Estimation of strength in different extra Watson–Crick hydrogen bonds in DNA double helices through quantum chemical studies

2 AUTHORS:

Debashree Bandyopadhyay
BITS Pilani, Hyderabad
12 PUBLICATIONS **49** CITATIONS

SEE PROFILE

Dhananjay Bhattacharyya
Saha Institute of Nuclear Physics
115 PUBLICATIONS **1,093** CITATIONS

SEE PROFILE

D. Bandyopadhyay[1]*
D. Bhattacharyya[2]
[1] Department of Chemistry,
Raja Peary Mohan College,
Uttarpara, Hooghly
PIN 712258, India

[2] Biophysics Division, Saha
Institute of Nuclear Physics,
1/AF Bidhannagar,
Kolkata 700064, India

Estimation of Strength in Different Extra Watson–Crick Hydrogen Bonds in DNA Double Helices Through Quantum Chemical Studies

Abstract: It was shown earlier, from database analysis, model building studies, and molecular dynamics simulations that formation of cross-strand bifurcated or Extra Watson–Crick hydrogen (EWC) bonds between successive base pairs may lead to extra rigidity to DNA double helices of certain sequences. The strengths of these hydrogen bonds are debatable, however, as they do not have standard linear geometry criterion. We have therefore carried out detailed ab initio quantum chemical studies using RHF/6-31G(2d,2p) and B3LYP/6-31G(2p,2d) basis sets to determine strengths of several bent hydrogen bonds with different donor and acceptors. Interaction energy calculations, corrected for the basis set superposition errors, suggest that N—H⋯O type bent EWC hydrogen bonds are possible along same strands or across the strands between successive base pairs, leading to significant stability (ca. 4–9 kcal/mol). The N—H⋯N and C—H⋯O type interactions, however, are not so stabilizing. Hence, consideration of EWC N—H⋯O H-bonds can lead to a better understanding of DNA sequence directed structural features.   © 2006 Wiley Periodicals, Inc. Biopolymers 83: 313–325, 2006

This article was originally published online as an accepted preprint. The "Published Online" date corresponds to the preprint version. You can request a copy of the preprint by emailing the Biopolymers editorial office at biopolymers@wiley.com

Keywords: extra Watson–Crick hydrogen bonds; cross-strand bifurcated hydrogen bonds; ab initio quantum chemical studies; strength of hydrogen bonds; DNA structural rigidity; pyramidal amino groups

INTRODUCTION

Hydrogen bonding (H-bond) between base pairs in a nucleic acid double helix is a unique phenomenon in biological chemistry. Its importance lies in its property to join two strands with the provision of base-pair opening, which allows genetic replication to occur with high fidelity. Hydrogen bond in its most general-

ized way is described as "any cohesive D—H···A interaction, where H carries partial positive charge and acceptor (A) carries a negative (partial or full electronic) charge, the charge on donor (D) is generally more negative than that on H."[1] These H-bonds in Watson–Crick (A:T, G:C) base pairs are the source of stability in DNA. H-bonds are also vital for various secondary structural motifs of protein, e.g., α-helix, β-sheet, β-turn, etc.

Crystal structure analysis[2] first proposed the possibility of cross-strand hydrogen bonds [CSH; also can be termed Extra Watson–Crick (EWC) H-bond] between adjacent base pairs, which leads to a large propeller twist and narrow minor groove in A-tract DNA. Subsequently, Yathindra and coworkers[3] had modeled 16 possible dinucleotides with varying structural parameters and found CSH formation in both grooves of d(GG)·d(CC), d(CA)·d(TG), in the minor groove of d(AG)·d(CT), and in the major groove of d(AA)·d(TT), and d(AC)·d(GT). We had postulated from crystal database analysis followed by model building studies that structural rigidity of some DNA sequences is associated with the formation of EWC H-bonds.[4] Lack of EWC H-bond formation in some polymeric sequence—namely, d(CTG)·d(CAG) and d(CCG)·d(CGG) repeats—might be one of the significant causes for their flexible character.[4,5] Further, the structural rigidity in d(AA)·d(TT) and the flexibility of d(CA)·d(TG) dinucleotide could be attributed to the presence and absence of the EWC H-bond in respective doublet steps.[6] Rapture of CSH bonds between successive adenine and thymine bases[7–9] was assumed to be the deciding factor for the premelting transition in poly(dA)·poly(dT) or A-tract DNA (duplexes with four or more adenine bases). A DNA crystal structure database survey[10] has established a significant population of cross-strand hydrogen-bonding interactions in non-Z-DNA (A-DNA, B-DNA), DNA–drug, DNA–protein complexes, and in unusual DNA forms, such as the quadruplex[11,12] following standard hydrogen-bonding criteria. These CSH bonds are generally longer in distance and smaller in angle values[10]—hence, relatively weak—because of their bent nature as compared to those of normal Watson–Crick pairs.[13]

Crystal structure analysis of DNA even with very high resolution is inadequate to determine positions of polar hydrogen atoms and the amino groups are generally assumed to be planar in most modeling studies. However, Sponer and coworkers[14,15] have shown by an extensive ab initio quantum chemical method that these amino groups of DNA are generally nonplanar and adopt a pyramidal geometry. Great importance of such pyramidal amino groups has also been reported recently from our laboratory.[16] Watson–Crick base

pairs, optimized by the MP2 method, also recently indicated nonplanar stable base pairs with pyramidal amino groups.[17] It was also experimentally verified by low-temperature infrared (IR) spectroscopy that adenine and cytosine amino groups are nonplanar.[18] Moreover, a recent analysis of protein structures determined by neutron diffraction method and X-ray crystallography with extremely high resolution[19,20] was able to determine positions of the polar and replaceable deuterium atoms, which indicated significant pyramidalization of the amide nitrogen. It was also shown that loss of an electron from the nitrogen lone-pair (involved in extended conjugation with an aromatic ring) caused complete disruption in pyramidal character of the amino group in the excited state.[20] All these studies indicate that consideration of planar amino groups of DNA bases is too simplistic. Usually the hydrogen-bond donors in DNA are primary or secondary amino nitrogen atoms; the acceptors are keto oxygen, amino, or imino nitrogen atoms; and H-bonds are of the form N—H···O or N—H···N. The existence of carbon as a H-bond donor in DNA was also reported from an analysis of the crystal database.[21–26] The C—H···O contacts in nucleic acid base pairs are less frequent and mostly exist as base–backbone contacts between purine C8—H or pyrimidine C6—H and the O5′ atom in the backbone when the glycosidic bonds are in the *anti* conformation.[27] The importance of a long C(2)—H(2)···O(2)=C contact in the A—U/T Watson—Crick and Hoogsteen base pair, corresponding to a third H-bond, was also reported.[28] This interaction was considered a weak H-bond by some group,[29] where the proton donor is a $(CH)^{\delta+}$ moiety instead of carbon atom. On the other hand, Hobza and coworkers[30] argued against the possibility of C—H···O contacts as true H-bonds in the A:U Watson–Crick/Hoogsteen base pair; instead, they called these "improper blue-shifting hydrogen bonds."[31]

