# 5-Hydroxyuracil can form stable base pairs with all four bases in a DNA duplex†

**Varatharasa Thiviyanathan, Anoma Somasunderam, David E. Volk and David G. Gorenstein***

*Received (in Cambridge, UK) 21st September 2004, Accepted 12th October 2004*
*First published as an Advance Article on the web 30th November 2004*
**DOI: 10.1039/b414474k**

**NMR studies, UV-monitored melting experiments, and *ab initio* calculations show that 5-hydroxyuracil, produced by the oxidative de-amination of cytosines by reactive oxygen species, can form stable base-pairs with dA, dG, dC and dT residues in a DNA duplex, providing a basis for the *in-vivo* incorporation of 5-hydroxyuracil during DNA replication.**

Oxidative DNA damage results from exposure to both endogenous and exogenous reactive oxygen species and has been implicated in a variety of diseases, including cancer, rheumatoid arthritis, and aging.[1] Oxidized cytosines contribute significantly to the GC to AT transition mutations in DNA, the most abundant base substitution mutation observed in aerobic organisms.[2] Several recent studies have established that the oxidized cytosine products are the major chemical precursors to the GC to AT transition mutations.[3] Oxidation of cytosine gives rise to 5,6-dihydroxy-5,6-dihydrocytosine (**1**), an unstable DNA lesion that either dehydrates or de-aminates or undergoes both reactions to form 4-amino-5-hydroxypyrimidine-2-one (5-OH–C) (**2**), 5,6-dihydroxy-5,6-dihydrouracil (**3**), or 5-hydroxy-2,4(1H,3H)-pyrimidinedione (5-OH–U) (**4**), respectively (Fig. 1).[4] Among the oxidized cytosine products, **4** shows the highest mutation frequency. Following oxidative damage to DNA, both **3** and **4** have been detected in human cells at levels comparable to those of 8-oxoguanine, a frequently observed lesion in DNA.[5] At least four different human enzymes are known to excise **4** lesions from DNA, suggesting that **4** is an important mutagenic lesion generated in significant amounts in the human genome and that it requires several back-up enzymes for its repair.[6]

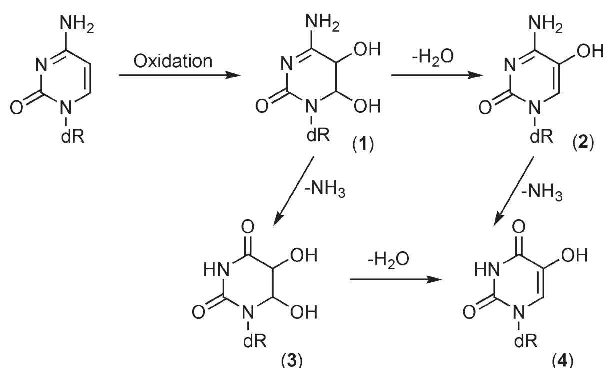


**Fig. 1** Oxidative products of cytosine.

Mammalian DNA polymerases are known to incorporate G, T, and A residues opposite to 5-OH–U[7] during DNA replication *in-vivo*, at different frequencies.[8] There is no evidence that C gets mispaired with 5-OH–U *in-vivo*. We used NMR, UV-melting and *ab initio* calculations to investigate the structures and relative stabilities of the base-pairs formed between the four DNA bases and 5-OH–U.

Compound **4** can undergo keto-enol tautomerism through a 1,3 hydrogen shift mechanism, the common form of tautomerism among carbonyl compounds containing an α-hydrogen. However, the enol form (at C-5 position) is favored because the enolic double bond is in conjugation, and this form is capable of forming an intra molecular hydrogen bond. The chemical structure of **4** shows that it is similar to that of thymidine and that it would form a normal base-pair with adenine.

*Ab initio* calculations[9] were used to investigate the interactions of the standard four DNA nucleosides (A, G, T and C) with 5-OH–U. The structures of the four nucleosides and the 5-OH–U consisted of the base and a deoxyribose ring in which methyl groups replaced the phosphate groups normally attached to the 3′ and 5′ oxygen atoms. The enthalpy changes for each base pair were calculated as the energy of the base pair minus the energies of the two nucleosides within the complex.

