NJC

From Fluorine Chemistry to Noncovalent Interactions: Celebrating Prof. Giuseppe Resnati Replacement of Hydrogen by Halogen Bonds within Nucleic Acid Base Pairs

Journal:	*New Journal of Chemistry*
Manuscript ID	NJ-ART-03-2025-000977.R1
Article Type:	Paper
Date Submitted by the Author:	23-Mar-2025
Complete List of Authors:	Amonov, Akhtam; Samarkand State University named after Sharof Rashidov, Department of Optics and Spectroscopy Scheiner, Steve; Utah State University, Department of Chemistry and Biochemistry

Replacement of Hydrogen by Halogen Bonds within Nucleic Acid Base Pairs

Dr. Akhtam Amonov[a] and Prof. Steve Scheiner[b,*]

[a]Department of Optics and Spectroscopy, Engineering Physics Institute
Samarkand State University
140104, University blv. 15, Samarkand
Uzbekistan

[b]Department of Chemistry and Biochemistry
Utah State University
Logan, Utah 84322-0300
USA

*email: steve.scheiner@usu.edu

Abstract

The effect of the replacement of the H-bonds (HBs) that hold nucleic acid base pairs together by halogen bonds (XBs) is examined by DFT calculations. When in its Hoogsteen configuration, XBs involving Cl reduce the interaction energy within the adenine-thymine (AT) pair by the most, followed by Br. I substitutions have little effect on the total binding. Similar reductions arise in the guanine-cytosine (GC) pair, although XBs involving I can enhance the interaction energy. With regard to Watson-Crick pairings, the replacement of two of the H atoms by I yields a stronger GC binding energy, while Cl and Br XBs weaken the interaction. Atoms in Molecules (AIM) analysis allows conclusions to be drawn concerning the contributions of each individual interaction, whether HB or XB, to the full energy.

Keywords: AIM; SAPT; Hoogsteen configuration; Watson-Crick configuration

INTRODUCTION

The hydrogen bond (HB) is arguably the most important and prevalent intermolecular force in all of chemistry and biology [1-8]. Our understanding of any process that occurs in water is predicated on a proper treatment of the hydrogen bonding between water molecules as well as HBs they might form with the reactants themselves. The structure and function of protein molecules is heavily dependent on the HBs between peptide groups, as well as those involving the sidechains of the various residues. Proton transfer within the confines of a HB is a prime ingredient of numerous enzymatic reactions [9-15]. The transmission of genetic information relies on the complementary H-bonding between nucleic acid bases and their proper pairing with one another. It is this latter phenomenon that has motivated scores of studies to better understand the details of the forces within these base pairs [16-20]. Questions that have been addressed include the overall strength of the bonding, how it is affected by the specific geometry of the pair, and how this bonding pattern resists mispairing. There has also been examination of the hypothesis that proton transfer within a base pair which would lead to a different set of tautomers might represent a form of mutation [21-23]. The recent developments that have caused researchers to rethink the possibility of CH··O HBs in biological contexts [24-28] have led to questioning whether such a bond might contribute to the bonding within the guanine-cytosine (GC) pair [29].

The last few decades have spawned a renaissance in the idea that a conventional HB has a number of closely related parallels [30-43]. The oldest and most widely studied of these interactions is the halogen bond (XB), whereby the central proton of the AH··D HB is replaced by a member of the halogen family, viz. Cl, Br, or I. Although the electronegative X does not have the benefit of an overall positive charge to attract a nucleophile as does the H within a HB, there is a small confined region around X that is positive. This so-called σ-hole lies along the extension of the A-X covalent bond, and is surrounded by an equator or negative potential, imparting to the XB a high preference for a linear configuration. Another factor stabilizing the XB is the transfer of charge from the approaching nucleophile's lone pair into the antibonding σ*(AX) orbital, in close parallel to the same charge transfer that contributes to a HB, which is in turn responsible for the well known red shift of its AH stretching frequency.

XBs are rather common in a wide range of biological systems, as are some of their closely related cousins such as chalcogen and pnicogen bonds [44-55]. Given the close similarities between HBs and XBs, it is natural that investigators have begun to study various phenomena where the former might be replaced by the latter. As one example, Yan et al found [56] that halogen bonds, specifically those between I and S, could promote a double helix structure, akin to that adopted by DNA. More recently, Lin et al used chalcogen bonding to assemble mimics of the β-sheet and α-helix structure so common to proteins [57].

Evolution has of course used H-bonding between bases as a means to guarantee the security of the genetic code within DNA. But the many similarities between the hydrogen and halogen bonding phenomena lead to the question as to what would happen if the former were replaced by the latter. Would there still be a highly complementary base pairing? How would such

substitutions affect the strength of the interaction between the bases? And given the fact that there are more than one HB within each base pair, what would be the effect of switching out one or the other, or perhaps all of them? And would such modified bases, with the larger halogen atoms, still be able to fit neatly into the double helix?

Although work on this question is still meager, there has been some activity in this direction. Seela et al [58] found that certain halogenated nucleotides enhanced the stability of parallel and antiparallel DNA, and better mismatch discrimination. Tawarada et al [59] evaluated the contribution of halogen bonding toward duplex stabilization, but the specific molecules considered were quite different from the normal bases. The way in which the various aspects of the normal base pairs might be influenced by the replacement of their HBs by halogen bonding was first addressed directly via quantum chemical calculations in 2012 [60]. Parker et al considered the Watson-Crick orientations of the G-C and A-T pairs. They first replaced all relevant bridging H atoms by X, and then reoptimized the geometries. In all cases, even with I which forms quite strong XBs, the pairing energy suffered a heavy deterioration, to only a fraction of that within the unsubstituted pairs. This energy also suffered if only one H is replaced at a time, leaving the remaining HBs in place. So the overall conclusion had been that halogenation weakened the interaction between base pairs in all cases.

However, the earlier work left some important questions unresolved. First with regard to the Watson-Crick configurations, the earlier work did not consider the possibility of replacing two of the three bridging H atoms by X within the GC dyad. The work described below shows how judicious choice of H→X mutations can actually strengthen the interaction. And in a wider perspective, there has been no assessment of how the replacement of H-bonding by XBs might influence the base-pair binding within the context of Hoogsteen pairing, another commonly observed motif. The quantum chemical calculations described below comprise a detailed evaluation of these substitutions for both the GC and AT pairs, considering all permutations of these replacements, and document several such combinations where the halogen-bonded structures are more strongly bound than their HB analogues. Another issue addressed here concerns the effects of placing a charge on one of the bases; how does such charge amplification of these bonds play into the comparisons between HBs and XBs. The calculations go beyond earlier work in elucidating the geometries and relative strengths of each individual HB/XB within the dyads, which are held together by multiple bonds, so as to better understand the importance of each, and how each substitution affects the geometry of its neighboring interactions. Also considered here for the first time is how well the substituted base pairs might fit into the space allotted to each pair within the double helix.

METHODS

The Gaussian 16 [61] suite of programs was employed to carry out the quantum chemical calculations, applying the DFT M06-2X functional [62] in conjunction with the triple-ζ def2-TZVP basis set. The pseudopotential placed on I by this basis takes into consideration certain relativistic effects. M06-2X has been repeatedly assessed to be one of the most accurate

functionals for H-bonding and related noncovalent interactions [63-71]. Geometries were fully optimized and characterized as true minima by harmonic vibrational analysis which yielded all positive frequencies. The interaction energy E_{int} is equal to the difference between the energy of the dyad and the sum of the energies of the two monomers in the geometry they adopt within the complex; E_{int} was corrected for basis set superposition error by the standard counterpoise prescription [72, 73]. Atoms in Molecules (AIM) bond paths and the properties of their associated critical points were evaluated via the AIMAll program [74]. Total interaction energies were decomposed into their contributing components by Symmetry-Adapted Perturbation Theory (SAPT) [75, 76] at the SAPT0 level through the PSI4 program [77].

