Estimating Strengths of Individual Hydrogen Bonds in RNA Base Pairs: Toward a Consensus between Different Computational Approaches

Antarip Halder,*†‡ Dhruv Data,‡ Preethi P. Seelam,†∗‖ Dhananjay Bhattacharyya,‡* and Abhijit Mitra*†

†Center for Computational Natural Sciences and Bioinformatics (CCNSB), International Institute of Information Technology (IIIT-H), Gachibowli, Hyderabad 500032, India
‡Computational Science Division, Saha Institute of Nuclear Physics (SINP), 1/AF, Bidhannagar, Kolkata 700064, India

Supporting Information

ABSTRACT: Noncoding RNA molecules are composed of a large variety of noncanonical base pairs that shape up their functionally competent folded structures. Each base pair is composed of at least two interbase hydrogen bonds (H-bonds). It is expected that the characteristic geometry and stability of different noncanonical base pairs are determined collectively by the properties of these interbase H-bonds. We have studied the ground-state electronic properties (using density functional theory (DFT) and DFT-D3-based methods) of all the 118 normal base pairs and 36 modified base pairs, belonging to 12 different geometric families (cis and trans of WW, WH, HH, WS, HS, and SS) that occur in a nonredundant set of high-resolution RNA crystal structures. Having addressed some of the limitations of the earlier approaches, we provide here a comprehensive compilation of the average energies of different types of interbase H-bonds ($E_{\text{HB}}$). We have also characterized each interbase H-bond using 13 different parameters to describe its geometry, charge distribution at its bond critical point (BCP), and $n\rightarrow\sigma^*$-type charge transfer from filled $\pi$ orbitals of the H-bond acceptor to the empty antibonding orbital of the H-bond donor. On the basis of the extent of their linear correlation with the H-bonding energy, we have shortlisted five parameters to model linear equations for predicting $E_{\text{HB}}$ values. They are (i) electron density at the BCP: $\rho$, (ii) its Laplacian: $\nabla^2\rho$, (iii) stabilization energy due to $n\rightarrow\sigma^*$-type charge transfer: $E(2)$, (iv) donor–hydrogen distance, and (v) hydrogen–acceptor distance. We have performed single variable and multivariable linear regression analysis over the normal base pairs and have modeled sets of linear relationships between these five parameters and $E_{\text{HB}}$. Performance testing of our model over the set of modified base pairs shows promising results, at least for the moderately strong H-bonds.

1. INTRODUCTION

Until recently, the functional roles of RNA were thought to be limited mainly because of its participation in gene expression in the form of the messenger RNAs, ribosomal RNAs, and transfer RNAs. However, in the past 2 decades, the ever-increasing discoveries of the catalytic and regulatory roles of non (protein)coding RNAs (ncRNA) have trashed the established rules to make way for new ones (see the Supporting Information). Recent discoveries of CRISPRs (clusters of regularly interspersed short palindromic repeats) and scanRNAs (small conjugation-specific RNA) suggest that ncRNAs can even intervene at the genome itself and modify DNA. To participate in these diverse cellular functions, RNA molecules have to fold into functionally competent structures, which are inherently complex. The understanding of how RNA, having essentially only four types of nucleobases [adenine (A), guanine (G), cytosine (C), and uracil (U)], can display such a variety and complexity is still far from complete. A detailed analysis of base pairing interactions, however, may provide significant points toward this understanding.

As shown in Figure 1, each nucleobase can be characterized by three edges—Watson–Crick (W), Hoogsteen (H), and Sugar (S). A base pair is formed when two bases interact noncovalently, each respectively involving one of these three edges, and form a considerably planar geometry with at least two interbase hydrogen bonds (H-bonds). Depending on the mutual orientation of the glycosidic bonds within a base pair, it can be annotated as either cis or trans. The cis-type interactions between the Watson–Crick edges of two complementary bases (G:C and A:U/T) are known as canonical base pairs, as they are predominantly found within the double helical stretches of both DNA and RNA. All other possible base pairs are categorized as noncanonical base pairs.

Received: December 31, 2018
Accepted: April 12, 2019
Published: April 23, 2019

DOI: 10.1021/acsomega.8b03689
ACS Omega 2019, 4, 7354−7368
Unlike the double-stranded DNA, the single-stranded RNA folds onto itself to form a large repertoire of noncanonical base pairs. On the basis of the identities of the interacting edges, Leontis and Westhof have organized all these base pairs into 12 geometric classes, viz., WW, WH, HH, HS, WS, and SS (cis and trans). It is theoretically possible to have a total of 144 different types of base pairing geometries. However, because of lack of complementary H-bonding network and other structural and functional constraints, only 118 types of base pairs are actually observed in available RNA crystal structures. These base pairing interactions lead to the formation of double helical stem regions (dominated by canonical base pairs) and loop regions (characterized by unpaired bases and noncanonical base pairs) in RNA. In a folded RNA, such stem and loop regions further interact with each other via tertiary interactions. Combination of such interconnected structural components is often observed recurrently as a conserved (in terms of both sequence and structure) modular unit in various RNA molecules and is known as an RNA motif. In this work, we have only focused on the interbase H-bonds.

Figure 1. (a) RNA bases have been characterized by three edges—Watson–Crick, Hoogsteen (C–H edge for pyrimidines), and Sugar. (b) In principle, any edge of a base can interact with any other edge of a second base. Depending on the mutual orientation of the two glycosidic bonds, the base pair can be annotated as either cis or trans.

For example, (a) studying the modulations in the IR spectra as a result of H-bond formation such as red shift of D–H stretching vibration frequency, increase in IR intensity, and so forth, (b) using experimental tools such as temperature-dependent field ionization mass spectrometry, calorimetry, or vibrational predissociation spectroscopy in order to calculate the dissociation energy of the hydrogen-bonded complex, and (c) studying the alteration in magnetic properties of the system owing to hydrogen bonding. Also, unlike covalent bonds, the correlation between the geometry of a H-bond (say, H-bond length) and its strength is not straightforward.

To overcome the limitations of the experimental methods and to evaluate the strength of individual H-bonds, numerous theoretical approaches have been implemented till date. Among them, the quantum theory of atoms in molecules (QTAIM) approach is arguably the most widely used theoretical tool to understand the H-bonding interaction. Several efforts have also been made to relate the QTAIM-based parameters with the strength of individual hydrogen bonds. For example, (a) the complex-derived Grabowski parameters relating geometric and topological parameters to the H-bonding strength, (b) the relationship given by Espinosa et al. providing H-bond dissociation energy ($D_h$) using the virial density ($V_r$), and (c) the linear relationship between intermolecular complex stabilization energy and electronic charge density ($\rho$) or it's Laplacian ($\nabla^2 \rho$) at the bond critical point (BCP) ($V_r(\rho = 0)$ have been proposed for different H-bonded molecular systems, including base pairs. Although extensively used, the QTAIM-based approaches have their limitations as discussed in the Supporting Information. However, it is to be noted that the theoretical basis of QTAIM is based on electron density which is an experimentally measurable quantity. Another such approach, based on experimentally measurable quantity, is the use of vibrational spectroscopy. Advanced quantum mechanics-based methods can replicate the experimentally
observed vibrational spectroscopy to a reasonable extent, at least for the nucleobases. The H-bond (D–H···A) formation enthalpy \((E_{\text{BH}})\) can be measured from the red shift in the D–H bond stretching frequency, which takes place because of H-bond formation. For intramolecular hydrogen bonds in a wide variety of biological systems, Nikolaenko and co-workers have established the correlation between these two theoretical approaches (QTAIM and vibrational spectroscopy) for estimating the H-bonding energy of individual H-bonds. Later, Hovorun and co-workers have implemented logansen's relationship to study the H-bonding energy of the intermolecular H-bonds present in unconventional DNA base pairs. Another effective and widely used approach is provided by natural bond orbital (NBO) analysis, which calculates the stabilization energy associated with the charge transfer from the filled \(\pi\) orbital of the H-bond acceptor \((A)\) to the corresponding empty antibonding orbital of the H-bond donor \((D–H)\). In addition to these, estimating H-bonding energies using different properties of the compliance matrix is another less frequently used, but promising, theoretical approach. Interestingly, all these different theoretical techniques have been implemented to extensively study the canonical base pairs in DNA (A:T/G:C) and RNA crystal structures provided by the HD-RNAS database. Gas-phase optimized geometries [B3LYP/6-31+G(d,p) level] of all the normal RNA base pairs are curated in the RNABP COGEST database. The hybrid GGA functional B3LYP is arguably the most widely used DFT functional for studying DNA and RNA base pairs and similar systems. Therefore, for the initial calculations, we have collected the B3LYP/6-31+G(d,p) level optimized geometries of different normal base pairs from the RNABP COGEST database.

2. METHODS

2.1. Geometry Optimization of Selected RNA Base Pairs. We have studied all the 118 different geometries of normal base pairs and 36 modified base pairs that occur in the nonredundant set of high-resolution (resolution cutoff = 3.5 Å) RNA crystal structures provided by the HD-RNAS database. Gas-phase optimized geometries [B3LYP/6-31+G(d,p) level] of all the normal RNA base pairs are curated in the RNABP COGEST database. The hybrid GGA functional B3LYP is arguably the most widely used DFT functional for studying DNA and RNA base pairs and similar systems. Therefore, for the initial calculations, we have collected the B3LYP/6-31+G(d,p) level optimized geometries of different normal base pairs from the RNABP COGEST database. Gas-phase optimized geometry of modified base pairs, at the same level of theory, has been taken from the supporting information of the recent article by Seelam et al. Hessian calculation has been performed over the optimized geometries of all the normal and modified base pairs. For some of the normal base pairs, Hessian calculation results in imaginary frequencies because the corresponding optimized geometries present in RNABP COGEST are obtained from constrained optimization. These systems have been reoptimized. We have found that some of the reoptimized geometries are also associated with imaginary frequencies (as listed in the Supporting Information), and we have removed those cases from our study. The remaining systems are composed of 17 WW, 17 WH, 12 HH, 22 WS, 22 HS, and 13 SS pairs. Out of this final set of 103 base pairs, optimized geometries of 96 base pairs are the same as the one given in the RNABP COGEST database and geometries of rest 7 base pairs are obtained after reoptimization. They are G:C W:W Trans, A:G H:H Cis, A:U(II) H:H Trans, C:G H:H Trans, A:A W:Trans, and G:rA H:S Cis. In order to incorporate the effects of dispersion interactions, they are further optimized using the B3LYP-D3(BJ) functional and the same basis set. In B3LYP-D3(BJ), the dispersion correction has been added explicitly by Grimme’s method (3rd order) with Becke–Johnson damping. All real frequencies were observed for B3LYP-D3(BJ) optimized systems on Hessian calculation. Similarly, for all the monomers, geometry optimization and subsequent Hessian calculation have been performed using both B3LYP and B3LYP-D3(BJ) functionals with the 6-31+G(d,p) basis set, respectively. Note that B3LYP and its dispersion-corrected version B3LYP-D3(BJ) are arguably the most universally used DFT functionals to study the nucleic acid-based systems. All the calculations were performed using Gaussian03 software.

