

## Quantum-Chemical Study of a Water-Mediated Uracil–Cytosine Base Pair

Maria Brandl, Michael Meyer, and Jürgen Sühnel\*

*Biocomputing, Institut für Molekulare Biotechnologie  
Beutenbergstrasse 11, D-07745 Jena, Germany*

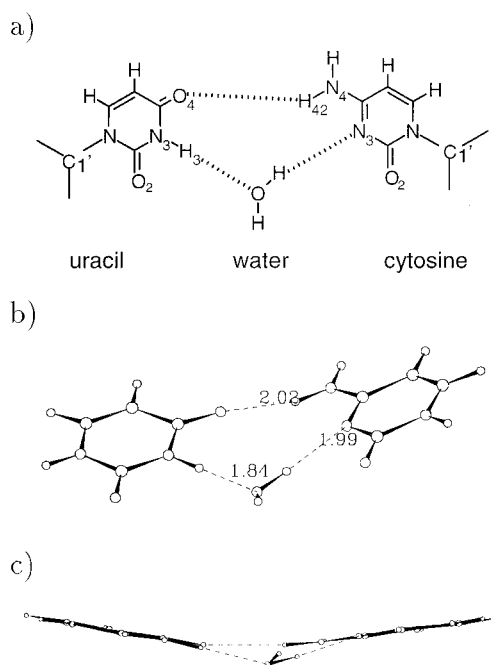
Received August 19, 1998

Hydrogen bonding within base pairs is one of the decisive factors that determine structure and stability of nucleic acids. It is well-known, especially from RNA structures, that in addition to Watson–Crick base pairs noncanonical pairs can be formed. All of them have at least two direct standard hydrogen bonds (H-bonds).<sup>1</sup> Recently, however, base pairs with only one or even no direct standard H-bond have been found. In these cases N–H···F,<sup>2,3</sup> C–H···O,<sup>4</sup> C–H···N<sup>5</sup> contacts, and water-mediated H-bonds<sup>6,7</sup> play a role. Thus far, it is not clear, whether the geometry of these complexes is a consequence of their intrinsic properties or enforced by backbone or stacking restraints exerted by the nucleic acid environment. Moreover, their stability remains to be assessed. Quantum-chemical ab initio calculations with the inclusion of correlation energy have provided a relatively consistent picture of Watson–Crick and noncanonical base pairs.<sup>8</sup> Therefore, these studies can be extended to the unusual complexes.

Until now water-mediated base pairs have only been detected in two RNA structures.<sup>6,7</sup> We have performed ab initio calculations on a water-mediated uracil–cytosine base pair (WUC). Our results indicate that the WUC complex is a structurally autonomous building block of RNA structure with a high degree of cooperativity.

In 1991 Holbrook et al. found that in the RNA duplex (r-GGACUUCGGUCC)<sub>2</sub> the central GU and UC mismatches do not form an internal loop, but rather a highly regular helix.<sup>6</sup> In the UC base pair H4 of cytosine is hydrogen-bonded to O4 of uracil, and H3 of uracil and N3 of cytosine are linked by H-bonds to a tightly associated water molecule (temperature factor, 11.7 Å<sup>2</sup>). A schematic drawing of the complex is shown in Figure 1a and selected experimental distances are listed in Table 1. The C1'(U)···C1'(C) distance is only slightly larger than the corresponding distances of Watson–Crick pairs in an ideal helix. Hence, incorporation of the WUC base pair leads only to minor distortions of the helix geometry.<sup>6</sup>

For the quantum-chemical calculations only the base parts of the nucleotides have been used in the same manner as in the studies of Šponer et al.<sup>8</sup> The geometry of the complex has been optimized at the HF/6-31G(d,p) level of theory. Energy minima of the optimized complex have been verified at the Hatree–Fock (HF) level by calculation of the Hessian. For the single point calculation of the interaction energy  $\Delta E^{\text{MP2}}$  electron correlation has been taken into account by means of second-order Møller–Plesset perturbation theory (MP2/6-31G(d,p)//HF/6-31G(d,p)). The interaction energies have been corrected for the basis set



**Figure 1.** Structure of the water-mediated UC (WUC) base pair: (a) chemical formula, (b) calculated geometry (top view), (c) calculated geometry (side view).