Despite the finding of numerous EWC hydrogen bond-like contacts, from crystal data as well as theoretical models, and their consequences on DNA structure and function, assessment of their strength in terms of energy is seldom done. Strengths of these bonding are of most importance in understanding the possible correlations between DNA structural rigidity and bent hydrogen-bond formation, and hence role of some DNA sequences in gene regulation. Because it is very difficult to design gas-phase experiments to study cross-strand H-bonds in DNA systems and condensed phase experiments do not provide unambiguous information, high-level quantum chemical calculations are most effective to study the nature and strengths of these unusual H-bonds.

For the first time we are reporting the binding strengths of different EWC H-bond like contacts in

six different systems. Initially, H-bond formation is indicated in those systems where the potential energy vs. internuclear separation curve reached a minimum. This is supported by additional changes in bond distances and bond angles due to hydrogen-bond formation. In addition, IR spectroscopic information and change in natural populations are also computed to support the identification of such bonds in experimental setting. We have carried out these calculations at Hartree–Fock (HF) and Density Functional Theory (DFT) using the 6-31G (2p,2d) basis set, as it was shown recently that these methods can describe interactions between biomolecular groups with sufficient accuracy.[15]

METHODS

Geometry of the Systems

We have studied six different systems to analyze the formation of EWC H-bonds involving N_p—H···O (interstrand), N_p—H···O (intrastrand), N_s—H···N_i, N_p—H—N_p, and C—H···O in the d(AA)·d(TT) doublet, and C—H···O in the d(GA)·d(TC) doublet (Table I), where N_p is the nitrogen atom belonging to the primary amino group, N_s is that belonging to the secondary amino group, and N_i is the imino nitrogen atom. The abbreviations of the types of H-bonds and the sources of the coordinates are given in Table I. Suitable N_s—H···N_i contact is not found in any pairs of bases with stacked geometry, which is decoupled from other stronger electrostatic interactions, like N—H···O. Hence, we have chosen two amino acids, glycine and histidine, stacked almost in same fashion as are DNA interstrand base pairs. These amino acids are blocked at both N- and C- terminal ends by a standard acetyl (COCH$_3$) group at the N-terminal end and a methyl (CH$_3$) group at the C- terminal end. Initial positions of the hydrogen atoms in all the systems are obtained from MOLDEN software[32] with conventional planar geometry. The induced pyramidal amounts were measured as an angle between the normal to the plane subtended by the two hydrogens and the nitrogen and the vector joining the nitrogen and bonded carbon.[16,20] The conventional improper torsion angles are also calculated for comparison.

Computational Methods

We have optimized structures of the supramolecules using the RHF/6-31G (2d,2p) basis set by GAMESS-US[35] with different D—A separation distances, where D is the H-bond donor heavy atom and A is the H-bond acceptor. Individual monomers are moved with respect to each other so that the distance between the H-bond donor and acceptor varies within the range of 1.5–15 Å. All these systems were geometry optimized, maintaining positional constraints to keep all the nonhydrogen atoms fixed in their initial position by IFREEZ option of GAMESS. Thus the relative arrangement of EWC base pairs remains unaltered during optimization.

We have carried out geometry optimizations of a few selected structures considering electron correlation effects by B3LYP/6-31G (2p,2d). We have also calculated basis set superposition error (BSSE) corrected interaction energies between the monomers of the supramolecule by MOROKUMA method using the HF/6-31G(2p,2d) method. Normal-mode vibrations of these optimized structures were evaluated by the B3LYP/6-31G(2p,2d) basis set using numerical Hessian, constraining degrees of freedom of the heavy atoms by IFREEZ. To standardize the theoretical IR calculations using numerical Hessian, we have also calculated analytical Hessians with fully geometry optimized A:T and G:C Watson–Crick base pairs. Natural Bond Orbital analysis was carried out using the NBO 3.1 suite.[36]
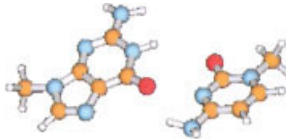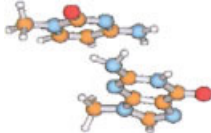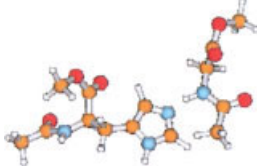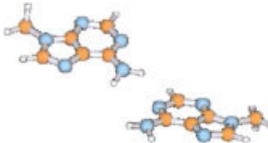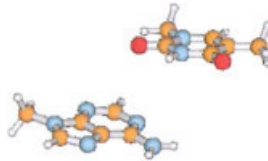
RESULTS AND DISCUSSION

Locating Suitable Orientations for Strongest Hydrogen-Bond Formation

Cross-strand bifurcated or EWC hydrogen bonds are not strong and unambiguous hydrogen bonds due to their nonstandard geometry. Previous studies indicated that these H-bonds are rather elongated and D—H···A angles also differ significantly from linearity. We have therefore first judged energy variations for different distances between D and A atoms both in optimized and unoptimized geometry from RHF/6-31G (2d,2p) basis set calculations. Comparison between optimized and unoptimized geometries gives a relative idea about energy improvement upon pyramidalization of the amino groups from conventional planar geometry. The variation in energy with D—A distances yield a minimum energy point only for a stabilized geometry, such as hydrogen-bond formation, as shown earlier by semiempirical calculation[4] irrespective of the blue-shift or red-shift nature of H-bonds.[37] In this study, we have observed such deflection points in energy curves for N_p—H···O (inter), N_p—H···O (intra), N_s—H···N_i, N_p—H···N_p, and C—H···O (AA) systems, out of which only two are shown in Figure 1. The N_p—H···O (inter) and N_p—H···O (intra) systems show the most stable geometry at a N···O separation of 2.9 and 3.6 Å, respectively. Minimum energy geometries in the N_s—H···N_i and N_p—H···N_p systems are observed at 3.6 and 3.7 Å of N···N separations, respectively. The C—H···O (AA) system is energetically most stable at 3.5 Å of the C···O separation.