2.2. Interaction Energy of a Base Pair. Interaction energy of a base pair AB has been calculated as the difference between the electronic energy of the base pair AB \((E^\text{E})\) and sum of the electronic energies of the individual bases \((E^\text{A} + E^\text{B})\). All the interaction energies are further corrected for zero-point vibrational energy \((ZPVE)\), basis set superposition error \((\text{BSSE})\) \((E^\text{BSSE})\), and deformation energy \((E^\text{fi})\), that is, \(E^\text{fi} = E^\text{E} - (E^\text{A} + E^\text{B}) + E^\text{BSSE} + E^\text{ZPVE}\). Further details on the method of interaction energy calculation are discussed in the Supporting Information.

2.3. Different Types of Interbase H-Bonds. Potential H-bond donor groups present in the nucleobases are primary amino groups \((N–H)\), secondary amino groups \((NH_2–H)\), hydroxyl group \((O–H)\), and C–H group. On the other hand,
potential H-bond acceptor groups present in the nucleobases are imino nitrogens (N_{III}), carbonyl oxygen (O_{C}), and hydroxyl oxygen (O_{OH}). Table 1 shows the count of different types of H-bonds studied in this work. Note that amino acceptor interactions observed in Hoogsteen bonds studied in this work. No amino acceptor interactions observed in Hoogsteen–Sugar base pairs are not included in this study as they occur only in a specific set of RNA base pairs.

### 2.4. H-Bonding Energy from logansen’s Relationship.

For all the D—H····A-type interbase H-bonds, we have calculated the red shift (Δν_{D—H}) of the D—H stretching vibrational frequency as Δν_{D—H} = ν_{D—H}^{free} − ν_{D—H}^{BP}, where ν_{D—H}^{free} and ν_{D—H}^{BP} are the vibrational frequencies corresponding to the D—H bond stretching in the isolated monomer and in the H-bonded base pair, respectively. As prescribed in the Gaussian03 manual, all the vibrational frequencies are scaled by a factor 0.96 to obtain the values consistent with experimental results. We have calculated the H-bonding energy (E_{HB}) from logansen’s relationship, which is

\[ E_{HB} [\text{kJ mol}^{-1}] = 0.33 \sqrt{\Delta \nu_{D—H} [\text{cm}^{-1}]} - 40 \]  

Note that logansen’s relationship allows us to estimate the H-bonding energy only for the red-shifted H-bonds with Δν_{D—H} > 40 cm⁻¹. In this context, it is also important to mention that earlier works that implement logansen’s relationship to evaluate the strength of interbase H-bonds have taken special care (partial deuteration of the O—H, N—H, and NH₂ groups involved in the interbase H-bonding) to minimize the effect of vibrational resonances. In this work, we have not taken any such measure. However, we confirm that in the numerically calculated IR spectra of all the base pairs, the symmetric stretching of each N—H, O—H, and C—H bond (that form interbase H-bonds) corresponds to a unique peak having high intensity (see Figure 3). Therefore, we can safely assume that the influences of vibrational resonance on our results are negligible.

### 2.5. Characterization of Interbase H-Bonds Using QTAIM Analysis.

The QTAIM analysis has been performed over the wave functions corresponding to the ground-state optimized geometry of the base pairs obtained at both B3LYP/6-31+g(d,p) and B3LYP-D3(BJ)/6-31+g(d,p) levels of theory. We have found that all the interbase H-bonds are associated with a bond path such that at the BCP, the slope of the gradient of the electron density distribution is positive (V_{\rho} > 0). We have calculated the following six topological parameters at the BCPs corresponding to the interbase H-bonds. They are (i) electron density (ρ), (ii) slope of the gradient of the electron density distribution (V_{\rho}), (iii) potential energy density (V), (iv) kinetic energy density (E), (v) total energy density (E_{tot} = V + E), and (vi) ellipticity (ε). These topological parameters are frequently used in the literature to estimate H-bonding strength. QTAIM calculations have been performed using standalone AIMALL package.

### 2.6. Characterization of Interbase H-Bonds Using NBO Analysis.

NBO analysis has been performed on the optimized geometries obtained using both B3LYP/6-31G + (d,p) and B3LYP-D3/6-31G+ (d,p) levels of theory. These calculations have been done using the NBO package as implemented in Gaussian03. The second-order perturbative energy has been estimated between the different Lewis-type “filled” (donor) and non-Lewis-type “empty” (acceptor) orbitals. These interactions are also referred to as “delocalization interactions” as they are a result of the donation of electron density from the localized NBOs of the idealized Lewis structure into the empty non-Lewis orbitals, thereby showing a departure from the idealized Lewis structure description. The second-order perturbative energy (E(2)) resulting due to delocalization can be estimated for a donor NBO (i) and acceptor NBO (j) as

\[ E(2) = \Delta(E_{ij}) = q_{i} \frac{F(i,j)^2}{E_{f} - E_{i}} \]  

where q_{i} is the donor orbital occupancy, E_{f} and E_{i} are diagonal elements (orbital energies), and F_{ij} is the off-diagonal NBO Fock matrix element. A higher value of E(2) represents a stronger H-bonding interaction. In this work, we will refer to the E(2) parameter as “charge-transfer-based parameter”.

### 2.7. Characterization of the Geometry of an Interbase H-Bond.

We have considered the geometry of a hydrogen bond as a triangle between the three atoms—donor (D), hydrogen (H), and acceptor (A) (see the Supporting Information). We have selected three distances (lengths of the three sides of the triangle DA, AH, and DH), the donor–hydrogen–acceptor angle (DHA), and the area (A_{DHA}) of the DHA triangle and its perimeter (S_{DHA}) to explore how the strength of the corresponding H-bonds depends on them.

### 2.8. Linear Regression Analysis.

For different H-bonds, we have calculated the Pearson correlation coefficient (r) between the E_{HB} values (say, dataset \{y_{1}, ..., y_{n}\}) and six topological parameters, six geometric parameters, and the stabilization energy E(2) obtained from NBO analysis (say, dataset \{x_{1}, ..., x_{n}\}). The Pearson correlation coefficient (r) between the two data sets \{x_{1}, ..., x_{n}\} and \{y_{1}, ..., y_{n}\} (having the average values \bar{x} and \bar{y}, respectively) is defined as

\[ r = \frac{\sum_{i=1}^{n} (x_{i} - \bar{x})(y_{i} - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_{i} - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_{i} - \bar{y})^2}} \]  

We have further performed single variable and multivariable linear regression analysis to establish linear relationships between the E_{HB} and other topology-based, geometry-based, or charge-transfer-based parameters, following the standard least-square fit-based approach. All the interbase H-bonds studied here contribute toward the stabilization of the base pairs containing them, and hence, their energies must be represented as a negative value. However, for the ease of discussion, in the following text, all the H-bonding energies are reported in terms of their magnitude and hence have the positive sign.
3. RESULTS AND DISCUSSION

It may be reiterated here at the outset that this approach toward estimating the average energy of a particular type of interbase H-bond has a limitation. It can be applied only to the H-bonds where the red shift in D—H bond stretching is \( > 40 \text{ cm}^{-1} \). As discussed elaborately in the following sections, C—H···O and C—H···N bonds found in RNA base pairs are usually associated with a small red shift or blue shift of the C—H bond stretching. Therefore, though improper C—H···O/N-type H-bonds are important factors that govern the geometries and stabilities of nucleic acid base pairs, we are not able to analyze their properties in detail. It may be noted that out of the remaining eight types of interbase H-bonds listed in Table 1, there are four bonds (serial numbers 5–8) that involve the sugar O2’ (either as donor or as acceptor, or both). Hence, they are found only in Sugar edge-mediated base pairs, that is, in the base pairs belonging to the WS, HS, or SS families. As shown later, \( E_{\text{HB}} \) values of these four types of H-bonds do not show strong correlation with any of the topology-based, charge-transfer-based, or geometry-based parameters. In contrast, the other four types of H-bonds (serial numbers 1–4 in Table 1) are found in all types of base pairs and their \( E_{\text{HB}} \) correlate well with some specific parameters studied in this work. Hence, majority of our discussions are focused on these four interbase H-bonds. They are as follows: N1—H···N\( _{\text{HB}} \), N1—H···O\( _{\text{O}} \), N1—H···O\( _{\text{O}} \), and N1—H···O\( _{\text{O}} \).

3.1. Average H-Bonding Energy of Different Types of Interbase H-Bonds. Calculating the average H-bonding energy (\( E_{\text{HB}} \)) for different types of interbase H-bonds that shape up the geometry and stability of noncanonical base pairs has been attempted earlier. There it has been assumed that the total interaction energy (\( E_{\text{HB}} \)) of a base pair is solely contributed by the interbase H-bonds present in them, that is, \( E_{\text{HB}} = \sum_{i=1}^{n} E_{i\text{HB}} \) where \( n \) is the number of H-bonds present between the two bases and \( E_{i\text{HB}} \) is the H-bonding energy of the interbase H-bond \( i \). Therefore, calculations of \( E_{\text{HB}} \) values of different base pairs give rise to a set of linear equations, which can further be solved to get the energetic contribution of each type of H-bonds (N—H···N, N—H···O, etc.). Another similar approach reported the H-bonding energies of the interbase H-bonds present in DNA base pairs. However, later investigations in this area have highlighted that in addition to interbase H-bonds, other factors, mainly charge delocalization, dipole–dipole interaction, London dispersion forces, and so forth, also contribute to the stability of base pairs. This means that for an RNA base pair, \( E_{\text{HB}} = E_{\text{HB}} - \sum E_{i\text{HB}} \) is a nonzero quantity, which varies with its composition. To illustrate this systematically, we have calculated the \( E_{i\text{HB}} \) values of individual interbase H-bonds using the Iogansen’s relationship as discussed in the Methods section and obtained the \( E_{\text{HB}} \) values for all the base pairs belonging to the WW, WH, and HH families and containing guanine. Note that out of all the four nucleobases, guanine has the highest dipole moment.97 Figure 2 clearly shows that \( E_{\text{HB}} \) values (the red bars) are significantly high, especially for high dipole moment base pairs such as G:C W:W Cis and G:G W:W Trans. For example, the interaction energy of the canonical G:C W:W cis base pairs is \(-22.8\) kcal mol\(^{-1}\) and the sum of the \( E_{\text{HB}} \) values of three interbase H-bonds present in it is \(-16.9\) kcal mol\(^{-1}\) at the B3LYP/6-31+G(d,p) level (where negative signs indicate stabilization because of bond formation). These observations underline the need to revisit the problem with a different approach.