**Table 1.** Calculated and Experimental Inter-Base Pair Distances ( $d_{\text{calc}}$ ,  $d_{\text{exp}}$ ) in the Water-Mediated (WUC) and Direct UC Base Pairs without Water (UC) and with One Additional Water (UC + H<sub>2</sub>O)

	WUC		UC	UC + H <sub>2</sub> O <sup>c</sup>
	$d_{\text{calc}}/\text{Å}$	$d_{\text{exp}}/\text{Å}^a$	$d_{\text{calc}}/\text{Å}$	$d_{\text{calc}}/\text{Å}$
O4(U)···H4(C)	2.02	1.8	1.97	1.96
O4(U)···N4(C)	2.98	2.7	2.97	2.96
H3(U)···O(H <sub>2</sub> O)	1.84	1.8		
N3(U)···O(H <sub>2</sub> O)	2.85	2.8		
H(H <sub>2</sub> O)···N3(C)	1.99			
O(H <sub>2</sub> O)···N3(C)	2.95	2.8		
H3(U)···N3(C)			2.14	2.17
N3(U)···N3(C)			3.14	3.16
C1'(U)···C1'(C)	11.69 <sup>b</sup>	11.7 <sup>b</sup>	8.79 <sup>b</sup>	8.86 <sup>b</sup>
O2(U)···O2(C)	6.79	6.8	3.73	3.95
H <sub>1</sub> (H <sub>2</sub> O)···O2(U)				2.27
O(H <sub>2</sub> O)···O2(U)				3.01
H <sub>2</sub> (H <sub>2</sub> O)···O2(C)				2.08
O(H <sub>2</sub> O)···O2(C)				2.98

<sup>a</sup> Experimental data according to Protein Data Bank entry: 255d<sup>6</sup>.

<sup>b</sup> Calculated from the optimized geometry assuming a N1–C1' distance of 1.48 Å. <sup>c</sup> Only data for one of the two conformations are given. The values for the other conformation are similar.

superposition error (BSSE) by the standard counterpoise method.<sup>9</sup> A further correction was done for the deformation energies  $\Delta E^{\text{DEF}}$ , which are defined as the energy differences between the geometry of the optimized monomers and the structures of the monomers adopted in the complex. Finally, the changes in zero-point vibrational energies  $\Delta E^{\text{ZPE}}$  upon complex formation calculated at the HF/6-31G(d,p) level have been taken into account. The resulting total interaction energies are denoted as  $\Delta E_0$ . All ab initio calculations have been performed with the GAUSSIAN 94 package.<sup>10</sup>

(9) Boys, S. B.; Bernardi, F. *Mol. Phys.* **1970**, *17*, 553–577.

(10) Gaussian 94; Revision E.1; Frisch, M. J. et al. Gaussian; Inc.: Pittsburgh PA; 1995.

(1) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1983.

(2) Moran, S.; Ren, X.-F.; Rumney, S.; Kool, E. T. *J. Am. Chem. Soc.* **1997**, *119*, 2056–2057.

(3) Meyer, M.; Sühnel, J. *J. Biomol. Struct. Dyn.* **1997**, *15*, 619–624.

(4) Wahl, M. C.; Rao, S. T.; Sundaralingam M. *Nat. Struct. Biol.* **1996**, *3*, 24–31.

(5) Brandl, M.; Lindauer, K.; Meyer, M.; Sühnel, J. *Theor. Chem. Acc.*, in press.

(6) Holbrook, S. R.; Cheong, C.; Tinoco, I. Jr.; Kim, S. H. *Nature* **1991**, *353*, 579–581.

(7) Correll, C. C.; Freeborn, B.; Moore, P. B.; Steitz, T. A. *Cell* **1997**, *91*, 705–712.

(8) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965–1974.