The presence of the minima on the energy curve only provides a guideline to the presence of the hydrogen-bonding possibility and also confirms that these H-bonds are naturally long. This possibility can be quantitatively measured from the energy depth of the most stable optimized structure, including BSSE

Table I   Description of Different Systems Studied to Emulate the Cross-Strand Hydrogen Bond in DNA Double Helices[a]

Type of H-Bond	Explanation of Atom Names	Source Structure	Residues	Donor and Acceptor Atom Names
N_p—H···O (inter)	N_p: primary amino nitrogen	PDB[33]: 1EN3	9 and 11	Donor: N4 (Cyt) Acceptor: O6 (Gua)
N_p—H···O (intra)	—	PDB: 460D	10 and 11	Donor: N2 (Gua) Acceptor: O2 (Cyt)
N_s—H···N_i	N_s: secondary amino nitrogen N_i: imino nitrogen	MOLDEN[32] generated model	Gly and His	Donor: N (Gly) Acceptor: $N\varepsilon$ (His)
N_p—H···N_p	N_p: primary amino nitrogen	B-DNA fiber model[34]	1 and 3 of d(AT)·d(AT)	Donor: N6 (Ade) Acceptor: N6 (Ade)
C—H···O (AA)	—	B-DNA fiber model	1 and 3 of d(AA)·d(TT)	Donor: C2 (Ade) Acceptor: O2 (Thy)
C—H···O (GA)	—	B-DNA fiber model	2 and 4 of d(GA)·d(TC)	Donor: C2 (Ade) Acceptor: O2 (Cyt)

[a] Most stable conformations of each system are shown here. PDB: Protein Data Bank.

corrections. The BSSE-corrected interaction energies using the HF/6-31G(2d,2p) basis set (GAMESS does not support BSSE calculation using DFT technique) with respect to the sum of component base energies for the most stable structures of N_p—H···O (inter), N_p—H···O (intra), N_s—H···N_i, and C—H···O (AA) systems are −9.73, −4.48, −1.06, and −0.25 kcal/ mol, respectively. We have not carried out the expen-
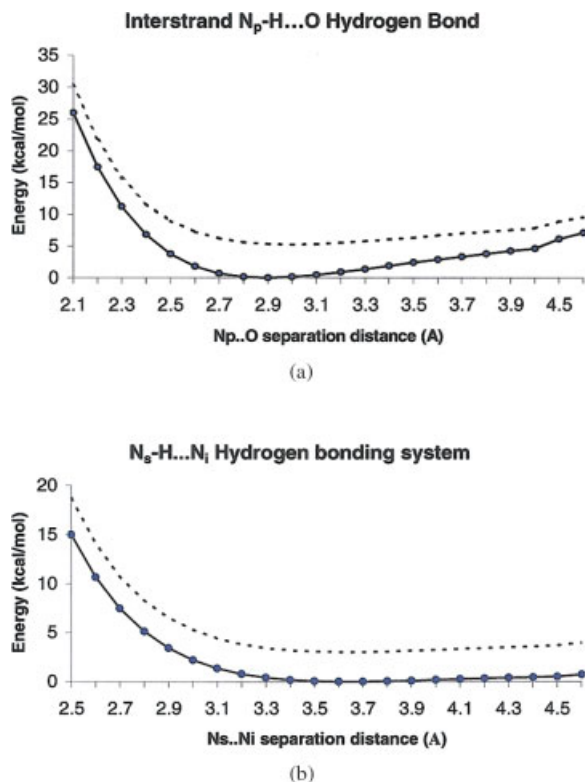
FIGURE 1 Variations of total energy as a function of distance between hydrogen-bonding nonhydrogen atoms in two representative systems, viz. (a) where the energy profile shows sufficient stabilization at a hydrogen-bonding distance and (b) where energy remains nearly constant over a range of distances. The solid lines with points represent energies from optimized geometries and corresponding energies with planar hydrogens for each structure are shown in the dotted lines.

sive BSSE-corrected interaction energy calculation for the C—H⋯O (GA) system where no energy minimum was obtained from the interaction energy curve. Interestingly, the N_p—H⋯N_p system produces a positive energy value (Table II) for the most stable structure also. This probably indicates a lack of sufficient energy stabilization in amino–amino contact geometry, as indicated earlier.[38]

Nonplanarity of amino hydrogen atoms are maximum in the primary amino group, it is profoundly reduced in secondary amino group, and finally, nonexistent in the hydrogen atom attached to the sp^2-hybridized ring carbon atom (Table III). Nonplanar movement of hydrogen atoms in an optimized structure brings these atoms closer to the hydrogen-bond acceptor (nitrogen or oxygen) atoms in the opposite base (H-bond acceptor), facilitating a shorter hydrogen-bonding distance and larger H-bond angle in EWC base pairs (Table II). These decreases in H⋯O

and the associated increases in D—H⋯A are most pronounced in the N_p—H⋯O (inter) and N_p—H⋯O (intra) systems. The N_p—H⋯N_p H-bonded system also exhibits a large decrease in the H⋯N distance and an increase in the N—H⋯N angle. However, a reverse effect, i.e., an increase in H⋯N_i and a decrease in the D—H⋯A angle is observed for the N_s—H⋯N_i system. As expected, such changes are very small in the C—H⋯O system because the hydrogen atom cannot move out of the plane formed by the conjugated ring atoms.

Study of IR Spectra for H-bonds in Watson–Crick and EWC Bases

Hydrogen-bond formation is most unambiguously determined by analysis of IR frequency and intensity changes in nucleic acids, and such changes have been found to be in very good agreement with the theoretically derived spectra.[39] As it is impossible to obtain experimental IR spectra of these EWC H-bonded complexes due to geometry restraints, we have analyzed the theoretical spectra using DFT studies. However, Raman spectra of poly(dA)·poly(dT) and poly(dA-dT)·poly(dA-dT) sequences in the premelting region have suggested the possibility of red-shift cross-strand N—H⋯O H-bonds.[40] It has been demonstrated recently that the increase in the basis set beyond the one we have used does not improve agreement between theoretical and experimental data.[18] We were, further, restricted to using the numerical Hessian method because of positional constraints of the heavy atoms. Hence, we have first compared the analytical and numerical Hessian methods for the two Watson–Crick base pairs (Table IV). The red shift of N—H-stretching frequencies in both A:T and G:C pairs is accompanied by a sharp increase in the I/I_o ratio (I_o is the intensity of a normal-mode vibration in the monomer and I is the intensity of the same vibration in complex). Hydrogen-bond formation involving amino H atoms is also sensitive to changes in the out-of-plane bending mode (or wagging) of the X—Y—N—H angle, i.e., amide V band (we shall refer it as ν_{wag}).[41] It is noteworthy that wagging motion in amino hydrogen atoms is only observed in individual bases, which disappears in H-bonded Watson–Crick base pairs. The hydrogen atoms involved in H-bond formation adopt positional restriction, which causes a blue shift in frequency of the out-of-plane hydrogen bending and concomitant reduction in its intensity (all the I/I_o values are less than one for out-of-plane bending motion of hydrogen atoms; Table IV). In numerical Hessian calculation, the 2-amino group of guanine