In this work, instead of calculating the interaction energy of a base pair for estimating the sum of all the interbase H-bonds present in it, we have calculated the H-bonding strength of individual H-bonds themselves (\( E_{i\text{HB}} \)) using Iogansen’s relationship. Therefore, the average H-bonding energy of a particular type of H-bond is given by the arithmetic mean of the \( E_{i\text{HB}} \) values of all its instances. Figure 3 illustrates the red shift in the D—H bond stretching frequency (\( \Delta v_{\text{D—H}} \)) because of D—H···A-type interbase H-bond formation in two canonical (A:U W:W Cis and G:C W:W Cis) and two noncanonical (A:U H:H Cis and G:U H:H Trans) base pairs. From these red shifts, the \( E_{i\text{HB}} \) values of individual H-bonds are calculated using Iogansen’s relationship, as discussed in the Methods section. Table 2 shows the average H-bonding energy (\( \bar{E}_{\text{HB}} \)) for different types of interbase H-bonds calculated using this approach. Note that earlier attempts toward calculating \( \bar{E}_{\text{HB}} \) values have not discriminated N—H···N (or N—H···O) type H-bonds on the basis of the chemical identity of the donor nitrogen. Our approach allows us to explore these varieties of interbase H-bonds. Interestingly, we have found that the average energy for N—H···N type H-bonds strongly depends on the hybridization state of the donor nitrogen. For N1—H···N\( _{\text{HB}} \)-type H-bonds (see Table 1 for the annotation of different H-bonds), the average H-bonding energy is 6.42 kcal mol\(^{-1}\), which is 2.11 kcal mol\(^{-1}\) higher than the average energy of the
NI−H···NIII-type H-bonds (4.31 kcal mol$^{-1}$). Similarly, the average H-bonding strength of the N−H···O hydrogen bonds depends on the hybridization states of both the donor nitrogen and acceptor oxygen atoms. The average energies of the NI−H···Oc, NII−H···NIII, and NII−H···Oc H-bonds are 2.85, 5.13, and 2.56 kcal mol$^{-1}$, respectively. The order of the mutual strengths of different interbase H-bonds can be explained on the basis of the mutual strengths of the corresponding H-bond

Figure 3. IR spectra (calculated at the B3LYP-D3(BJ)/6-31+G(d,p) level) of four nucleobases (A) adenine, (B) uracil, (C) guanine, and (D) cytosine; two canonical base pairs (E) A:U W:W Cis and (F) G:C W:W Cis; and two noncanonical base pairs (G) A:U H:H Cis and (H) G:U W:H Trans. For the nucleobases, the orange arrow and the green arrow point to the frequencies corresponding to symmetric stretching of the N−H bonds of the primary amino and secondary amino groups, respectively. Interbase H-bonds of the base pairs are shown in broken line. Frequencies corresponding to symmetric stretching of the N−H (or C−H) bond in interbase H-bonds are pointed using different colored arrows for different types of H-bonds: NI−H···Oc (blue), NII−H···NIII (green), NII−H···Oc (cyan), C−H···N (orange), and C−H···O (black).
Table 2. Average Hydrogen Bonding Energy \( E_{\text{HB}} \) of Different Interbase Hydrogen Bonds Obtained Using B3LYP and B3LYP-D3 Functionals Is Reported in kcal mol\(^{-1}\)

| name | type of base pair | B3LYP | B3LYP-D3(BJ) |
|------|------------------|-------|--------------|
| N_\text{II}−H−N_{\text{III}} | all | 4.31 (1.23) | 4.24 (1.35) |
|  | NS | 4.35 (1.38) | 4.40 (1.52) |
|  | S | 4.26 (1.02) | 4.00 (1.04) |
| N_{\text{II}}−H−N_{\text{III}} | all | 6.42 (0.94) | 6.37 (1.16) |
|  | NS | 6.74 (0.84) | 6.81 (0.92) |
|  | S | 5.66 (0.76) | 5.21 (1.01) |
| N_\text{II}−H−O_\text{c} | all | 2.85 (1.52) | 3.09 (1.46) |
|  | NS | 3.52 (1.51) | 3.75 (1.48) |
|  | S | 2.25 (1.28) | 2.57 (1.25) |
| N_{\text{II}}−H−O_{\text{c}} | all | 5.13 (1.50) | 5.29 (1.39) |
|  | NS | 5.56 (0.81) | 5.80 (0.83) |
|  | S | 4.33 (2.16) | 4.40 (1.74) |
| N_\text{II}−H−O_\text{b} | all | 2.56 (1.62) | 2.65 (1.38) |
|  | NS | 5.24 (1.24) | 5.34 (1.19) |
|  | S | 4.09 (0.75) | 4.79 (1.01) |
| O_{\text{H}}−H−N_{\text{III}} | all | 3.33 (1.34) | 2.76 (1.1) |

“Values reported within parenthesis represent corresponding standard deviation. All base pairs studied, "S": base pairs of the WS, HS, and SS families, "NS": base pairs of the WW, WH, and IH families.

donors and acceptors. The gas-phase acidity of the H-bond donor sites and gas-phase basicity (or proton affinity) of the H-bond acceptor sites of different nucleobases provide an overall estimate for their mutual H-bonding strengths. For nucleobases, a detailed and accurate estimation of the gas-phase acidity and gas-phase basicity of different polynucleotides has been reported by Lee and co-workers.\(^{99,99}\) It suggests that secondary amino groups (N_\text{II}−H) are usually more acidic and hence stronger H-bond donors than primary amino groups (N_\text{II}−H). Similarly, the exocyclic carboxylic oxygen groups (O_{\text{c}}) are usually less basic and hence weaker H-bond acceptors, compared to the imino nitrogen (N_{\text{III}}) atoms of the purine/pyrimidine rings. This explains why the average H-bonding energy of N_\text{II}−H−N_{\text{III}} bonds is higher than that of N_{\text{II}}−H−O_{\text{c}} and N_{\text{II}}−H−O_{\text{b}} bonds (Table 2). Also, it justifies the trend that \( E_{\text{HB}}(N_{\text{II}}−H−O_{\text{b}}) > E_{\text{HB}}(N_{\text{II}}−H−O_{\text{c}}) \). On gas-phase geometry optimization, the G:C W:H Trans base pair deviates from the reverse Watson−Crick geometry to form a bifurcated geometry, whereas the G:G W:H Cis base pair converges to a G:G W:W Trans geometry. Further studies show that they require support either from the formation of higher-order interactions or from the buildup of positive charge environment at the Hoogsteen edge of the guanine residue to stabilize their respective native geometries. The positive charge buildup may take place in the form of metal ion coordination or post-transcriptional modification, even on nucleobase protonation. These observations demonstrate how the geometries and stabilities of different canonical base pairs are determined by their constituent interbase H-bonds.

Another important trend, as shown in Table 2, is that the average H-bonding energy of the interbase H-bonds of WS, HS, and SS base pairs is smaller than the same present in WW, WH, and HH base pairs. This trend is consistent with the fact that the base pairing interactions forming via the sugar edge are comparatively flexible than others and are therefore unstable under QM-based gas-phase geometry optimization. We have also found that the influences of the dispersion interactions on the individual H-bonding energies are small (| \( E_{\text{HB}}(\text{B3LYP}) − E_{\text{HB}}(\text{B3LYP-D3(BJ)}) | < 0.25 \text{ kcal mol}^{-1} \)).
Table 3. Pearson Correlation Coefficients between \( E_{\text{HB}} \) and Different QTAIM, NBO, and Geometric Parameters (at B3LYP/6-31+G(d,p) Level) for Eight Different Types of Interbase H-Bonds Studied in This Work

| sl. | type of HB | type of BP | \( \rho \)  | \( \nabla \rho \) | \( V \)  | \( G \)  | \( H_{\text{rot}} \) | \( E(2) \) | \( D\text{-A} \) | \( H\text{-A} \) | \( D\text{-H} \) | \( \angle D\text{-H}\text{-A} \) | \( S_{\text{DHEA}} \) | \( \Delta_{\text{DHEA}} \) |
|-----|------------|------------|--------|---------|--------|--------|-------------|--------|--------|--------|--------|---------------------|--------|----------|
| 1   | \( N_\text{H-}N_{\text{H}} \) | all        | 0.89 (0.91) | 0.85 (0.90) | -0.88 | 0.87 (0.89) | -0.89 | 0.90 (0.92) | -0.74 | -0.88 | 0.92 (0.95) | 0.75 (0.73) | -0.85 | -0.77 |
|     |            | NS         | 0.96 (0.96) | 0.93 (0.92) | -0.95 | 0.94 (0.93) | -0.89 | 0.97 (0.94) | -0.82 | -0.94 | 0.97 (0.97) | 0.87 (0.82) | -0.91 | -0.88 |
|     |            | S          | 0.76 (0.89) | 0.74 (0.91) | -0.75 | 0.75 (0.90) | -0.65 | 0.80 (0.93) | -0.69 | -0.64 | 0.80 (0.91) | 0.52 (0.60) | -0.66 | -0.54 |
| 2   | \( N_\text{H-}N_{\text{H}} \) | all        | 0.83 (0.88) | 0.72 (0.80) | -0.79 | 0.76 (0.82) | -0.86 | 0.87 (0.91) | -0.77 | -0.82 | 0.99 (0.99) | 0.70 (0.65) | -0.78 | -0.71 |
|     |            | NS         | 0.76 (0.79) | 0.56 (0.60) | -0.71 | 0.65 (0.67) | -0.92 | 0.79 (0.82) | -0.61 | -0.69 | 0.99 (0.99) | 0.66 (0.67) | -0.63 | -0.66 |
|     |            | S          | 0.98 (0.96) | 0.97 (0.93) | -0.96 | 0.97 (0.93) | -0.86 | 0.93 (0.95) | -0.98 | -0.98 | 0.99 (0.99) | 0.53 (0.46) | -0.98 | -0.56 |
| 3   | \( N_\text{H-}O_\text{H} \) | all        | 0.93 (0.95) | 0.84 (0.83) | -0.91 | 0.88 (0.88) | -0.24 | 0.93 (0.93) | -0.91 | -0.89 | 0.95 (0.95) | 0.95 (0.75) | -0.92 | -0.73 |
|     |            | NS         | 0.95 (0.97) | 0.96 (0.89) | -0.94 | 0.95 (0.90) | 0.12 (0.41) | 0.98 (0.93) | -0.91 | -0.90 | 0.95 (0.95) | 0.95 (0.75) | -0.92 | -0.73 |
|     |            | S          | 0.94 (0.93) | 0.87 (0.85) | -0.93 | 0.91 (0.89) | -0.07 | 0.87 (0.91) | -0.57 | -0.88 | 0.96 (0.97) | 0.47 (0.71) | -0.87 | -0.48 |
| 4   | \( N_\text{H-}O_\text{H} \) | all        | 0.84 (0.89) | 0.78 (0.83) | -0.78 | 0.78 (0.83) | -0.16 | 0.90 (0.94) | -0.62 | -0.81 | 0.96 (0.98) | 0.76 (0.79) | -0.73 | -0.78 |
|     |            | NS         | 0.74 (0.75) | 0.68 (0.71) | -0.66 | 0.67 (0.69) | -0.38 | 0.85 (0.85) | -0.69 | -0.79 | 0.98 (0.97) | 0.23 (0.11) | -0.71 | -0.26 |
|     |            | S          | 0.87 (0.93) | 0.87 (0.93) | -0.84 | 0.86 (0.92) | 0.42 (0.45) | 0.93 (0.97) | -0.69 | -0.81 | 0.98 (0.99) | 0.84 (0.91) | -0.76 | -0.86 |
| 5   | \( N_\text{H-}O_\text{H} \) | all        | 0.39 (0.17) | 0.48 (0.31) | -0.41 | 0.44 (0.23) | -0.25 | 0.31 (0.10) | -0.80 | -0.63 | 0.18 (0.21) | 0.04 (-0.15) | 0.18 | -0.08 |
|     |            | NS         | 0.54 (0.60) | 0.43 (0.49) | -0.55 | 0.51 (0.57) | -0.61 | 0.55 (0.63) | -0.48 | -0.46 | 0.63 (0.64) | -0.10 (0.26) | 0.46 | -0.27 |
| 6   | \( O_\text{H-}N_{\text{H}} \) | all        | 0.46 (0.64) | 0.38 (0.61) | -0.43 | 0.40 (0.63) | -0.67 | 0.64 (0.60) | -0.44 | -0.47 | 0.41 (0.66) | 0.73 (0.66) | -0.58 | -0.65 |
| 7   | \( O_\text{H-}O_\text{H} \) | all        | -0.45 (0.04) | -0.51 (0.07) | -0.49 (-0.01) | -0.50 | 0.34 (0.35) | -0.34 | 0.64 (0.04) | 0.50 (-0.01) | -0.01 | -0.09 | 0.58 (-0.01) | 0.22 (0.05) |

