Quantum-Chemical Analysis of C-H...O and C-H...N Interactions in RNA Base Pairs - H-Bond Versus Anti-H-Bond Pattern

Maria Brandl, Michael Meyer[1], and Jürgen Sühnel*

Biocomputing, Institut für Molekulare Biotechnologie, Postfach 100813, D-07708 Jena / Germany

[1]present address: Konrad-Zuse-Zentrum für Informationstechnik, Takustr.7, D-14195 Berlin-Dahlem, Germany

Abstract

Geometries and interaction energies of unusual UU and AA base pairs with one standard hydrogen bond (H-bond) and additional C-H...O or C-H...N contacts have been determined by quantum-chemical methods taking into account electron correlation. Whereas the C-H bond length in the UU C-H...O contact increases upon complex formation (H-bond pattern), the C-H bond of the AA C-H....N interaction is shortened (anti-H-bond pattern). The same properties are found for model complexes between U or A and formaldehyde that have intermolecular C-H...acceptor contacts but no standard H-bonds. Both the C-H...acceptor H-bond and anti-H-bond interactions are attractive. A possible influence of the donor CH group charge distribution on the interaction pattern is discussed.

Phone: +49-3641-656200; Fax: +49-3641-656210; E-mail: jsuehnel@imb-jena.de.

Introduction

For a long time it has been assumed that at least two direct standard hydrogen bonds (H-bonds) are necessary for base pair formation in nucleic acid structures (1-3). Recently, however, base pairs that are linked by only one or even no direct standard H-bond have been detected. In some cases N-H...F (4,5), C-H...O (6) and C-H...N (7) contacts may replace conventional H-bonds. Moreover, water has been found capable of bridging H-bond donor/acceptor sites (8-11). In other cases only one H-bond occurs between the bases without additional base-base interactions (12).

Based on a geometrical analysis of small molecules C-H...O contacts have been considered as weak H-bonds (13). A comprehensive review on the properties of C-H...O contacts in small molecules has been presented by Steiner (14). Interactions of this type have also been claimed to be a long-neglected stabilizing force in biopolymers (15). They have been found in proteins (16), RNA (7,17) and DNA (18,19).

At least for biopolymer structures, the majority of these analyses was based on a comparison of experimental H...O and H...N distances with the respective sums of van der Waals radii. It is not unlikely, however, that short C-H...O/N contacts are enforced by strong neighbouring interactions and do not contribute to the overall stability of biopolymers. More direct information can be obtained from the change of the C-H bond length and of the related stretching vibration. H-bonds are typically characterized by an elongation of the covalent bond between the heavy atom and hydrogen in the donor group and a red shift of its stretching vibration (20). Therefore, it came as a surprise when it was first noticed that the C-H bond length decreases in an intramolecular C-H...O interaction in 2-(methylthio)ethanol (21). More recently, complexes between C-H groups that are perpendicularly oriented towards aromatic ring systems have been investigated by quantum-chemical methods and double-resonance infrared depletion spectroscopy (22,23). For a benzene C-H group within a T-shaped benzene-dimer, for methane and chloroform C-H groups interacting with benzene and for a chloroform/fluorobenzene complex quantum-chemical calculations including electron correlation have predicted shortenings of the covalent C-H bond length upon complex formation and blue shifts of the respective stretching vibrations. For the complex between chloroform and fluorobenzene the calculated frequency shifts could be confirmed by infrared spectroscopy. Meanwhile, C-H bond length contraction has also been found in calculations on dimeric and trimeric complexes between donor formaldehyde and water or ammonia (24) and in complexes between methane or fluoro substituted methanes and water, methanol or formaldehyde as acceptor (25). Finally, based on quantum-chemical calculations at the Hartree-Fock level without consideration of electron correlation effects the C2-H2(A)...O2(U) interaction in the Watson-Crick AU base pair has been suggested to exhibit anti-H-bond character, too (26). There are, however, also two C-H...π interactions for which the C-H bond length increases upon complex formation, namely within the complexes of HCCH and HCN with benzene (22, 27).

As decreases in C-H bond lengths and blue shifts of the corresponding stretching vibration frequencies are contrary to the features of conventional H-bonds, interactions exhibiting such characteristics have been classified as anti-H-bonds (22). Moreover, it has been claimed that the anti-H-bond concept is of a general nature and should also be of relevance to biological macromolecules (22,23,26). From an electron density topology analysis it has been concluded that the anti-H-bond pattern originates from the specific redistribution of electron density in the C-H bond upon complexation. It has further been claimed that contrary to standard H-bonds anti-H-bond interactions are primarily governed by dispersion forces (22,27). Other authors have arrived at the conclusion that C-H...O interactions should be considered as true H-bonds in spite of the shortening of the C-H bond upon complex formation found in almost all cases studied thus far (25).

In order to investigate the properties of unusual base pairs with only one standard H-bond and additional C-H...O or C-H...N contacts and to gain deeper insight into the nature of C-H...O and C-H...N interactions we have performed quantum-chemical calculations on a UU and an AA base pair. The so-called Calcutta UU base pair has been identified by Wahl et al. (15) in the overhanging portion of the RNA hexamer r(UUCGCG) (Protein Data Bank (PDB) code: 1osu; (28)). The two uracil bases are linked by a standard H-bond between H3 of U2 and O4 of U1. In addition, H5 of U1 forms a C-H...O contact to O4 of U2 (Figure 1a). The second system is an AA pair that we have identified in a systematic search for C-H...O or C-H...N contacts in RNA structures (7). It occurs in the J4/5 helical segment of the *Tetrahymena thermophila* self-splicing intron (PDB code: 1gid; base pairs A113-A207, A114-A206; (29) and in $\mathrm{tRNA}^{\mathrm{Glu}}$/$\mathrm{tRNA}^{\mathrm{Glu}}$-synthethase complexes (PDB codes: 1gtr, 1gts, 1qrs, 1qru, 1qrt; base pair A13-A22, (8,30)). In this base pair H62 of A2 forms an H-bond to N3 of A1 and a C-H...N contact is observed between H2 of A2 and N7 of A1 (Figure 1b).