**Table 2.** Interaction Energies (kcal/mol) in the Water-Mediated (WUC) and Direct UC Base Pairs without Water (UC) and with One Additional Water (UC+ H<sub>2</sub>O)

	WUC	UC	UC + H <sub>2</sub> O
$\Delta E^{\text{HF}a}$	-20.71	-10.78	-18.37
$\Delta E^{\text{MP2}a}$	-22.79 <sup>f</sup>	-11.95 <sup>f</sup>	-20.00 <sup>f</sup>
$\Delta E^{\text{MP2(UC)}b}$	-6.00		-11.82
$\Delta E^{\text{MP2(CW)}b}$	-7.72		-5.88
$\Delta E^{\text{MP2(UW)}b}$	-6.19		-2.50
$\Delta E^3c$	-2.88		0.20
$\Delta E^{\text{DEF(U)}d}$	0.63	0.76	0.91
$\Delta E^{\text{DEF(C)}d}$	0.52	0.34	0.57
$\Delta E^{\text{DEF(H2O)}d}$	0.08		0.14
$\Delta E^{\text{T}} = \Delta E^{\text{MP2}} + \Delta E^{\text{DEF}}$	-21.56	-10.85	-18.38
$\Delta E^{\text{ZPE}e}$	3.32	1.26	3.44
$\Delta E_0 = \Delta E^{\text{T}} + \Delta E^{\text{ZPE}}$	-18.24	-9.59	-14.94

<sup>a</sup> Hartree–Fock and MP2 interaction energy. <sup>b</sup> Pairwise MP2 interaction energy. <sup>c</sup> Three-body term. <sup>d</sup>  $\Delta E^{\text{DEF}}$ : deformation energies. <sup>e</sup>  $\Delta E^{\text{ZPE}}$ : change in zero-point energy upon complex formation, scaling factor: 0.9. <sup>f</sup> Energies of the minimized complexes (Hartrees): WUC: -883.73081; UC: -807.49060; UC + H<sub>2</sub>O: -883.72926

Figures 1b and 1c illustrate the calculated equilibrium geometry of the base complex and selected inter-base pair distances are listed in Table 1. The three H-bonds of the crystal structure are conserved in the optimized complex. The C1'(U)–C1'(C) distance, that has been calculated from the optimized structure by adopting a C1'–N1 distance of 1.48 Å, is 11.7 Å. This value is in agreement with the crystal structure distance. The complete WUC pair deviates significantly from a common plane (Figure 1). The angle between the planes of the bases is approximately 18° and mainly determined by the buckle component. Cytosine and water are located on different sides with respect to the plane defined by uracil. This deviation from planarity is obviously due to the sp<sup>3</sup> hybridization of the water oxygen atom and should be typical for those water-mediated base pairs in which water is acting both as acceptor and donor. Recent ab initio calculations on an isocytosine–cytosine base pair with a standard Watson–Crick H-bond pattern have predicted that the inclusion of at least six water molecules as a model for the first solvation sphere leads to a similar geometric distortion.<sup>11</sup>

The total interaction energy  $\Delta E_0$  of the WUC base pair corrected for the zero-point energy contribution is -18.24 kcal/mol (Table 2). This value as well as the electron correlation contribution at the MP2 level and deformation energies are in the range of interaction energies found for other canonical and noncanonical base pairs.<sup>8</sup>

The interaction energy  $\Delta E^{\text{MP2}}$  can be understood as the sum of three pairwise dimer contributions and a three-body term  $\Delta E^3$ , that accounts for cooperative effects.<sup>12</sup>

$$\Delta E^{\text{MP2(WUC)}} = \Delta E^{\text{MP2(UC)}} + \Delta E^{\text{MP2(CW)}} + \Delta E^{\text{MP2(UW)}} + \Delta E^3$$

All interaction energies have been calculated for the optimized geometry of the WUC complex and have been corrected for the basis set superposition error in the trimer-centered basis set.  $\Delta E^3$  has a value of -2.88 kcal/mol, which indicates strong cooperativity (Table 2).

