

# Homopairing Possibilities of the DNA Base Thymine and the RNA Base Uracil: An ab Initio Density Functional Theory Study

R. E. A. Kelly\* and L. N. Kantorovich

Department of Physics, School of Physical Sciences and Engineering, King's College London, Strand, London WC2R 2LS, United Kingdom.

Received: September 29, 2005; In Final Form: December 6, 2005

All planar homopairings of the DNA base thymine and the RNA base uracil are reported for the first time in this study. Using the idea of *binding sites* discussed in our previous work (Kelly et al. *J. Phys. Chem. B* 2005, 109, 11933; *J. Phys. Chem. B* 2005, 109, 22045) and ab initio density functional theory, we predict and relax 10 thymine and 10 uracil homopairs. The stabilization energies of the homopairs vary from just below zero to  $-0.82$  eV. The results on the pair geometry and energetics are compared with those available in the literature. The collected data on all planar thymine and uracil homopairs can be used to construct the thymine and uracil superstructures seen experimentally on various surfaces.

## 1. Introduction

The DNA and RNA bases molecules (adenine, cytosine, guanine, and either thymine or uracil for DNA and RNA, respectively) exhibit an ability to form nanoscale assemblies spontaneously on many surfaces, as has been demonstrated using in situ experimental imaging techniques such as scanning tunneling microscopy (STM; see refs 1–17 and references therein) and atomic force microscopy (AFM).<sup>3,4,18,19</sup> The importance of self-assembly mechanisms responsible for the creation of the superstructures cannot be underestimated, as these are thought to be very fundamental for the evolution process before complex genetic organisms were created.<sup>12</sup> Note that the monolayers rather than bulk crystal structures are the preferred choice for prebiotic evolution due to the vast number of adsorbate possibilities that may be available. The possibility for these molecules to form one- and two-dimensional superstructures is also interesting from a nanotechnological perspective.<sup>20</sup>

All of the DNA and RNA bases have shown the ability to form monolayers at the solid–liquid interface (see ref 12 and references therein). However, recent research has shown that by studying in situ experiments using inert surfaces and performing the experiments in ultra-high-vacuum (UHV) conditions more functionally varied structures form on the relevant substrate. Good examples of this are the copper (111),<sup>5,8,10,9,21,16</sup> the copper (110),<sup>22</sup> and the gold (111)<sup>17,23</sup> surfaces.

Unfortunately, at present, the molecular resolution from STM experiments can give little *atomistic* information about the superstructures formed. The detailed molecular packing or atomic resolution is rarely seen in STM, so theoretical calculations must be performed to obtain the actual atomistic structures. Semiempirical or semiclassical calculations have been previously used to determine the superstructures.<sup>16,24</sup> However, the difference between ab initio calculations and semiclassical or semiempirical techniques is too large to justify the usage of approximate techniques.<sup>25,26,27</sup>

The DNA base thymine was seen to form two-dimensional islands on the copper (111) surface.<sup>5,10</sup> The base uracil forms

trimers at low coverages, while at higher coverages larger hexagonal superstructures are formed.<sup>8</sup> It is interesting to note that STM experiments on the highly oriented pyrolytic graphite (HOPG) and molybdenum disulfide surfaces performed in air showed that both uracil and thymine self-assemble into two-dimensional arrays or monolayers.<sup>7,28</sup>

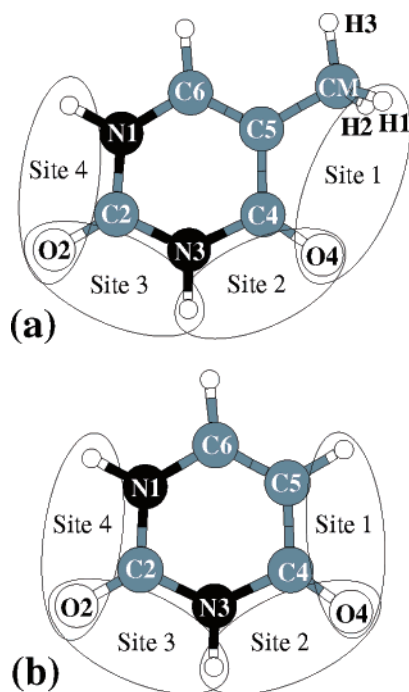
There are not many theoretical calculations on these superstructures reported in the literature. For instance, possible structural models were suggested for the uracil trimer<sup>8</sup> and the uracil monolayer in air.<sup>28</sup> A thymine superstructure on the copper (111) surface was also considered in ref 16, where only three thymine homodimers were used. However, in our previous work on adenine,<sup>26</sup> cytosine, and guanine<sup>27</sup> homopairs, we found that many distinct planar possibilities should be considered, and this can be done in a systematic way. From the 21 possible adenine homopairs<sup>26</sup> we found thousands of possible monolayer structures,<sup>29</sup> many of which have similar lattice vectors. We stress that in the two-dimensional (2D) case the most energetically favorable structures are not necessarily formed using only the most stable pairs; less stable pairs also may be involved to accommodate additional bonding between nearest molecules.

In this paper, in section 3 we present all possible planar pairs of the nucleic acid bases thymine and uracil using an ab initio method based on density functional theory (DFT). This is outlined in section 2. To our knowledge, no previous work has presented a complete analysis of all planar possibilities of thymine and uracil homopairs. The results of this study should serve as a natural basis for the understanding of the superstructures containing thymine and uracil bases.