RESULTS

1. Hoogsteen Pairing

Adenine-Thymine

The geometry of the fully optimized A-T pair in its Hoogsteen configuration is exhibited in Fig 1a. The shortest HB is the NH··N bond in its center, labeled as site 2 in Fig 1a, where R(H··N)=1.771 Å. Somewhat longer is the NH··O bond at the top, labeled site 1, with R=1.993 Å. Both of these bonds are fairly linear with θ(NH··X) angles of 174.5° and 169.4°, respectively. The geometry of the CH··O configuration at the bottom is suggestive of a HB, but one that is somewhat long and bent, what with R(H··O)=2.703 Å and θ(CH··O)=122.5°. Nevertheless, AIM analysis does confirm a CH··O HB, albeit a weak one with a density at the bond critical point of only 0.0058 au as shown in Fig 1b. The strength of this bond is dwarfed by the two conventional HBs, particularly the central NH··N HB with a ρ_{BCP} of 0.0444 au, followed by 0.0207 au for the upper and longer NH··O HB. Both geometrical and density topological features then suggest it is the central NH··N HB which dominates this Hoogsteen A-T coupling, although the upper NH··O is certainly important as well.

The effects of replacing each of the NH protons by a Cl center is illustrated in Fig 2a-2c. Fig 2a shows that placing a Cl in position 1 allows a very nearly linear NCl··O alignment, conducive to a XB. But this substitution disrupts the NH··N HB at site 2 to some extent, lengthening H··N from 1.771 to 2.056 Å, while also making this bond much less linear, cutting θ(NH··N) down to 159°. On the other hand, the realignment caused by the Cl substitution strengthens the CH··O, contracting the H··O distance by 0.5 Å. The effects upon the AIM densities in Fig 2d echo the geometric trends: The CH··O density rises from 0.0058 to 0.0153 au. The central NH··N density is cut in half by this lengthening. The XB in site 1 has a density only slightly smaller than that in the unsubstituted system. Overall, then, the major effect of this H→Cl mutation is the severe weakening of the central NH··N HB, with only minor compensation from CH··O. It is therefore not surprising that Table 1 shows that this replacement reduces the total interaction by 2.6 kcal/mol form 15.48 to 12.87 kcal/mol.

Positioning a Cl atom at site 2 has some very different effects on the geometry. For one thing, Fig 2b shows the complete dissociation of the lower CH··O HB. The NH··O HB in site 1 is weakened a bit. Its near linearity in Fig 1a is distorted to a NH··O angle of only 138°, and the

H··O distance is stretched from 1.993 to 2.218 Å. On the positive side is a favorable linear arrangement of the NCl··N XB at site 2 with a relatively short distance of 2.722 Å. The AIM diagram in Fig 2e confirms the assumptions made on the basis of geometry. There is no CH··O bond path, and the NH··O density is reduced from 0.0207 to 0.0131 au. The new NCl··N XB is reasonably strong with a density of 0.0211. Inspection of Fig 2e presents a new facet to the intermolecular bonding. There is a bond path leading from the upper NH_2 of the adenine to the Cl center of thymine. With a density of 0.0100 au, this additional NH··Cl HB must be included in the constellation of contributing intermolecular forces. Due to the general weakening of the bonds, the placement of a Cl atom at site 2 reduces the total A-T interaction energy from 15.48 to 8.63 kcal/mol.

The combined effect of substituting the H centers at sites 1 and 2 on the geometry is visible in Fig 2c. Compared to Fig 2b, the site 2 NCl··N largely maintains its integrity after this second Cl addition. On the other hand, the NCl··O XB at site 1 is stretched by 0.3 Å. The AIM diagram of Fig 2f reflects these themes. The site 2 density is largely unaltered, while that at site 1 is appreciably degraded. Consequently, this double Cl substitution leads to the most weakly bonded dyad, with an interaction energy down to 6.04 kcal/mol.

Many of these same general trends play out for substitution by Br and I. As is evident from Figs 3 and 4, replacement of the H in site 1 shortens and strengthens the CH··O HB to about 2.21 Å, and with a density of 0.015 au in all cases. At the same time, the central NH··N HB at site 2 is degraded by a stretch and imposed nonlinearity. The progressively larger X atom causes this HB to weaken accordingly, getting longer, and with its density shrinking from 0.0222 for Cl, down to 0.0156 for I. The site 1 XB is comparable in all cases, with a bond length of some 2.71 Å, and reasonably close to linear. But there is a strengthening of the XB in the sequence Cl < Br < I as the density rises from 0.0182 to 0.0264 au. The latter strengthening is sufficient to raise the interaction energy of the dyad from 12.87 kcal/mol to 15.79 kcal/mol. Note that this last quantity is actually a little larger than the 15.48 kcal/mol for the unsubstituted A-T dimer, so it would be fair to say that this I substitution at site 1 stabilizes the system to a small extent.

In all three cases, the X substitution at site 2 destroys the CH··O HB. The site 1 NH··O HB is deformed and its density sinks accordingly, to the point that AIM finds no such HB at all for X=I. In partial compensation, in all cases, there is an auxiliary NH··X XB from the site 1 NH_2 to the X atom at site 2. The density of the latter rises from 0.0100 for Cl up to 0.0142 for I. The heart of the interaction for these substituted dimers is the central site 2 NX··N XB. This bond is close to linear in all cases, which would appear to be major factor driving the geometry. And the strength of this bond grows quickly in the Cl < Br < I sequence. Even though the X atom is growing larger, the intermolecular N··X distance shrinks in this same order from 2.722 Å down to 2.550 Å for I. The BCP density confirms this growing strength, rising from 0.0211 for Cl up to 0.0427 for I. The site 2 substitutions all result in reduced interaction energies relative to those at site 1. These decreases may be attributed in large part to the near loss of the site 1 NH··O HB, as well as the complete loss of the CH··O HB. It is only for X=I that the site 2 NI··N XB is strong enough to nearly make up for these other bond weakenings.

Most weakly bound of all are those structures where the H atoms at both sites 1 and 2 are replaced by halogens. As in the case of Cl, the two XBs at sites 1 and 2 are simply unable to replace the two HBs in the unsubstituted A-T pair. This distinction is true even for I with its strong halogen bonding ability. As may be seen in Fig 4f, the BCP densities of these two XBs are equal to 0.0120 and 0.0250. These values are considerably smaller than the densities of the two HBs in Fig 1b of 0.0207 and 0.0444.

It is worth inquiring how well the BCP densities reflect the bond strengths. One means of considering this question relates the sum of the various intermolecular BCP densities to the overall dyad interaction energies, which is a composite of all of these interactions. Of course, it would not be correct to claim that all of the interaction arises from two or three specific noncovalent bonds. There are also other contributors, like for example the London dispersion attraction which is spread out over the entire domains of the molecules. Similarly, the electrostatic energy between the electrostatic potentials of the two monomers is due to more than just the coulombic attractions between a few isolated atomic centers.

But nevertheless, the sums of the BCP densities of all intermolecular bond paths are listed in Table 2. These quantities do in fact bear a strong resemblance to the total interaction energies in Table 1. The values become larger from left to right as one progresses from Cl to Br to I. Site 1 substitutions yield the largest quantities, followed by site 2, and then by the combination of both. It would thus not be unreasonable to consider the BCP densities as semiquantitative measures of the individual bond strengths, so as to draw conclusions as to the relative contributions of each.

Guanine-Cytosine

The Hoogsteen pairing of guanine with cytosine leads to the interesting additional factor of an overall positive charge on the C [78]. The optimized geometry of this G-C^+ dimer is exhibited in Fig 5a. The charge amplification of this system leads to a large interaction energy of 46.31 kcal/mol, three times that in the neutral AT pair. There are two key sites containing a NH··O (site 1) and NH··N (site 2) HB, each rather short, and close to linear. Fig 5b shows the two have fairly large BCP densities, particularly site 2. This pair also contains what appears to be a significant, if weak CH··O HB with a density of only 0.0060 au.

The geometries that result when the H atoms are replaced by Cl are showcased in Fig 6a-6c. The site 1 substitution replaces the NH··O HB by a NCl··O XB, with a comparable density. But the reorientation of the two molecules weakens the site 2 HB, lengthening it by 0.2 Å and reducing its ρ to only 0.0350. There is also a mild enhancement of the CH··O HB. The sum of the BCP densities is reduced by this site 1 substitution, as is the interaction energy. As one changes the site 1 atom from Cl to Br and then to I, Figs 7 and 8 document how the XB is progressively strengthened, which is countered by a continuous lengthening and weakening of the site 2 HB. But the former XB strengthening is the dominant influence, causing a progressive rise in both the interaction energy and the density sum.