Quantum-chemical calculations have been shown to produce reasonably good estimates for the geometries and interaction energies of standard base pairs involving at least two H-bonds (31). In order to get reliable results for base pairs with weak interactions electron correlation has to be taken into account. This has been done using second order Møller-Plesset perturbational theory (MP2) for both the geometry optimizations and interaction energy calculations.

In the Calcutta and in the adenine-adenine base pairs the C-H...acceptor contacts are located next to standard H-bonds. In order to separate the effects stemming from the C-H...O or C-H...N contacts from those arising from the neighboring standard H-bonds, we have performed additional calculations on model systems where the adenine and uracil bearing the acceptor site of the C-H...O/N contact were replaced by formaldehyde (Figure 2a, b).

Density functional theory (DFT) methods have been reported to produce good results for the geometries of standard base pairs but to fail for stacking interactions (32). We have shown that the DFT geometry of the Watson-Crick AT pair is in better agreement with experimental structural data obtained from 21 high-resolution DNA structures than the geometry calculated at the Hartree-Fock level (5). In order to find out, whether these methods are also reliable for the weaker C-H...acceptor contacts, the calculations for each complex have been repeated with the DFT method and the results are compared. In addition, the computationally less expensive DFT approach has been used for frequency calculations.

After having submitted this work for publication, we became aware of a quantum-chemical study on C-H...O contacts in the Watson-Crick AU and in two different UU base pairs that includes also the Calcutta UU pair (32). Our results obtained for this base pair are in full agreement with the data reported by Hobza et al (33). A more detailed comparison is presented in the Results and Discussion section. In ref. (33) the term anti-H-bond, originally coined by these authors, has been replaced by blue-shifting H-bond. In this work we use the terms H-bond and anti-H-bond pattern for describing the features of these different interaction types.

Methods

In order to investigate the geometries, interaction energies and spectroscopic properties of the base pairs and of related model systems, quantum-chemical ab initio calculations have been carried out. All quantum-chemical calculations have been performed using Gaussian 94 (34). As in the studies of Šponer et al. (31) only the base parts of the nucleotides have been used and ribose C1' carbons have been replaced by hydrogens. The geometries of the base pairs have been optimized with inclusion of electron correlation according to second order Møller-Plesset perturbational theory (MP2(FC); FC – frozen core) and DFT methods using Becke's three parameter hybrid method and the Lee-Yang-Parr correlation functional (B3LYP). For both types of calculations the 6-31G(d,p) basis set has been used. In order to achieve the accuracy that is required for the description of weak interactions all optimizations have been performed using the "Tight" set of convergence criteria. The energy minima obtained from B3LYP/6-31G(d,p) optimizations have been checked by calculations of the Hessian. Interaction energies ΔE have been corrected for the basis set superposition error by the standard counterpoise method (35). A further correction was done for the deformation energies (ΔE^{DEF}), which are defined as the energy differences between the geometries of the isolated bases and the geometries of the bases in the complex. The total interaction energy ΔE^{T} is thus given by $\Delta E^{T}=\Delta E+\Delta E^{DEF}$. For comparison, the DFT interaction energies have been corrected by the changes in zero-point vibrational energies ΔE^{ZPE}. These changes have been scaled with the factor 0.9804 (36). The quantity $\Delta E_0 = \Delta E^{T} + \Delta E^{ZPE}$ corresponds to the enthalpy at 0 K. Our discussion is, however, primarily focused on the ΔE^{T} values. Computationally expensive calculations of vibrational frequencies have been performed at the B3LYP/6-31G(d,p)//B3LYP/6-31G(d,p) level only. The vibrational frequencies obtained that way have been scaled with the factor 0.9613. The scaling factors for both zero-point energies and frequencies have been taken from ref. (36). The charge distribution of the systems under study has been determined by means of a Mulliken population analysis.

Results and Discussion

The Calcutta UU base pair

Geometry optimization of the Calcutta base pair at the MP2(FC)/6-31G(d,p) and B3LYP/6-31G(d,p) levels led to rather similar planar geometries. The root mean square deviation (RMSD) for the two structures is 0.01 Å and the N3-H3(U1)...O4(U2) and C5-H5(U2)...O4(U1) distances differ by only 0.02 and 0.01 Å, respectively (Table I, Figure 1a). With an RMSD value of 0.18 Å the MP2 optimized geometry of the isolated base pair deviates only slightly from the geometry of the Calcutta base pair within the X-ray structure (PDB code: 1osu; resolution: 1.40 Å), and with 2.19 Å the calculated H...O length of the C5-H5(U2)...O4(U1) contact is within the experimental error of the X-ray bond length (2.24 Å). Both experimental and calculated C-H...O contact distances are substantially lower than 2.7 Å, which is the sum of the van der Waals radii of H and O (13). This indicates, that the experimental geometry of the base pair is governed by the standard H-bond interaction and the C-H...O contact alone. Other effects, such as interactions with the backbone and stacking do not effectively change the base pair geometry.