## 2. Methods

Ab initio DFT calculations were performed using the SIESTA method<sup>30,31,32</sup> described previously in our work on adenine,<sup>26</sup> cytosine, and guanine<sup>27</sup> homopairs. The SIESTA method employs localized numerical atomic orbital basis sets, norm-conserving pseudopotentials, and periodic boundary conditions. We use the Perdew, Becke, and Ernzerhof (PBE)<sup>33</sup> exchange–correlation functional, which was found to be more than suitable for DNA base pairs.<sup>26</sup> The double- $\zeta$  plus polarization (DZP)

\* Author to whom correspondence should be addressed. E-mail: ross.kelly@kcl.ac.uk.



**Figure 1.** (a) Thymine and (b) uracil molecules in configurations T and U. The configurations  $\bar{T}$  and  $\bar{U}$  are obtained by flipping T and U, respectively, in the molecular plane. The possible bonding sites that can participate in forming a double hydrogen bond in homopairs are explicitly indicated. The atom label nomenclature used is the same as that in refs 34 and 35 for thymine and uracil, respectively.

basis set was used for all calculations defined by an energy cutoff of 10 meV.<sup>32</sup>

Atomic relaxation of all atoms in our calculations was complete when the force on each atom was not greater than 0.01 eV/Å. A large cell size was preferred to avoid unphysical interaction between adjacent periodic images, and therefore only one (Gamma)  $k$ -point was required for our calculations. Note that, completely opposite to the case of a plane wave basis set implemented in other DFT codes, increasing the cell size does not really affect the SIESTA performance due to the localized basis set used.

Some properties of the thymine and uracil molecules, shown in Figure 1, such as the intramolecular bond lengths and angles, compare well with previous DFT<sup>36</sup> and quantum chemical (QC) calculations.<sup>37</sup> For thymine, bond lengths calculated using plane wave DFT calculations compare to within 1.3% on average. For thymine and uracil bond lengths calculated using the Hartree–Fock (HF) theory, we find a slightly larger average deviation of 2.2%. The difference in bond angles between our work and other theoretical calculations<sup>36,37</sup> is under 0.5% on average. Note that our bond lengths are consistently larger than those in other calculations. We do not compare our geometries with experimental data measured in the crystal phases since that comparison may be misleading due to charge redistribution, which directly affects each individual monomer geometry.<sup>38</sup>

The dipole moment of thymine was found to be 4.2 D, in good agreement with the DFT result of 4.5 D<sup>36</sup> and the QC result<sup>39</sup> of 4.3 D. Uracil was found to have a similar dipole moment of 4.1 D, which is also in good agreement with the previous QC result of 4.4 D.<sup>39</sup>

To analyze the stability of each homopair, we calculate the stabilization (or binding) energy,  $E_{\text{stab}}$ , along with some other useful energies (see, for example, refs 26 and 40). The stabilization energy is defined as the total energy of the relaxed system (a pair) minus the energy of each relaxed subsystem

(each relaxed monomer). Similarly, the interaction energy,  $E_{\text{int}}$ , is defined as the total energy of the relaxed system minus the energy of each subsystem (each monomer) calculated in the geometry of the pair. The difference between these two quantities is known as the deformation energy,  $E_{\text{def}} = E_{\text{stab}} - E_{\text{int}}$ , which is positive and represents an energy loss from both subsystems when they form the relaxed combined system.

It should be stressed that, due to the incomplete localized basis set used in the calculations, the stabilization energy should be calculated with care. We have used the basis set superposition error (BSSE)<sup>41</sup> correction to obtain accurate binding energies.

### 3. Results

**3.1. Thymine and Uracil: Binding Sites.** Our approach to determine various superstructure possibilities is based on the idea that the binding there is determined by all possible individual homodimers. To find the latter in the cases of thymine and uracil, we rely on our approach reported earlier on adenine pairs in ref 26, where the notion of binding sites (or sites for short) was introduced. The point is that two molecules bind together if two closely positioned (adjacent in most cases) intermolecular hydrogen bonds can be formed between a site of one molecule and that of the other. In fact, the approach is very general and can be applied to any system where at least two hydrogen bonds can be formed between two molecules. For example, we have considered all of the cytosine and guanine homopairs<sup>27</sup> using this method.