Table II Estimating Geometrical and Electronic Properties at the Minima of the Potential Energy Curve $r_{(D—A)}$ in Six Different Systems to Characterize the Formation of H-Bonds[a]

Systems	$r_{(D—A)}$ (Å)	$\Delta r_{D—H}$ (mÅ)	$r_{H\cdots A}$ Optimized (Å)	$\Delta r_{H\cdots A}$ (Å)	$\langle D—H\cdots A\rangle$ Planar (°)	$\langle D—H\cdots A\rangle$ Optimized (°)	ΔE (kcal/mol)	qD	qH	qA	qCT	$n(\sigma^*_{DH})$	$E(n—\sigma^*)$ (kcal/mol)	Pol (σ_{DH}) N%; H%
N_p—H···O (inter)	2.9	5.95	2.138	−0.440	98.25	130.29	−9.73	−0.801 (−0.826)	0.446 (0.427)	−0.652 (−0.486)	0.015	0.026 (0.011)	1.8, 2.65[b]	73.21 (71.72) 26.79 (28.28)
N_p—H···O (intra)	3.6	1.45	2.771	−0.262	116.15	139.41	−4.48	−0.830 (−0.855)	0.445 (0.428)	−0.651 (−0.632)	0.006	0.013 (0.008)	0.06	72.63 (71.59) 27.37 (28.41)
N_s—H···N	3.6	−2.07	3.241	−0.011	104.25	102.86	−1.06	−0.642 (−0.651)	0.429 (0.429)	−0.491 (−0.485)	0.003	0.017 (0.017)	—	71.99 (71.93) 28.01 (28.07)
N_p—H···N_p	3.7	0.73	3.188	−0.237	66.51	112.95	0.15	−0.816 (−0.823)	0.432 (0.423)	−0.826 (−0.815)	0.001	0.012 (0.011)	0.12	72.02 (71.50)
C—H···O (AA)	3.5	0.41	2.767	−0.020	122.92	124.58	−0.25	0.251 (0.216)	0.222 (0.227)	−0.630 (−0.626)	0.002	0.029 (0.020)	0.12, 0.28[c]	61.62 (61.67) 38.38 (38.33)

[a] Minus signs indicate decrease in parameters during H-bond formation. The ΔE values are BSSE corrected interaction energies. The last seven columns are the results obtained from NBO analysis on EWC systems. All the values in parentheses are from monomers, which constitute the H-bonded complexes. The "q" denotes natural charges on respective atoms. The "qCT" is the net charge transfer between H-bonded components (determined by natural population analysis). The n denotes the population in each orbital). "Pol" is extent of polarization in respective bond.

[b] $\pi_{C=O}$–$\sigma^*_{N—H}$·

[c] $\sigma_{C=O}$–$\sigma_{N—H}$·

Table III   Improper Dihedrals in Different Systems at the Most Stable Structures Around the Donor (D) and Hydrogen (H) Bonds[a]

System	Angle	B3LYP/6-31G(2d, 2p) Improper	B3LYP/6-31G(2d, 2p) pyr Method[16,20]
N—H⋯O (inter) (2.9 Å)	H41—H42—N4—C4	−135.9°	39.71°
N—H⋯O (intra) (3.6 Å)	H21—H22—N2—C2	−136.6°	38.70°
N—H⋯N1 (3.6 Å)	H—Cα—N—C′	178.9°	−0.96°
N—H⋯N2 (3.7 Å)	H61—H62—N6—C6	−147.64°	28.25°
	H61—H62—N6—C6	−147.11°	28.01°
C—H⋯O (AA) (3.5 Å)	H2—C2—N3—N1	−178.6°	0.5°

[a] Values obtained from unambiguous method to evaluate pyramidalization are shown in the last column. Improper dihedral angle is 180° before optimization in all the systems.

Table IV   Frequency Δν and Intensity Changes (I/I_0) Observed in Stretching and Wagging Motions Due to Hydrogen-Bond Formation in Watson–Crick Base Pairs with Respect to Their Individual Components[a]

System	Mode of Vibration[b]	Free Optimization Δν	Free Optimization I/I_0	Constraint Optimization Δν	Constraint Optimization I/I_0
Watson–Crick H-bonding systems					
A:T H-Bonds	N6—H61 stretching	−58	6.47	−10	6.49
Involving	N3—H3 stretching	−308	13.10	−351	20.93
(i) N6—H61	Out-of-plane bend	+62	0.35	+88	0.75
(ii) N3—H3	H62 (Amino wagging)				
G:C H-bonds	N2—H21 stretching	−67	14.55	−137	17.15
Involving	N4—H41 stretching	−207	4.27	−209	9.48
(i) N2—H21	N1—H1 stretching	−204	15.26	−242	12.03
(ii) N4—H41	Out-of-plane bend H22 (Gua amino wagging)	−328	0.84	−282	0.63
(iii) N1—H1	Out-of-plane bend H42 (Cyt amino wagging)	+72	0.13	+43	0.75
EWC H-bonding systems					
N$_p$—H⋯O (inter)	N4—H41 stretching			−43	11.26
	Out-of-plane bend H41 (Amino wagging)			150	4.25
N$_p$—H⋯O (intra)	Asymmetric amino stretching[c] (Symmetric amino stretching and N1—H1 stretching)			—	—
	Amino wagging + Out-of-plane bend H1			20	1.48
N$_s$—H⋯N$_i$	N—H stretching			87	0.9
	Out-of-plane bend H			—	—
N$_p$—H⋯N$_p$	N6—H61 stretching[c] (symmetric and asymmetric amino stretching)			—	—
	Amino wagging			6	1.70
C—H⋯O (AA)	C2—H2 Stretching			—6	0.14
	Out-of-plane bend H2			24	2.20

[a] Frequencies are reported in cm^{-1}. Negative values indicate red shift in frequency and positive values correspond to blue shift. Primary amino hydrogen atoms marked with 1 are involved in hydrogen bond (e.g., H61, H21) formation.
[b] Vibrational modes in the parentheses belong to individual monomers, which differ from those in complexes.
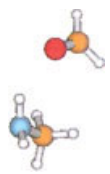[c] Two different modes of vibration appear in monomer and complex, which are not comparable.

shows a rather unusual conversion of wagging motion in the isolated state ($\nu = 580$ cm^{-1}) to the twisting motion in the G:C pair ($n = 640$ cm^{-1}), along with conversion of twisting motion in the isolated state ($\nu = 351$ cm^{-1}) to wagging motion in G:C pair ($\nu = 297$ cm^{-1}). A similar feature was also noted in analytical Hessian calculations for this base pair. These comparisons indicate that frequencies and intensities obtained from numerical Hessian calculations are very similar to those obtained from analytical Hessian and hence, in turn, to the experimental values.