The results from these calculations, presented in Table 1, suggest that 5-OH–U can form stable base-pairs with all four standard DNA bases, with G forming the most stable base pair. The base-pairs with A, C and T are considerably less stable, with C forming the least stable base pair. Due to the repulsive interaction between the cytosine amino group and the 5-OH–U O4 oxygen atom in the 5-OHU:C base pair, the cytosine is twisted 40° relative to the 5-OH–U base. Thus, within a duplex DNA environment, the interaction between these two bases would be significantly less stable than calculated due to the energy required to accommodate this twisted base pair.

To experimentally investigate the formation of base-pairing of the DNA bases with 5-OH–U, NMR spectroscopy and

**Table 1** RHF Enthalpies of 5-OH–U interactions using the 6–311g (d,p) basis set

| Base pair | $\Delta G$/kcal mol$^{-1}$ |
|---|---|
| 5-OH–U:G | −13.8 |
| 5-OH–U:A | −10.9 |
| 5-OH–U:T | −9.9 |
| 5-OH–U:C | −9.7 |

† Electronic supplementary information (ESI) available: Chemical shift assignments for the imino protons, UV melting curves for all four DNA duplexes, and structures of the possible tautomeric forms, and the proposed hydrogen bonding of 5-OH–U with A, C, G and T. See http://www.rsc.org/suppdata/cc/b4/b414474k/
*david@nmr.utmb.edu

 This journal is © The Royal Society of Chemistry 2005

UV-monitored thermal melting experiments were done on the self-complementary DNA duplexes, d(CGCXAATTU*GCG)$_2$, (where U* is the 5-OH–U residue and X = A, T, C or G). This sequence is related to the *EcoR1* sequence that has been extensively studied by NMR spectroscopy, crystallography and computational methods.[10] Imino proton signals from the 1D NMR spectra‡ obtained for all four DNA duplexes are shown in Fig. 2. Downfield shifted, intense, sharp signals observed for the imino protons indicate that these protons are involved in hydrogen bonding.[11] For the 5-OH–U:A, 5-OH–U:T and 5-OH–U:G duplexes, the imino proton signals were sharp and intense. For the 5-OH–U:C duplex, the imino proton signals are downfield and intense, however, the linewidths are broader compared to those observed in the other three duplexes, suggesting that although 5-OH–U and C could form stable base-pairs, the stability of the 5-OH–U:C base pair is lower compared to the other three base-pairs involving 5-OH–U. In all four duplexes, the 1D NMR spectra collected at different temperatures did not show any evidence for the presence of multiple conformers, further supporting that the 5-OH–U residue can form stable base-pairs in the interior of the duplex.

Phosphorus chemical shifts, strongly influenced by the backbone torsional angles of the DNA duplex, are good indicators of local structural perturbations in DNA duplexes.[12] One-dimensional $^{31}P$ NMR spectra[13] were collected on all four DNA duplexes. Although the assignments of individual phosphorus signals was not possible due to severe overlap of the signals, all signals were seen within a narrow 0.9 ppm range. This observation shows that all residues in the DNA duplexes adopt the normal $B_I$ conformation,[12] providing further evidence that base-pairs formed between 5-OH–U and the four DNA bases do not cause significant structural perturbations in the DNA backbone. We compared the $^{31}P$ NMR spectra of the 5-OH–U:A containing duplex to that of a DNA duplex with the same sequence but with a U:A pair. The $^{31}P$ NMR spectra of both DNA duplexes were identical.

Two-dimensional NOESY spectra were collected on all four DNA duplexes, and the assignments of NOE cross peaks were achieved using established procedures.[11] In the 2D spectra, the intensities of the intra-residue NOE cross peaks between the H8/H6 aromatic proton and the H1′ sugar proton is lower than the intensities of the intra-residue cross peaks between the H5 and H6 protons in cytosines. This shows that all the residues, including the 5-OH–U and the base-paired residue, adopt the *anti* conformation about the glycosidic bond, normally observed in DNA duplexes. The *syn-anti* switch is determined primarily by the hydrogen bonding pattern of the nucleotides, with the presence of bulky substitutions in the base portion of the nucleotides playing a lesser role.[14] The additional hydroxyl group at the fifth position in 5-OH–U does not change the arrangement of hydrogen bond donor-acceptor groups in the Watson–Crick edge of this residue, and this pattern is similar to that observed in a U or a T residue.