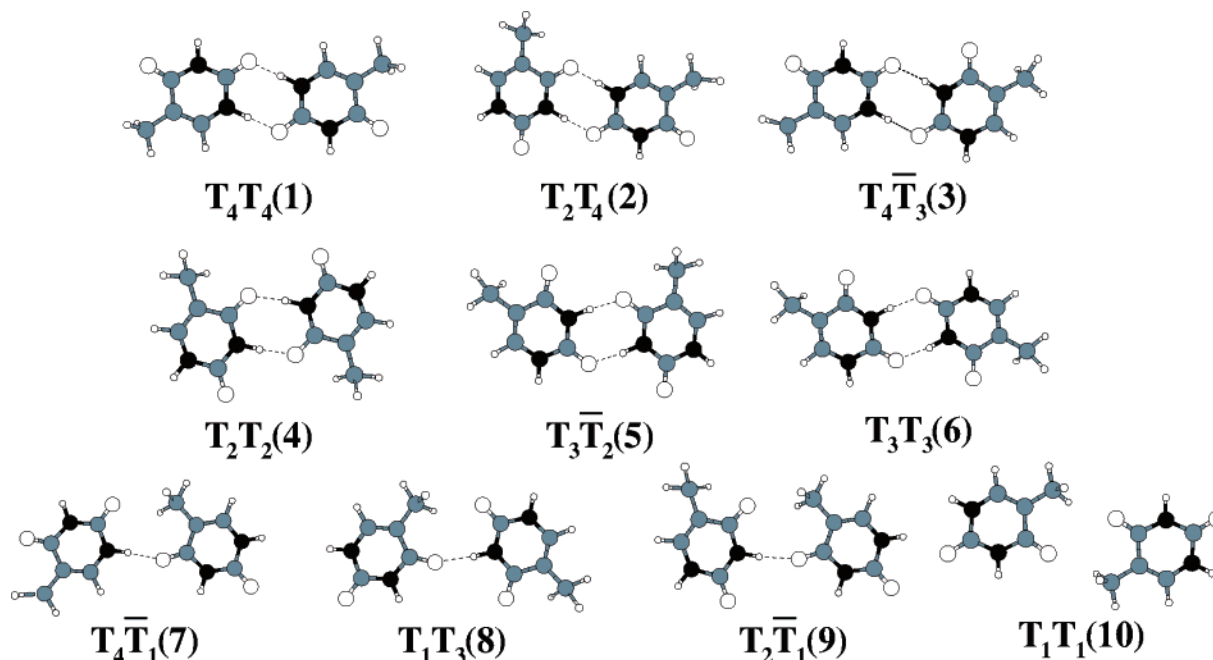
The orientations of the molecules shown in Figures 1a and 1b will be denoted T and U, respectively, while their nonchiral counterparts obtained by flipping the molecules in the molecular plane will be denoted  $\bar{T}$  and  $\bar{U}$ . Following the notation from our previous studies,<sup>26,27</sup> all sites for each molecule will be numbered, and the relevant sites involved in the bonding of every pair will be indicated explicitly, e.g.,  $T_4\bar{T}_4$  or  $U_3\bar{U}_4$ .

The thymine and uracil molecules (Figures 1a and 1b, respectively) have an almost round shape, so bonding sites can only be made from adjacent exposed hydrogen or acceptor atoms. We will discuss the uracil molecule (Figure 1b) first of all.

In total there are six atoms that can form a hydrogen bond: four exposed hydrogens and two exposed oxygen atoms. Therefore, potentially there are six sites that can form a double hydrogen bond with another molecule. Note, however, since in this paper only homopairs are considered, the total number of sites reduces to four, each having one donor and one acceptor. Indeed, two sites containing two hydrogen atoms each (i.e., two donors) cannot find their counterpart with two acceptors in the other uracil molecule. It is useful to have in mind, however, that these two donor sites could make possible bonds in heterogeneous pairs, e.g., between uracil and guanine or cytosine.

The four sites indicated in Figure 1b can bond together to create  $(5 \times 4)/2 = 10$  possible uracil homodimers. The order of hydrogen and acceptor atoms places a chiral constraint on the bonding, so some site–site combinations require one molecule to be flipped.

For thymine, shown in Figure 1a, the total number of exposed atoms that have the ability to form a hydrogen bond is eight: four exposed planar hydrogens (not labeled), two nonplanar hydrogens (labeled H1 and H2), and two exposed oxygen atoms. In total there are seven sites: (i) three with one hydrogen (donor) and one acceptor (sites 2–4), (ii) one (site 1) with one acceptor and two out-of-plane hydrogens, (iii) two with adjacent hydrogens, and (iv) one with one planar hydrogen and two out-of-



**Figure 2.** Relaxed thymine homopairs in order of stability (shown by the number in the round brackets). The notation  $T_nT_m$  corresponds to a thymine dimer with chiral monomers, while  $T_n\bar{T}_m$  corresponds to a pair with nonchiral monomers. The indices  $n$  and  $m$  indicate explicitly the sites of the two molecules (as adopted in Figure 1) engaged in the two hydrogen bonds in each pair.

plane hydrogens. The latter three sites are not shown in Figure 1a since, similar to the uracil homopairs, only the first four mentioned sites are functional for thymine homopairs, since the last three require sites with two adjacent acceptors not present in the molecule.

The three sites 2–4 contain only one hydrogen and one acceptor in each site, so that only combinations of these sites can bond together. The H donors and O acceptors in each site alternate, chirality is imposed on the bonding, and therefore, only one pairing possibility is possible for each site–site combination. Thus, there are  $(4 \times 3)/2 = 6$  homopair possibilities here.

The site 1 of the thymine molecule, which has two nonplanar hydrogens from the methyl group and an oxygen acceptor atom, is a little more complex than the other site combinations. Due to the fact that site 1 can accommodate either of the two nonplanar hydrogen atoms, it can, in fact, be split into two sites, giving rise to eight additional homopairs that could also be considered (some will be symmetry equivalent). However if these pairs were relaxed, they would most probably relax to nonplanar configurations that would be outside the scope of this study. Instead we decided to try to relax the homopairs involving site 1 in the planar configuration to see if the molecules would still stay within the plane after the relaxation. Therefore only four planar additional thymine homopairs involving the site 1 will be considered ( $T_1T_1$ ,  $T_2\bar{T}_1$ ,  $T_1T_3$ , and  $T_4\bar{T}_1$ ), 10 possible thymine homodimers altogether. If a nonplanar state is preferred for the thymine homopairs with the site 1 bonding, then this should become clear after the relaxation.

**3.2. Thymine and Uracil Homodimers: Energetics.** The relaxed geometries of all planar homopairs of thymine and uracil are shown in Figures 2 and 3. All 10 thymine and uracil pairs relaxed including four chiral centrosymmetric pairs for each molecule ( $T_nT_n$  and  $U_nU_n$ ,  $n = 1-4$ ) built using the same sites on each monomer, each with the  $C_2$  rotational symmetry. The geometry of all pairs is planar for all pairs reported here. We define the pair planar when the rings of both molecules are approximately within the same plane. Even the thymine pairs

with site 1 involved in the bonding, all have relaxed to planar configurations.