In the case of the inter N—H···O EWC system, a 43-cm^{-1} red shift is observed in the N4—H41-stretching frequency with an eleven-fold increase in intensity in the complex with respect to those observed in the monomer; this data is consistent with the observed increase in N—H bond length (Table II). The situation becomes complicated in N_p—H···O (intra) and N_p—H···N_p systems where comparison of N—H bond stretching is not straightforward as the same mode of vibration does not exist in the monomer and the complex (Table IV). Particularly, in the N_p—H···O (intra) system there is no way to cross-examine the N4—H42 bond stretching from IR data—hence we shall only rely on the geometry of the system. The N_s—H···N_i system shows a large blue shift in N—H stretching frequency with no detectable out-of-plane bending of H atom in N—H bond. The C—H···O (AA) system shows a red shift by 6 cm^{-1} in C—H bond stretching accompanied by diminished intensity value with respect to monomer. The red shift of C—H bond correlates with bond length increase but very small I/I_o ratio does not indicate strong H-bond.

Assessment of H-Bonds in Different Systems on the Basis of NBO Analysis

The results obtained from geometrical and IR data lead us to screen the possible H-bonds and to proceed further toward their characterization in EWC systems. In this section, we mainly concentrate on population changes due to H-bond formation as obtained from NBO analysis (Table II). Charge transfer hyperconjugative $n(Y)$–σ^*(D—H) interactions, as revealed by NBO analysis, are of critical importance in H-bonded complexes, although sometimes it is counterbalanced by the Pauli repulsion (as a result of occupied–occupied orbital interaction) when orbital interaction is very small, as in case of a improper blue-shift H-bond.[37] Here, σ^*(D—H) is a Lewis (hyperconjugative) acceptor or Lewis acid and a $n(Y)$ is Lewis (hyperconjugative) donor or Lewis base.[42] Other analytical methods to dissect H-bonds also rely on overall charge

Table V   NBO Property Changes in Formaldehyde–Methyl Amine Complex; Model to EWC Inter N—H···O H-Bonded System (Values in Parentheses Correspond to Those in Monomers)

Parameters Analyzed	
$E_{n-\sigma}$ (kcal/mol)	1.66
q_{CT}	0.00028
$n(\sigma_{N—H})$	1.987 (1.990)
$n(\sigma^*_{N—H})$	0.019 (0.011)
sp^n (N—H)	2.70 (2.81)
%s on N	26.96 (26.21)
pol ($\sigma_{N—H}$) N%	70.17 (69.68)
H%	29.83 (30.32)
q_N	−0.919 (−0.920)
q_H	0.391 (0.384)
q_O	−0.512 (−0.496)

transfer from acid to base in order to determine H-bond formation.[43] Thus, the energies $E_{(n-\sigma^*)}$ of hyperconjugative interactions[44] are used as a measure to judge the H-bond formation. In the case of a bent EWC H-bond in the N_p—H···O (inter) system, this value is as small as 1.8 kcal/mol. However, when symmetry-allowed $\pi_{C=O}$–$\sigma^*_{N—H}$ hyperconjugative interaction is considered, the total energy associated with the charge transfer to the $\sigma^*_{N—H}$ orbital (4.46kcal/mol) is within the 3–5 kcal/mol threshold for red-shift H-bond formation.[42] The net charge transfer qCT (0.015 au) is also in the range of red-shift H-bonds, as described earlier in small molecular complexes (Table 5 of Ref. 37). The net increase in the $\sigma^*_{N—H}$ population after complex formation is significant (0.015 au). The $\sigma_{N—H}$ bond is more polarized in the complex system compared to the monomer (i.e., percentage of electron density on hydrogen atom decreases with an associated increase on N atom), which indicates possible formation of an H-bond in the complex.[45] The $\sigma_{N—H}$ bond polarization leads to an increase in positive charge on H, a decrease in negative charge on the donor atom (N), and increase in negative charge on the acceptor atom (O), which is in agreement with earlier observations during H-bond formation.[42] The changes in natural charges[46] in donor (D), hydrogen, and acceptor (A) atoms of different EWC systems are summarized in Table II. An increase in charge on donor H atom upon H-bond formation was also noted earlier to explain dihydrogen bond.[47]

To understand the effect of delocalization through an aromatic ring on H-bond formation, we have modeled a smaller system (replacing cytosine by methylamine and guanine by formaldehyde) with the same optimized H-bonded geometry (N—H distance: 2.138 Å; N—H···O angle: 130.29°; and amino pyramidalization: 39.71°). The results from NBO analysis for this model system (Table V) is comparable to that of the EWC N—H···O system, except in a few cases, such as the decrease of net charge transfer between the individual monomers (0.00028) and reduction in hyperconjugative energy, which corresponds only to the $n_{(Y)}-\sigma^*_{(N—H)}$ interaction with zero $\pi_{C—O}-\sigma^*_{(N—H)}$ interaction. The last two observations suggest that H-bond formation in the EWC system is favored by the delocalization of charge over the aromatic ring. The interesting point to note here is that the EWC H-bond is not merely confined within donor–hydrogen–acceptor atoms but it participates in delocalizing the charge over all the conjugated atoms. The most pronounced effect of charge reorganization in the complex is observed on the polar oxygen and nitrogen atoms of guanine and remote carbonyl oxygen on cytosine (Figure 2a), which can only be interpreted in terms of charge delocalization during H-bond formation. This delocalization of the $\sigma^*_{(N—H)}$ population over the molecule resists the N atom to rehybridize its $\sigma_{(N—H)}$ orbital to increase the s-character in the EWC H-bond. On the other hand, in the model formaldehyde–methylamine complex, the nitrogen atom reorganizes its hybridization state from sp$^{2.71}$ (monomer) to sp$^{2.66}$ (complex) increasing slightly its percentage in s-character as observed previously.[42] The red-shift nature of EWC inter H-bond can be reinforced from the fact that no change in hybridization of the N—H bond was detected upon complex formation. Weinhold and coworkers[42] showed that for improper H-bonding to occur, X—H bond should undergo significant rehybridization. All the above effects along with the geometrical changes such as N—H bond lengthening in the complex with 43 cm^{-1} red shift in IR stretching frequency, etc., can attribute red-shift or true H-bond nature to N—H···O in EWC inter G—C complexes.