*Values given in parenthesis represent the Pearson correlation coefficients calculated at the B3LYP-D3/6-31+G(d,p) level of theory.*
literature studies, which underscore the importance of dispersion interactions in stabilizing the base pairs of WS,74,75 HS,77,79 and SS73 families. As mentioned in the Methods section, for each of the interbase H-bonds, we have calculated 13 different parameters based on their geometry, topology of the corresponding BCP, and associated charge transfer. We have tried to find out the correlation between the H-bonding energy and these 13 parameters.

3.2. How Well Do the Topology-Based, Geometry-Based, and Charge-Transfer-Based Parameters Correlate with the H-Bonding Energies? Pearson correlation coefficients ($r$) calculated between the H-bonding energies obtained from vibrational spectroscopy ($E_{\text{HB}}$) and different geometry-based, topology-based, and charge-transfer-based parameters are shown in Table 3. It may be noted that though the trends observed are similar, there are noticeable differences between the values obtained using B3LYP and B3LYP-D3(BJ) functionals. In most of the cases, the Pearson correlation coefficient ($r$) is better with the values obtained using the dispersion-corrected B3LYP-D3(BJ) functional and has hence been used for our detailed analysis. Further to note is that we have limited our discussion below to the H-bond types 1–4 as listed in Table 3. As explained at the beginning of the section, H-bond types 5–8 are excluded from the discussion because they show poor correlation between $E_{\text{HB}}$ and other parameters.

For these four types of hydrogen bonds, some of the parameters show remarkably good linear correlation ($|r| \geq 0.9$). Out of them, the best correlation with the $E_{\text{HB}}$ values is displayed by the $E(2)$ values obtained from NBO calculations. $E(2)$ values correspond to the stabilization energy because of $n \rightarrow \sigma^*$-type charge transfer from the $\pi$ orbital of the H-bond acceptor to the corresponding antibonding orbital of the H-bond donor ($N$–$H$, $O$–$H$, or $C$–$H$ bond). Therefore, our results suggest that for H-bonds with red-shifted $D$–$H$ bonds, the extent of $n \rightarrow \sigma^*$ charger transfer determines the H-bonding strength. This is in line with the justification for the electronic basis of H-bonding given by Alabugin and coworkers.109 We further observe that correlation between $E(2)$ and $E_{\text{HB}}$ is the maximum for the two strongest H-bonds found in the studied RNA base pairs (Table 2). They are (i) $N_I$–$H$–···$N_{II}$ bonds ($r = 0.94$ for H-bonds observed in base pairs of the WW, WH, and HH families) and (ii) $N_{II}$–$H$–···$O_{III}$ bonds ($r = 0.97$ for H-bonds observed in the base pairs of the WS, HS, and SS families).

The set of geometry-based parameters consists of three interatomic distances ($D$–$A$, $D$–$H$, and $H$–$A$); the angle described by H-bond donor, hydrogen, and H-bond acceptor ($\angle \text{DHA}$); the perimeter of the DHA triangle ($S_{\text{DHA}}$) and its area ($\Delta_{\text{DHA}}$). As discussed earlier,100,109 $n \rightarrow \sigma^*$-type charge transfer from the $\pi$ orbital of the H-bond acceptor to the antibonding orbital of the H-bond donor results into elongation of the corresponding $\sigma$ bond ($N$–$H$, $O$–$H$, or $C$–$H$). Therefore, the change in $D$–$H$ distance due to H-bonding is inherently connected to the concomitant red shift. This rationalizes the observation that out of the three interatomic distances, $D$–$H$ shows exceptionally good correlation ($|r| \geq 0.98$) with the H-bonding strength of all the four types of H-bonds (Table 3). Such weakening of the $D$–$H$ $\sigma$ bonds is linked to the attractive interaction between the acceptor and hydrogen atoms. Consequently, we have found that the extent of weakening of the $N$–$H$/$O$–$H$ bonds is also strongly correlated with the H-bonding energies calculated from logansen’s relationship ($r = 0.95$ for all the H-bonds present in all the base pairs of the WW family). Hence, the $H$–$A$ distances are significantly well correlated ($|r| \geq 0.85$) with the H-bonding strength of all the four types of H-bonds. This also explains why the perimeter of the DHA triangle $S_{\text{DHA}}$ shows reasonably good correlation with the corresponding H-bonding strengths. We further note that consistent with the recent reports,116 the correlations between $E_{\text{HB}}$ and $D$–$A$ distances are comparatively poor ($|r| < 0.8$) for all kinds of H-bonds. It is therefore intuitive that the strength of an interbase H-bond ($D$–$H$–···$A$) does not depend strongly on the corresponding $\angle \text{DHA}$ and the geometry-based parameters derived from that (e.g., the area of the DHA triangle, $\Delta_{\text{DHA}}$). Hence, as shown in Table 3, $\angle \text{DHA}$ and $\Delta_{\text{DHA}}$ display comparatively poor correlation ($|r| < 0.8$) with the corresponding H-bonding strengths ($E_{\text{HB}}$).

Unlike the charge-transfer-based and geometry-based parameters, the physical origin of the relationship between the H-bonding energies and topological parameters calculated at the corresponding BCPs is not well understood. Out of the six topological parameters thus for each interbase H-bond, the electron density ($\rho$) shows good correlation ($0.95 \geq |r| \geq 0.88$) with the $E_{\text{HB}}$ values for all the four interbase H-bonds. On the other hand, the ellipticity ($\epsilon$) shows poor correlation with $E_{\text{HB}}$ for all kinds of interbase H-bonds studied and has therefore not been included in Table 3. Performances of the other four topological parameters depend on the nature of the H-bond. For example, the Laplacian of the electron density ($\nabla^2 \rho$) is well correlated with the $E_{\text{HB}}$ values corresponding to the H-bonds having primary amino group as the donor ($N_6$ of adenine, N4 of cytosine, and N2 of guanine), but for the H-bonds having secondary amino groups as the donor ($N_1$ of guanine and N3 of uracil) the correlation is comparatively poor. Similar trend is also observed for the potential energy density ($V$) and kinetic energy density ($G$). Table 3 suggests that the poor correlation is explicitly due to those $N_{II}$–$H$–···$N_{II}$ and $N_{II}$–$H$–···$O_{III}$ bonds which are present exclusively in the base pairs belonging to the WW, WH, and HH families. Interestingly, the Pearson correlation coefficient between $E_{\text{HB}}$ and the total energy density ($E_{\text{tot}}$) shows remarkably different trends for $N$–$H$–···$N$ and $N$–$H$–···$O$ H-bonds, respectively. The correlation is good for the $N$–$H$–···$N$ bonds ($r = 0.89$ and $-0.86$ for $N_I$–$H$–···$N_{II}$ and $N_{II}$–$H$–···$N_{II}$ bonds, respectively), whereas it is extremely poor for $N$–$H$–···$O$ bonds ($r = -0.24$ and $0.16$ for $N_I$–$H$–···$O_I$ and $N_{II}$–$H$–···$O_{III}$ bonds, respectively).

On the basis of the above analysis, we may infer that the three parameters $E(2)$, $\rho$, and $D$–$H$, covering three different aspects (charge transfer, topology, and geometry, respectively) display remarkably good ($|r| > 0.9$) correlation with the H-bonding energy of all the four different types of H-bonds, especially when the dispersion corrections are taken into consideration. Among the remaining 10 parameters, only $V^2\rho$ and $H$–$A$ are the two parameters, which show good correlation ($0.9 < |r| < 0.85$) with $E_{\text{HB}}$ for all the four different types of H-bonds. Therefore, these five parameters are shortlisted as potential candidates for modeling linear relationships with $E_{\text{HB}}$. These models will be useful to estimate the H-bonding energy of individual H-bonds.

3.3. Modeling Linear Relationship between $E_{\text{HB}}$ and Other Topology-Based, Charge-Transfer-Based, and Geometry-Based Parameters. We have modeled linear equations linking the above-mentioned topology-based, charge-transfer-based, and geometry-based parameters with the H-bonding energy calculated from vibrational spectroscopy
(EHB). To benchmark the efficiency of our models, we have selected a set of 36 modified base pairs. They occur naturally in a nonredundant set of high-resolution RNA crystal structures (HDRNAs\textsuperscript{83}). Their occurrence contexts, structures, and possible contribution in the structural dynamics of functional RNAs have been analyzed recently by Seelam et al.\textsuperscript{83} The modified base pairs constitute an ideal test set for our model because their overall physicochemical properties are reasonably close to our training set (normal base pairs), but at the same time, they possess a significant amount of chemical variations. Both the bases are modified in 2 of these 36 modified base pairs (m\textsuperscript{2}G:Um W:WC and Cm:Gm W:WC), whereas the rest of the base pairs are composed of one normal base and one modified base. Among these 36 base pairs, 12 base pairs have modification in the sugar moiety in the form of methylation at O2\textsuperscript{2} position. Others display different nucleobase modifications, such as m\textsuperscript{2}G (2 base pairs), s\textsuperscript{8}U (3 base pairs), m\textsuperscript{C}C (3 base pairs), m\textsuperscript{4}U (3 base pairs), m\textsuperscript{5}G (2 base pairs), D (3 base pairs), m\textsuperscript{2}G (2 base pairs), m\textsuperscript{2}A (1 base pair), and Ψ (5 base pairs). Note that we have removed the base pairs containing m\textsuperscript{C} modification as they are charged systems. In the rest 34 base pairs, we have identified 56 hydrogen bonds which are associated with a red shift greater than 40 cm\textsuperscript{-1}, and therefore, the corresponding hydrogen bonding energies can be calculated using Logansen’s relationship. Among them, 6 are of N\textsubscript{1}−H···N\textsubscript{3} type, 16 are of N\textsubscript{2}−H···N\textsubscript{3} type, 21 are of N\textsubscript{1}−H···O\textsubscript{c} type, and 13 are of N\textsubscript{2}−H···O\textsubscript{c} type.