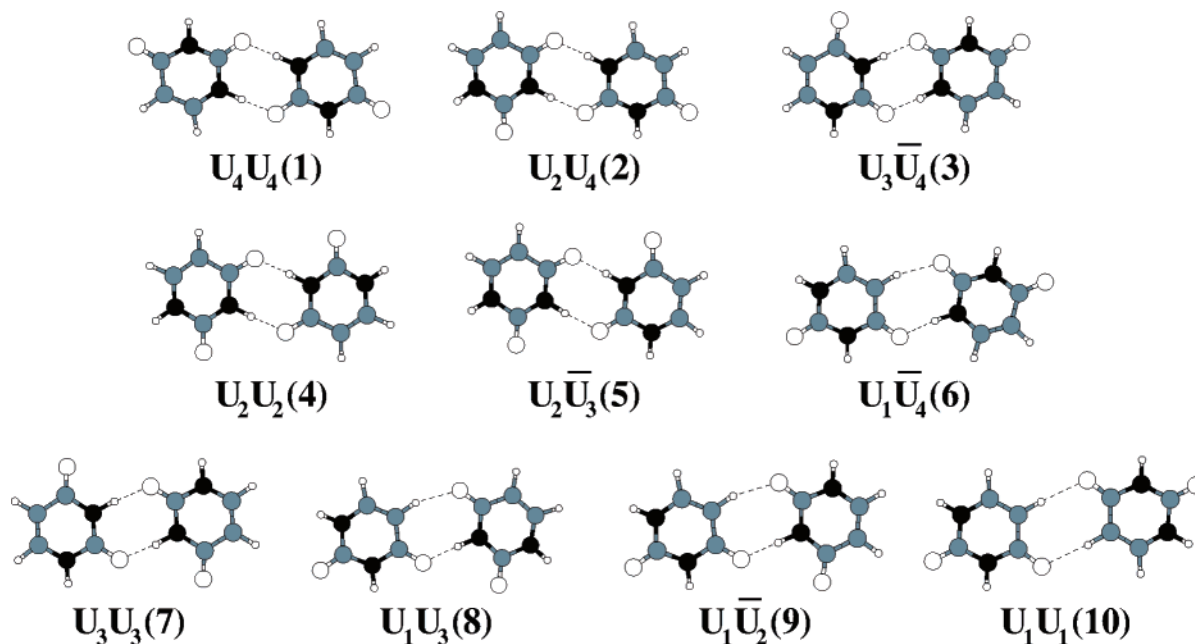
The stabilization, interaction, deformation, and BSSE energies of all the thymine and uracil homodimers are given in Table 1. Note that in all cases the BSSE corrections are essential to obtain realistic stabilization energies.

The most stable thymine pair,  $T_4T_4$ , and uracil pair,  $U_4U_4$ , are both centrosymmetric, like the most stable adenine,<sup>26</sup> cytosine, and guanine<sup>27</sup> homopairs. In total there are six thymine and seven uracil pairs with similarly high stability (below  $-0.5$  eV).

Previous theoretical studies of thymine and uracil homopairs are very scarce. To our knowledge, the most stable thymine pair has never previously been studied, and many of the thymine and uracil pairs are reported for the first time in this study. Previous results from the literature are presented in Table 1. The recent benchmark ab initio study<sup>35</sup> on DNA base pairs does not include any thymine–thymine base pairs and only two uracil–uracil pairs. One can see from the Table 1 that the agreement in energies with these two uracil pairs is very good. The agreement with the MP2 results relaxed to the HF geometry<sup>40,43</sup> is also very good.

We see that the semiempirical method<sup>16</sup> performs poorly when compared with the ab initio results, which has been concluded in the previous studies.<sup>25,26,27</sup> The energies are largely underestimated, and even though the energetic ordering is similar in this case, there is no sizable energy difference between the three referenced semiempirical results.

Note that 11 uracil homodimers have been previously reported,<sup>42</sup> seven of which are planar, three are stacked, and one is “T-shaped”. In this study, geometrical identifications of possible dimers were performed using a semiempirical technique. Ab initio HF relaxations were then applied to all of the identified homopairs followed by post priori MP2 calculations to evaluate the correlation energy. However we find three more possible planar pairs that all have significant stabilization energies. The agreement between our results and the stabilization energies presented in the last column may be considered as