In the intrastrand EWC complex, the minimum on the potential energy curve (at 3.5 Å of N···O separation) corresponds to a large H···O separation (2.77 Å), which is sufficient to overcome the Pauli repulsion caused by the occupied–occupied orbital interaction—thus accounting for the N4—H42 bond lengthening (Figure 1b). In our EWC intra system, it is the large N···O separation at potential minima that helps to overcome the repulsion though the orbital interaction is not so strong [vide small qCT and very small $E(n-\sigma)$ value in Table II]. The increase in the N4—H42 distance in intra EWC N—H···O system might be a consequence of a decrease in the N4—H41 distance (1.0088 Å in the complex and 1.0106 Å in the monomer) arising out of the geometrical proximity of H41 and H1 atoms in the complex (2.31 Å) with respect to the monomer (2.39 Å). However, the change in the $\sigma^*_{N4—H42}$ population is significant. The changes in natural charges on donor (N), acceptor (O), and hydrogen atoms and the polarization on the $\sigma^*_{N4—H42}$ bond are also in the desired direction towards the formation of the H-bond in the complex. Moreover, the total interaction energy in the N—H···O EWC intra G—C complex, −4.48 kcal/mol, is within the specified range of H-bonds[25] and it is also comparable to that of HOH···OH$_2$ interactions (−4.80 kcal/mol; with B3LYP/6-31+G** basis set).[48] The red-shift H-bond formation in the EWC intra G—C complex cannot be discarded, even considering the limitation of the poor hyperconjugative $n_{(Y)}-\sigma^*_{N—H}$ interaction energy. Results of H$_2$···Y (Y = OH$_2$, OMe$_2$, and Cl$^-$) systems studied earlier indicated red-shift H-bonds despite their poor hyperconjugative interaction energy values, which were explained in terms of the incapability of the H—H bond to rehybridize.[42] Similarly, the donor N atom in the EWC intra G—C system remains almost invariant towards its rehybridization in the complex (sp$^{2.7}$) with respect to the monomer (sp$^{2.74}$), which might be a consequence of charge delocalization over the whole system (Figure 2b), and this complex shows a significant increase in the N—H bond length. Therefore, the N—H···O (intra) EWC can be characterized as a red-shift H-bond though it is rather weak as compared to normal Watson–Crick H-bonds.[49]

The possibility of H-bond formation in N_s/N_p—H···N_i/N_p complexes, here studied, are ruled out on the basis of interaction energy, geometry, and orbital interaction. In the C—H···O (AA) system, C—H bond lengthening (0.0041 Å) and associated decrease in stretching frequency (6 cm^{-1}) can be explained in terms of smaller repulsive interactions due to a large separation distance between H-bond donor and acceptor atoms, similar to that in EWC intra G—C complex. Net charge transfer, energy change in hyperconjugative interaction $E(n-\sigma^*)$, population increase in $\sigma^*_{C—H}$ orbital all are small in magnitude and quite consistent with very small interaction energy value (−0.25 kcal/mol). From these data it is not clear whether C—H···O (AA) can be designated as a true red-shift H-bond or not. This interaction cannot even be supported as a blue-shift H-bond from the viewpoint of C—H bond stretching, red shift in bond stretching frequency and decrease in the percentage of s-character on the carbon atom of $\sigma_{C—H}$ bond in the complex (sp$^{2.18}$; % s-character 31.45) with respect to that in the monomer (sp$^{1.81}$; % s-character 35.52).
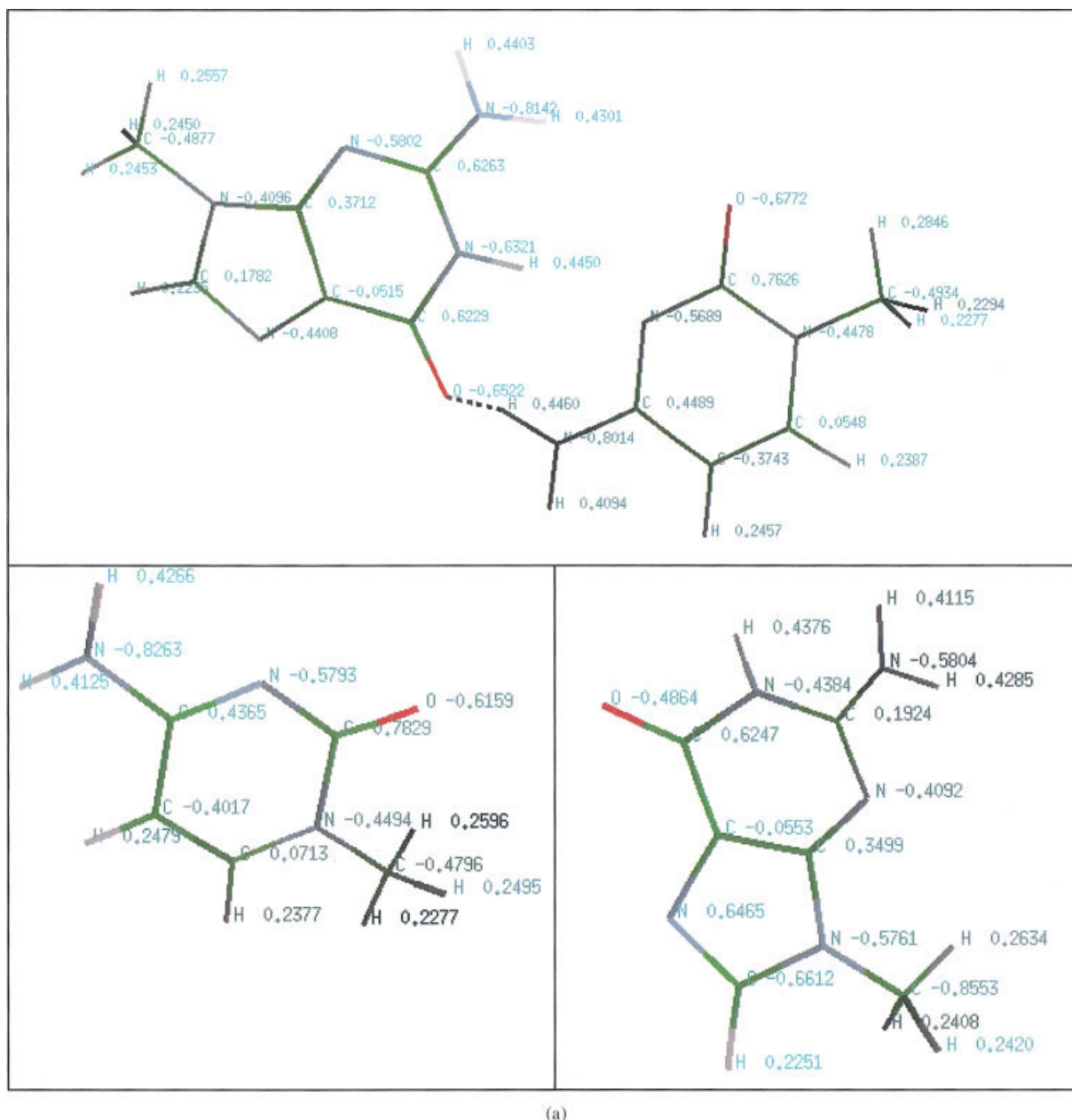
FIGURE 2 The Natural Charges[46] on each of the atoms are shown for (a) N—H⋯O (inter) EWC H-bonded systems and (b) in N—H⋯O (intra) EWC H-bonded systems. Upper panels in (a) and (b) are for the complex systems while the charges on their individual monomers are shown in the lower panels: Cytosine is at the left in the lower panel and guanine is at the right in the lower panel of both (a) and (b).

