H:S cis base pairs, wherever O2′ interaction is involved, O2′ always act as a donor atom. In case of base pairs of H:S trans family, O2′ acts as an acceptor atom in case of A:A ans A:C geoemtries, and as a donor atom in case of A:G, C:C and C:A base pairs. Thus the dual ability of the 2′-OH group of the base interacting through it sugar edge, provides additional stabilization to these base pairs.

Occasionally, a third hydrogen bond also contributes to the stabilization of studied H:S cis and trans binding patterns. Two of the five studied nucleoside-base systems (called as base pairs for simplicity) of H:S trans family, viz. A:A H:S trans and A:G H:S trans, possess three hydrogen bonds. But the nature of hydrogen bonding is different in all these three cases. In case of A:A H:S cis, a strong hydrogen bond involving N-H...N and a weaker hydrogen bond involving C-H...N interaction stabilizes the interbase contacts. An additional hydrogen bond of N-H...O type, involving O2′ interaction is also present. In case of A:G, both the interbase hydrogen bonds involve N-H...N interactions. The O2′ provides additional stabilization by forming a hydrogen bond with N6-H of adenine.

Due to very weak electron density of hydrogen electrons obtained using X-ray crystallography, there is an ambiguity in assignment of position of hydrogen atoms in crystal structures of macromolecules. As a result, the correct assignment of donor-acceptor tendencies to particular atoms forming hydrogen bonds is an exceedingly difficult task, and is writ with many errors. The optimized geometries of the base pairs can correctly describe the intermolecular interactions in such well defined intrinsically stable building blocks of RNA. For example, in case of

Table 1. Selected geometrical parameters of studied base pairs

Family	Base Pair	Bond	D-A	A-H	D-H	∠D-H-A
H:S cis	A:A	O2′ H(A1)...N7	2.79	1.88	0.98	153.4
		N6-H(A2)...N3	3.00	2.08	1.02	160.0
	A:C	O2′-H(C)...N7	2.82	1.87	0.99	161.5
		N6-H(A)...O2	2.86	1.85	1.02	172.4
	C:C	C5-H(C1)...O2	3.23	2.49	1.08	124.5
		N4-H(C1)...O2	2.88	1.93	1.02	154.5
	A:U	N6-H(A)...O2	3.01	2.01	1.01	170.7
		C1′-H(U)...N7	3.30	2.45	1.09	133.6
	C:U	N4-H(C)...O2	3.10	2.13	1.01	162.8
		C5-H(C)...O2′	3.52	2.46	1.08	167.0
H:S trans	A:A	C2-H(A1)N7	3.48	2.48	1.09	152.5
		N6-H(A2)...O2′	3.01	2.12	1.01	145.6
		N6-H(A2)...N3	3.08	2.09	1.02	165.0
	A:C	N6-H(A)...O2	2.98	2.24	1.01	129.8
		N6-H(A)...O2′	3.17	2.16	1.02	167.5
	C:A	O2′-H(A)...N4	2.89	1.92	0.98	165.3
		N4-H(C)...N3	3.03	2.02	1.03	172.0
	C:C	N4-H(C)...O2	2.86	1.92	1.02	153.1
		O2′-H(C1)...N4	2.89	1.92	0.98	170.4
	A:G	N2-H (G)...N7	3.04	2.03	1.02	170.5
		O2′-H(G)...N6	2.93	1.98	0.98	162.6
		N6-H(A)...N3	3.18	2.20	1.02	160.5

C:C H:S cis, C:A H:S cis, C:C H:S trans and A:G H:S trans geometries, the O2′ is considered as an acceptor atom by Leontis *et al.* [4], on the basis of crystal structure analysis. But the full optimization of the crystal geometry clearly shows the oxygen atom behaving as a donor in these cases.

Some hydrogen bonds are not assigned in the crystal geometry of these base pairs. But the optimized geometries unambiguously show the presence of these hydrogen bonds. For example, O2′ interaction is not considered in crystal geometries of A:C and C:U H:S cis base pairs. Similarly, the hydrogen bonding interaction of C5-H with O2 is not considered in case of C:C H:S cis geometries. Similarly, the weak interaction of C1′-H with N7 is not considered in crystal geometry of A:U H:S cis. In case of A:A H:S cis, the O2′-H...N7 interaction can be considered as a hydrogen bond according to geometric criteria. But the results obtained using NBO and AIM analysis do not support this assumption (discussed later). Thus, the nature of this interaction remains ambiguous. Thus, the accurate strength and interaction in such hydrogen bonded complexes can be assigned only using optimized geometries.