**Figure 3.** Relaxed uracil homopairs in order of stability. Notations are analogous to those in Figure 2.

**TABLE 1: Energetics of the Thymine and Uracil Homopairs (in eV)<sup>a</sup>**

pair	PBE				$E_{\text{stab}}$ (literature)				
	$E_{\text{BSSE}}$	$E_{\text{int}}$	$E_{\text{def}}$	$E_{\text{stab}}$	ref 35	ref 40	ref 16	ref 42	ref 43
T <sub>4</sub> T <sub>4</sub> (1)	0.18	−1.00	0.18	−0.82					
T <sub>2</sub> T <sub>4</sub> (2)	0.18	−0.84	0.15	−0.69					
T <sub>4</sub> T <sub>3</sub> (3)	0.18	−0.78	0.13	−0.65					
T <sub>2</sub> T <sub>2</sub> (4)	0.18	−0.68	0.12	−0.57		−0.43	−0.27		
T <sub>3</sub> T <sub>2</sub> (5)	0.18	−0.65	0.11	−0.54		−0.43	−0.27		
T <sub>3</sub> T <sub>3</sub> (6)	0.17	−0.60	0.09	−0.51		−0.43	−0.28		
T <sub>4</sub> T <sub>1</sub> (7)	0.13	−0.41	0.02	−0.39					
T <sub>1</sub> T <sub>3</sub> (8)	0.13	−0.26	0.02	−0.24					
T <sub>2</sub> T <sub>1</sub> (9)	0.14	−0.26	0.02	−0.24					
T <sub>1</sub> T <sub>1</sub> (10)	0.11	−0.03	0.00	−0.03					
U <sub>4</sub> U <sub>4</sub> (1)	0.18	−0.96	0.15	−0.71				−0.69	
U <sub>2</sub> U <sub>4</sub> (2)	0.18	−0.84	0.13	−0.63				−0.55	
U <sub>3</sub> U <sub>4</sub> (3)	0.17	−0.75	0.12	−0.58				−0.54	
U <sub>2</sub> U <sub>2</sub> (4)	0.18	−0.70	0.12	−0.58				−0.45	
U <sub>2</sub> U <sub>3</sub> (5)	0.17	−0.63	0.10	−0.53	−0.49			−0.46	
U <sub>1</sub> U <sub>4</sub> (6)	0.13	−0.59	0.06	−0.53				−0.46	−0.50
U <sub>3</sub> U <sub>3</sub> (7)	0.17	−0.58	0.08	−0.50				−0.46	
U <sub>1</sub> U <sub>2</sub> (8)	0.13	−0.43	0.04	−0.39	−0.39				
U <sub>1</sub> U <sub>3</sub> (9)	0.13	−0.40	0.04	−0.36					
U <sub>1</sub> U <sub>1</sub> (10)	0.08	−0.27	0.00	−0.27					

<sup>a</sup>  $E_{\text{int}}$ ,  $E_{\text{def}}$ , and  $E_{\text{stab}}$  are the interaction, deformation, and stabilization energies, respectively, while  $E_{\text{BSSE}}$  corresponds to the BSSE correction. Results of our calculations are compared with other available work presented in the last four columns.

reasonable, taking into account the different levels of theory used in ref 42 to perform geometry relaxations and calculate the stabilization energies. Note that energies, calculated in this way, strictly speaking, do not correspond to the real energy minimum.

Please note that in this work we did not perform vibrational analysis for any structure as this would be extremely time-consuming. Therefore, care should be taken in considering the relaxed structures as local minima on the potential energy surface.

**3.3. Thymine and Uracil Homodimers: Geometry.** The stability of a dimer can be assessed by its geometry as well.<sup>44,40</sup> Each bond can be characterized geometrically by the donor hydrogen—acceptor distance and the corresponding angle along the bond. The strong hydrogen bonds prefer to be shorter and linear. However, because of the particular rigid shapes of the molecules and the fact that each bond has to accommodate many

secondary interactions<sup>44</sup> (the relaxed geometry is said to be of a cooperative nature<sup>45</sup>), the bonds in practice are not straight and of different strengths.

The distances for the two or three hydrogen bonds and the angles associated with these are given in Table 2. Similarly to the adenine, cytosine, and guanine homopairs, the geometry usually does follow the trend explained in the last paragraph. The cooperativity can be easily seen, especially in the strongest pairs. Note that because the symmetry of centrosymmetric pairs is broken due to weak interaction with periodic images, the bond lengths and angles of centrosymmetric pairs were slightly different, and the averaged results are presented in Table 2. The deviation from the perfect symmetry was found to be within 1.4% and 0.8% for bond lengths and angles, respectively, similar to the adenine, cytosine, and guanine homopairs.

The available data from the literature, also presented in Table 2 for comparison, are similar to our results. The agreement is



TABLE 2: Geometrical Characteristics of the Hydrogen Bonds in Thymine and Uracil Homopairs<sup>a</sup>