The result of a site 2 replacement breaks any CH··O HB. But more importantly, it also weakens the site 1 NH··O HB, particularly for the two heavier X atoms. But Br and I also yield

a substantial XB. While this bond is weaker as compared to the original HB for Cl, it is quite as strong for both Br and I. This trend is reflected to some degree in both the interaction energy and density sums. In fact, the site 2 substitution by I leaves the G-C^+ complex slightly more strongly bound than the unsubstituted original. Another factor which influences these concepts is the secondary XB which appears for the 2-substituted Br and I cases between the O on G and the new X atom. This supplement adds some 0.01 au to the total density sum and helps stabilize these dyads.

Many of these same patterns are repeated if both sites are substituted, and the two HBs are replaced by XBs. Replacement of the two HBs by Cl XBs deteriorates the overall bonding, but the analogous I substitutions yield the opposite effect of an enhancement of the binding. Indeed, the most strongly bound G-C^+ complex is that with both H centers replaced by I, with an interaction energy of 52.7 kcal/mol. But it should be understood that it is the site-2 bonding that dominates this interaction in most instances. That is, whether both sites 1 and 2 involve HBs, or if both represent XBs, the density of the site 2 interaction is considerably larger than that of site 1. This same dominance of the second site applies also when the bonding is mixed: when HB occurs at site 1 and there is a XB present as site 2.

Tautomers

This pairing of a neutral G with a cationic C^+ also introduces another factor into the equation. The migration of a proton from one neutral molecule to another would be highly disfavored energetically as it would introduce a great deal of charge separation into a cation-anion system. On the other hand, this same migration of H^+ from C^+ to G simply replaces one ion-neutral G-C^+ with another of this type: G^+-C. Indeed, calculations suggest this particular transfer of the site 2 proton raises the energy by only 2.53 kcal/mol.

These same sorts of transfers can arise in the halogenated pairs. After site 1 halogenation, the same transfer of the site 2 proton lowers the energy by 3.67 kcal/mol for Cl, and 1.12 kcal/mol for Br. This order of stability reverses for I where the original G-C^+ tautomer is more stable by 3.62 kcal/mol. The barriers separating these tautomers are of variable height. In the case of Cl, the barrier for G-$C^+ \rightarrow G^+$-C transfer is 2.9 kcal/mol, rises to 5.6 kcal/mol for Br, while it reaches 10.3 kcal/mol for I for which this transfer is endothermic. For the sake of comparison, this barrier is 2.7 kcal/mol in the unhalogenated case.

It is not only the H atoms whose position between the two units is of interest. The heavier halogen atoms do not always strictly adhere to one subunit or the other. When Br is placed in site 2, Fig 7b shows its distance from the N of G is only 0.3 Å longer than its bondlength with N on the C^+ unit. When a Br is added to site 1 as well, there is a shift in the site 2 Br in Fig 7c so that the former distance is 0.43 Å shorter than the latter, so this configuration may be better thought of as a Br-transferred G^+-C. (It is for this reason that the numbers on the bond paths in Fig 7f consider the site 2 Br to be located on the G species on the left.) The situation is similar when Br is replaced by I. The site 2 I is nearly half-transferred in Fig 8b, with its bondlength to

the N of C only 0.013 Å shorter than its distance from the N of G. Its shift in this direction is amplified when an I is placed at site 1, latter distance is shorter than the former by 0.10 Å.

2. Watson-Crick Pairing

Guanine-Cytosine

The Watson-Crick orientation of the guanine-cytosine pair contains three strong HBs, as may be seen in Fig 9a. All of these HB lengths are well below 2 Å, and very close to linearity. The AIM BCP densities are all sizable, between 0.0258 and 0.0369 au. Although there is some variation from one HB to the next, none can be considered as dominant. The sum of these three densities is 0.0943 au, and they make for a large interaction energy of 29.95 kcal/mol, as contained in Table 3. This quantity is larger than any of the Hoogsteen AT pairs, halogenated or not, but smaller than the charge-amplified G-C^+ pairs.

Parker et al [60] had earlier shown that halogenation of any one of these three sites weakens the total interaction. Their optimized geometries made it evident that substitution at either site 1 or 3 disrupts the site 2 HB, while the site 3 HB is essentially destroyed by a site 2 substitution. The weakening is particularly severe if all three sites are halogenated simultaneously. These authors noted similarly universal interaction weakening caused by halogenation of either, or both, of the two sites within the A-T pair.

Given that halogenation at any site seems to disrupt the site 2 HB of G-C, the question arises as to whether the loss of this central HB might be compensated by two XBs, at sites 1 and 3. The optimized geometries of such complexes are delineated in Fig 10a-c, from which it may be seen that the two XBs are both very nearly linear following substitution by Cl, Br, or I. The NH··N distance is quite long, as expected, all longer than 3.3 Å, so no stability is expected from that quarter. The AIM diagrams in Fig 10d-f, verify that there are substantial XBs at sites 1 and 3. The AIM densities of these bonds are nearly equal to one another and vary from 0.020 for Cl up to 0.037 for I. These quantities rival the 0.026 - 0.037 au range of the three HBs in the nonsubstituted dyads in Fig 9b. There is also a weak NH··X HB to the site 3 X atom, which may contribute a bit to the overall energetics, based on its density which varies from 0.007 up to 0.012 au.

The energetics listed in Table 3 are consistent with the dependence of the binding purely upon halogen bonding, as the interaction energy climbs quickly with the Cl < Br < I sequence. The lower half of Table 3 is consistent with that observation, as the bond critical point densities rise accordingly. It is when the halogen size has increased up to I that the two NX··O XBs are able to successfully rival the three HBs within the unsubstituted dyad. The interaction energy of 32.41 kcal/mol for the diiodinated complex is several kcal/mol higher than that for the plain G-C pair. In summary then, the halogenation of any of the three HB sites of the GC pair reduces the complexation energy. The exception to this rule is the 1,3 I-replacement for which the two XBs are comparable to, and even cumulatively stronger than, the three HBs in the original dimer.

It might be added at this point that selected replacement of proton donor groups such as NH by other groups such as SeF which might replace a HB by a chalcogen bond rather than a XB [79]

found a similar effect, that most such replacements degraded the interaction energy. Unfortunately, these calculations did not consider chalcogen atoms below the third row of the periodic table, stopping at Se. There was thus no data generated pertaining to Te, the fourth-row chalcogen lying adjacent to I.

Given the sensitivity of the energetics of these systems to the nature of the halogen atom, it would be interesting to examine how each individual force comes into play. Table 4 contains the SAPT breakdown of each total interaction energy into its separate components. The electrostatic term (ES) represents the coulombic energy between the charge distributions of the two monomers, prior to their mutual perturbation of one another. The induction energy (IND) encapsulates how these perturbations stabilize the complex, and these attractive forces are complemented by the London dispersion energy (DISP). The exchange energy (EX) is loosely connected with the classical steric repulsion which prevents the collapse of the system into a single unit.

The quantities listed in Table 4 show the replacement of H by Cl or Br reduces most of these quantities, while they are all very substantially raised by iodination. In particular, this 1,3 I-substitution nearly doubles both IND and DISP. Within the context of XBs, all components rise as the X atom grows in size. The fact that these components are smaller for X=Cl or Br as compared to H can be attributed in part to the longer interatomic distances to the O electron donor upon halogenation. It is the very strong XB formed by the I that is able to overcome this limitation. The last three columns of Table 4 refer to the percentage contribution of each component to the total of the three attractive terms. The results show that ES is most important for the HBs, slightly less so for the three XBs. The dispersion percentage is rather low for the unsubstituted HBs at only 14%, but is larger for the XBs, between 19% and 26%. Even though the absolute value of DISP rises with size of X, the more dramatic increase of both ES and IND causes the percentage contribution of DISP to diminish. Given the data in Table 4, the greater binding energy of the diiodinated GC pair has several causes. Although its exchange repulsion is much larger than the unsubstituted dyad, this destabilization is more than made up for by enhancements of the ES, IND, and DISP by 8.5, 12.6, and 9.2 kcal/mol, respectively.

The SAPT decompositions of the neutral Hoogsteen AT complexes are displayed in Table 5. Results generally reflect many of the same trends as in the WC combinations. ES comprises somewhere between 49 and 59% of the total attractive elements. IND and DISP are competitive with one another. The latter is the larger of the two with the only exceptions being the unsubstituted pair and the situation where I occupies site 2.

