

## Do C—H···O Hydrogen Bonds Contribute to the Stability of Nucleic Acid Base Pairs?

BY GORDON A. LEONARD

*Department of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, England*

KATHERINE MCAULEY-HECHT AND TOM BROWN

*Department of Chemistry, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JJ, Scotland*

AND WILLIAM N. HUNTER\*

*Department of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, England*

(Received 18 January 1994; accepted 3 May 1994)

### Abstract

The possible formation of inter-base C—H···O hydrogen bonds in A·T, A·U and certain non-Watson–Crick base pairs is examined. A geometrical analysis in conjunction with implications for the thermodynamic stability of the base pairs suggests that C—H···O hydrogen bonds could form in nucleic acid base pairs. They may alleviate destabilizing interactions that would arise if an unsatisfied hydrogen-bond acceptor were present and mediate secondary hydrogen-bonding effects in these base pairs.

### 1. Introduction

Studies on the nature of non-Watson–Crick base-pair formation have characterized a number of hydrogen-bonding patterns that contribute to base-pair stability (Brown, Hunter & Leonard, 1993). X-ray analyses of the structures of A(*anti*)·O8G(*syn*) base-pairs and G(*anti*)· $\epsilon$ dA(*syn*) base pairs in DNA dodecamer structures where O8G = 7,8-dihydro-8-oxoguanine;  $\epsilon$ dA = 1,*N*<sup>6</sup>-ethenoadenine, (McAuley-Hecht *et al.*, 1994) have highlighted the possibility of C—H···O hydrogen-bond formation on the minor groove sides of certain nucleic acid base pairs including both the Watson–Crick and Hoogsteen conformations of the A·T or A·U pair (Fig. 1).

The existence of C—H···O hydrogen bonds has been the subject of controversy (Desiraju, 1991; Donohue, 1968). However, it has become widely accepted that they do form and contribute to the stability of many structures including those of biological significance (Desiraju, 1991; Jeffrey & Maluszynska, 1982; Jeffrey & Saenger, 1991; Steiner

& Saenger, 1993; Taylor & Kennard, 1982; Wiberg, Waldron, Schulte & Saunders, 1991). Nevertheless, to the best of our knowledge they have never been mentioned in the context of nucleic acid base-pair formation.

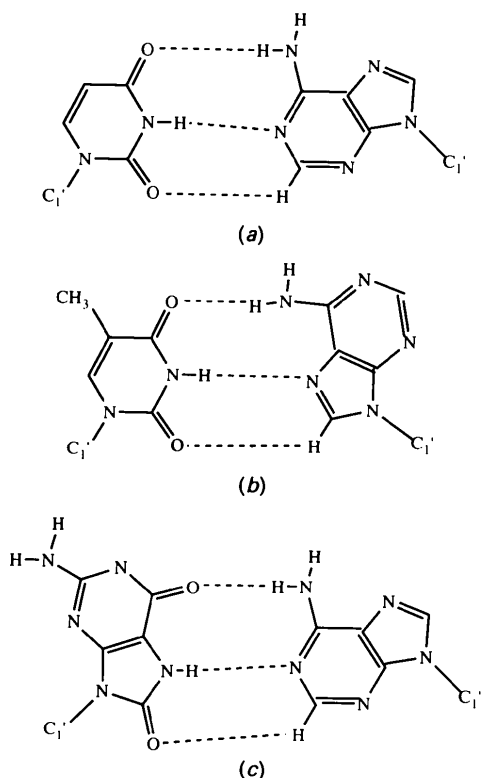


Fig. 1. Representations of (a) the Watson–Crick A·U base pair; (b) the A·T Hoogsteen base pair; (c) the A(*anti*)·O8G(*syn*) base pairs found in the structure of d[CGCAAATT(O8G)GCG]. In all cases the hydrogen-bonding scheme shows the C—H···O bond on the minor-groove side of the base pairs.

\* Author for correspondence.