Analysis of charge density $(\rho(r_c))$ and the associated Laplacian of charge density at the hydrogen bond critical points (HBCP) provides a good measure of predicting the strengths of hydrogen bonds [18]. In general, the values of charge density and the Laplacian of charge density lie between 0.002-0.34 and

0.016-0.13 a.u respectively. The values for these two parameters lie well within these proposed limits in these studied complexes. The values suggest that C-H...O and C-H...N bonds are much weaker than other hydrogen bonds involving O and N as donor atoms. However, the values of ρ are very close to each other in case of O-H...N, N-H...N and N-H...O hydrogen bonds. Hence the relative strength of these bonds cannot be predicted with accuracy. The O2′-H...N7 interaction in case of A:A H:S cis geometry, although satisfying the geometrical criterion for hydrogen bonding (Table 1) cannot be regarded as a proper hydrogen bond, due to comparatively much weaker electron density at HBCP as compared to similar hydrogen bonds in other studied complexes. This is further substantiated by the absence of interaction between nonbonding orbital of N7 and antibonding orbital of O2′-H bond in NBO analysis.

Table 2. Results of AIM and NBO analysis for studied base pairs

Family	Base Pair	Bond	ρ	$\nabla^2\rho$	n_a	$\sigma^*_{(D-H)}$	$E(2)$
H:S cis	A:A	O2′H(A1)...N7	0.005	0.040	1.89257	-	-
		N6-H(A2)...N3	0.021	0.091	1.88982	0.04161	13.60
	A:C	O2′-H(C)...N7	0.024	0.136	1.89277	0.05501	19.35
		N6-H(A)O2	0.022	0.124	1.95478	0.03889	10.89
	C:C	C5-H(C1)...O2	0.009	0.034	1.96017	0.01435	0.09
		N4-H(C1)...O2	0.026	0.079	1.96017	0.03512	5.75
	A:U	N6-H(A)...O2	0.022	0.062	1.99974	0.02777	4.52
		C1′-H(U)...N7	0.012	0.035	1.99944	0.03875	2.10
	C:U	N4-H(C)...O2	-	-	1.95618	0.02103	3.86
		C5-H(C)...O2′	-	-	1.97102	0.01934	1.95
HS trans	A:A	C2-H(A1)N7	0.012	0.031	1.92191	0.02952	3.13
		N6-H(A2)...O2′	0.019	0.053	1.70220	0.02465	5.48
		N6-H(A2)...N3	0.023	0.056	1.70220	0.03249	10.80
	A:C	N6-H(A)...O2	0.013	0.048	0.98520	0.00583	1.38
		N6-H(A)...O2′	0.018	0.046	0.98632	0.01758	2.68
	C:A	O2′-H(A)...N4	0.024	0.123	1.97160	0.04718	16.18
		N4-H(C)...N3	0.021	0.091	1.89000	0.05030	15.82
	C:C	N4-H(C)...O2	0.026	0.084	1.89162	0.04798	9.44
		N4-H(C)...O2	0.033	0.073	1.77412	0.03168	17.41
	A:G	N2-H(G)...N7	0.025	0.062	0.95337	0.01929	6.23
		O2′-H(G)...N6	0.029	0.065	0.88855	0.01994	6.89
		N6-H(A)...N3	0.018	0.045	0.95236	0.01466	3.86

We had earlier shown a strong linear relationship between sum of electron densities at HBCP and the binding energy of base pairs. Using the same criterion, the strength of binding between these base pairs can be estimated as: A:G H:S trans > C:C H:S trans >A:A H:S trans > A:C H:S cis > C:A H:S trans > C:C H:S cis > A:U H:S cis > A:C H:S cis > A:A H:S cis. The strength of binding could not be predicted in case of C:U H:S trans geometry due to very high nonplanarity of this systemand presence of two critical points between some atoms.

NBO analysis is carried out in order to understand the physical nature of intermolecular contacts in these base pair complexes. The charge transfer interaction between nonbonding orbital of the acceptor and the antibonding σ^* orbital on thedonor-hydrogen bond (D-H bond) are of crucial importance in hydrogen bonded complexes. The occupancies of nonbonding orbitalsof acceptor atoms (n_a) and antibonding orbital of D-H bond (σ^* D-H) are listed in Table 2. The occupancies of σ^* (C-H), σ^* (N-H) and σ^* (O-H) bonds lie between 0.01435-0.02952, 0.00583-0.05031 and 0.01994-0.05501 e (e stands for one electronic charge) respectively, showing greatest charge transfer in hydrogen bonds where O2$'$ acts as a donor atom. Moreover, the charge transfer stabilization energy E(2) is also greatest in case of O2$'$-H...N bonds. These observations further substantiate the importance of O2$'$ interactions in the stabilization of these base pairing interactions. As discussed earlier, no charge transfer interaction between n_a and σ^* (D-H) orbitals is apparent from NBO analysis in case of O2$'$-H...N7 interaction in A:A H:S cis base pair. Thus, it cannot be regarded as a hydrogen bond.