In terms of comparison with other calculations in the literature, Šponer et al [80] had computed interaction energies of various base pairs with complete basis set extrapolation at the RI-MP2 level, with higher-order corrections. Optimization of the WC GC and AT pairs yielded interaction energies of 28.8 and 15.4 kcal/mol, respectively. A more recent calculation of the GC pairwise energy [81] is 27.1 kcal/mol. These quantities agree quite well with our own computed value of 29.9 kcal/mol for the WC GC interaction energy. This work complemented earlier calculations [82] wherein the Hoogsteen arrangement of the methylated AT pair was bound

by 14.8 kcal/mol, again quite close to our value of 15.5 kcal/mol. On another note, a very recent decomposition of the interaction energy of the Hoogsteen AT pair [83], albeit with a different decomposition scheme, evaluated components very much like those derived here.

SUMMARY

Replacement of the bridging H atoms of the various base pairs by halogens generally leads to a reduction in the interaction energies within the dyad. The XBs that result from this substitution follow the expected pattern of increasing strength with larger size halogen: Cl < Br < I. Consequently, the Cl substitutions weaken the interactions the most and I the least. Indeed, the I substitutions have only a very mild reducing effect in general and there are certain cases where the interaction is in fact strengthened.

Within the Hoogsteen A-T pair, substitution at site 1 produces the largest reduction, and site 2 a bit less. But the replacement of H by X at both these sites leads to a smaller reduction than either site individually. On a granular level, it is the site 2 HB that is strongest within the A-T pair, and it is only I that can replace this HB with an equally strong XB. The overall binding is amplified by the charge within the G-C^+ Hoogsteen pair. The pattern is similar here in that the substitution of the hydrogens at sites 1 and 2 together provides stronger binding than either site individually, with site 1 replacement effecting the largest weakening of the binding. The I XBs are sufficiently potent that substitution of the site 2 H by I, with or without the same site 1 mutation, provides a more strongly bound complex than in the original G-C^+. Here too it is the site 2 HB that provides the lion's share of the binding. And once again, it is only I that can match the strength of the original HB with a NI··N XB.

The three separate HBs in the Watson-Crick GC pair are roughly equal contributors to the total interaction energy. This quantity is lessened by substitutions at any of these three HB sites by a halogen atom. In fact, this weakening is exacerbated for I, despite its capability to engage in a strong XB. Replacement at all three sites simultaneously is particularly detrimental. On the other hand, substitution at sites 1 and 3 has a much more varied effect. This pattern eliminates the site 2 HB, replacing the other two HBs by XBs of comparable strength to one another. In the case of I, these two XBs are strong enough that they are able to compensate for the loss of the site 2 HB, and the 1,3 diiodinated G-C pair is more strongly bound than the native pair.

Of course, all of the substitutions considered here have certain repercussions upon the geometry of each dyad. Perhaps most important is the ability of each of these pairs to fit into the DNA double helix space allotted to it. One can use as a barometer of the extent of the full base pair the distance between the N atoms that are attached to the sugar species within the biopolymer. These N atoms are indicated by the green asterix in Figs 1 and 5. The distance between them is contained in Table 6 for each of the dyads considered here. It may be noted for example, that this distance within the optimized Hoogsteen A-T dyad is 6.652 Å. Halogenation at site 1 has only a small effect, reducing this distance by about 0.2 Å, since site 1 is rather removed from the starred N atoms. It is the site 2 substitution that very substantially elongates this distance by more than 2 Å, as that region is closer to the pertinent N atoms of attachment to

the sugars. This elongation is mitigated a bit by a simultaneous replacement of the site 1 HB by a XB, and the ensuing swivel of the two molecules relative to one another. The next segment of Table 6 follows very similar trends for the GC^+ pair. The optimized N··N distance in the native dimer is 6.603 Å, which is shortened a bit for site 1 substitution. Again, it is the site 2 substitution which causes a roughly 2 Å stretch, which is not as severe if site 1 is replaced as well. As may be seen in the final two rows of Table 6, there is also a 1.5 Å expansion in the Watson-Crick GC pair when the H atoms at sites 1 and 3 are replaced by halogen atoms. It should thus be understood that even though certain of these substitutions with I might strengthen the interaction, they would require a distortion of the double helix to allow room for the substituted dyad which takes up more space.

Author Contributions

AA: bulk of calculations, initial analysis, data curation

SS: conceptualization, Supervision of calculations, final analysis, manuscript preparation, funding acquisition

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1954310.

Conflict of Interest

The authors declare no conflict of interest.

Supporting Information

The Supporting Information is available free of charge. It contains the atomic coordinates of the dyads.

REFERENCES

1. G. C. Pimentel and A. L. McClellan, *The Hydrogen Bond*, Freeman, San Francisco, 1960.
2. S. N. Vinogradov and R. H. Linnell, *Hydrogen Bonding*, Van Nostrand-Reinhold, New York, 1971.
3. M. D. Joesten and L. J. Schaad, *Hydrogen Bonding*, Marcel Dekker, New York, 1974.
4. E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, H. G. Kjaergaard, A. C. Legon, B. Mennucci and D. J. Nesbitt, *Pure Appl. Chem.*, 2011, **83**, 1637-1641.
5. S. J. Grabowski, ed. *Hydrogen Bonding - New Insights*, Springer, Dordrecht, Netherlands, 2006.
6. A. Chand, D. K. Sahoo, A. Rana, S. Jena and H. S. Biswal, *Acc. Chem. Res.*, 2020, **53**, 1580-1592.
7. K. K. Mishra, S. K. Singh, S. Kumar, G. Singh, B. Sarkar, M. S. Madhusudhan and A. Das, *J. Phys. Chem. A*, 2019, **123**, 5995-6002.

8. H. S. Biswal and S. Wategaonkar, *J. Chem. Phys.*, 2011, **135**, 134306.
9. K. Kamiya, M. Boero, M. Tateno, K. Shiraishi and A. Oshiyama, *J. Am. Chem. Soc.*, 2007, **129**, 9663-9673.
10. Y. Kim and K.-H. Ahn, *Theor. Chem. Acc.*, 2001, **106**, 171-177.
11. E. A. Hillenbrand and S. Scheiner, *J. Am. Chem. Soc.*, 1984, **106**, 6266-6273.
12. B. A. Feniouk, D. A. Cherepanov, W. Junge and A. Y. Mulkidjanian, *FEBS Lett.*, 1999, **445**, 409-414.
13. C. Tu, M. Qian, J. N. Earnhardt, P. J. Laipis and D. N. Silverman, *Biophys. J.*, 1998, **74**, 3182-3189.
14. Z. Latajka, Y. Bouteiller and S. Scheiner, *Chem. Phys. Lett.*, 1995, **234**, 159-164.
15. S. Scheiner, *J. Chem. Phys.*, 1982, **77**, 4039-4050.
16. N. J. Thornton and T. van Mourik, *Int. J. Mol. Sci.*, 2020, **21**, 6571.
17. S. C. C. van der Lubbe and C. Fonseca Guerra, *Chem. Eur. J.*, 2017, **23**, 10249-10253.
18. Y. Shi, W. Jiang, Z. Zhang and Z. Wang, *New J. Chem.*, 2017, **41**, 12104-12109.
19. H. Szatylowicz, A. Jezierska and N. Sadlej-Sosnowska, *Struct. Chem.*, 2016, **27**, 367-376.
20. O. Khakshoor, S. E. Wheeler, K. N. Houk and E. T. Kool, *J. Am. Chem. Soc.*, 2012, **134**, 3154-3163.
21. O. h. O. Brovarets and D. M. Hovorun, *J. Biomol. Struct. Dyn.*, 2014, **32**, 127-154.
22. J. Florián and J. Leszczynski, *J. Am. Chem. Soc.*, 1996, **118**, 3010-3017.
23. S. Scheiner and C. W. Kern, *Chem. Phys. Lett.*, 1978, **57**, 331-333.
24. S. M. Anderson, B. K. Mueller, E. J. Lange and A. Senes, *J. Am. Chem. Soc.*, 2017, **139**, 15774-15783.
25. E. N. Nikolova, R. L. Stanfield, H. J. Dyson and P. E. Wright, *Biochem.*, 2018, **57**, 2109-2120.
26. A. Balaceanu, M. Pasi, P. D. Dans, A. Hospital, R. Lavery and M. Orozco, *J. Phys. Chem. Lett.*, 2017, **8**, 21-28.
27. J. Dickerhoff, B. Appel, S. Müller and K. Weisz, *Angew. Chem. Int. Ed.*, 2016, **55**, 15162-15165.
28. G. Y. Huang, O. O. Gerlits, M. P. Blakeley, B. Sankaran, A. Y. Kovalevsky and C. Kim, *Biochem.*, 2014, **53**, 6725-6727.
29. H. Szatyłowicz and N. Sadlej-Sosnowska, *J. Chem. Infor. Model.*, 2010, **50**, 2151-2161.
30. E. Gougoula, C. Medcraft, I. Alkorta, N. R. Walker and A. C. Legon, *J. Chem. Phys.*, 2019, **150**, 084307.
31. I. Alkorta and A. Legon, *Molecules*, 2018, **23**, 2250.
32. S. J. Grabowski, *Struct. Chem.*, 2019, **30**, 1141-1152.
33. S. Grabowski, *Molecules*, 2018, **23**, 1183.
34. S. Scheiner, *Phys. Chem. Chem. Phys.*, 2024, **26**, 27382-27394.
35. A. Franconetti and A. Frontera, *Chem. Eur. J.*, 2019, **25**, 6007-6013.
36. A. Frontera and A. Bauzá, *Chem. Eur. J.*, 2018, **24**, 16582-16587.
37. J. S. Murray and P. Politzer, *J. Mol. Model.*, 2019, **25**, 101.
38. T. Clark, J. S. Murray and P. Politzer, *Phys. Chem. Chem. Phys.*, 2018, **20**, 30076-30082.
39. K. E. Riley and K.-A. Tran, *Faraday Disc.*, 2017, **203**, 47-60.
40. W. Zierkiewicz, M. Michalczyk, R. Wysokiński and S. Scheiner, *Molecules*, 2019, **24**, 376.
41. W. Dong, Y. Wang, J. Cheng, X. Yang and Q. Li, *Mol. Phys.*, 2019, **117**, 251-259.