First, we have performed a single variable linear regression for all the four hydrogen bonds and modeled the equations (y = Ax + B) relating E\textsubscript{HB} with each of the five parameters (ρ, $\nabla^2\rho$, E(2), D−H, and H−A), respectively, following the conventional least-square fit approach. The slope (A) and the y-intercept (B) of the modeled equations are listed in Table 4. Similar linear equations for other seven parameters (V, G, H, D−A, ΔDHA, S\textsubscript{DHA}, and Δ\textsubscript{DHA}) are shown in Table S3 of the Supporting Information. We have implemented these single variable equations to predict the E\textsubscript{HB} values of the interbase hydrogen bonds of the modified base pairs. Mean square error (MSE) and RMSE values (Table 5) calculated between the predicted E\textsubscript{HB} and expected E\textsubscript{HB} values (as calculated from the vibrational spectra of modified base pairs) show that all the RMSE of all the single variable models varies between 0.4 and 0.8 kcal mol\textsuperscript{-1}. As expected, the model based on D−H distance gives the best result (RMSE = 0.41 kcal mol\textsuperscript{-1}), followed by the models based on E(2) and ρ. It is important to note that the MSE and RMSE values of these single variable models are independent of the strength of the H-bonds (Table 5). Hence, in the worst-case scenario (E(2) ≈ 2.0 kcal mol\textsuperscript{-1}), the mean error in predicting the H-bonding energies will be as high as 20%. On the other hand, in the best-case scenario (E(2) ≈ 8.0 kcal mol\textsuperscript{-1}), the mean error will be only 5%.

To improve the performance of our model, we have performed multiple linear regression analysis with the three parameters ρ, E(2), and D−H (we call it a three-parameter model) and with all the five parameters (we call it a five-parameter model). Using a similar least-square fit approach, we model the following two sets of linear equations;

Three-parameter model: E\textsubscript{HB} = C\textsubscript{0} + C\textsubscript{1} × ρ + C\textsubscript{2} × E(2) + C\textsubscript{3} × D−H

- N\textsubscript{1}−H···N\textsubscript{3} type: E\textsubscript{HB} = -135.1 + 11.2ρ + 0.08E(2) + 135.0D−H

Table 4. Linear Relationships (y = Ax + B) between H-Bonding Strength and the Individual Parameters, Which Have High Pearson Correlation Coefficient (r) for All the Four Types of H-Bonds in Table 3, Were Derived from Single Variable Linear Regression Analysis\textsuperscript{a}

| H-bond type | independent variable (x) | slope (A) | intercept (B) |
|-------------|--------------------------|-----------|---------------|
| N\textsubscript{1}−H···N\textsubscript{3} | ρ | 190.34 ± 14.68 | -0.60 ± 0.38 |
| | $\nabla^2\rho$ | 90.09 ± 7.60 | -1.35 ± 0.48 |
| | E(2) | 0.22 ± 0.02 | 1.33 ± 0.22 |
| | H−A | -6.90 ± 0.81 | 18.44 ± 1.67 |
| | D−H | -210.20 ± 11.88 | -210.78 ± 12.16 |
| N\textsubscript{2}−H···N\textsubscript{3} | ρ | 135.45 ± 21.24 | 1.89 ± 0.72 |
| | $\nabla^2\rho$ | 57.20 ± 12.58 | 1.98 ± 0.99 |
| | E(2) | 0.13 ± 0.02 | 3.65 ± 0.39 |
| | H−A | -9.90 ± 1.58 | 25.42 ± 3.04 |
| | D−H | 150.20 ± 5.17 | -149.41 ± 5.36 |
| N\textsubscript{1}−H···O\textsubscript{c} | ρ | 0.26 ± 0.02 | -0.14 ± 0.24 |
| | $\nabla^2\rho$ | 251.43 ± 13.83 | -3.42 ± 0.37 |
| | E(2) | 75.03 ± 8.71 | -2.71 ± 0.69 |
| | H−A | -12.32 ± 1.10 | 27.04 ± 2.14 |
| | D−H | 253.08 ± 13.98 | -254.87 ± 14.25 |
| N\textsubscript{2}−H···O\textsubscript{c} | ρ | 0.18 ± 0.01 | 1.88 ± 0.27 |
| | $\nabla^2\rho$ | 145.50 ± 16.13 | 0.60 ± 0.54 |
| | E(2) | 46.48 ± 6.76 | 0.88 ± 0.66 |
| | H−A | -8.57 ± 1.17 | 21.28 ± 2.21 |
| | D−H | 212.44 ± 9.71 | -213.70 ± 10.01 |

\textsuperscript{a}For this, two topological parameters (ρ, $\nabla^2\rho$), one charge-transfer-based parameter (E(2)), and two geometry-based parameters (H−A distance and D−H distance) were considered as independent variables (x) and E\textsubscript{HB} was considered as the scalar dependent variable (y). The values for the slope (A) and y-intercept (B), along with their respective standard deviations, are tabulated here. Units of different parameters are as follows: E\textsubscript{HB} [kcal mol\textsuperscript{-1}], ρ [a.u., 1 a.u. = e\textsubscript{Bohr}], $\nabla^2\rho$ [a.u., 1 a.u. = e\textsubscript{Bohr}\textsuperscript{2}], E(2) [kcal mol\textsuperscript{-1}], H−A [Å] and D−H [Å].

Table 5. MSE and RMSE Values (in kcal mol\textsuperscript{-1}) between the Set of Expected E\textsubscript{HB} Values and Set of Predicted E\textsubscript{HB} Values from Different Single Parameter and Multiparameter Models\textsuperscript{a}

| parameter | MSE | RMSE |
|-----------|-----|------|
| ρ | 0.32 | 0.56 | 0.58 |
| $\nabla^2\rho$ | 0.59 | 0.77 | 0.79 |
| E(2) | 0.20 | 0.45 | 0.49 |
| H−A | 0.44 | 0.66 | 0.69 |
| D−H | 0.17 | 0.41 | 0.41 |
| 3-P model | 0.54 | 0.74 | 0.55 |
| 5-P model | 0.12 | 0.34 | 0.30 |

\textsuperscript{a}“Weak” and “strong” correspond to H-bonds with E\textsubscript{HB} < 4 and 15 kcal mol\textsuperscript{-1} ≥ E\textsubscript{HB} ≥ 4 kcal mol\textsuperscript{-1}, respectively.

- N\textsubscript{1}−H···N\textsubscript{3}: E\textsubscript{HB} = -121.9 - 15.5ρ + 0.04E(2) + 123.3D−H
- N\textsubscript{2}−H···N\textsubscript{3}: E\textsubscript{HB} = -112.6 + 103.6ρ + 0.05E(2) + 110.2D−H
- N\textsubscript{1}−H···O\textsubscript{c}: E\textsubscript{HB} = -142.4 + 27.6ρ + 0.04E(2) + 141.7D−H

Five-parameter model: E\textsubscript{HB} = C\textsubscript{0} + C\textsubscript{1} × ρ + C\textsubscript{2} × $\nabla^2\rho$ + C\textsubscript{3} × E(2) + C\textsubscript{4} × H−A + C\textsubscript{5} × D−H
• \(N_{II} \rightarrow H \cdot \cdot \cdot N_{III}: E_{HB} = -110.2 + 227.8\rho - 27.9\sqrt{\rho} + 0.066E(2) + 4.4HA + 98.3DH\)
• \(N_{II} \rightarrow H \cdot \cdot \cdot N_{III}: E_{HB} = -87.4 + 4.5\rho - 37.2\sqrt{\rho} + 0.03E(2) - 6.7HA + 104.9DH\)
• \(N_{II} \rightarrow H \cdot \cdot \cdot O_{C} : E_{HB} = -107.95 + 108.0\rho - 13.9\sqrt{\rho} + 0.05E(2) - 2.2HA + 110.8DH\)
• \(N_{II} \rightarrow H \cdot \cdot \cdot O_{C} : E_{HB} = -143.2 + 44.3\rho - 10.0\sqrt{\rho} + 0.03E(2) - 1.26HA + 145.4DH\)

Comparison of the performance of the five-parameter model (5-P model) and the three-parameter model (3-P model) is illustrated in Figure 4. Interestingly, our analyses show that the performance of the three-parameter model (RMSE = 0.55 kcal mol\(^{-1}\)) is better than all the single parameter models, except the D–H-based model (RMSE = 0.41 kcal mol\(^{-1}\)). At the same time, performance of the five-parameter model (RMSE = 0.3 kcal mol\(^{-1}\)) is even better than that of the D-H-based single parameter model. It is also noteworthy that the corresponding MSE and RMSE values of the 3-P and 5-P models are dependent on the strength of the H-bonds. Table 5 shows that in comparison to the weak H-bonds (\(E_{HB} < 4\) kcal mol\(^{-1}\)), their performance is significantly better for the moderately strong H-bonds (15 \(\geq E_{HB} \geq 4\) kcal mol\(^{-1}\)). Therefore, for the worst-case scenario (for H-bonds with \(E_{HB} \approx 2\) kcal mol\(^{-1}\)), the 5-P model is expected to predict the H-bonding energies with an average error of 17.5%. On the other hand, for the best-case scenario (for H-bonds with \(E_{HB} \approx 8\) kcal mol\(^{-1}\)), the 5-P model is expected to predict the H-bonding energies with an average error of only 3.5%. Importantly, in an average case scenario (\(E_{HB} \approx 5\) kcal mol\(^{-1}\)), the 5-P model is expected to perform reasonably well with an average error of only 6%.