4 Conclusion

In this paper, we present the first *ab-initio* studies on base pairs belonging to cis and trans H:S families. In addition to characterization of hydrogen bonds, these calculations refine the crystallographic data and evaluate the intrinsic stabilities of these base pairs. All the studied structures do not show any difference in edge interaction and glycosidic bond orientation, as compared to their respective crystal geometries. Hence, they can be considered as well defined intrinsically stable building blocks of RNA. The calculations thus, unambiguously show that hydrogen bonding is the most important stabilizing interaction in these noncanonical base pairs. In addition to assignment of correct donor-acceptor interacions, these calculations also highlight the importance of interactions inviolving sugar O2$'$ in RNA base pairing. Previous studies [4] on detection of base pairs in RNA structures considered the oxygen atom of the sugar hydroxyl group as hydrogen bond acceptor only. We find that in many cases, this hydroxyl group acts as a hydrogen bond donor. Recently it was also indicated from statistical analysis that O-H...N is a very strong hydrogen bond [19]. This indicates the need to revisit the problem of base pair finding by automated tools such as BPFind [11] or X3DNA [20]. Further studies regarding evaluation of interaction energies and complete quantum chemcial characterization and physicochemical impact of change in edge orietnation on optimization in other members of these RNA base pairing families are in progress.

References

1. Černy, J., Hobza, P.: Non-covalent interactions in biomacromolecules. Phys. Chem. Chem. Phys. 9, 5291–5303 (2007)
2. Šponer, J., Hobza, P.: Molecular interactions of nucleic acid bases. A review of quantum-chemical studies. Collect; Czech Chem. Commun. 68, 2231–2282 (2003)
3. Leontis, N.B., Westhof, E.: Geometric nomenclature and classification of RNA base pairs. RNA 7, 499–512 (2001)
4. Leontis, N.B., Stombaugh, J., Westhof, E.: The non-Watson-Crick base pairs and their associated isostericity matrices. Nucleic Acids Res. 30, 3497–3531 (2002)
5. Šponer, J.E., Spackov, N., Pulhanek, P., Lesczynski, J., Šponer, J.: Non-Watson-Crick Base Pairing in RNA. Quantum chemical analysis of the cis Watson-Crick/Sugar Edge base pair family. J. Phys. Chem. A 109, 2292–2301 (2005)
6. Das, R., Baker, D.: Automated de novo prediction of native-like RNA tertiary structures. Proc. Natl. Acad. Sci. 104, 14664–14669
7. Bhattacharyya, D., Koripella, S.C., Mitra, A., Rajendran, V.B., Sinha, B.: Theoretical analysis of noncanonical base pairing interactions in RNA molecules. J. Biosci. 32, 809–825 (2007)
8. Sharma, P., Mitra, A., Sharma, S., Singh, H.: Basepairing in RNA: A computational study of structural aspects and interaction energies. J. Chem. Sci. 119, 525–531 (2007)
9. Sharma, P., Mitra, A., Sharma, S., Singh, H., Bhattacharyya, D.: Quantum chemical studies of noncanonical RNA base pairs: The trans Watson-Crick: Watson-Crick family. J. Biomol. Struct. Dyn. (in press)
10. Frisch, M.J., et al.: Gaussian03, revision B.05; Gaussian, Inc.: Pittsburgh, PA (2003)
11. Das, J., Mukherjee, S., Mitra, A., Bhattacharyya, D.: Noncanonical base pairs and higher order structures in nucleic acids: Crystal structure database analysis. J. Biomol. Struct. Dyn. 24, 149–162 (2006)
12. Šponer, J.E., Spackova, L.J., Šponer, J.: Principles of RNA Base Pairing: Structures and Energies of the Trans Watson-Crick/Sugar Edge Base Pairs. J. Phys. Chem. B 109, 11399–11410 (2005)
13. Becke, A.D.: Density-functional thermochemistry. III. The role of exact exchange. J. Chem. Phys. 98, 5648–5652 (1993)
14. Lee, C., Yang, P.R.G.: Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys. Rev. B 37, 785–789 (1988)
15. Šponer, J., Jurecka, P., Hobza, P.: Accurate Interaction Energies of Hydrogen-Bonded Nucleic Acid Base Pairs. J. Am. Chem. Soc. 126, 10142–10151 (2004)
16. Reed, A.E., Weinstock, R.B., Weinhold, F.J.: Natural population analysis. J. Chem. Phys. 83, 735–746 (1985)
17. Bader, R.F.W.: Atoms in molecules. A Quantum Theory. The Clarendon Press, Oxford (1990)
18. Kolandaivel, P., Nirmala, V.: Study of proper and improper hydrogen bonding using Bader's atoms in molecules (AIM) theory and NBO analysis. J. Mol. Struct. 694, 33–38 (2004)
19. Roy, A., Bhattacharyya, M., Bhattacharyya, D.: Structures, Stability and Dynamics of Canonical and Noncanonical base pairs: Quantum Chemical Studies. J. Phys. Chem. B (accepted)
20. Lu, X.J., Olson, W.K.: 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids. Res. 31, 5108–5121 (2003)