42. M. Esrafili and P. Mousavian, *Molecules*, 2018, **23**, 2642.
43. S. Scheiner and U. Adhikari, *J. Phys. Chem. A*, 2011, **115**, 11101-11110.
44. P. Auffinger, F. A. Hays, E. Westhof and P. S. Ho, *Proc. Nat. Acad. Sci., USA*, 2004, **101**, 16789-16794.
45. F. T. Burling and B. M. Goldstein, *J. Am. Chem. Soc.*, 1992, **114**, 2313-2320.
46. M. Iwaoka and S. Tomoda, *J. Am. Chem. Soc.*, 1994, **116**, 2557-2561.
47. A. R. Viguera and L. Serrano, *Biochem.*, 1995, **34**, 8771-8779.
48. O. Carugo, *Biol. Chem.*, 1999, **380**, 495-498.
49. D. Pal and P. Chakrabarti, *J. Biomol. Struct. Dyn.*, 2001, **19**, 115-128.
50. P. Metrangolo and G. Resnati, *Chem. Eur. J.*, 2001, **7**, 2511-2519.
51. M. Iwaoka, S. Takemoto, M. Okada and S. Tomoda, *Bull. Chem. Soc. Jpn.*, 2002, **75**, 1611-1625.
52. M. U. Engelhardt, M. O. Zimmermann, M. Dammann, J. Stahlecker, A. Poso, T. Kronenberger, C. Kunick, T. Stehle and F. M. Boeckler, *J. Chem. Theory Comput.*, 2024, **20**, 10716-10730.
53. G. Renno, D. Chen, Q.-X. Zhang, R. M. Gomila, A. Frontera, N. Sakai, T. R. Ward and S. Matile, *Angew. Chem. Int. Ed.*, 2024, **63**, e202411347.
54. J. Li, L. Zhou, Z. Han, L. Wu, J. Zhang, W. Zhu and Z. Xu, *J. Med. Chem.*, 2024, **67**, 4782-4792.
55. R. M. Gomila and A. Frontera, *Chem. Asian J.*, 2024, **n/a**, e202400081.
56. X. Yan, K. Zou, J. Cao, X. Li, Z. Zhao, Z. Li, A. Wu, W. Liang, Y. Mo and Y. Jiang, *Nat. Comm.*, 2019, **10**, 3610.
57. K. Lin, P. Weng, Y. Qi, J. Teng, Z. Lei and X. Yan, *ACS Applied Materials & Interfaces*, 2024.
58. F. Seela, X. Peng and H. Li, *J. Am. Chem. Soc.*, 2005, **127**, 7739-7751.
59. R. Tawarada, K. Seio and M. Sekine, *J. Org. Chem.*, 2008, **73**, 383-390.
60. A. J. Parker, J. Stewart, K. J. Donald and C. A. Parish, *J. Am. Chem. Soc.*, 2012, **134**, 5165-5172.
61. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, Wallingford, CT2016.
62. Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215-241.
63. G. Paytakov, T. Dinadayalane and J. Leszczynski, *J. Phys. Chem. A*, 2015, **119**, 1190-1200.
64. B. S. D. R. Vamhindi and A. Karton, *Chem. Phys.*, 2017, **493**, 12-19.
65. R. Podeszwa and K. Szalewicz, *J. Chem. Phys.*, 2012, **136**, 161102.

66. S. Karthikeyan, V. Ramanathan and B. K. Mishra, *J. Phys. Chem. A*, 2013, **117**, 6687-6694.
67. M. Majumder, B. K. Mishra and N. Sathyamurthy, *Chem. Phys.*, 2013, **557**, 59-65.
68. M. A. Vincent and I. H. Hillier, *Phys. Chem. Chem. Phys.*, 2011, **13**, 4388-4392.
69. A. D. Boese, *ChemPhysChem.*, 2015, **16**, 978-985.
70. M. Walker, A. J. A. Harvey, A. Sen and C. E. H. Dessent, *J. Phys. Chem. A*, 2013, **117**, 12590-12600.
71. L. F. Molnar, X. He, B. Wang and K. M. Merz, *J. Chem. Phys.*, 2009, **131**, 065102.
72. S. F. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553-566.
73. Z. Latajka and S. Scheiner, *J. Chem. Phys.*, 1987, **87**, 1194-1204.
74. T. A. Keith, TK Gristmill Software, Overland Park KS2013.
75. K. Szalewicz and B. Jeziorski, in *Molecular Interactions. From Van der Waals to Strongly Bound Complexes*, ed. S. Scheiner, Wiley, New York1997, pp. 3-43.
76. R. Moszynski, P. E. S. Wormer, B. Jeziorski and A. van der Avoird, *J. Chem. Phys.*, 1995, **103**, 8058-8074.
77. D. G. A. Smith, L. A. Burns, A. C. Simmonett, R. M. Parrish, M. C. Schieber, R. Galvelis, P. Kraus, H. Kruse, R. D. Remigio, A. Alenaizan, A. M. James, S. Lehtola, J. P. Misiewicz, M. Scheurer, R. A. Shaw, J. B. Schriber, Y. Xie, Z. L. Glick, D. A. Sirianni, J. S. O'Brien, J. M. Waldrop, A. Kumar, E. G. Hohenstein, B. P. Pritchard, B. R. Brooks, H. F. SchaeferIII, A. Y. Sokolov, K. Patkowski, A. E. DePrinceIII, U. Bozkaya, R. A. King, F. A. Evangelista, J. M. Turney, T. D. Crawford and C. D. Sherrill, *J. Chem. Phys.*, 2020, **152**, 184108.
78. A. L. Stelling, A. Y. Liu, W. Zeng, R. Salinas, M. A. Schumacher and H. M. Al-Hashimi, *Angew. Chem. Int. Ed.*, 2019, **58**, 12010-12013.
79. K. D. Sharma, P. Kathuria, S. D. Wetmore and P. Sharma, *Phys. Chem. Chem. Phys.*, 2020, **22**, 23754-23765.
80. J. Šponer, P. Jurečka and P. Hobza, *J. Am. Chem. Soc.*, 2004, **126**, 10142-10151.
81. R. Oliva, L. Cavallo and A. Tramontano, *Nucleic Acid Res.*, 2006, **34**, 865-879.
82. P. Jurečka and P. Hobza, *J. Am. Chem. Soc.*, 2003, **125**, 15608-15613.
83. Y. Xiao, F.-Y. Zhang, M.-J. Ma, S.-J. Lu, Y.-J. Zhang, X.-Y. Zhao, Y.-X. Wang, K. Zhang and H.-L. Xu, *Chem. Phys.*, 2025, **588**, 112465.