pair	hydrogen bond	PBE		ref 40		ref 35
		length (Å)	angle (deg)	length (Å)	angle (deg)	length (Å)
T <sub>4</sub> T <sub>4</sub> (1)	N1—H···O2	2.69	178.6			
	N1—H···O2	2.69	178.6			
T <sub>2</sub> T <sub>4</sub> (2)	N3—H···O2	2.75	175.9			
	N1—H···O4	2.70	178.0			
T <sub>3</sub> T <sub>4</sub> (3)	N3—H···O2	2.76	174.3			
	N1—H···O2	2.71	177.7			
T <sub>2</sub> T <sub>2</sub> (4)	N3—H···O4	2.72	176.5	2.98	167.4	
	N3—H···O4	2.72	176.5	2.98	167.4	
T <sub>2</sub> T <sub>3</sub> (5)	N3—H···O2	2.76	175.1	2.97	167.3	
	N3—H···O4	2.76	174.3	2.98	166.2	
T <sub>3</sub> T <sub>3</sub> (6)	N3—H···O2	2.76	173.8	2.98	166.2	
	N3—H···O2	2.76	173.8	2.98	166.2	
T <sub>1</sub> T <sub>4</sub> (7)	CM—H1···O2	3.25	108.2			
	CM—H2···O2	3.25	108.4			
	N1—H···O4	2.82	170.5			
T <sub>1</sub> T <sub>3</sub> (8)	CM—H1···O2	3.22	108.2			
	CM—H2···O2	2.88	108.3			
	N3—H···O4	2.88	176.0			
T <sub>1</sub> T <sub>2</sub> (9)	CM—H1···O4	3.17	107.9			
	CM—H2···O4	3.17	108.1			
	N3—H···O4	2.93	175.9			
T <sub>1</sub> T <sub>1</sub> (10)	CM—H1···O4	3.22	108.3			
	CM—H2—O4	3.22	108.3			
	CM—H1—O4	3.22	108.3			
	CM—H2—O4	3.22	108.3			
U <sub>4</sub> U <sub>4</sub> (1)	N1—H···O2	2.71	177.9			
	N1—H···O2	2.71	177.9			
U <sub>2</sub> U <sub>4</sub> (2)	N3—H···O2	2.76	175.8			
	N1—H···O4	2.71	178.7			
U <sub>3</sub> U <sub>4</sub> (3)	N3—H···O2	2.72	176.7			
	N1—H···O2	2.75	176.8			
U <sub>2</sub> U <sub>2</sub> (4)	N3—H···O4	2.72	176.4			
	N3—H···O4	2.72	176.4			
U <sub>2</sub> U <sub>3</sub> (5)	N3—H···O2	2.73	177.0			2.83
	N3—H···O4	2.80	175.0			2.83
U <sub>1</sub> U <sub>4</sub> (6)	C5—H···O2	3.14	164.5			
	N1—H···O4	2.76	173.8			
U <sub>3</sub> U <sub>3</sub> (7)	N3—H···O2	2.77	173.4			
	N3—H···O2	2.77	173.4			
U <sub>1</sub> U <sub>2</sub> (8)	C5—H···O4	3.17	161.6			3.21
	N3—H···O4	2.80	177.5			2.87
U <sub>1</sub> U <sub>3</sub> (9)	C5—H···O2	3.13	161.6			
	N3—H···O4	2.80	179.9			
U <sub>1</sub> U <sub>1</sub> (10)	C5—H···O4	3.29	175.7			
	C5—H···O4	3.29	175.7			

<sup>a</sup> Each bond is presented via the distance between the H donor atom of one molecule and the corresponding acceptor (N or O atom) of another. Also, the angle between the X—H atoms of the first molecule and the O atom of another is shown, where X, depending on the particular dimer, is either C or N. The numbering of atoms is identical to that adopted in Figure 1. The bond lengths and angles of centrosymmetric pairs are averaged. In the last three columns results of our calculations are compared with those found in refs 40 and 35 using the HF and MP2 geometry optimization methods, respectively.

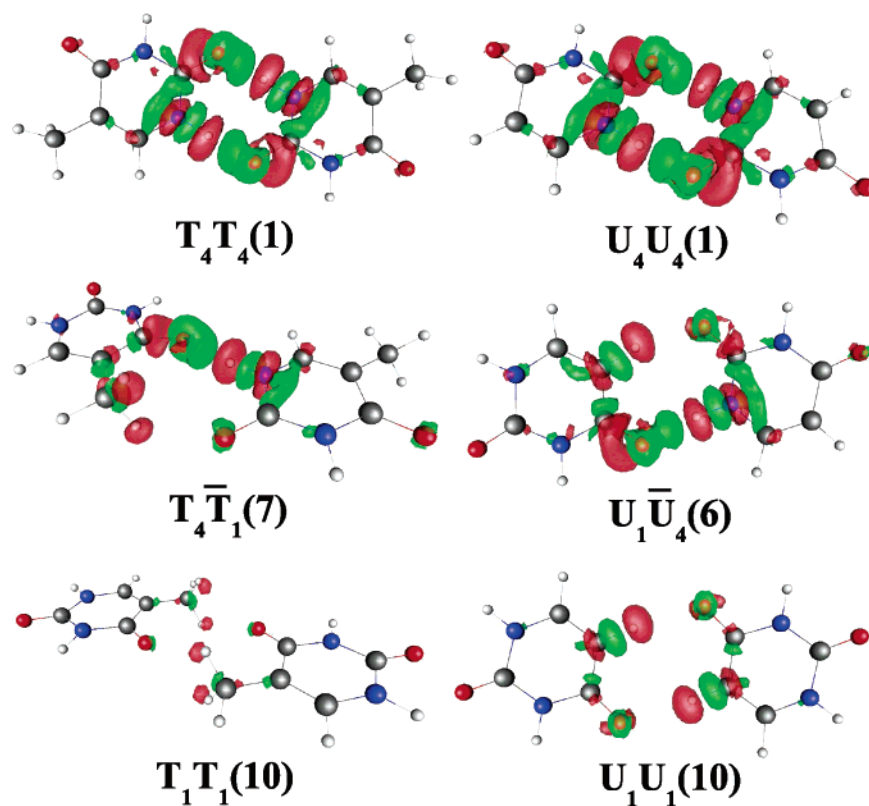
especially good with the recent benchmark calculations<sup>35</sup> in which a high-level method for the geometry optimization was used. The less recent study by the same authors<sup>40</sup> also shows similar results although the geometry optimization was only performed at the Hartree–Fock level there.

It is interesting to note that some of the bonding patterns are very similar, for example, U<sub>1</sub>U<sub>4</sub>(6) and U<sub>1</sub>U<sub>2</sub>(8), which both have one clear C—H—O and one N—H—O bond. However the stabilities of the two pairs are significantly different. This is because, as was discussed in refs 44 and 34, secondary interactions associated with other atoms of the molecules not involved in the direct hydrogen bonding are also important.