CONCLUSION

Stability of the DNA structure stands on the strength of Watson–Crick H-bonds. Additional cross-strand or same-strand bifurcated H-bonds, which we have determined here as EWC H-bonds, also exist in DNA structures. These EWC H-bonds, due to their geometrical constraints in the double-helical context, deviate from normal linear H-bond geometry with slightly elongated A⋯H bond lengths, which make them weaker in strength. Our examination of different properties has established significant strengths of the N—H⋯O H-bond across (inter) and along (intra) the double-helical strands. The importance of these H-bonds was already realized in order to explain the stability of different DNA sequences, which were only described in terms of geometrical criteria. Here we have succeeded in characterizing them as H-bonds,
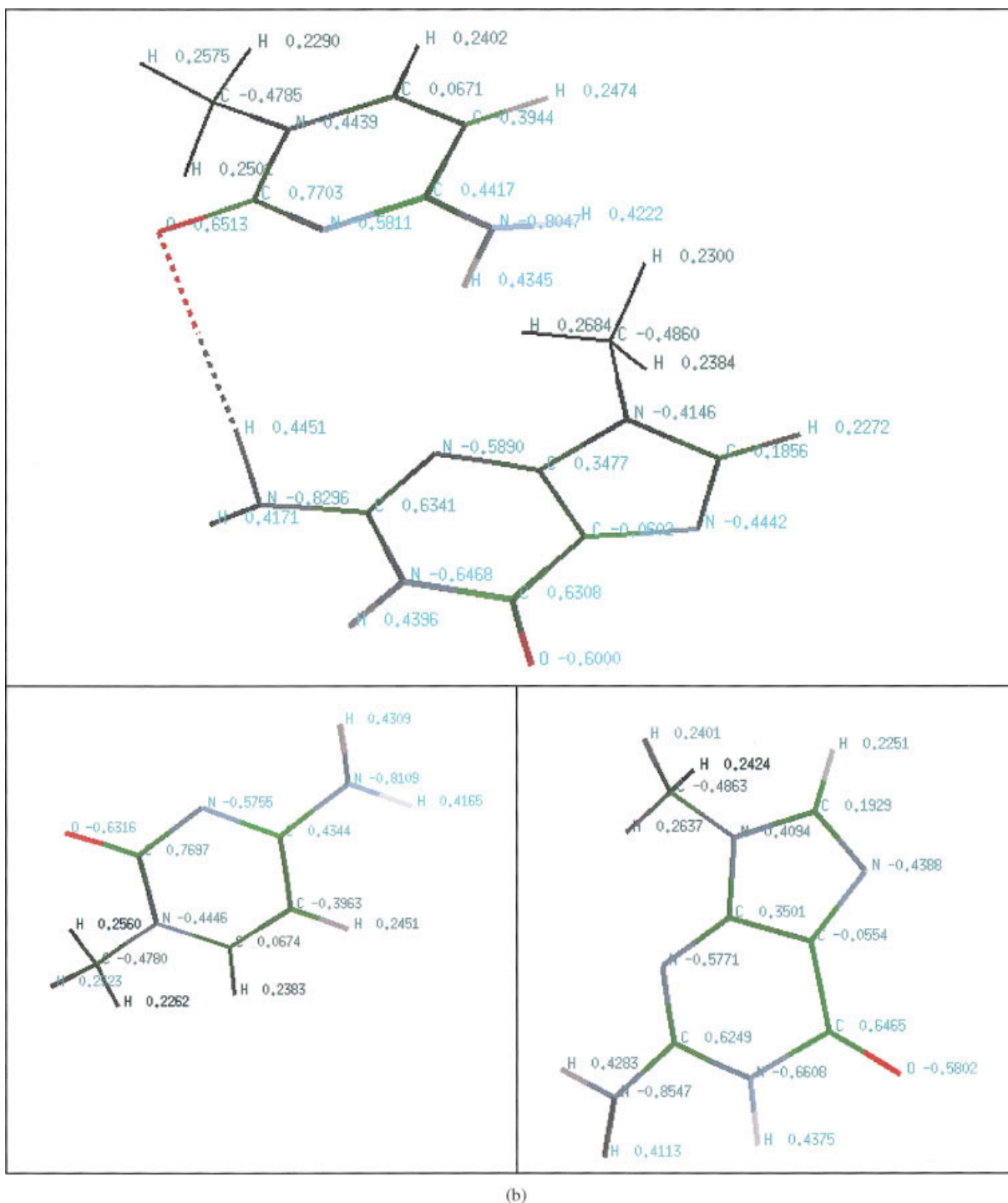
(b)

FIGURE 2 (*Continued from the previous page.*)

distinguishing them from mere electrostatic attractions. However, systems like N_p—H···N_p, N_s—H···N_i and C—H···O exhibit an insignificant amount of property changes, which makes them too insensitive towards their characterization as H-bonds. The more realistic way to analyze these EWC H-bonds is in the proper double-helical DNA context, which is quantum mechanically tedious, unless one treats the reaction center using quantum mechanics and the rest of the atoms by molecular mechanics. Considering the

strengths of the six different types of hydrogen bonds, it appears that N—H···O (inter) hydrogen bonds are the strongest and may possibly significantly alter the structure and rigidity of some dinucleotide steps. Such hydrogen bonds are possible in the major groove of d(AA)·d(TT), d(GG)·d(CC), d(AC)·d(GT), and d(CA)·d(TG) steps as well in the minor groove of d(GG)·d(CC) and d(AG)·d(CT) steps. It was earlier postulated that these H-bonds may be the major factor in sequence-directed DNA structural rigidity which abstain poly(dA)·poly(dT) or poly(dG)·poly(dC) sequences form nucleosome formation.[4,6] In the purine–purine steps, these hydrogen bonds can be stronger due to the equivalent size of the stacked bases, while in the mixed steps, such as d(AC)·d(GT), these may demand modified translational parameters, such as shift or slide. Presumably d(CA)·d(TG) steps do not generally adopt regular B-DNA-like structure, with near-zero inter base-pair translational parameters, in order to facilitate such H-bonds. They are hence more flexible. Obviously, there are some other competing factors also, such as hydrogen bonds between the bases and spine of hydration, entropic factor, etc., but their analysis is beyond scope of the present study.