To investigate the relative stabilities of the base-pairs involving the 5-OH–U in a DNA duplex, thermal melting of the DNA duplexes was monitored by UV absorbance at 260 nm. The sequences of the DNA duplexes used for melting studies are the same as used for the NMR studies. Melting profiles obtained for all four 5-OH–U containing DNA duplexes followed a sigmoidal curve with flat base-lines at both ends, typically observed for the two-state model.[15] The melting profiles also indicate the formation of stable base-pairs in the DNA duplexes. For the 5-OH–U:A duplex, where the base-pairing is exactly as the Watson–Crick type observed in a U:A pair, the melting temperature at low salt concentrations (100 mM NaCl), is significantly lower compared to that of the U:A containing duplex with identical sequence. However, at high salt concentrations (1 M NaCl), the melting temperatures of the two duplexes are essentially the same. Even though the hydrogen bondings between the bases in each strand are essentially the same for both duplexes, the 5-OH–U containing duplex has an additional –OH group in the major groove of the DNA. The DNA repair enzymes involved in excising the 5-OH–U lesions from damaged DNA may also recognize the

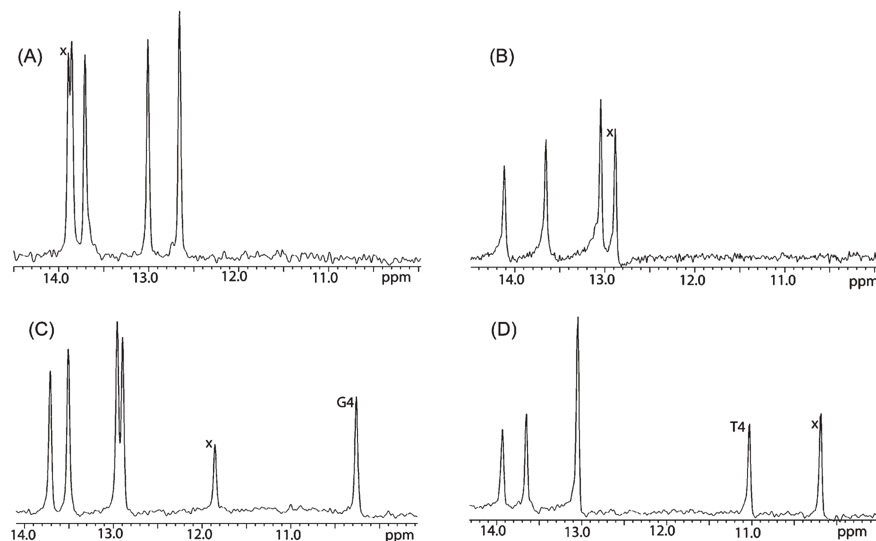


**Fig. 2** Imino proton region of the 1D proton NMR spectra. The imino proton signal of the 5-OH–U is marked with (x). (A) 5-OH–U:A duplex; (B) 5-OH–U:C duplex; (C) 5-OH–U:G duplex. The imino proton signal of G4 is marked; (D) 5-OH–U:T duplex. The imino proton on T4 is marked.

additional –OH group in the major groove of the DNA. Presently, we are probing the detailed tertiary structures of the 5-OH–U containing base-pairs in all four DNA duplexes and the structure of a complex formed between the 5-OH–U and the zinc finger motif of the DNA repair enzyme known to excise 5-OH–U from the damaged DNA.

The work was supported by the NIEHS center grant ES06676, the NIH grant AI27744, and the Welch Foundation (H-1296).