Initially, it was thought that a hydrogen bond could be formed only if the donor-atom-acceptor-atom distance was less than the sum of their van der Waals radii (Hamilton & Ibers, 1968). It has since been pointed out that the interaction between the donor and the acceptor is mainly electrostatic in nature (Umeyama & Morokuma, 1977). Thus, the energy of a hydrogen bond diminishes linearly with increasing distance and, when considering C—H...O hydrogen bonds, we are no longer constrained to a maximum distance of say 2.6–2.9 Å. This range represents the sum of the van der Waals radii of C and H using values taken from a variety of sources (*e.g.* Muller, 1991). A further potential barrier to the formation of this type of hydrogen bond is the apparent charge on the H atom itself. Recently, Steiner & Saenger (1993) defined a hydrogen bond as 'any cohesive interaction X—H...Y, where H carries a positive and Y a negative (partial or full) charge, and the charge on X is more negative than on H'. This would seem to preclude the formation of hydrogen bonds by the majority of nucleotide base C—H groups as although the partial charge on the H atoms are positive (Saenger, 1984) in some cases (and this includes adenine H2) the charge on the C atom is more positive. However, it has been shown (Wiberg, Waldron, Schulte & Saunders, 1991) that C—H...O interactions reverse a C<sup>δ+</sup>—H<sup>δ-</sup> bond polarization in alkanes and it is likely that a similar effect would occur in other systems. For adenine H2 this suggests that the partial charge would increase from the positive value of 0.07 (Saenger, 1984) under the influence of a nearby O atom. We suggest that C—H...O hydrogen bonds can occur in base pairs. The question we seek to address is whether or not there is geometrical evidence to support the idea and to stimulate further studies, theoretical and experimental, in this area.

## 2. General considerations

Of note are two structures of nucleic acid base pairs determined at atomic resolution in which putative C—H...O hydrogen-bonding interactions may occur. These are the structures of ApU (Seeman, Rosenberg, Suddath, Kim & Rich, 1976), which was solved using X-ray diffraction and contains Watson-Crick A·U base-pairs (Fig. 1*a*), and the 1-methylthymine-9-methyl adenine base pair (Frey, Koetzle, Lehmann & Hamilton, 1973) solved by neutron diffraction and which contains Hoogsteen base pairs (Fig. 1*b*). Although for the ApU structure the positions of the H atoms were calculated, with a bond length of 1.08 Å, rather than directly observed, each structure contains quite accurate H-atom positions. In each case the primary interactions between paired bases are clearly attributed to the two N—H...O hydrogen

Table 1. H...O and C...O distances (Å), C=O...H angles (°) for A·T and A(anti)·O8G(syn) base pairs

Base pair	H...O	C...O	C=O...H
A·U*	2.8	3.6	130
A·T†	3.0	3.7	121
A·T‡	2.7	3.4	122
A(anti)·O8G(syn)§	2.8	3.5	122

\* Watson-Crick conformation (Seeman *et al.*, 1976).

† Hoogsteen conformation. From the structure of the 9-methyladenine-1-methylthymine base pair (Frey *et al.*, 1973).

‡ Average values from the four Watson-Crick base pairs in the structure of d[CGCAAATT(O8G)GCG] (McAuley-Hecht *et al.*, 1994). H-atom positions calculated as described in the text assuming a C—H bond length of 1.08 Å. The range of C...O distances is 3.2–3.5 Å.

§ Average of values, both of 3.5 Å, found in two base pairs (McAuley-Hecht *et al.*, 1994).

bonds. These two strong interactions are the major determinant of base-pair stability and serve to align the carbonyl and C—H groups.

Table 1 shows values of the H...O and C...O distances and the C=O...H angles observed in these structures. The distances are in good agreement with C—H...O hydrogen-bond lengths previously discussed in the context of small-molecule crystallography (*e.g.* Desiraju, Kashino, Coombs & Glusker, 1993). C=O...H angles, with values near to 120°, are also consistent with hydrogen-bond formation given that the lone-pair deformation density in C=O groups is generally resolved into two lobes in the directions predicted for *sp*<sup>2</sup> hybridization (Olovsson, 1982). This results in certain base pairs having one of the carbonyl lone pairs on one base directed towards a C—H H atom on the other base of the pair (Fig. 2). It has been demonstrated that hydrogen bonds involving carbonyl O atoms tend to form along the conventionally viewed *sp*<sup>2</sup> lone-pair directions (Taylor, Kennard & Versichel, 1983) and, hence, the C=O...H angles are consistent with the