According both MP2 and DFT calculations, the C5-H5 bond of uracil1 in the complex is by 0.002 Å longer than that of isolated uracil (Table I). The bond elongation at the DFT level is accompanied by a 34 cm^{-1} decrease (red shift) in the wavenumber of the stretching vibration

(Table V). The N3-H3 bond length involved in the standard H-bond increases by 0.014 Å (MP2) and 0.017 Å (DFT). Summarising, the C-H...O contact in the Calcutta UU base pair is predicted to fulfil both the van der Waals and the spectroscopic criteria expected for H-bonds, but the effects are an order of magnitude smaller than for the N3-H3...O4 H-bond.

The B3LYP/6-31G(d,p)//B3LYP/6-31G(d,p) interaction energy ΔE^T corrected for the basis set superposition error and deformation energies is -7.57 kcal/mol. This is smaller than the interaction energies found for base pairs linked by two H-bonds, which are between -9 and -24 kcal/mol (31). As expected, the interaction energy provided by a C-H...O contact is much smaller than the related energy for a standard H-bond. With -7.56 kcal/mol the MP2 ΔE^T value is practically identical and this further justifies the use of the DFT method for the frequency calculations (Table I).

Kratochvil et al. (37) have performed a quantum-chemical study on low-energy conformations of H-bonded and stacked UU base pairs including a structure with a C-H...O contact. In the latter complex the N1-H1 group is involved in an inter-base H-bond. However, in RNA structures only the N3-H3 group is available for H-bonding because N1 is linked to the sugar-phosphate backbone.

Our results on the Calcutta UU pair at the MP2(FC)/6-31G(d,p) level can be directly compared to the data reported by Hobza et al. (33). From a comparison of Tables I in both papers it turns out that the numerical data given are practically identical. For example, the interaction energies ΔE are –8.13 and – 8.1 kcal/mol, respectively. The calculated distances O4...N3, and C5...O4 have values of 2.913/2.91 Å, and 3.236/3.23 Å, where the data from ref. (33) are given first. Finally, the bond length changes found on base pair formation are

+0.014/+0.014 Å for the N3-H3 bond and +0.002/+0.003 Å for the C5-H5 bond. In ref. (33) the vibrational analysis has not been performed for the Calcutta UU pair. The red shift found for the symmetrical stretching vibration of the C5-H5 bond of a different UU base pair is 34 cm^{-1} at the MP2 level. We find a red shift of 34 cm^{-1} for the Calcutta pair and of 32 cm^{-1} for the uracil/formaldehyde complex.

The AA base pair

The geometry of the AA base pair has been optimized using the same methods as described above for the Calcutta base pair (Table II). B3LYP optimization resulted in a geometry that is essentially planar, but exhibits a significant non-planarity of the amino group not involved in hydrogen bonding. By contrast, in the structure obtained after MP2 optimization both the two amino groups of the adenines and as a consequence the whole complex are essentially non-planar.

As shown by quantum-chemical calculations of Šponer et al., many hydrogen bonded nucleic acid base pairs are intrinsically nonplanar, which is mainly due to the sp^3 hybridisation of the nitrogen atoms of their amino group and secondary electrostatic interactions (38). Note, however, that for several reasons this property is difficult to detect in experimental nucleic acid structures. In X-ray studies crystal effects may enforce a planarization of the amino group. Moreover, hydrogen atom coordinates are usually not determined. This means that any deviation from planarity can only be monitored by the shift of the nitrogen atom out of the plane defined by the aromatic ring. Finally, NMR structure determination is often done adopting force fields that do not take into account a non-planarity of the exocyclic amino group.

With an RMSD value of 0.14 Å the B3LYP geometry deviates more from the MP2 value than in the case of the Calcutta base pair. In the MP2 structure the angle between the planes of the two bases is 14° and mainly determined by the buckle component. As in case of the Calcutta pair, B3LYP and MP2 optimizations produce very similar results for the lengths of the standard H-bond and of the C-H....N interaction. In the MP2 structure the H2(A2)...N7(A1) distance in C-H...N has a length of 2.41 Å (Table II), which is larger than the H5(U2)...O4(U1) length in the C-H...O contact of the Calcutta base pair, but nevertheless considerably shorter than the sum of van der Waals radii of hydrogen and nitrogen (2.7 Å) (13). The calculated MP2 distance for the standard H-bond H6(A1)...N3(A2) is 2.07 Å.

The AA base pair has been observed twice in the structure of the group I intron ribozyme (29), and in tRNA^{Glu}/tRNA^{Glu}-synthetase complexes (8,30). Within these structures the length of the standard H-bond H6(A1)...N3(A2) is between 1.7 Å and 2.0 Å and the H2(A2)....N7(A1) distance of the C-H...N contact ranges between 1.9 Å and 2.9 Å. As X-ray structures containing the AA base pair are rather poorly resolved, comparisons between experimental and calculated geometries are not possible. The small interaction energy of AA as compared to Calcutta suggests that the geometry of AA can be more easily affected by the surrounding nucleic acid than that of the Calcutta base pair.

For the standard H-bond the typical elongation of the N-H bond length is observed upon H-bond formation (MP2: 0.010 Å, DFT: 0.013 Å), whereas for the C-H...N contact both MP2 and DFT calculations yield a shortening of the C-H bond length by about 0.002 Å (Table II). Consequently, the DFT frequency calculations indicate a slight blue shift (14 cm^{-1}) of the C2-H2 stretching vibration (isolated A: 3055 cm^{-1}, AA complex: 3069 cm^{-1}, Table V). This means that the H...N contact in the AA pair exhibits an anti-H-bond pattern. At first sight this is in line with results obtained by Shishkin et al. (26) for the C2-H2...O2 interaction in the AT base pair. These authors have found that the C-H stretching vibration shows a blue shift upon base pair formation, whereas no bond length change occurred in three significant digits. However, more recent MP2 calculations on the related AU pair have shown no change of the C2-H2 bond length and only a marginal frequency shift of about 2 cm^{-1} (33). According to our view, it is not unexpected that in the AU complex the effects for the C2-H2...O2 interaction are small because its H...O distance is rather long.