Note that for all pairs involving the site 1 containing the methyl group (T<sub>1</sub>T<sub>1</sub>, T<sub>2</sub>T<sub>1</sub>, T<sub>1</sub>T<sub>3</sub>, and T<sub>4</sub>T<sub>1</sub>), the binding energies are quite small; see Table 1. At the same time, the methyl group hydrogens involved in the bonding remained out of the plane in the relaxed pairs. A question arises if the bonding can be

enhanced by rotating the group so that perfectly planar hydrogen bonds can be formed with a single hydrogen atom. We calculated the potential energy surface of an isolated molecule with respect to the rotation of the three hydrogens in the methyl group about the C5—CM axis and found a very small ( $\leq 0.05$  eV) energy barrier for this. We also performed geometry relaxations of the four pairs in which the methyl hydrogens involved in the bonding were rotated to facilitate the favorable linear bond arrangement. However, the obtained structures were less stable than the geometries presented before. This observation can be explained by the well-known fact that the methyl group hydrogens are bad donors and the group does not usually participate in the hydrogen bonding. We shall come to this point again in the next section when considering the charge difference plots of selected pairs.

**3.4. Thymine and Uracil Homodimers: Electronic Density.** The interaction densities (defined with respect to the



**Figure 4.** Interaction density of thymine (on the left-hand side) and uracil (right) homopairs with the electron density difference of  $\pm 0.01$  electrons. The green surfaces correspond to regions with a positive electron density difference (i.e., charge excess areas), while red areas correspond to regions with a negative electron density difference (charge depletion areas).

isolated monomers in the geometry of the pair<sup>26</sup>) of the strongest, weakest, and a medium stability pair of each base are shown in Figure 4. The interaction densities of the uracil  $U_4U_4$  and the thymine  $T_4T_4$  homopairs all show two characteristic kebab-like structures expected for the double hydrogen bonding between the molecules.<sup>26,27</sup> In the pair  $U_1\bar{U}_4$  one bond is much stronger than the other, while the  $U_1U_1$  pair has two weak hydrogen bonds. Vibrational analysis, atoms-in-molecules Bader analysis, and natural bond orbital analysis were performed for the  $U_1\bar{U}_4$  pair by Hobza et al.,<sup>43</sup> and it was concluded that there is one strong and one weaker hydrogen bond.

The remaining pairs  $T_4\bar{T}_1$  and  $T_1T_1$  have slightly more complex charge redistributions due to the out-of-plane hydrogens of the methyl group participating in one ( $T_4\bar{T}_1$ ) or two ( $T_1T_1$ ) bonds. The bonds involving the methyl group can be considered as weak hydrogen bonds since very small charge redistribution is seen. Interestingly, both the C atom and more than one H atom of the methyl group participate in the bonding.

Thus, as follows from the discussion above, the interaction charge densities for all of the homopairings are seen to correlate strongly with the stability of the particular pair. This is in agreement with our studies on other DNA base molecules.<sup>26,27</sup>

Another effect that can be of importance for more complicated structures involving more than two molecules is also noticed in our interaction charge densities. Indeed, oxygens not directly involved in the hydrogen bonding are seen to have been depleted from some of the electron charge, indicating that their ability to accept extra charge from another molecule is probably enhanced. This charge redistribution may be the origin of the resonance-assisted hydrogen bonding (RAHD) known to play some role in binding superstructures (see ref 17 and references therein).

#### 4. Conclusions

In total 10 possible planar homopairings for each of the nucleic acid bases thymine and uracil calculated using a DFT method are reported in this study. The stability of each pair varies from  $-0.03$  to  $-0.81$  eV. The strongest homopairs for either of the bases are centrosymmetric, the same conclusion as was drawn in the cases of adenine, cytosine, and guanine homopairs.<sup>26,27</sup> There is only one pair with a very low stability (less than 0.2 eV),  $T_1T_1$ , which we attribute to the methyl groups involved in the two bonds. All thymine pairs that involve one or two methyl groups in the bonding have much smaller binding energies.

We find a very good agreement between our results and the recent benchmark ab initio calculations of DNA base pairs.<sup>35</sup> A much worse comparison can be made with the previous semiempirical results.<sup>16</sup>

All studied homopairs relaxed to planar configurations. The deformation energies have been found to be small, so these molecules can be considered as approximately rigid in the pairs.

We also analyzed the interaction density for all homopairs. Again, as expected, stronger pairs show more charge redistribution than weaker pairs. We also find that many pairs have a charge redistribution on the remote oxygen atoms not directly involved in the hydrogen bonds; a similar effect was reported previously in ref 27.

A thorough study of all planar thymine and uracil homopairs reported in this paper should give a profound basis for a better understanding of their superstructures formed on various crystal surfaces. This work is now in progress and will be reported elsewhere.

**Acknowledgment.** We thank Roberto Otero, Maya Stöck, and Flemming Besenbacher for many useful discussions. We

acknowledge the computer time on the HPCx supercomputer provided via the Materials Chemistry Consortium and the technical support provided by the HPCx supercomputer helpdesk. R.E.A.K. is also grateful to the EPSRC for financial support (Grant No. GR/P01427/01).