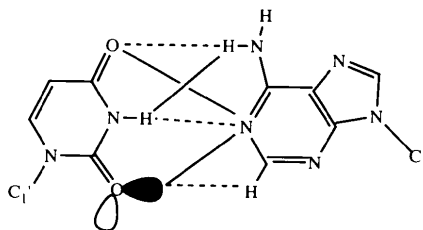


Fig. 2. A representation of a A·U or A·T base pairs showing the destabilization that would occur if the minor-groove carbonyl O atom of the pyrimidine (O2) were not involved in a C—H...O hydrogen bond. Primary hydrogen bonds are shown as dashed lines and secondary interactions as full lines. Formation of the third hydrogen bond would balance out any destabilization caused by the secondary interaction between thymine O2 and the N1 of the purine base. Note that, as mentioned in the text, the shaded *sp*<sup>2</sup> orbital points towards the C—H group on the purine and, thus, we expect the C=O...H angles for C—H...O hydrogen-bond formation to be near to 120°.

formation of C—H...O hydrogen bonds in these base pairs.

The software package *SYBYL* (Tripos Associates, St Louis, Missouri, USA) was used to generate ideal H-atom positions (C—H bond length 1.08 Å) for the structure of d[CGCAAATT(O8G)GCG] (McAuley-Hecht *et al.*, 1994). This structure is at an effective resolution of 2.5 Å, with an *R* factor of 16.8%. Restrained least-squares methods, using geometrical information from accurate small-molecule analyses, were employed in the refinement. Weak restraints were placed on normal Watson–Crick hydrogen bonds and so the values presented for the A·T pairs in Table 1 represents a reference set for what is considered as normal A·T base-pair geometry. No restraints were placed on the positioning of the A(*anti*)·O8G(*syn*) base pairs. Table 1 provides the relevant distances and angles for potential C—H...O hydrogen-bond formation in the A(*anti*)·O8G(*syn*) and for reference the A·T base pairs found in this dodecamer structure. It is clear that on the basis of geometrical considerations then C—H...O hydrogen bonds could also form in these base pairs (Figs. 1a, 1c) and that A·T and A·U base pairs could be depicted with three inter-base hydrogen bonds rather than the traditional two.

Whilst we can postulate the formation of inter-base C—H...O hydrogen bonds in some nucleic acid base pairs on purely geometric grounds there is a compelling thermodynamic reason to invoke their formation. It has long been appreciated (Crick, 1966) that the presence of an unfulfilled hydrogen-bond donor in the minor groove of certain DNA purine:purine wobble base pairs would have a large destabilizing influence. In the absence of C—H...O hydrogen bonds the base pairs we have been discussing would have, on the minor-groove side, an unfulfilled hydrogen-bond acceptor. As is the case of unfilled hydrogen-bond donors this situation would also cause the destabilization of the base pair as the secondary hydrogen-bonding interactions (Pranata, Wierschke & Jorgenson, 1991; Fig. 2) in which it could participate would not be balanced by a stabilizing hydrogen bond. Using optimized potentials for liquid-simulation methods the optimized energy of interaction of the Watson–Crick conformation of the 9-methyladenine·1-methylthymine base pair was calculated to be  $-43.5 \text{ kJ mol}^{-1}$  (Pranata, Wierschke & Jorgenson, 1991) compared to the experimentally determined value of  $-54.4 \text{ kJ mol}^{-1}$  (Yanson, Teplitsky & Sukhodub, 1979). If the contribution of an additional C—H...O hydrogen bond, (about  $8.37 \text{ kJ mol}^{-1}$  according to Desiraju, 1991) were to be included in these calculations the theoretical energy would be in extremely good agreement with the experimentally observed value. This seems a further reason to invoke the

presence of inter-base C—H...O hydrogen bonds in some nucleic acid base pairs. The exact contribution from C—H...O hydrogen bonds in these nucleic acid base pairs is not known. We anticipate that it would be much less than even  $4.18 \text{ kJ mol}^{-1}$  given the distances quoted in Table 1 and may vary depending on such factors as, for example, the surrounding sequence. Note, however, we judge that the important effect is not so much the contribution from a weak hydrogen bond as the alleviation of the destabilizing effect of having an unsatisfied hydrogen-bond acceptor.