In contrast to the Calcutta base pair, the B3LYP/6-31G(d,p)//B3LYP/6-31G(d,p) ΔE^T value of -4.99 kcal/mol differs somewhat from the respective MP2/6-31G(d,p)//MP2/6-31G(d,p) value of -5.39 kcal/mol, which is likely to be due to the geometry differences described above (Table II). In spite of the larger size of the complex the interaction energy of the AA pair is smaller than that for the Calcutta pair. This can probably be attributed to the reduced strengths of N-H...N and C-H...N contacts as compared to the N-H...O and C-H...O interactions.

The uracil/ formaldehyde and adenine/formaldehyde model complexes

In order to study potential effects of neighbouring standard H-bonds and of the different acceptor atoms O and N in the C-H...O and C-H...N contacts the base pairs have been simplified by replacing the base containing the acceptor atom by formaldehyde (FA). The uracil/formaldehyde and adenine/formaldehyde model complexes (Figure 2) have been subjected to the same calculations as the base pairs. This way, both differences caused by different acceptor elements oxygen and nitrogen and influences of the neighbouring standard hydrogen bonds could be excluded.

Similar to the base pair optimisations MP2 and B3LYP methods produced rather similar results for the simplified model systems. With 2.29 Å, the H5(U)...O(FA) distance within the C5-H5(U)...O(FA) contact is slightly longer (0.1 Å) than its counterpart in the Calcutta base pair (Table III). This can be accounted for by the lack of a neighbouring standard H-bond. With a 0.002 Å elongation of the C5-H5 bond at the MP2 level and a 32 cm^{-1} red shift of the stretching frequency upon complex formation the C-H...O contact has rather similar properties as the one in the Calcutta base pair (Tables III, V). This clearly shows that the H-bond pattern observed for the C-H...O contact in the Calcutta UU base pair is an intrinsic property and is not enforced by the neighbourhood of a standard H-bond. The uracil/FA complex contains a second C-H...O contact between the C-H group of formaldehyde and O4 of uracil. In this case the C-H bond length within formaldehyde is decreased on complex formation by 0.004 Å and the DFT stretching frequency is shifted by 67 cm^{-1} towards larger wavenumbers. Hence, the uracil/formaldehyde complex contains both an H-bond and an anti-H-bond C-H...O interaction. Linked by C-H...O contacts only this complex offers the opportunity to estimate the order of magnitude of interaction energies of C-H...O contacts. With –3.68 kcal/mol and –3.62 kcal/mol the ΔE^T values from the MP2 and B3LYP calculations are again in very good agreement. The MP2 and B3LYP ΔE^T data as well as the ΔE_0 value at the B3LYP level (-2.37 kcal/mol) indicate that the interaction energy is larger than expected for two van der Waals contacts (Table III).

As in the uracil case the MP2 and B3LYP results for the adenine/formaldehyde complex are in excellent agreement (Table IV). In this complex both C-H...O/N contacts show shortenings of the C-H bond length and blue shifts of the related stretching vibration characteristic of the anti-H-bond pattern. With 0.001 Å at both the MP2 and the B3LYP levels and a B3LYP blue shift of 15 cm^{-1} the effects on the C2-H2 bond are considerably less pronounced than those on the formaldehyde C-H group (shortening of 0.004-0.005 Å and blue shift of 48 cm^{-1}) (Tables IV, V). With -1.71 and -1.36 kcal/mol the ΔE^T values at the MP2 and B3LYP levels (B3LYP ΔE_0 = -0.50 kcal/mol) are rather small, however. This is in line with the long C-H...O and C-H...N distances of 2.48 and 2.46 Å, respectively, found in the adenine/formaldehyde complex. The fact that the C-H...O distance is even slightly longer than the corresponding C-H...N distance in the AA base pair indicates that the character of this interaction mainly depends on the features of the donor group and only to a minor extent on the acceptor atom.

Possible influence of carbon charge on the interaction pattern

From the results obtained for the base/formaldehyde complexes it can be concluded that the neighbouring standard H-bonds have only a minor effect on the C-H...O/N interactions. In passing from the UU pair to the uracil/formaldehyde complex where the C5-H5...O4 interaction is replaced by the C5-H5...O (formaldehyde) contact the H...O distances are elongated by 0.1 Å at the MP2 level. In the adenine/formaldehyde complex the AA C2-H2...N7 contact is replaced by the C2-H2...O (formaldehyde) interaction and with 0.07 Å, the H...N/O elongation is somewhat smaller.

Both in the base pairs and in the base/formaldehyde complexes the C5-H(U) bond was elongated upon C-H...O/N contact formation, whereas the C2-H2(A) and C-H(FA) distances are shortened. The obvious difference between these donor groups is the partial charge at the carbon atom, which is negative for C5(U) (Mulliken analysis: -0.34), but positive for C2(A) (+0.34) and C(FA) (+0.24). It should be noted, that any scheme of assigning charges to nuclei is arbitrary to some extent. In particular, it is well-known that the absolute values of Mulliken charges are extremely sensitive to the basis set used for the calculation. However, the following discussion is based on charge differences in a qualitative sense only. In addition, the Mulliken charges for uracil and adenine are in qualitative agreement with the RESP (restrained electrostatic potential fit) charges used in the current AMBER force field and derived by a completely different procedure (39). Charge distributions are only slightly altered when the complexes are formed (data not shown).