We have also analyzed the performances of these two multivariable linear regression models for different types of H-bonds separately (Tables S5 and S6 in the Supporting Information) and found that performances of both the models are reasonably good (variance score or \(R^2 > 0.9\)) for all the four types of interbase H-bonds (\(N_{II} \rightarrow H \cdot \cdot \cdot N_{III}, N_{II} \rightarrow H \cdot \cdot \cdot N_{III}, N_{II} \rightarrow H \cdot \cdot \cdot O_{C}\) and \(N_{II} \rightarrow H \cdot \cdot \cdot O_{C}\)). Consistent with the trends discussed above, their performances are exceptionally good (\(R^2 = 1\)) for the strongest \(N_{III} \rightarrow H \cdot \cdot \cdot N_{III}\) bonds. To understand the relative significance of the individual parameters in these two models, importance analysis\(^{111}\) was carried out. In keeping with the usual procedure,\(^{112}\) the importance value is estimated as the difference in the corresponding \(R^2\) values, when the corresponding variable is replaced with its average value (see Tables S5 and S6). As expected, the distance between donor and hydrogen atoms (D–H) turns out to be the most significant parameter (importance \(\sim 0.5\)), specially for \(N_{II} \rightarrow H \cdot \cdot \cdot O_{C}\) type H-bonds.

There are several earlier reports attempting to express H-bonding energies as linear functions of different individual parameters, such as \(E(2), \rho, V, \sqrt{\rho}\), and so forth. The novelty in our multivariable models lies in the inclusion of additional parameters, belonging to different domains such as charge transfer, topology, and geometry, which may be capturing additional information regarding system properties. While the respective reliabilities of earlier and our models can be seen in terms of the extent of correlation between predicted values and experimental data, there are problems when it comes to rating or comparing them. That is because none of the earlier approaches have considered the subclassification of the different H-bond classes, in terms of the bonding contexts of the donor and acceptor atoms. However, despite the demonstrated reliability of our multivariable models, caution must be exercised before applying them universally to any system of interest. Rather, it is advisable to validate the approach for other H-bonded systems before carrying out any predictive applications.

4. CONCLUSIONS

In this work, we have come up with a comprehensive estimate of the average H-bonding energies (\(E_{HB}\)) of different interbase hydrogen bonds that constitute the canonical and non-canoncal RNA base pairs. Depending on the chemical identity of the H-bond donor and acceptor atoms, all these interbase hydrogen bonds can be classified into 10 categories, out of which 2 are associated with a small red shift (<40 cm\(^{-1}\)) of the D–H bond stretching and hence are not part of the detailed analysis (C–H–O and C–H–N). \(E_{HB}\) values for these eight types of interbase H-bonds have been calculated from their vibrational spectra obtained from computationally expensive Hessian calculation. The average \(E_{HB}\) values fall within the range of 2.5–7 kcal mol\(^{-1}\) and follow the order: \(N_{II} \rightarrow H \cdot \cdot \cdot N_{III} > N_{II} \rightarrow H \cdot \cdot \cdot O_{C} \geq O_{II} \rightarrow H \cdot \cdot \cdot N_{III} > N_{II} \rightarrow H \cdot \cdot \cdot N_{III} \geq O_{II} \rightarrow H \cdot \cdot \cdot O_{C} > N_{II} \rightarrow H \cdot \cdot \cdot O_{C} \geq O_{II} \rightarrow H \cdot \cdot \cdot O_{C} > N_{II} \rightarrow H \cdot \cdot \cdot O_{C}\). We have studied the linear correlation of \(E_{HB}\) with 13 other frequently used H-bond strength descriptors that can be obtained from computationally inexpensive methods (QTAIM, NBO, etc.). The four H-bonds, which occur exclusively within the Sugar edge-mediated base pairs, do not show strong correlation with any of the parameters studied in this work. For the other four interbase H-bonds (also they are the strongest four interbase H-bonds), we have found (a) three parameters \(E(2), \rho, \) and D–H that belong to three different domains (charge transfer, topology, and geometry, respectively) and display remarkably good (\(|r| > 0.9\)) linear correlation and (b) two parameters \(V, \sqrt{\rho}\) and H–A that show reasonably good linear correlation (0.9 \(< |r| < 0.85\)), especially when the dispersion corrections are taken into consideration. Therefore, we have performed single variable and multiple variable regression analysis to establish linear relationships between the H-bond strength descriptors.
obtained from computationally inexpensive methods with the $E_{\text{HB}}$ values obtained from accurate but computationally expensive methods. We have tested the performance of our linear models over an independent set of base pairs having modified nucleobases. We have obtained promising results, at least with the model composed of five parameters (the 5-P model). In an average case scenario (H-bonds with $E_{\text{HB}} \approx 5$ kcal mol$^{-1}$), it can predict the strength of an interbase H-bond with a 94% accuracy. At the same time, in the best-case scenario (H-bonds with $E_{\text{HB}} \approx 8$ kcal mol$^{-1}$), the accuracy is as good as 96.5%. These predictive models show relatively poor accuracy for the weak H-bonds ($E_{\text{HB}} < 4.0$ kcal mol$^{-1}$). In this regard, it may be noted that depending on the chemical identity of the donor (D) and acceptor (A) groups of a H-bond (D–H···A), the shift in D–H symmetric stretching frequency leads to a continuum of behavior (red shift to no shift to blue shift), where the zero shift occurs around $E_{\text{HB}} \leq 4.0$ kcal mol$^{-1}$. Therefore, any interpretation at this energy range is bound to be error-prone. Overall, these linear models provide the scope to calculate the $E_{\text{HB}}$ values of any individual interbase H-bond.

We expect that our efforts will be useful in exploring the intriguing world of H-bonding as our predictive models provide reliable estimates for H-bond strengths with computationally inexpensive calculations. On the other hand, our analysis of the strengths of individual interbase H-bonds provides effective inputs for designing coarse-grain-based force fields for RNA and other nucleic acid-based systems. These consolidated information are also helpful in analyzing the trajectories obtained from molecular dynamics simulations of DNA/RNA and similar systems.

### ASSOCIATED CONTENT

5 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.8b03689.

Limitations of different computational approaches proposed for estimating the H-bonding energy, computational details of interaction energy calculation, list of RNA base pairs studied in this work, details of $E_{\text{HB}}$ calculation from logangens’s relationship (range of observed red shifts), average values of different parameters (AIM, NBO, geometry-based), and details about the performances of the single variable and multivariable linear regression-based models (PDF)

Cartesian coordinates of the optimized geometries of the natural and modified base pairs and optimized geometries (ZIP)

### AUTHOR INFORMATION

**Corresponding Authors**

*E-mail: antarip.halder@research.iit.ac.in* (A.H.).

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

**ORCID**

Antarip Halder: 0000-0002-3921-5207

Dhruv Data: 0000-0002-5356-4905

Dhananjay Bhattacharyya: 0000-0001-9227-6490

**Present Addresses**

$^4$Solid State and Structural Chemistry Unit (SSCU), Indian Institute of Science (IISc), Bangalore 560012, India.

$^3$Department of Chemistry and Biochemistry, University of Lethbridge, 4401 University Drive West, Lethbridge, Alberta, Canada T1K 3M4.

**Notes**

The authors declare no competing financial interest.

### ACKNOWLEDGMENTS

A.H. acknowledges CSIR for SRF support. A.M. and A.H. thank DBT, Government of India, project BT/PR-148715/PBD/16/903/2010 for partial funding and financial support. A.M. and D.B. thank DBT, Government of India, project BT/PR-11429/BID/07/271/2008 for supporting computational infrastructure.

### ADDITIONAL NOTES

$a \varepsilon = (\lambda_1/\lambda_2) - 1$, where $\lambda_1, \lambda_2,$ and $\lambda_3$ are the eigenvectors of the Hessian matrix.

One can estimate the change in energy of the N–H/O–H bond because of H-bond formation by evaluating $E_{\text{N–H–H bond}} = \frac{1}{2}k_1x_1^2 - \frac{1}{2}k_2x_2^2$. Here, $k_1$ and $x_1$ represent the force constant and bond length of the D–H bond of the isolated nucleobase, respectively. On base pairing and consequent H-bond formation, the force constant and bond length of the D–H bond change to $k_1$ and $x_\rho$, respectively.

### REFERENCES

(1) Nelson, D. L.; Lehninger, A. L.; Cox, M. M. Lehninger principles of biochemistry, 5th ed.; W. H. Freeman: New York, 2008.

(2) Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. Self-splicing RNA: autocleavage and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. Cell 1982, 31, 147–157.

(3) RNA and the Regulation of gene expression: A hidden layer of complexity, 1st ed.; Morris, K. V., Ed.; Caister Academic Press: Norfolk, U.K., 2008.

(4) Eddy, S. R. Non-coding RNA genes and the modern RNA world. Nat. Rev. Genet. 2001, 2, 919–929.

(5) Cech, T. R.; Steitz, J. A. The Noncoding RNA Revolution-Trashing Old Rules to Forge New Ones. Cell 2014, 157, 77–94.

(6) Barrangou, R.; Fremaux, C.; Deveau, H.; Richards, M.; Boyaval, P.; Moineau, S.; Romero, D. A.; Horvath, P. CRISPR provides acquired resistance against viruses in prokaryotes. Science 2007, 315, 1709–1712.

(7) Nowacki, M.; Vijayan, V.; Zhou, Y.; Schotanus, K.; Doak, T. G.; Landweber, L. F. RNA-mediated epigenetic programming of a genome-rearrangement pathway. Nature 2007, 451, 153–158.

(8) Leontis, N. B.; Westhoff, E. Geometric nomenclature and classification of RNA base pairs. RNA 2001, 7, 499–512.

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

(10) Hendrix, D. K.; Brenner, S. E.; Holbrook, S. R. RNA structural motifs: building blocks of a modular biomolecule. Q. Rev. Biophys. 2005, 38, 221–243.

(11) Halder, S.; Bhattacharyya, D. RNA structure and dynamics: a base pairing perspective. Prog. Biophys. Mol. Biol. 2013, 113, 264–283.

(12) Zirbel, C. L.; Ønner, J. E.; Ønner, J.; Stombaugh, J.; Leontis, N. B. Classification and energetics of the base-phosphate interactions in RNA. Nucleic Acids Res. 2009, 37, 4898–4918.