Table 1. Interaction energies ($-E_{int}$, kcal/mol) of Hoogsteen paired dimers, with and without halogen substitutions at indicated sites.

	Cl	Br	I
AT	15.48		
1	12.87	13.90	15.79
2	8.63	10.98	15.21
both	6.04	8.38	12.09
GC^+	46.31		
1	37.51	38.19	44.54
2	35.02	40.97	47.22
both	28.82	30.68	52.71

Table 2. Sums of intermolecular bond critical point densities (10^{-4} au) of Hoogsteen paired dimers, with and without halogen substitutions at indicated sites.

	Cl	Br	I
AT	709		
1	557	641	668
2	442	529	569
both	376	433	515
GC^+	1045		
1	844	993	1001
2	667	988	928
both	655	979	912

Table 3. Interaction energies ($-E_{int}$, kcal/mol) and sums of intermolecular bond critical point densities (10^{-4} au) for Watson-Crick paired G-C dimers, with and without halogen substitutions at sites 1 and 3.

	H	Cl	Br	I
$-E_{int}$	29.95	11.91	17.43	32.41
$\Sigma\rho_{BCP}$	943	465	584	865

Table 4. SAPT components (kcal/mol) for Watson-Crick paired GC dimers, with and without halogen substitutions at sites 1 and 3.

	ES	EX	IND	DISP	%ES	%IND	%DISP
H	-44.77	37.99	-18.17	-10.62	60.9	24.7	14.4
Cl	-18.16	20.21	-7.02	-8.71	53.6	20.7	25.7
Br	-28.19	31.62	-12.33	-11.90	53.8	23.5	22.7
I	-53.23	63.36	-30.79	-19.85	51.2	29.6	19.1

Table 5. SAPT components (kcal/mol) for Hoogsteen paired AT dimers, with and without halogen substitutions at indicated sites.

	ES	EX	IND	DISP	%ES	%IND	%DISP
H	-28.19	28.56	-10.83	-8.53	59.3	22.8	17.9
Cl1	-21.16	21.56	-6.65	-8.84	57.7	18.1	24.1
Cl2	-14.82	16.93	-4.95	-7.01	55.3	18.5	26.2
Cl1,2	-10.45	14.64	-3.65	-7.34	48.7	17.0	34.2
Br1	-23.56	24.25	-7.33	-9.75	58.0	18.0	24.0
Br2	-22.30	26.92	-8.61	-9.32	55.4	21.4	23.2
Br1,2	-15.97	21.12	-5.39	-9.72	51.4	17.3	31.3
I1	-27.69	29.66	-9.49	-11.34	57.1	19.5	23.4
I2	-33.97	44.85	-17.65	-13.69	52.0	27.0	21.0
I1,2	-23.22	32.77	-10.36	-14.14	48.7	21.7	29.6

Table 6. Distance (Å) between N atoms attached to sugars in DNA.

	Cl	Br	I
Hoogsteen			
AT	6.652		
1	6.412	6.422	6.467
2	8.747	8.851	8.951
both	8.333	8.529	8.842
GC^+	6.603		
1	6.388	6.439	6.517
2	8.612	8.559	8.840
both	8.128	8.092	8.425
WC			
GC	8.980		
1,3	10.354	10.501	10.699

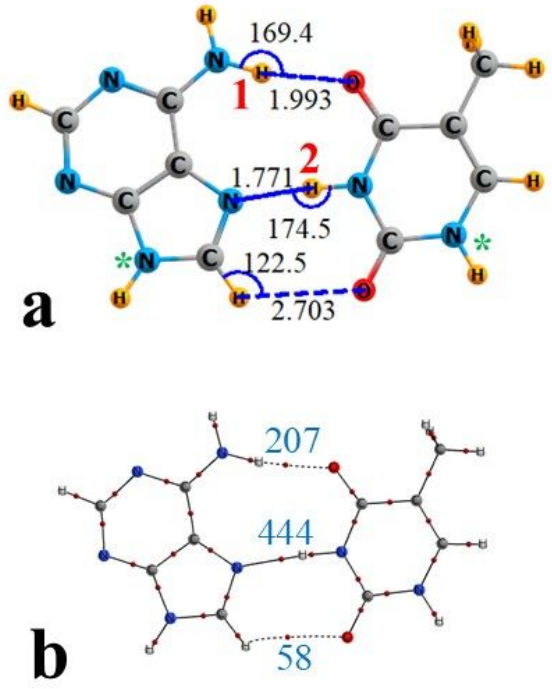

Fig 1. a) Geometry of Hoogsteen A-T pair, indicating the numbering of the two HB sites in red. Distances in Å, angles in degs. Green asterix indicates N atom attached to sugar in nucleotide. b) AIM diagram with bond critical point densities shown as blue numbers in units of 10^{-4} au.

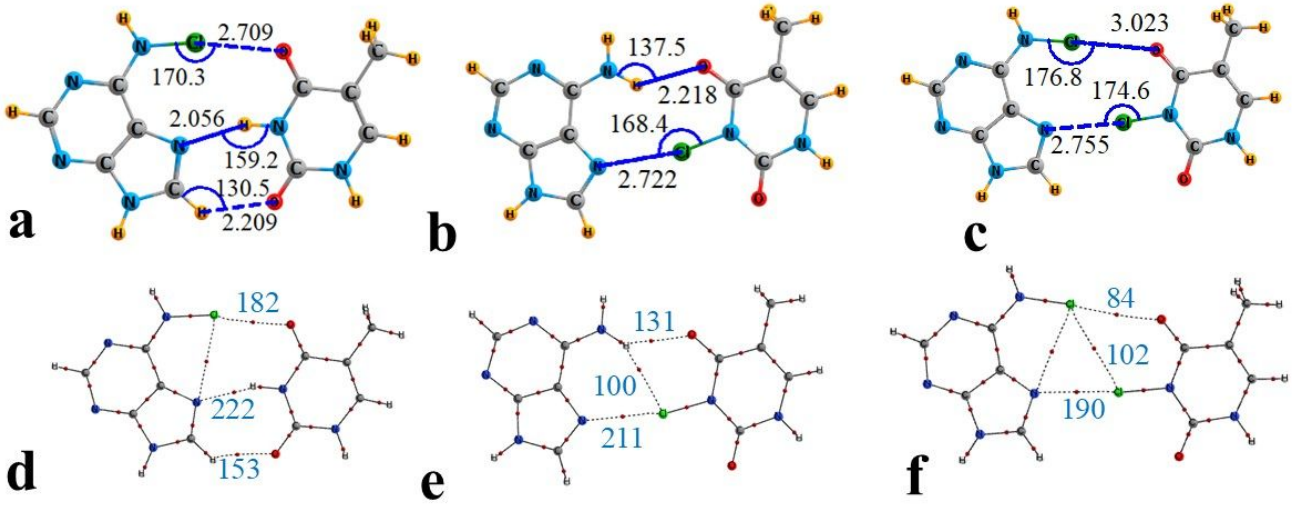

Fig 2. Geometry of Hoogsteen A-T pair, chlorinated at a) site 1, b) site 2, c) sites 1 and 2. d-f) AIM diagrams with bond critical point densities shown as blue numbers in units of 10^{-4} au.