In the following we want to describe an extremely simplistic model that might explain the C-H bond length compression or elongation. Let us represent the C-H...O/N interaction as three point charges with negatively charged O/N acceptor atoms, a positive H and a carbon atom that bears either a positive or a negative charge. Let us further assume that the H...O/N distance is kept fixed and then estimate the Coulomb electrostatic interaction energy as function of the C-H distance. Then, for a negatively charged carbon the interaction energy should increase when the C-H bond length is increased and decrease when it is shortened. By contrast, the opposite changes, namely decrease in interaction energy upon bond elongation and increase upon shortening would be expected for a positively charged donor atom. This model does definitely not tell us the whole story. On the other hand, it works at least for all C-H...O/N interactions in the UU and AA base pairs and the base/formaldehyde complexes and may stimulate the analysis of other systems.

A shortening of the C-H bond has been found in complexes between CH_4, CH_3F, CH_2F_2, CHF_3 and water, methanol or formaldehyde as acceptor (25), in dimeric or trimeric complexes between formaldehyde and water or ammonia (24) and within complexes involving C-H bonds and π-systems (22,23,27). The only examples where a C-H bond elongation upon complex formation has been observed thus far are the HCN/benzene and HCCH/benzene complexes (22,27). Gu et al. have pointed out that the carbon charge in CH_4, CH_3F, CH_2F_2 and CH_3F is respectively -0.915, -0.168, +0.436, and +0.944 and, nevertheless, in all cases the C-H bond is shortened upon complexation with water, methanol or formaldehyde. Therefore, these authors have concluded that electrostatics alone cannot account for the observed effects. Rather, an energy decomposition analysis for CH_2F_2/OH_2 and H_2O/OH_2 has shown that the various energy components behave in a similar way for C-H bond contraction and O-H bond stretching. Obviously, the forces pushing towards contraction are slightly larger than the elongation forces for the C-H case, whereas the opposite is true for the for the OH. Interestingly, however, in the work of Gu et al. (25) C-H bond shortenings increase as the carbon charge gets less negative or more positive. Together with our own results this indicates that donor-acceptor electrostatics is certainly relevant even though it is definitely not the only force that has to be taken into account.

Conclusions

The interaction energies stabilising the unusual UU and AA base pairs with one standard H-bond and an additional C-H...O or C-H...N contact are low in comparison with base pairs linked by two standard H-bonds. Both the C-H...N contact in the adenine-adenine base pair and the C-H...O contact in the Calcutta base pair exhibit H...O and H...N distances that are significantly shorter than the sums of the respective van der Waals radii. Whereas the C-H bond involved in the C-H...O contact of the Calcutta base pair becomes longer upon formation of the complex, the C-H group participating in the C-H...N interaction in the AA complex is shortened. Consequently, the frequency of the stretching vibration in the Calcutta base pair is shifted to smaller wavenumbers on complex formation, whereas that of the C-H...N contact is increased. This means that the C-H...O contact in UU exhibits an H-bond pattern, whereas the C-H...N interaction in AA is of the anti-H-bond type. In order to exclude neighbor effects of standard H-bonds model complexes of adenine and uracil with formaldehyde that contain only C-H...O/N intermolecular contacts have been studied. The data obtained for these systems are in agreement with the results obtained for the base pairs. These results suggest a preference for negatively charged donor carbon atoms to form H-bond like interactions and a preference for positively charged donors to form anti-H-bond like contacts. MP2 and density functional B3LYP geometries and interaction energies UU and AA base complexes are similar. Therefore, the computationally less expensive DFT approach can be applied to base complexes with weak C-H...acceptor interactions.

We are grateful to the Thüringer Ministerium für Wissenchaft, Forschung und Kultur for financial support and A. Jabs for helpful discussions.

References

1. G. Dirheimer, G. Keith, P. Dumas, and E. Westhof, Primary, secondary, and tertiary structures of tRNA. In tRNA, Structure, biosynthesis, and function. D. Söll and U. RajBhandary, editors. American Society for Microbiology, Washington. 93-126 (1995).
2. I. Tinoco, Jr, Structures of base pairs involving at least two hydrogen bonds. In The RNA world. R.F. Gesteland, and J.F. Atkins, editors. Cold Spring Harbor Laboratory Press, 603-607 (1993).
3. G.A. Jeffrey and W. Saenger, Hydrogen Bonding in Biological Structures. Springer-Verlag, Berlin. 247-268 (1991).

4. S. Moran, X.-F. Ren, S. Rumney, and E.T. Kool, *J. Am. Chem. Soc. 119,* 2056-2057 (1997).
5. M. Meyer and J. Sühnel, *J. Biomol. Struct. Dyn. 15,* 619-624 (1997).
6. M.C. Wahl, S.T. Rao, and M. Sundaralingam, *Nat. Struct. Biol. 3,* 24-31 (1996).
7. M. Brandl, K. Lindauer, M. Meyer, and J. Sühnel, *Theoret. Chem. Acc. 101,* 103-113 (1999).
8. M.A. Rould, J.J. Perona, and T.A. Steitz, *Nature 352,* 213-218 (1991).
9. S.R. Holbrook, C. Cheong, I. Tinoco Jr., and S.H. Kim, *Nature 353,* 579-581 (1991).
10. C.C. Correll, B. Freeborn, P.B. Moore, and T.A. Steitz, *Cell 91,* 705-712 (1997).
11. C.C. Correll, I.G. Wool, and A. Munishkin, *J. Mol. Biol. 292*, 275-287 (1999).
12. P. Auffinger and E. Westhof, *J. Mol. Biol. 292,* 467-483 (1999).
13. R. Taylor and O. Kennard, *J. Am. Chem. Soc. 104,* 5063-5070 (1982).
14. T. Steiner, *Chem. Commun.* 727-734 (1997).
15. M.C. Wahl and M. Sundaralingam, *Trends Biochem. Sci.* 22, 97-102 (1997).
16. Z.S. Derewenda, L. Lee, and U. Derewenda, *J. Mol. Biol. 252,* 248-262 (1995).
17. P. Auffinger and E. Westhof, *J. Mol. Biol. 274,* 54-63 (1997).
18. G.A. Leonard, K. Mc-Auley-Hecht, T. Brown, and W.N. Hunter, *Acta Cryst. D51,* 136-139 (1995).
19. I. Berger, M. Egli, and A. Rich, *Proc. Natl. Acad. USA 93*, 12116-12121 (1996).
20. G.A. Jeffrey and W. Saenger, Hydrogen Bonding in Biological Structures, Springer-Verlag, Berlin. 50, 95 (1991).
21. H. Yoshida, T. Harada, T. Murase, K. Ohno and H. Matsuura, *J. Phys. Chem. A 101,* 1731-1737 (1997).
22. P. Hobza, V. Špirko, H.L. Selzle, and E.W. Schlag, *J. Phys. Chem. A 102,* 2501-2504 (1998).