**Varatharasa Thiviyanathan, Anoma Somasunderam, David E. Volk and David G. Gorenstein***
*Department of Human Biological Chemistry & Genetics, University of Texas Medical Branch, Galveston, TX 77555, USA.*
*E-mail: david@nmr.utmb.edu; Fax: 409 747 6850; Tel: 409 747 6800*

## Notes and references

‡ The proton NMR data were collected at 15 °C and 750 MHz on a Varian UnityPlus instrument, and the signals were referenced to an external sample of DSS at 0.0 ppm.


1 (*a*) G. Mitra, G. T. Pauly, R. Kumar, G. K. Pei, S. H. Hughes, R. C. Moschel and M. Barbacid, *Proc. Natl. Acad. Sci. USA*, 1989, **86**, 8650; (*b*) R. Goth and M. F. Rajewsky, *Proc. Natl. Acad. Sci. USA*, 1974, **71**, 639.

2 (*a*) R. M. Schaaper, B. N. Danforth and B. W. Glickman, *J. Mol. Biol.*, 1986, **189**, 273–284; (*b*) E. C. Moraes, S. M. Keyse and R. W. Tyrrell, *Carcinogenesis*, 1990, **11**, 283; (*c*) L. K. Tkeshelashvili, T. Mcbride, K. Spence and L. A. Loeb, *J. Biol. Chem.*, 1991, **266**, 6401.

3 (*a*) A. A. Purmal, Y. W. Kow and S. S. Wallace, *Nucleic Acids Res.*, 1994, **22**, 3930–3935; (*b*) D. I. Feig, L. C. Sowers and L. A. Loeb, *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 6609–6613; (*c*) D. A. Kreutzer and J. M. Essigmann, *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 3578–3582.

4 M. Dizdaroglu, E. Holwitt, M. P. Hagan and W. F. Blakey, *Biochem J.*, 1986, **235**, 531–536.

5 J. R. Wagner, C. C. Hu and B. N. Ames, *Proc. Natl. Acad. Sci. USA*, 1992, **89**, 3380–3384.

6 (*a*) T. K. Hazra, T. Izumi, Y. W. Kow and S. Mitra, *Carcinogenesis*, 2003, **24**, 155–157; (*b*) T. K. Hazra, T. Izumi, I. Boldogh, B. Imhoff, Y. W. Kow, P. Jaruga, M. Dizdaroglu and S. Mitra, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 3523–3528.

7 The term '5-OH–U' is used when we refer to the 5-hydroxyuracil residue as a nucleotide in DNA.

8 A. Vaisman and R. Woodgate, *EMBO J.*, 2001, **20**, 6520–6529.

9 *Ab initio* calculations were performed using Gaussian98, revision A.11.3 at the RHF level of theory. BSSE energy corrections were not performed.

10 (*a*) R. E. Dickerson and H. R. Drew, *J. Mol. Biol.*, 1981, **149**, 761–786; (*b*) B. Wellenzohn, W. Flader, R. H. Winger, A. Hallbrucker, E. Mayer and K. R. Leidel, *Nucleic Acids Res.*, 2001, **29**, 5036–5043.

11 K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986.

12 (*a*) V. A. Roongta, C. R. Jones and D. G. Gorenstein, *Biochemistry*, 1990, **29**, 5245–5258; (*b*) C. Karslake, M. V. Botuyan and D. G. Gorenstein, *Biochemistry*, 1992, 1849–1858.

13 Phosphorus NMR data were collected at 240 MHz on a Varian UnityPlus-600 instrument. $^{31}$P chemical shifts were referenced to TMP at 0.0 ppm, which resonates 3.45 ppm downfield of 85% phosphoric acid.

14 V. Thiviyanathan, A. Somasunderam, T. K. Hazra, S. Mitra and D. G. Gorenstein, *J. Mol. Biol.*, 2003, **325**, 433–442.

15 (*a*) J. D. Puglisi and I. Tinoco, *Methods Enzymo.*, 1989, **80**, 304–325; (*b*) C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry; Techniques for the Study of Biological Structure and Function*, W. H. Freeman, New York, 1990.



This journal is © The Royal Society of Chemistry 2005