The potential stabilization derived from C—H...O hydrogen bonds in base-pair associations may also be important in RNA structure. There are many examples of base interactions involving more than two bases or modified bases in structures as diverse as tRNA and TAR RNA (see Gesteland & Atkins, 1993) where these interactions may occur. We suggest that the inclusion of terms relating to such hydrogen bonds may enhance the quality of molecular modelling studies directed towards nucleic acid structure. Before such features can be realistically incorporated in modelling, further investigations in this area are warranted. Experimental and theoretical methods should be applied to more rigorously characterize the C—H...O interaction in nucleic acid base pairs.

We thank the Wellcome Trust and the Science and Engineering Council of the United Kingdom for support.

#### References

- BROWN, T., HUNTER, W. N. & LEONARD, G. A. (1993). *Chem. Br.* **29**, 484–488.
- CRICK, F. H. C. (1966). *J. Mol. Biol.* **19**, 548–555.
- DESIRAJU, G. R. (1991). *Acc. Chem. Res.* **24**, 290–296.
- DESIRAJU, G. R., KASHINO, S., COOMBS, M. M. & GLUSKER, J. P. (1993). *Acta Cryst.* **B49**, 880–892.
- DONOHUE, J. (1968). *Structural Chemistry and Molecular Biology*, edited by A. RICH & N. DAVIDSON, pp. 459–463. San Francisco: W. H. Freeman.
- FREY, M. N., KOETZLE, T. F., LEHMANN, M. S. & HAMILTON, W. C. (1973). *J. Chem. Phys.* **59**, 915–924.
- GESTELAND, R. F. & ATKINS, J. F. (1993). *The RNA World*. Cold Spring Harbour Laboratory Press.
- HAMILTON, W. C. & IBERS, J. A. (1968). *Hydrogen Bonding in Solids*. New York: W. A. Benjamin.
- JEFFREY, G. A. & MALUSZYNSKA, H. (1982). *Int. J. Biol. Macromol.* **4**, 173–185.
- JEFFREY, G. A. & SAENGER, W. (1991). *Hydrogen Bonding in Macromolecular Structures*. Berlin: Springer.
- MCAULEY-HECHT, K., GIBSON, N., LEONARD, G. A., WATSON, W. P., HUNTER, W. N. & BROWN, T. (1995). *Biochemistry*. In the press.
- MULLER, U. (1991). *Inorganic Structural Chemistry*. New York: John Wiley.
- OLOVSSON, I. (1982). *Croat. Chem. Acta*, **55**, 171–190.
- PRANATA, J., WIERSCHKE, S. G. & JORGENSEN, W. L. (1991). *J. Am. Chem. Soc.* **113**, 2810–2819.

- SAENGER, W. (1984). *Principles of Nucleic Acid Structure*. Berlin: Springer.
- SEEMAN, N. C., ROSENBERG, J. M., SUDDATH, F. L., KIM, J. J. P. & RICH, A. (1976). *J. Mol. Biol.* **104**, 109–144.
- STEINER, T. & SAENGER, W. (1993). *J. Am. Chem. Soc.* **115**, 4540–4547.
- TAYLOR, R. & KENNARD, O. (1982). *J. Am. Chem. Soc.* **104**, 5063–5070.
- TAYLOR, R., KENNARD, O. & VERSICHEL, W. (1983). *J. Am. Chem. Soc.* **105**, 5761–5766.
- UMEYAMA, H. & MOROKUMA, K. J. (1977). *J. Am. Chem. Soc.* **99**, 1316–1332.
- WIBERG, K. B., WALDRON, R. F., SCHULTE, G. & SAUNDERS, M. (1991). *J. Am. Chem. Soc.* **113**, 971–977.
- YANSON, I. K., TEPLITSKY, A. B. & SUKHODUB, L. F. (1979). *Biopolymers*, **18**, 1149–1170.