23. P. Hobza, V. Špirko, Z. Havlas, K. Buchhold, B. Reimann, H.-D. Barth, and B. Brutschy, *Chem. Phys. Lett. 299,* 180-186 (1999).

24. M. Masella, and J.-P. Flament, *J. Chem. Phys. 110,* 7245-7255 (1999).

25. Y. Gu, T. Kar, and S. Scheiner, *J. Am. Chem. Soc. 121,* 9411-9422 (1999).

26. O.V. Shishkin, J. Šponer, and P. Hobza, *J. Mol. Struct. 477,* 15-21 (1999).

27. E. Cubero, M. Orozco, P. Hobza, and F. J. Luque, *J. Phys. Chem. A 103,* 6394-6401 (1999).

28. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne, *Nucleic Acids Res. 28,* 235-242 (2000).

29. J.H. Cate, A.R. Gooding, E. Podell, K. Zhou, B.L. Golden, C.E. Kundrot, T.R. Cech, and J.A. Doudna, *Science 273,* 1678-1685 (1996).

30. J.G. Arnez and T.A. Steitz, *Biochemistry 35,* 14725-14733 (1996).

31. J. Šponer, J. Lezczynski, and P. Hobza, *J. Phys. Chem. 100,* 1965-1974 (1996).

32. J. Šponer, J. Lezczynski, and P. Hobza, *J. Comput. Chem. 17,* 841-849 (1996).

33. P. Hobza, J. Šponer, E. Cubero, M. Orozco, and F.J. Luque *J. Phys. Chem. 104,* 6286-6292 (2000).

34. M.J. Frisch, et al. Gaussian 94, revision E.1. Gaussian, Inc., Pittsburgh (1995).

35. S.B. Boys and F. Bernardi, *Mol. Phys. 17,* 553-577 (1970).

36. J.B. Foresman and Æ Frisch, Exploring chemistry with electronic structure methods, Gaussian Inc., Pittsburgh (1996).

37. M. Kratochvil, O. Engkvist, J. Sponer, P. Jungwirth, and P. Hobza, *J. Phys. Chem. A 102,* 6921-6926 (1998).

38. J. Šponer, J. Floriàn, P. Hobza and J. Lezczynski, *J. Biomol. Struc. Dyn. 13,* 827-833 (1996).

39. W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.A. Kollman, *J. Am. Chem. Soc. 117*, 5179-5197 (1995).

Table I.

Geometrical parameters (distance d (Å), angle a (°)) and interaction energies (kcal/mol) of the Calcutta base pair (U1U2) calculated by MP2 and DFT methods.

	MP2 [a)]	DFT/B3LYP [a)]	Exp [b)]
d[H3(U2)...O4(U1)]	1.89	1.87	1.78
d[N3(U2)...O4(U1)]	2.91	2.90	2.76
a[N3(U2) H3(U2) O4(U1)]	171	171	158
d[H5(U1)...O4(U2)]	2.19	2.19	2.24
d[C5(U1)...O4(U2)]	3.23	3.23	3.27
a[C5(U1) H5(U1) O4(U2)]	150	151	157
d[N3-H3(U2)]	1.026	1.030	
d[N3-H3(U)]	1.012	1.013	
d[N3-H3(U2) - d[N3-H3(U)]	+ 0.014	+ 0.017	
d[C5-H5(U1)]	1.080	1.084	
d[C5-H5(U)]	1.077	1.082	
d[C5-H5(U1)] - d[C5-H5(U)]	+ 0.002	+ 0.002	
ΔE	-8.13	-8.25	
$\Delta E^{DEF(U1)}$	0.36	0.44	
$\Delta E^{DEF(U2)}$	0.21	0.24	
$\Delta E^{T} = \Delta E + \Delta E^{DEF}$	-7.56	-7.57	
ΔE^{ZPE}	-	0.87	
$\Delta E_0 = \Delta E^{T} + \Delta E^{ZPE}$	-	-6.70	

[a] Basis set: 6-31G(d,p)

[b] Experimental data: PDB code: 1osu.[6]

Table II.

Geometrical parameters (distance d (Å), angle a (°)) and interaction energies (kcal/mol) in the adenine-adenine base pair (A1A2) calculated by MP2 and DFT methods.