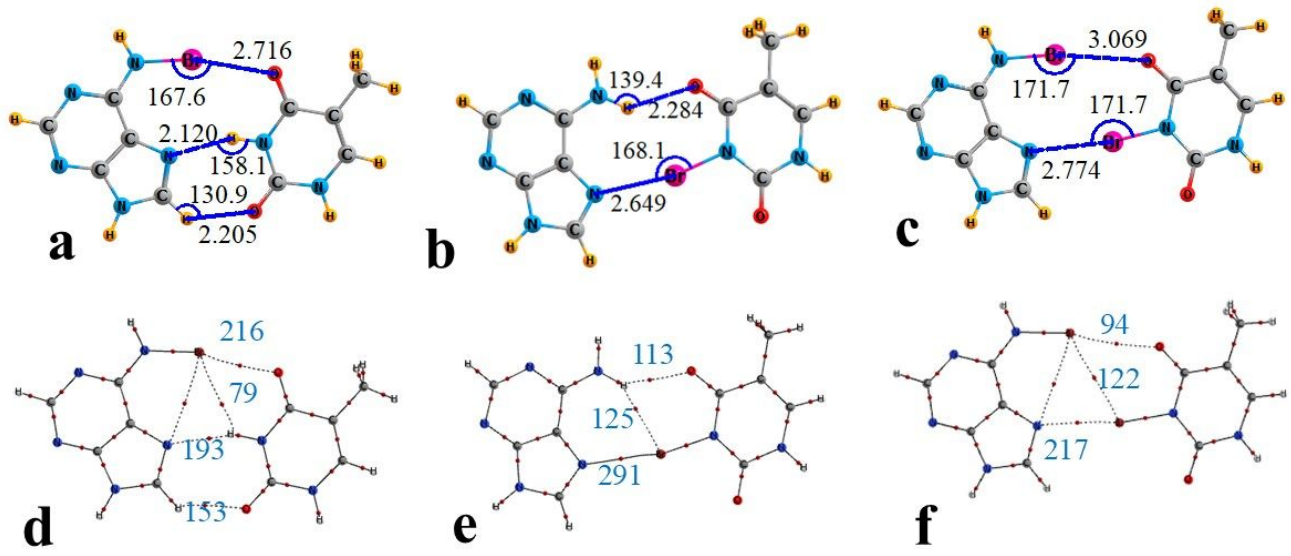

Fig 3. Geometry of Hoogsteen A-T pair, brominated at a) site 1, b) site 2, c) sites 1 and 2. d-f) AIM diagrams with bond critical point densities shown as blue numbers in units of 10^{-4} au.

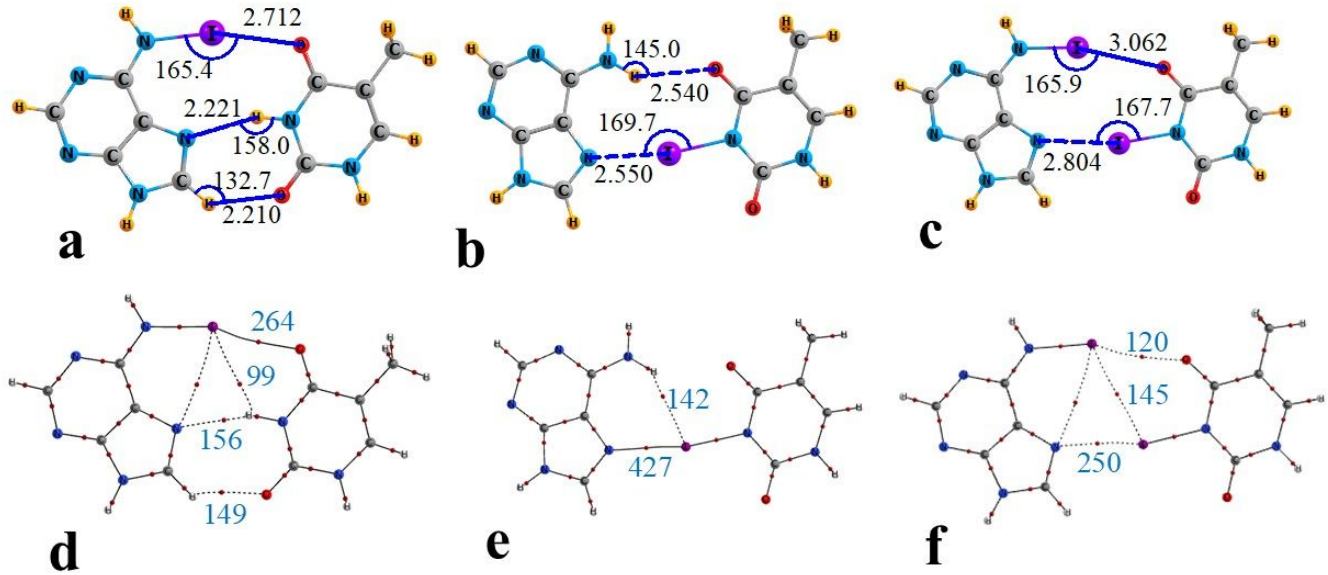

Fig 4. Geometry of Hoogsteen A-T pair, iodinated at a) site 1, b) site 2, c) sites 1 and 2. d-f) AIM diagrams with bond critical point densities shown as blue numbers in units of 10^{-4} au.

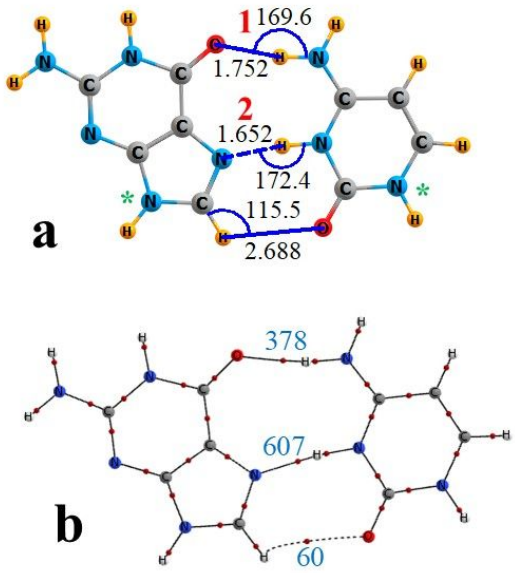

Fig 5. a) Geometry of Hoogsteen G-C$^+$ pair, indicating the numbering of the two HB sites in red. Distances in Å, angles in degs. Green asterix indicates N atom attached to sugar in nucleotide. b) AIM diagram with bond critical point densities shown as blue numbers in units of 10^{-4} au.

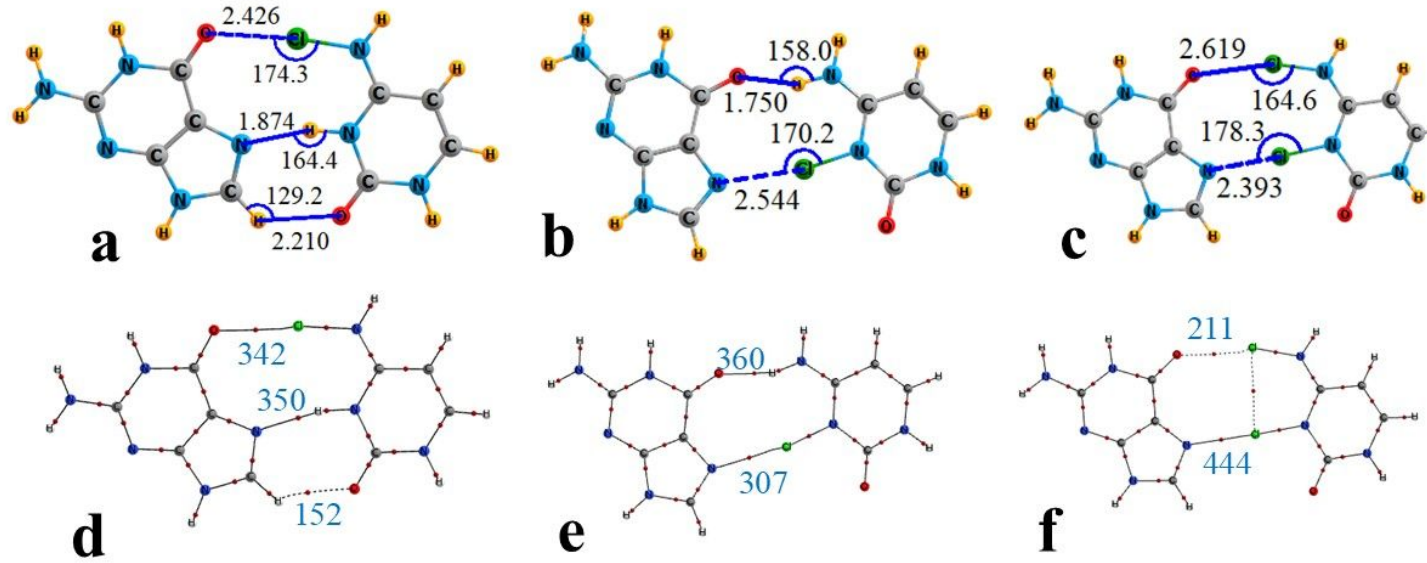

Fig 6. Geometry of Hoogsteen G-C$^+$ pair, chlorinated at a) site 1, b) site 2, c) sites 1 and 2. d-f) AIM diagrams with bond critical point densities shown as blue numbers in units of 10^{-4} au.