REFERENCES

1. Steiner, T.; Saenger, W. J Am Chem Soc 1993 115, 4540–4547.
2. Nelson, H. C. M.; Finch, T.; Luisi, B. R.; Klug, A. Nature 1987, 330, 221–226.
3. Mohan, S.; Yathindra, N. J Biomol Struct Dynam 1991, 9, 113–126.
4. Bhattacharyya, D.; Kundu, S.; Thakur, A. R.; Majumdar, R. J Biomol Struct Dyam 1999, 17, 289–300.
5. Bacolla, A.; Gellibolian, R.; Shimizu, M.; Amirhaeri, S.; Kang, S.; Ohshima, K.; Larson, J. E.; Harvey, S. C.; Stollar, B. D.; Wells, R. D. J Biol Chem 1997, 272, 16783–16792.
6. Bandyopadhyay, D.; Bhattacharyya, D. J Biomol Struct Dynam 2002, 19, 659–667.
7. Austin, R.; Breslauer, K.; Chan, S.; Hogan, M.; Kessler, D.; Ojemann, J.; Passner, J.; Wiles, N. C. Biochemistry 1990, 29, 6161.
8. Premilat, S.; Albiser, G. J Mol Biol 1997, 274, 64–71.
9. Chan, S. S.; Austin, R. H.; Mukherji, I.; Spiro, T. G. Biophys J 1997, 72, 1512–1520.
10. Suhnel, J. Bioploymers (Nucleic Acid Sci) 2002, 61, 32–51.
11. Meyer, M.; Steinke, T.; Brandl, M.; Suhnel, J. J Comp Chem 2001, 22, 109–124.
12. Meyer, M.; Brandl, M.; Suhnel, J. J Phys Chem A 2001, 105, 8223–8225.
13. Seeman, N. C.; Rosenberg, J. M.; Suddath, F. L.; Kim, J. J. P.; Rich, A. J Mol Biol 1976, 104, 1090–1144.
14. Sponer, J.; Kypr, J, Int J Biol Macromol 1994, 16, 3–6.
15. Sponer, J.; Burel, R; Hobza, P. J Biomol Struct Dynam 1994, 11, 1357–1376.
16. Mukherjee, S.; Majumder, S.; Bhattacharyya, D. J Phys Chem B 2005, 109, 10484–10492.
17. Danilov, V. I.; Anisinov, V. M. J Biomol Struct Dynam 2005, 22, 471–482.
18. Dong, F.; Miller, R. E. Science 2002, 298, 1227–1230.
19. Dasgupta, A. K.; Majumdar, R.; Bhattacharyya, D. Ind J BioChem Biophys 2004, 42, 49–55.
20. Bhattacharyya, D.; Sen, K.; Mukherjee, S. Ind J Chem 2006, 45A, 58–67.
21. Sutor, D. J. J Chem Soc 1963, 1105–1110.
22. Sundaralingam, M. Acta Cryst 1966, 21, 495–508.
23. Wahl, M. C.; Rao, S. T.; Sundaralingam, M. Nature Struct Biol 1996, 3, 24–31.
24. Ghosh, A.; Bansal, M. J Mol Biol 1999, 294, 1149–1158.
25. Desiraju, G.; Steiner, T. The Weak Hydrogen Bond in Structural Chemistry and Biology; Oxford University Press: New York, 1999.
26. Sarkhel, S.; Desiraju, G. R. Proteins Struct Funct Genet 2004, 54, 247–259.
27. Wahl, M. C.; Sundaralingam, M. TIBS 1997, 22, 97–102.
28. Leonard, M.-H.; Hunter, B. Acta Cryst D 1995, 51, 136–139.
29. Starikov, B.; Steiner, T. Acta Cryst D, 1997, 53, 345–347.
30. Hobza, P.; Sponer, J.; Cubero, E.; Orozco, M.; Luque, F. J. J Phys Chem B 2000, 104, 6286–6292.
31. Hobza, P. Phys Chem Chem Phys 2001, 3, 2555–2556.
32. Schaftenaar, G.; Noordik, J. H. J Comput-Aided Mol Design 2000, 14, 123–134.
33. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliard, G.; Bhat, T. N.; Weissig, H.; Shindyalov, N.; Buorne, P. E. Nucleic Acids Res 2000, 28, 235–242.
34. Arnott, S.; Chandrasekhar, R. J Biomol Struct Dynam 1996, 13, 1015–1027.
35. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jenson, J. J.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A., Jr. J Comput Chem 1993, 14, 1347–1363.

36. Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F. NBO 3.1.

37. Sponer, J. R.; Leszczynski, Hobza, P. J Biomol Struct Dynam 1996, 14, 117–136.

38. Li, X.; Liu, L.; Schlegel, H. B. J Am Chem Soc 2002, 124, 9639–9647.

39. Santamaria, R.; Charro, E.; Zaccarias, A.; Castro, M. J Comp Chem 1999, 20, 511–530.

40. Movileanu, L.; Benevides, J. M.; Thomas, G. J., Jr. Biopolymers 2002, 63, 181–194.

41. Rozenberg, M.; et al. Spectrochimia Acta A 2003, 59, 3253–3266.

42. Alabugin, I. V.; Manoharan, M.; Peabody, S.; Weinhold, F. J Am Chem Soc 2003, 125, 5973–5987.

43. Koch, U.; Popelier, P. L. A. J Phys Chem 1995, 99, 9747–9754.

44. Reed, A. E.; Curtiss, L. A.; Weinhold, F. Chem Rev, 1988, 88, 899–926.

45. Liu, S. Y.; Dykstra, C. E. J Phys Chem 1986, 90, 3097–3103.

46. Reed, A. E.; Weinstock, R. B.; Weinhold, F. J Chem Phys 1985, 73, 735–746.

47. Popelier, P. L. A. J Phys Chem A 1998, 102, 1873–1878.

48. Gu, Y.; Kar, T.; Scheiner, S. J Am Chem Soc 1999, 121, 9411–9422.

49. Gould, I. R.; Kollman, P. A. J Am Chem Soc 1994, 116, 2493–2499.

Reviewing Editor: J. A. McCammon