	MP2 [a)]	DFT/B3LYP [a)]
d[H62(A2)...N3(A1)]	2.07	2.05
d[N6(A2)...N3(A1)]	3.08	3.06
a[N6(A2) H62(A2) N3(A1)]	179	180
d[H2(A1)...N7(A2)]	2.41	2.43
d[C2(A1)...N7(A2)]	3.39	3.42
a[C2(A1) H2(A1) N7(A2)]	162	161
d[N6-H62(A2)]	1.017	1.020
d[N6-H6(A)]	1.007	1.007
d[N6-H62(A2)] – d[N6-H6(A)]	+ 0.010	+ 0.013
d[C2-H2(A1)]	1.082	1.087
d[C2-H2(A)]	1.084	1.088
d[C2-H2(A1)] - d[C2-H2(A)]	- 0.002	- 0.001
ΔE	-5.96	-5.58
$\Delta E^{DEF(A1)}$	0.43	0.49
$\Delta E^{DEF(A2)}$	0.12	0.11
$\Delta E^{T} = \Delta E + \Delta E^{DEF}$	-5.39	-4.99
ΔE^{ZPE}	-	0.65
$\Delta E_0 = \Delta E^{T} + \Delta E^{ZPE}$	-	-4.34

[a] Basis set: 6-31G(d,p)

Table III.

Geometrical parameters (distance d (Å), angle a (°)) and interaction energies (kcal/mol) in the uracil-formaldehyde complex (UFA) calculated by MP2 and DFT methods.

	MP2 [a)]	DFT/B3LYP [a)]
d[H(FA)...O4(U)]	2.33	2.31
d[C(FA)...O4(U)]	3.36	3.32
a[C(FA) H(FA) O4(U)]	155	151
d[H5(U)...O(FA)]	2.29	2.29
d[C5(U)...O(FA)]	3.34	3.33
a[C5(U) H5(U) O4(FA)]	163	160
d[C-H(FA^{UFA})]	1.096	1.105
d[C-H(FA)]	1.100	1.110
d[C-H(FA^{UFA})]- d[C-H(FA)]	-0.004	-0.005
d[C5-H5(U^{UFA})]	1.079	1.084
d[C5-H5 (U)]	1.077	1.082
d[C5-H5(U^{UFA})]- d[C5-H5(U)]	+0.003	+0.002
ΔE	-3.83	-3.80
$\Delta E^{DEF(U)}$	0.06	0.08
$\Delta E^{DEF(FA)}$	0.09	0.10
$\Delta E^{T}= \Delta E+\Delta E^{DEF}$	-3.68	-3.62
ΔE^{ZPE}	-	1.25
$\Delta E_0= \Delta E^{T}+\Delta E^{ZPE}$	-	-2.37

[a] Basis set: 6-31G(d,p)

Table IV

Geometrical parameters (distance d (Å), angle a (°)) and interaction energies (kcal/mol) in the adenine/formaldehyde complex calculated by MP2 and DFT methods.

	MP2 [a)]	B3LYP [a)]
d[H(FA)...N3(A)]	2.46	2.47
d[C(FA)...N3(A)]	3.40	3.39
a[H(FA) C(FA) N3(A)]	143	140
d[H2(A)..O(FA)]	2.48	2.52
d[C2(A)...O(FA)]	3.43	3.45
a[H2(A) C2(A) O(FA)]	146	142
d[C-H(FA^{AFA})]	1.097	1.106
d[C-H(FA)]	1.100	1.110
d[C-H(FA^{AFA})]- d[C-H(FA)]	-0.003	-0.004
d[C2-H2 (A^{AFA})]	1.083	1.087
d[C2-H2 (A)]	1.084	1.088
d[C2-H2 (A)] – d[C2-H2 (A^{AFA})]	-0.001	-0.001
ΔE	-1.79	-1.44
$\Delta E^{DEF(A)}$	0.03	0.01
$\Delta E^{DEF(FA)}$	0.06	0.06
$\Delta E^{T} = \Delta E + \Delta E^{DEF}$	-1.71	-1.36
ΔE^{ZPE}	-	0.86
$\Delta E_0 = \Delta E^{T} + \Delta E^{ZPE}$	-	-0.50

a) Basis set: 6-31G(d,p)

Table V

Wavenumbers (cm^{-1}) of the stretching vibration for selected C-H groups in adenine, uracil, the AA and UU base pairs and the formaldehyde/base complexes (B3LYP/6-31G(d,p))