(13) Yureno, Y. P.; Zhuravskivy, R. O.; Samjlenko, S. P.; Hovorun, D. M. Intramolecular CH...O Hydrogen Bonds in the AI and BI DNA-like Conformers of Canonical Nucleosides and their Watson-Crick Pairs. Quantum Chemical and AIM Analysis. J. Biomol. Struct. Dyn. 2011, 29, S1–65.
(14) Jeffrey, G. A.; Saenger, W. Hydrogen Bonding in Biological Structures, 1st ed.; Springer-Verlag: Berlin, Heidelberg, 1991.
(15) Müller-Dethlefs, K.; Hobza, P. Noncovalent interactions: a challenge for experiment and theory. Chem. Rev. 2000, 100, 143–168.
(16) Buck, U.; Huisken, F. Infrared spectroscopy of size-selected water and methanol clusters. Chem. Rev. 2000, 100, 3863–3890.
(17) Neusser, H. J.; Sigkow, K. High-resolution ultraviolet spectroscopy of neutral and ionic clusters: hydrogen bonding and the external heavy atom effect. Chem. Rev. 2000, 100, 3921–3942.
(18) Logansen, A. V. Direct proportionality of the hydrogen bonding energy and the intensification of the stretching (νH) vibration in infrared spectra. Spectrochim. Acta, Part A 1999, 55, 1585–1612.
(19) Sukhodub, L. F. Interactions and hydration of nucleic acid bases in a vacuum. Experimental study. Chem. Rev. 1987, 87, 589–606.
(20) Arnett, E. M.; Joris, L.; Mitchell, E.; Murty, T. S. S. R.; Gorrie, T. M.; Schleyer, P. v. R. Hydrogen-bonded complex formation. III. Thermodynamics of complexing by infrared spectroscopy and calorimetry. J. Am. Chem. Soc. 1970, 92, 2365–2377.
(21) Jelesekrov, I.; Bossard, H. R. Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. J. Mol. Recognit. 1999, 12, 3–18.
(22) Solomonov, B. N.; Novikov, V. B.; Varfolomeev, M. A.; Klimovitskii, A. E. Calorimetric determination of hydrogen-bonding enthalpy for neat aliphatic alcohols. J. Phys. Org. Chem. 2005, 18, 1132–1137.
(23) Rocher-Casterline, B. E.; Chen, L. C.; Mollner, A. K.; Reisler, H. Communication: Determination of the bond dissociation energy (D0) of the water dimer, (H2O)2, by velocity map imaging. J. Chem. Phys. 2011, 134, 211101.
(24) Flieg, H.; Lehtonen, O.; Sundholm, D.; Kaila, V. R. I. Hydrogen-bond strengths by magnetically induced currents. Phys. Chem. Chem. Phys. 2011, 13, 434–437.
(25) Parthasarathi, R.; Subramanian, V. Hydrogen Bonding—New Insights, 1st ed.; Grabowski, S. J., Ed.; Springer: Dordrecht, The Netherlands, 2006, Chapter 1, pp 1–50.
(26) Bader, R. F. W. Atoms in Molecules: A Quantum Theory, 1st ed.; Oxford Science: Oxford, 1990.
(27) Bader, R. F. W. A quantum theory of molecular structure and its applications. Chem. Rev. 1991, 91, 893–928.
(28) Grabowski, S. J. A new measure of hydrogen bonding strength: ab initio and atoms in molecules studies. Chem. Phys. Lett. 2001, 338, 361–366.
(29) Espinoso, E.; Molins, E.; Lecomte, C. Hydrogen bond strengths revealed by topological analyses of experimentally observed electron densities. Chem. Phys. Lett. 1998, 285, 170–173.
(30) Parthasarathi, R.; Subramanian, V.; Sathyamurthy, N. Electron Density Topography, NMR, and NBO Analysis of Water Clusters. Synth. React. Inorg., Met-Org., Nano-Met. Chem. 2008, 38, 18–27.
(31) Parthasarathi, R.; Subramanian, V.; Sathyamurthy, N. Hydrogen Bonding in Phenol, Water, and Phenol—Water Clusters. J. Phys. Chem. A 2005, 109, 843–850.
(32) Parthasarathi, R.; Subramanian, V.; Sathyamurthy, N. Hydrogen bonding in protonated water clusters: an atoms-in-molecules perspective. J. Phys. Chem. A 2007, 111, 13287–13290.
(33) Parthasarathi, R.; Subramanian, V.; Sathyamurthy, N. Hydrogen bonding without borders: an atoms-in-molecules perspective. J. Phys. Chem. A 2006, 110, 3349–3351.
(34) Rai, S.; Singh, H. Electronic structure theory based study of proline interacting with gold nano clusters. J. Mol. Model. 2013, 19, 4099–4109.
(35) Parthasarathi, R.; Amutha, R.; Subramanian, V.; Nair, B. U.; Ramasami, T. Bader’s and Reactivity Descriptors’ Analysis of DNA Base Pairs. J. Phys. Chem. A 2004, 108, 3817–3828.
(36) Grabowski, S. J. 1-Electron delocalisation for intramolecular resonance assisted hydrogen bonds. J. Phys. Org. Chem. 2003, 16, 797–802.
(37) Sharma, P.; Mitra, A.; Sharma, S.; Singh, H.; Bhattacharyya, D. Quantum Chemical Studies of Structures and Binding in Non-canonical RNA Base pairs: The Trans Watson-Crick:Watson-Crick Family. J. Biomol. Struct. Dyn. 2008, 25, 709–732.
(38) Ebrahimi, A.; Habibi Khorassani, S. M.; Delaram, H. Estimation of individual binding energies in some dimers involving multiple hydrogen bonds using topological properties of electron charge density. Chem. Phys. 2009, 365, 18–23.
(39) Mata, I.; Molins, E.; Alkorta, I.; Espinoso, E. The Paradox of Hydrogen-Bonded Anion-Anion Aggregates in Oxoanions: A Fundamental Electrostatic Problem Explained in Terms of Electrophilic–Nucleophilic Interactions. J. Phys. Chem. A 2014, 119, 183–194.
(40) Rai, S.; Singh, H.; Priyakumar, U. D. Binding to gold nanoclusters alters the hydrogen bonding interactions and electronic properties of canonical and size-expanded DNA base pairs. RSC Adv. 2015, 5, 49408–49419.
(41) Alkorta, I.; Mata, I.; Molins, E.; Espinoso, E. Charged versus Neutral Hydrogen-Bonded Complexes: Is There a Difference in the Nature of the Hydrogen Bonds? Chem. - Eur. J. 2016, 22, 9226–9234.
(42) Balamurugan, K.; Prakash, M.; Subramanian, V. Theoretical Insights into the Role of Water Molecules in the Guanidinium-Based Protein Denaturation Process in Specific to Aromatic Amines. J. Phys. Chem. B 2012, 116, 2191–2202.
(43) Gould, I. R.; Vincent, M. A.; Hillier, I. H. A theoretical study of the infrared spectrum of uracil. J. Chem. Soc., Perkin Trans. 2 1992, 69–71.
(44) Kwiatkowski, J. S.; Leszczynski, J. Molecular structure and vibrational IR spectra of cytosine and its thio and seleno analogues by density functional theory and conventional ab initio calculations. J. Phys. Chem. 1996, 100, 941–953.
(45) Nowak, M. J.; Lapinski, L.; Kwiatkowski, J. S.; Leszczynski, J. Molecular Structure and Infrared Spectra of Adenine. Experimental Matrix Isolation and Density Functional Theory Study of Adenine+SN Isotopomers. J. Phys. Chem. 1996, 100, 3527–3534.
(46) Lopes, R. P.; Valero, R.; Tomkinson, J.; Marques, M. P. M.; Batista de Carvalho, L. A. E. Applying vibrational spectroscopy to the study of nucleobases - adenine as a case-study. New J. Chem. 2013, 37, 2691–2699.
(47) Lopes, R. P.; Marques, M. P. M.; Valero, R.; Tomkinson, J.; de Carvalho, L. A. E. B. Guanine: a combined study using vibrational spectroscopy and theoretical methods. Spectroscopy 2012, 27, 273–292.
(48) Nikolaienko, T. Y.; Bulavin, L. A.; Hovorun, D. M. Bridging QTAIM with vibrational spectroscopy: The energy of intramolecular hydrogen bonds in DNA-related biomolecules. Phys. Chem. Chem. Phys. 2012, 14, 7441–7447.
(49) Brovarets’, O. O.; Hovorun, D. M. Does the G G\*syn DNA mismatch contain canonical and rare tautomers of the guanine tautomerise through the DPT? A QM/QTAIM microstructural study. Mol. Phys. 2014, 112, 3033–3046.
(50) Brovarets’, O. O.; Tsipua, K. S.; Hovorun, D. M. Unexpected A-T(WC)-A(T(rWC))/A(T(rH))-A(T(rWC)) conformational transitions between the classical A-T DNA base pairs: A QM/QTAIM comprehensive study. Int. J. Quantum Chem. 2018, 118, No. e25674.
(51) Brovarets’, O. O.; Tsipua, K. S.; Hovorun, D. M. Non-dissociative structural transitions of the Watson-Crick and reverse Watson-Crick A-T DNA base pairs into the Hoogsteen and reverse Hoogsteen forms. Sci. Rep. 2018, 8, 10371.
(52) Brovarets’, O. O.; Tsipua, K. S.; Hovorun, D. M. Novel pathway for mutagenic tautomerization of classical A-T DNA base pairs via sequential proton transfer through quasi-orthogonal transition states: A QM/QTAIM investigation. PLoS One 2018, 13, No. e0199044.
(53) Weinhold, F. Nature of H-bonding in clusters, liquids, and enzymes: an ab initio, natural bond orbital perspective. J. Mol. Struct. 1997, 398–399, 181–197.
Acid Interactions in Watson-Crick Disposition. Topological and Electric Field Analyses of Cation-Cation Nucleic Acid Interactions of NBO energies of individual hydrogen bonds in nucleic acid base pairs of natural nucleobases: a thorough quantum-chemical study. \( J. \text{Chem. Phys. Chem. A} \text{ 2016, 27, 367–376.} \)

(59) Foster, J. P.; Weinhold, F. Natural hybrid orbitals. \( J. \text{Am. Chem. Soc. 1980, 102, 7211–7218.} \)

(60) Weinhold, F.; Landis, C. R. Natural Bond Orbitals and Extensions of Localized Bonding Concepts. \( \text{Chem. Educ. Res. Pract. 2001, 2, 91–104.} \)

(61) Grunenberg, J. Direct Assessment of Intermolecular Forces in Watson–Crick Base Pairs Using Theoretical Compliance Constants. \( J. \text{Am. Chem. Soc. 2004, 126, 16310–16311.} \)

(62) Pandey, S. K.; Manogaran, D.; Manogaran, S.; Schaefer, H. F. Quantification of Hydrogen Bond Strength Based on Interaction Coordinates: A New Approach. \( J. \text{Phys. Chem. A 2017, 121, 6090–6103.} \)

(63) Gould, I. R.; Kollman, P. A. Theoretical investigation of the hydrogen bond strengths in guanine-cytosine and adenine-thymine base pairs. \( J. \text{Am. Chem. Soc. 1994, 116, 2493–2499.} \)

(64) Asemsio, A.; Kobko, N.; Dannenberg, J. J. Cooperative Hydrogen-Bonding in Adenine–Thymine and Guanine–Cytosine Base Pairs. Density Functional Theory and Moller–Plesset Molecular Orbital Study. \( J. \text{Phys. Chem. A 2003, 107, 6441–6443.} \)