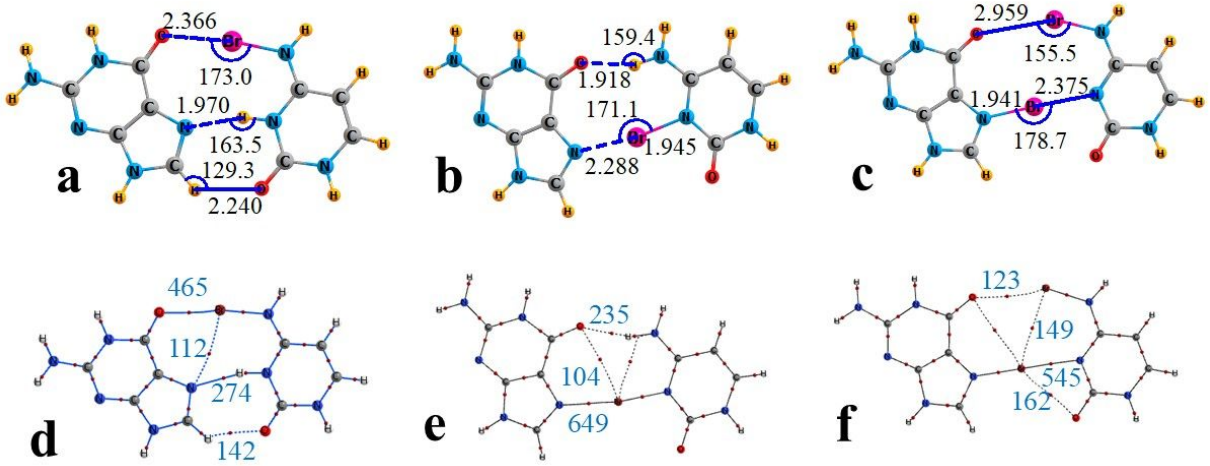

Fig 7. Geometry of Hoogsteen G-C$^+$ pair, brominated at a) site 1, b) site 2, c) sites 1 and 2. d-f) AIM diagrams with bond critical point densities shown as blue numbers in units of 10^{-4} au.

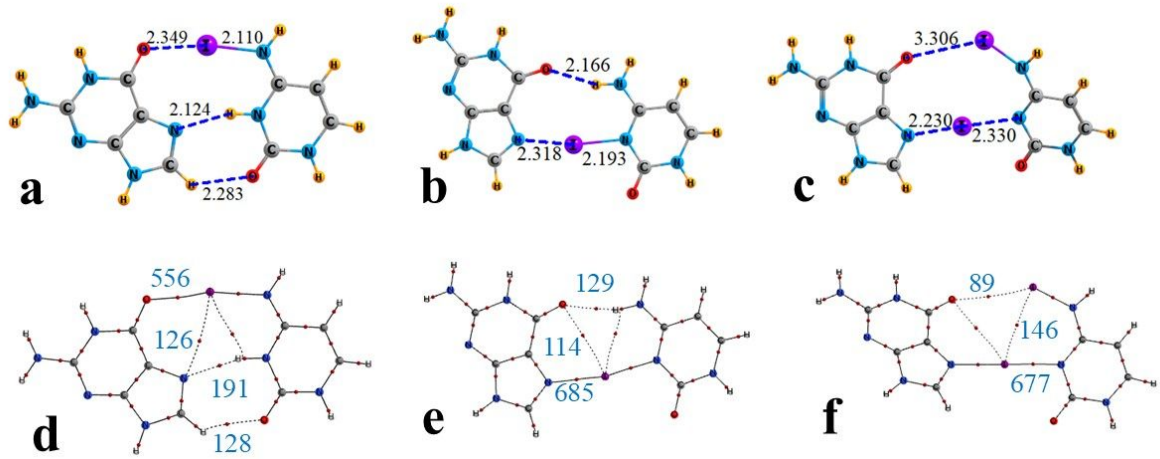

Fig 8. Geometry of Hoogsteen G-C$^+$ pair, iodinated at a) site 1, b) site 2, c) sites 1 and 2. d-f) AIM diagrams with bond critical point densities shown as blue numbers in units of 10^{-4} au.

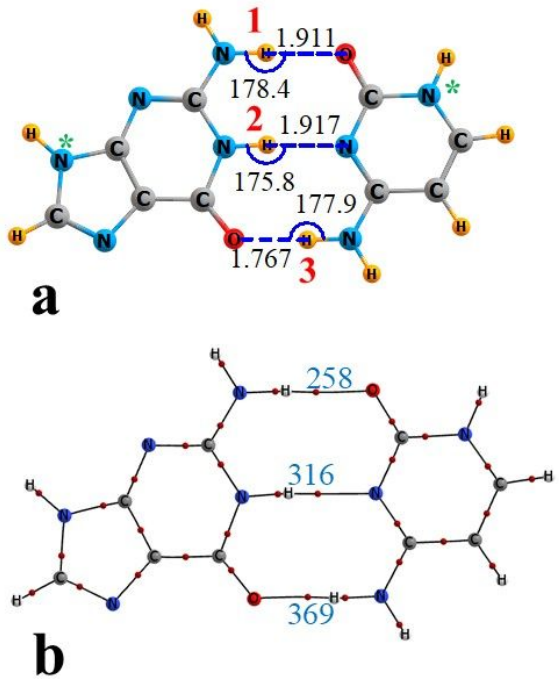

Fig 9. a) Geometry of Watson-Crick G-C pair, indicating the numbering of the three HB sites in red. Distances in Å, angles in degs. Green asterix indicates N atom attached to sugar in nucleotide. b) AIM diagram with bond critical point densities shown as blue numbers in units of 10^{-4} au.

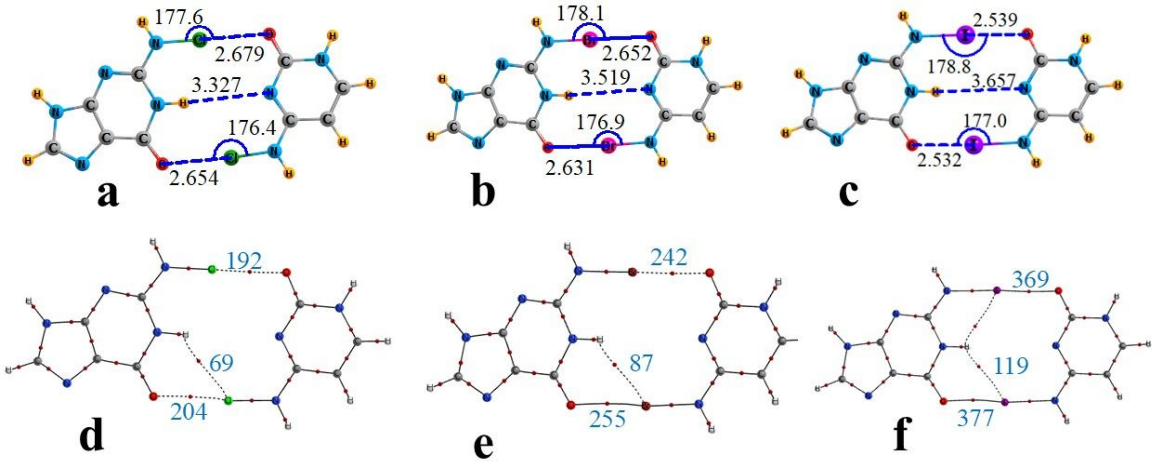

Fig 10. Geometry of Watson-Crick G-C pair, halogenated at sites 1 and 3 by a) Cl, b) Br, c) I. Distances in Å, angles in degs. d-f) AIM diagrams with bond critical point densities shown as blue numbers in units of 10^{-4} au.

In addition to the coordinates included in the Supporting Information, other data are available from the authors upon request.