For comparison we have investigated UC base pairs directly linked by two H-bonds between H4(C)–O4(U) and H3(U)–N3(C). We found the interaction energy  $\Delta E_0$  of the water-free pair to be -10.22 kcal/mol. This is consistent with the value of -9.5 kcal/mol calculated by Šponer et al.<sup>8</sup> for the analogous CT base pair. Repulsion between O2(C) and O4(U) leads to a propeller

twist of about 25°. In the direct UC base pair the C1'(U)–C1'(C) distance is 8.79 Å and thus about 2 Å smaller than those of canonical base pairs. Therefore, the direct UC base pair fits much worse into A-RNA than the water-mediated WUC complex does. After addition of water to the complex we obtained two conformations in which water bridges O2 of uracil and O2 of cytosine and is located either above or below an assumed common base pair plane. The additional water does not significantly change the overall geometry of the UC pair. With -15.32 kcal/mol (Table 2) and -14.94 kcal/mol the interaction energies of these two conformations are very similar and significantly less negative than that of the WUC complex. This shows that the WUC pair adopts a unique structure with a maximum enthalpic gain.

Our results indicate that the WUC base pair is structurally autonomous and that it is the preferred conformation of UC mismatches given they are not affected by any other forces of the nucleic acid environment in RNA. We do know from the crystal structure of Holbrook et al. that the WUC pair indeed occurs in this special environment.<sup>6</sup> We do not know yet whether this is also the case if the UC pair is flanked by other base pairs in an RNA helix or by other RNA structure elements. Tandem UC mismatches in RNA octamers with adjacent Watson–Crick base pairs have been shown to lack the imino resonance signal of N3 of uracil which suggests that the imino proton is rapidly exchanging with water and forms no direct H-bond with the other base.<sup>13</sup> This is a further indication that in this structure UC mismatches do not form direct UC base pairs with two standard H-bonds.

Recently, an unusual UC base pair has been found in the central hairpin of the HDV antigenomic ribozyme structure solved by NMR spectroscopy.<sup>14</sup> It has only one H-bond between H4(C) and O2(U), and the optimized geometry of the isolated base pair is completely different from the geometry within the nucleic acid structure. Therefore, it is very likely that the geometry of this base pair is primarily governed by restraints from the nucleic acid structure.

For three-dimensional RNA structures the examples mentioned are the only UC pairs known thus far. However, UC mismatches are predicted to occur in helical regions of secondary structure models.<sup>13,15</sup>

Thermodynamic studies on the stabilities of RNA octamers (absorbance versus temperature melting curves) have shown that UC tandem mismatches flanked by Watson–Crick pairs are destabilizing the nucleic acid duplex.<sup>13</sup> In light of these data the large calculated gas-phase interaction energy for the WUC pair seems to be surprising, given that one would directly correlate the calculated interaction energies with the thermodynamic data. Yet, for several reasons such a correlation is not possible. As suggested by one of the reviewers it is likely that removing water from the bulk and incorporating it into the WUC pair is unfavorable in the entropic sense. Moreover, the thermodynamic data represent effective quantities, which take into account both intrinsic base-pairing properties and effects originating from stacking and backbone interactions. However, these studies do not yield any information on the geometry of the base pairs. Therefore, the thermodynamic measurements have to be supplemented by other experimental and theoretical studies.<sup>16</sup> We hope that our results can contribute to that aim.

**Acknowledgment.** We thank Peter Slickers, Christoph Schneider, and Dr. Andreas Jabs for fruitful discussions and the TMWFK for financial support.

JA9829923

(13) Wu, M.; McDowell, J. A.; Turner, D. H. *Biochemistry* **1995**, *34*, 3204–3211.

(14) Kolk, M. H.; Heus, H. A.; Hilbers, C. W. *EMBO J.* **1997**, *16*, 3685–3692.

(15) Gutell, R. R. *Nucleic Acids Res.* **1994**, *22*, 3502–3507.

(16) Turner, D. H. *Curr. Opin. Struct. Biol.* **1996**, *6*, 299–304.

(11) Zhanpeisov, N. U.; Leszczynski, J. *Int. J. Quantum Chem.* **1998**, *69*, 37–47.

(12) Šponer, J.; Burda, J. V.; Mejzlík, P.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1997**, *14*, 613–628.