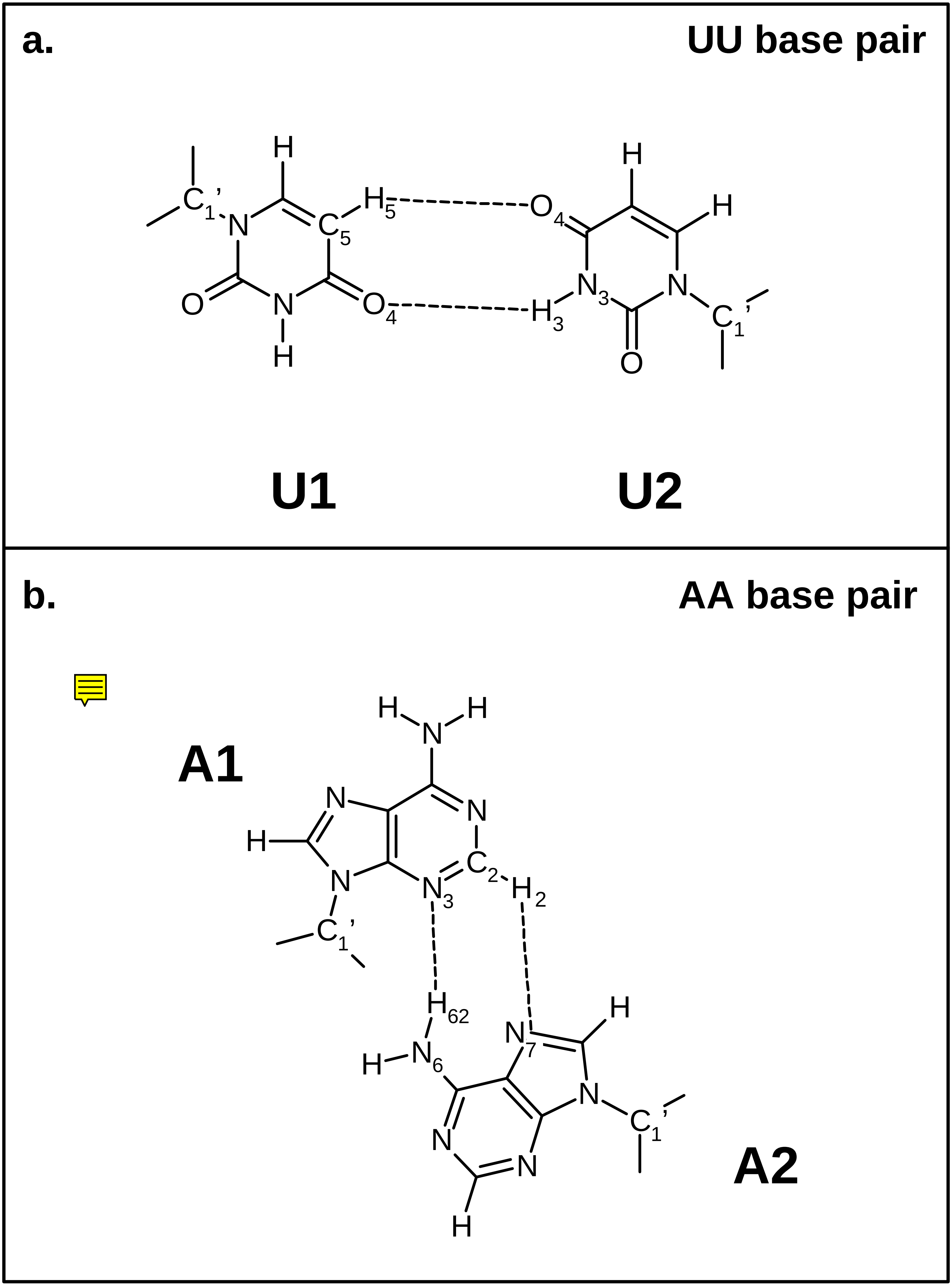
C-H donor groups	Wavenumber [a)]	Wavenumber difference
C5-H5(U)	3137	
C5-H5(UFA)	3105	
C5-H5(U1U2)	3103	
C2-H2(A)	3055	
C2-H2(AFA)	3070	
C2-H2(A1A2)	3069	
C-H(FA)	2841	
C-H(AFA)	2889	
C-H(UFA)	2908	
C5-H5(U1U2) - C5-H5(U)		- 34
C5-H5(UFA) - C5-H5(U)		- 32
C2-H2(A1A2) - C2-H2(A)		+ 14
C2-H2(AFA) - C2-H2(A)		+ 15
C-H(UFA) - C-H(FA)		+ 67
C-H(AFA) - C-H(FA)		+ 48

[a)] scaling factor: 0.9613

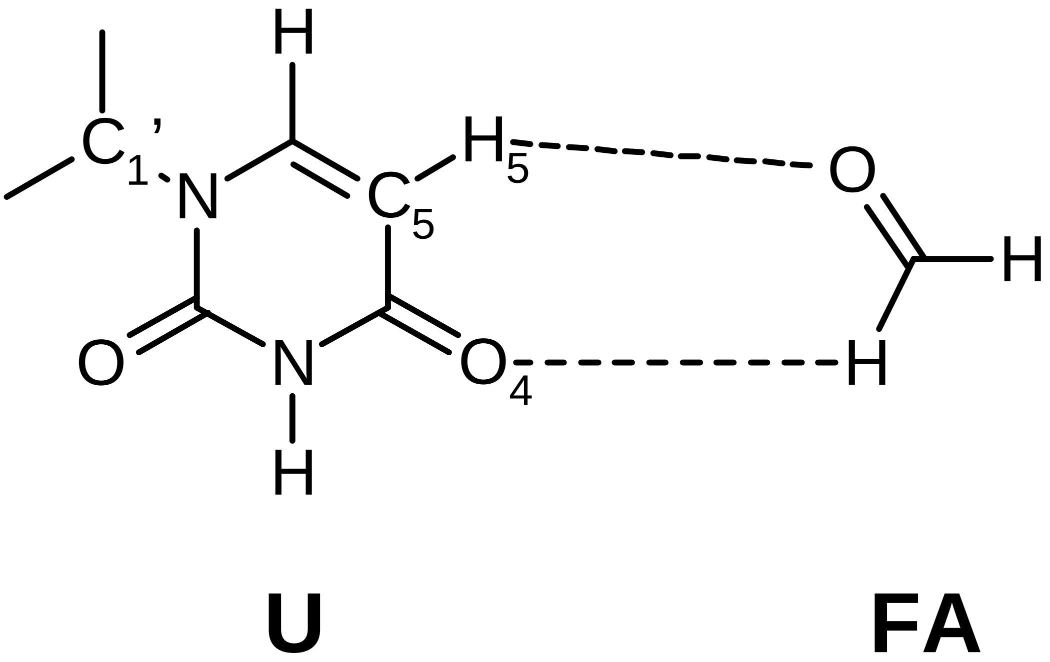
Figure Captions.

Figure 1: Structural formula of the UU (a.) , of the AA base pairs (b.).

Figure 2: Structural formula of the uracil/formaldehyde (UFA) (a.) and of the adenine/formaldehyde (AFA) (b.) complexes.


a.
UU base pair
U1
U2
b.
AA base pair
A1
A2

a. uracil / formaldehyde complex (UFA)



b. adenine / formaldehyde complex (AFA)