(65) Matta, C. F.; Castillo, N.; Boyd, R. J. Extended Weak Bonding Interactions in DNA: \( \pi \)-Stacking (Base–Base), Base–Backbone, and Backbone–Backbone Interactions†. \( J. \text{Phys. Chem. B 2006, 110, 563–578.} \)

(66) Brovarets’ O. O.; Hovorun, D. M. New structural hypostases of the A-T and G-C Watson-Crick DNA base pairs caused by their mutagenic tautomerisation in a wobbly manner: a QM/QTAIM prediction. \( \text{RSC Adv. 2015, 5, 99594–99605.} \)

(67) Alkorta, I.; Mata, I.; Molins, E.; Espinosa, E. Energetic, Topological and Electric Field Analyses of Cation-Cation Nucleic Acid Interactions in Watson-Crick Disposition. \( \text{ChemPhysChem 2019, 20, 148–158.} \)

(68) Ray, S. S.; Halder, S.; Kaypee, S.; Bhattacharyya, D. HD-RNAS: An automated hierarchical database of RNA structures. \( \text{Front. Genet. 2012, 3, 59.} \)

(69) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. \( \text{Phys. Rev. A 1988, 38, 3098.} \)

(70) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. \( \text{Phys. Rev. B 1988, 37, 785.} \)

(71) Michelh, B.; Savin, A.; Stoll, H.; Preuss, H. Results obtained with the correlation energy density functionals of Becke and Lee, Yang and Parr. \( \text{Chem. Phys. Lett. 1989, 157, 200–206.} \)

(72) Swart, M.; Fonseca Guerra, C.; Bickelhaupt, F. M. Hydrogen bonds of RNA are stronger than those of DNA, but NMR monitors only presence of methyl substituent in uracil/thymine. \( \text{J. Am. Chem. Soc. 2004, 126, 16718–16719.} \)

(73) Šponer, J. E.; Leszczyński, J.; Sychrovský, V.; Šponer, J. Sugar Edge/Sugar Edge Base Pairs in RNA: Stabilities and Structures from Quantum Chemical Calculations. \( \text{J. Phys. Chem. B 2005, 109, 18680–18689.} \)

(74) Šponer, J.; Špačková, N.; Kulhánek, P.; Leszczynski, J.; Šponer, J. Non-Watson-Crick base pairing in RNA. quantum chemical analysis of the cis Watson-Crick/sugar edge base pair family. \( \text{J. Phys. Chem. A 2005, 109, 2392–2401.} \)

(75) Šponer, J.; Špačková, N.; Leszczynski, J.; Šponer, J. Principles of RNA base pairing: structures and energies of the trans Watson-Crick/sugar edge base pairs. \( \text{J. Phys. Chem. B 2005, 109, 11399–11410.} \)

(76) Bhattacharyya, D.; Koripella, S. C.; Mitra, A.; Rajendran, V. B.; Sinha, B. Theoretical analysis of noncanonical base pairing interactions in RNA molecules. \( \text{J. Biosci. 2007, 32, 809–825.} \)

(77) Mładek, A.; Sharma, P.; Mitra, A.; Bhattacharyya, D.; Šponer, J.; Šponer, J. E. Trans Hoogsteen/Sugar Edge Base Pairing in RNA. Structures, Energies, and Stabilities from Quantum Chemical Calculations. \( \text{J. Phys. Chem. B 2009, 113, 1743–1755.} \)

(78) Sharma, P.; Chawla, M.; Sharma, S.; Mitra, A. On the role of Hoogsteen:Hoogsteen interactions in RNA: Ab initio investigations of structures and energies. \( \text{RNA 2010, 16, 942–957.} \)

(79) Sharma, P.; Šponer, J. E.; Šponer, J.; Sharma, S.; Bhattacharyya, D.; Mitra, A. On the cis Hoogsteen:Sugar Edge Family of Base Pairs in Platforms and Triplets-Quantum Chemical Insights into RNA Structural Biology. \( \text{J. Phys. Chem. B 2010, 114, 3307–3320.} \)

(80) Mładek, A.; Šponer, J. E.; Kulhánek, P.; Lu, X.-J.; Olson, W. K.; Šponer, J. Understanding the sequence preference of recurrent RNA building blocks using quantum chemistry: the intranucle RNA dinucleotide platform. \( \text{J. Chem. Theory Comput. 2011, 8, 335–347.} \)

(81) Sharma, P.; Lait, L. A.; Wetmore, S. D. Exploring the limits of nucleobase expansion: computational design of naphthohomologated (xx-) purines and comparison to the natural and xDNA purines. \( \text{Chem. Phys. Chem. 2013, 15, 15538–15549.} \)

(82) Sharma, P.; Lait, L. A.; Wetmore, S. D. yDNA versus yyDNA pyrimidines: computational analysis of the effects of unidirectional ring expansion on the preferred sugar-base orientation, hydrogen-bonding interactions and stacking abilities. \( \text{Phys. Chem. Chem. Phys. 2013, 15, 2443–2448.} \)

(83) Seelam, P. P.; Sharma, P.; Mitra, A. Structural landscape of base pairs containing post-transcriptional modifications in RNA. \( \text{RNA 2017, 23, 847–859.} \)

(84) Grimme, S.; Antony, J.; Ehrlisch, S.; Krrieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. \( \text{J. Chem. Phys. 2010, 132, 154104.} \)

(85) Grimme, S.; Ehrlisch, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. \( \text{J. Comput. Chem. 2011, 32, 1456–1465.} \)

(86) Azarhazin, E.; Izadyar, M.; Housaindokht, M. R. Molecular dynamic simulation and DFT study on the Drug-DNA interaction; Crocin as an anti-cancer and DNA nanostructure model. \( \text{J. Biomol. Struct. Dyn. 2018, 36, 1063–1074.} \)

(87) Kaur, S.; Sharma, P. Cyanocacetaldehyde as a building block for prebiotic formation of pyrimidines. \( \text{Int. J. Quantum Chem. 2019, No. e25886.} \)

(88) Frisch, M. J.; et al. Gaussian 03, Revision E.01; Gaussian, Inc.: Wallingford, CT, 2004.

(89) Keith, T. AIMAll, Version 11.12.19, 2011.

(90) Reed, A. E.; Carpenter, J. E.; Glendening, E. D.; Weinhold, F. NBO, Version 3; Gaussian, Inc.: Wallingford, CT, 2004.

(91) Brovarets’, O. O.; Yurenko, Y. P.; Hovorun, D. M. The significant role of the intermolecular CH···O/N hydrogen bonds in governing the biologically important pairs of the DNA and RNA modified bases: a comprehensive theoretical investigation. \( \text{J. Biomol. Struct. Dyn. 2015, 33, 1624–1625.} \)

(92) Brovarets’, O. O.; Yurenko, Y. P.; Hovorun, D. M. Intermolecular CH···O/N H-bonds in the biologically important pairs of natural nucleobases: a thorough quantum-chemical study. \( \text{J. Biomol. Struct. Dyn. 2014, 32, 993–1022.} \)
(93) Roy, A.; Panigrahi, S.; Bhattacharyya, M.; Bhattacharyya, D. Structure, stability, and dynamics of canonical and noncanonical base pairs: quantum chemical studies. J. Phys. Chem. B 2008, 112, 3786–3796.

(94) Dong, H.; Hua, W.; Li, S. Estimation on the Individual Hydrogen-Bond Strength in Molecules with Multiple Hydrogen Bonds. J. Phys. Chem. A 2007, 111, 2941–2945.

(95) Anatole von Lilienfeld, O.; Tkatchenko, A. Two- and three-body interatomic dispersion energy contributions to binding in molecules and solids. J. Chem. Phys. 2010, 132, 234109.

(96) Hobza, P. Calculations on Noncovalent Interactions and Databases of Benchmark Interaction Energies. Acc. Chem. Res. 2012, 45, 663–672.

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

(98) Liu, M.; Li, T.; Amegayibor, F. S.; Cardoso, D. S.; Fu, Y.; Lee, J. K. Gas-phase thermochemical properties of pyrimidine nucleobases. J. Org. Chem. 2008, 73, 9283–9291.

(99) Zhachkina, A.; Liu, M.; Sun, X.; Amegayibor, F. S.; Lee, J. K. Gas-Phase Thermochemical Properties of the Damaged BaseO6-Methylguanine versus Adenine and Guanine. J. Org. Chem. 2009, 74, 7429–7440.

(100) Joseph, J.; Jemmis, E. D. Red-, Blue-, or No-Shift in Hydrogen Bonds: A Unified Explanation. J. Am. Chem. Soc. 2007, 129, 4620–4632.

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

(102) Halder, A.; Vemuri, S.; Roy, R.; Katuri, J.; Bhattacharyya, D.; Mitra, A. Evidence for Hidden Involvement of N3-Protonated Guanine in RNA Structure and Function. ACS Omega 2019, 4, 699–709.

(103) Chawla, M.; Abdel-Azeim, S.; Oliva, R.; Cavallo, L. Higher order structural effects stabilizing the reverse Watson–Crick Guanine-Cytosine base pair in functional RNAs. Nucleic Acids Res. 2014, 42, 714.

(104) Oliva, R.; Cavallo, L. Frequency and Effect of the Binding of Mg2+, Mn2+, and Co2+Ions on the Guanine Base in Watson–Crick and Reverse Watson–Crick Base Pairs. J. Phys. Chem. B 2009, 113, 15670–15678.

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

(106) Halder, A.; Roy, R.; Bhattacharyya, D.; Mitra, A. Consequences of Mg2+ binding on the geometry and stability of RNA base pairs. Phys. Chem. Chem. Phys. 2018, 20, 21934–21948.

(107) Oliva, R.; Tramontano, A.; Cavallo, L. Mg2+ binding and archaeosine modification stabilize the G15 C48 Levitt base pair in tRNAs. RNA 2007, 13, 1427–1436.

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

(109) Alabugin, I. V.; Manoharan, M.; Peabody, S.; Weinhold, F. Electronic Basis of Improper Hydrogen Bonding: A Subtle Balance of Hyperconjugation and Rehybridization. J. Am. Chem. Soc. 2003, 125, 5973–5987.

(110) Hansen, P. E.; Spanget-Larsen, J. NMR and IR investigations of strong intramolecular hydrogen bonds. Molecules 2017, 22, 552.

(111) Tonidandel, S.; LeBreton, J. M. Relative Importance Analysis: A Useful Supplement to Regression Analysis. J. Bus. Psychol. 2011, 26, 1–9.

(112) Bothra, N.; Rai, S.; Pati, S. K. Tailoring Ca2Mn2O5 Based Perovskites for Improved Oxygen Evolution Reaction. ACS Appl. Energy Mater. 2018, 1, 6312–6319.

(113) Sen, S.; Patwari, G. N. Electrostatics and Dispersion in X-H···Y (X = C, N, O; Y = N, O) Hydrogen Bonds and Their Role in X-H Vibrational Frequency Shifts. ACS Omega 2018, 3, 18518–18527.