## References and Notes

- (1) Srinivasan, R.; Murphy, J. C.; Fainchtein, R.; Pattabiraman, N. *Ultramicroscopy* **1991**, *312*, 293.
- (2) Tao, N. J.; DeRose, J. A.; Lindsay, S. M. *J. Phys. Chem.* **1993**, *97*, 910.
- (3) Tao, N. J.; Shi, Z. *J. Phys. Chem.* **1994**, *98*, 7422.
- (4) Tao, N. J.; Shi, Z. *J. Phys. Chem.* **1994**, *98*, 1464.
- (5) Tanaka, H.; Nakagawa, T.; Kawai, T. *Surf. Sci.* **1996**, *364*, L575.
- (6) Sowerby, S. J.; Heckl, W. M.; Petersen, G. B. *J. Mol. Evol.* **1996**, *43*, 419.
- (7) Sowerby, S. J.; Petersen, G. B. *J. Electroanal. Chem.* **1997**, *433*, 85.
- (8) Nakagawa, T.; Tanaka, H.; Kawai, T. *Surf. Sci.* **1997**, *370*, L144.
- (9) Furukawa, M.; Tanaka, H.; Kawai, T. *Surf. Sci.* **1997**, *392*, L33.
- (10) Kawai, T.; Tanaka, H.; Nakagawa, T. *Surf. Sci.* **1997**, *386*, 124.
- (11) Freund, J. E.; Edelwirth, M.; Krobek, P.; Heckl, W. M. *Phys. Rev. B* **1997**, *55*, 5394.
- (12) Sowerby, S. J.; Heckl, W. M. *Origins Life Evol. Biosphere* **1998**, *28*, 283.
- (13) Cavallini, M.; Aloisi, G.; Bracali, M.; Guidelli, R. *J. Electroanal. Chem.* **1998**, *444*, 75.
- (14) Sowerby, S. J.; Edelwirth, M.; Heckl, W. M. *J. Phys. Chem. B* **1998**, *102*, 5914.
- (15) Furukawa, M.; Tanaka, H.; Kawai, T. *Surf. Sci.* **2000**, *445*, 1.
- (16) Furukawa, M.; Tanaka, H.; Kawai, T. *J. Chem. Phys.* **2001**, *115*, 3419.
- (17) Otero, R.; Schöck, M.; Molina, L. M.; Laegsgaard, E.; Stensgaard, I.; Hammer, B.; Besenbacher, F. *Angew. Chem., Int. Ed.* **2005**, *44*, 2270.
- (18) Uchihashi, T.; Okada, T.; Sugawara, Y.; Yokoyama, K.; Morita, S. *Phys. Rev. B* **1999**, *60*, 8309.
- (19) Uchihashi, T.; Ishida, T.; Komiyama, M.; Ashino, M.; Sugawara, Y.; Mizutani, W.; Yokoyama, K.; Morita, S.; Tokumoto, H.; Ishikawa, M. *Appl. Surf. Sci.* **2000**, *157*, 244.
- (20) De Feyter, S.; De Schryver, C. *Chem. Soc. Rev.* **2003**, *32*, 139.
- (21) Furukawa, M.; Tanaka, H.; Sugiura, K.; Sakata, Y.; Kawai, T. *Surf. Sci.* **2000**, *445*, L58.
- (22) Chen, Q.; Frankel, D. J.; Richardson, N. V. *Langmuir* **2002**, *18*, 3219.
- (23) Otero, R.; Schöck, M.; Besenbacher, F. Private communication.
- (24) Edelwirth, M.; Freund, J.; Sowerby, S. J.; Heckl, W. M. *Surf. Sci.* **1998**, *417*, 201.
- (25) Hobza, P.; Kabelac, M.; Sponer, J.; Mejzlik, P.; Vondrasek, J. *J. Comput. Chem.* **1997**, *18*, 1136.
- (26) Kelly, R. E. A.; Lee, Y. J.; Kantorovich, L. N. *J. Phys. Chem. B* **2005**, *109*, 11933.
- (27) Kelly, R. E. A.; Lee, Y. J.; Kantorovich, L. N. *J. Phys. Chem. B* **2005**, *109*, 22045.
- (28) Sowerby, S. J.; Edelwirth, M.; Heckl, W. M. *Appl. Phys. A* **1998**, *66*, S649.
- (29) Kelly, R. E. A.; Kantorovich, L. N. *Surf. Sci.* **2005**, *589*, 139.
- (30) Ordejon, P.; Artacho, E.; Soler, J. M. *Phys. Rev. B* **1996**, *53*, R10441.
- (31) Soler, J. M.; Artacho, E.; Gale, J. D.; Garcia, A.; Junquera, J.; Ordejon, P.; Sanchez-Portal, D. *J. Phys.: Condens. Matter* **2002**, *14*, 2745.
- (32) Artacho, E.; Sanchez-Portal, D.; Ordejon, P.; Garcia, A.; Soler, J. M. *Phys. Status Solidi B* **1999**, *215*, 809.
- (33) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.
- (34) Sponer, J.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 3161.
- (35) Sponer, J.; Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142.
- (36) Preuss, M.; Schmidt, W. G.; Seino, K.; Furthmüller, J.; Bechstedt, F. *J. Comput. Chem.* **2004**, *25*, 112.
- (37) Shukla, M. K.; Mishra, P. C. *Chem. Phys.* **1999**, *240*, 319.
- (38) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *J. Am. Chem. Soc.* **2000**, *122*, 4117.
- (39) Sponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers* **2001**, *61*, 3.
- (40) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965.
- (41) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
- (42) Kratochvíl, M.; Engkvist, O.; Sponer, J.; Jungwirth, P.; Hobza, P. *J. Phys. Chem. A* **1998**, *102*, 6921.
- (43) Hobza, P.; Sponer, J.; Cubero, E.; Orozco, M.; Javier Luque, F. J. *Phys. Chem. B* **2000**, *104*, 6286.
- (44) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc.* **1987**, *109*, 1302.
- (45) Asensio, A.; Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A* **2003**, *107*, 6441.