Hydrogen bonding of 3- and 5-methyl-6-aminouracil with natural DNA bases†

Gábor Paragi,*[a] Emília Szájli,[b] Ferenc Bogár,[a] Lajos Kovács,[c] Célia Fonseca Guerra[d] and F. Matthias Bickelhaupt*[d]

Received (in Montpellier, France) 3rd March 2008, Accepted 12th June 2008
First published as an Advance Article on the web 11th August 2008
DOI: 10.1039/b803593h

We have theoretically analyzed the hydrogen bonding of two artificial nucleobases (3- and 5-methyl-6-aminouracil) with the natural DNA bases using the generalized gradient approximation (GGA) of density functional theory at BP86/TZ2P level. The analysis of the monomers provides the possibility to distinguish the different active parts of molecules and the interactions with natural nucleobases have been determined. Another purpose of this work is to clarify the relative importance of electrostatic interaction *vs.* orbital interaction in the hydrogen bonds between the artificial base and the natural DNA base. At variance with widespread belief, the orbital interaction component in these hydrogen bonds is found to contribute about 40% of the attractive interactions and is thus of the same order of magnitude as the electrostatic component, which provides the remaining attraction. According to our theoretical results both candidates are potential artificial nucleobases for incorporation in DNA.

Introduction

The design of new, artificial nucleobases is one of the premier tasks in nucleic acid research due to their potential role in regulating gene expression.[1] Until the late 1980s, the investigation of the interaction of such new compounds with natural bases was possible only through experimental techniques. In view of the fact that high throughput synthetic techniques (*i.e.*, combinatorial chemistry) are scarcely available in this field, the development of a new nucleobase is a very time consuming process.

From the beginning of the 1990s, the increasing computer capacities and the more sophisticated program codes provide new possibilities to narrow down suitable (classes of) candidate molecules, in advance of the costly experimental explorations. Although the computational model systems are often strongly simplified compared to the species used in experiments, it turned out that in certain cases they are really well applicable.[2,3]

Previously, we reported about the capabilities of the new, artificial nucleobases 3-methyl-6-aminouracil (**1a**) and 5-methyl-6-aminouracil (**2b**, see Fig. 1) to form duplex and triplex complexes with natural DNA bases.[4] Here, species **1a** and **1b** served as models for more complex 3- and 5-substituted 6-aminouracils, respectively. Although the applied level of theory was the simple exchange only Hartree–Fock method in the previous work, the binding preference order corresponded well to the results of present higher level investigations.

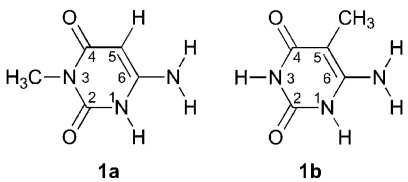


Fig. 1 3-Methyl-6-aminouracil (**1a**) and 5-methyl-6-aminouracil (**1b**).

In the present paper, we present a more sophisticated description of the chemical behaviour of the candidate molecules **1a** and **1b** in various Watson–Crick and other types of complexes with natural DNA bases, using the Amsterdam Density Functional (ADF) program[5] at BP86/TZ2P.[6] This level of theory was recently shown to perform very satisfactorily in describing the stability and geometry of simple as well as multiply hydrogen-bonded systems.[3a,b,g] The main purpose is the analysis of the different possible base-pair arrangements (*e.g.* Watson–Crick or Hoogsteen type and their reverse forms) of the new compounds where H-bonds play a central role in the formation of complexes. The MO analysis of the monomers let us explain the differences of the interaction energies in those cases where the structure of the complexes seems to be similar. In this way, one can distinguish active and less active parts of the molecules.

The article is structured in the following way: after a brief summary of the theoretical and computational background we introduce the investigated monomers and their duplex

[a] *Supramolecular and Nanostructured Materials Research Group of the Hungarian Academy of Sciences at the University of Szeged, Dóm tér 8, H-6720 Szeged, Hungary. E-mail: paragi@sol.cc.u-szeged.hu; Fax: 36 62 545971; Tel: 36 62 544593*
[b] *Proteomics Research Group at the Biological Research Center of the Hungarian Academy of Science, Temesvári krt. 62, H-6726 Szeged, Hungary*
[c] *University of Szeged, Department of Medicinal Chemistry, Dóm tér 8., H-6720 Szeged, Hungary*
[d] *Theoretische Chemie and Amsterdam Center for Multiscale Modeling, Scheikundig Laboratorium der Vrije Universiteit, De Boelelaan 1083, NL-1081 HV Amsterdam, The Netherlands. E-mail: FM.Bickelhaupt@few.vu.nl; Fax: +31 20 5987617*

† Electronic supplementary information (ESI) available: Fig. S1: Total bonding energy of 3-methyl-6-aminouracil and 5-methyl-6-aminouracil with natural nucleobases. Table S1: Cartesian coordinates of all calculated duplex systems. See DOI: 10.1039/b803593h

arrangements. First the electronic structure of monomers, then the general results of the duplex calculations will be analyzed. Finally, we consider the incorporation of these artificial bases into natural DNA.

Theoretical background

The Kohn–Sham (KS) molecular orbital formulation of density functional theory has been successfully used from the beginning of the seventies. The KS method yields in principle exact energies. In practice, with the available approximate exchange and correlation functionals, rather accurate total energies can be computed. Thus, we have the special situation that a seemingly one-particle model (an MO method), in principle, completely accounts for the bonding energy. The overall bond energy ΔE between two bases in a base pair is analyzed using a quantitative bond energy decomposition scheme:[7,8]

$$\Delta E = \Delta E_{\text{prep}} + \Delta V_{\text{elstat}} + \Delta E_{\text{Pauli}} + \sum_{\Gamma} \Delta E_{\text{oi}}^{\Gamma} \qquad (1)$$

Here ΔE_{prep} is the amount of energy required to deform the separate bases from their equilibrium structure to the geometry that they acquire in the pair. ΔV_{elstat} corresponds to the classical electrostatic interaction between the unperturbed charge distributions of the prepared (*i.e.* deformed) bases and is usually attractive. The Pauli repulsion ΔE_{Pauli} comprises the destabilizing interactions between occupied orbitals and is responsible for the steric repulsion. Finally, the last term in eqn (1) is the orbital interaction energy ΔE_{oi} which can be further decomposed into the contributions from each irreducible representation Γ of the interacting system using the extended transition state scheme developed by Ziegler and Rauk.[7,8] This further decomposition is most informative in the case of clear σ–π (or A′–A″) separation, *e.g.*, in our systems with C_s symmetry. This has previously provided valuable insight into H-bonded systems.[3f]

The orbital interaction ΔE_{oi} in any MO model, and therefore also in Kohn–Sham theory, accounts for charge transfer (*i.e.*, donor–acceptor interactions between occupied orbitals in one moiety with unoccupied orbitals of the other, including the HOMO–LUMO interactions) and polarization (empty/occupied orbital mixing on one fragment due to the presence of another fragment).

Computational method

All calculations were performed using density functional theory (DFT) as implemented in the Amsterdam Density Functional (ADF) program developed by Baerends and others.[5] The Becke exchange functional was used in combination with the correlation functional suggested by Perdew.[6] The MOs were expanded using triple-ζ basis set augmented with two polarization functions (TZ2P).[9] This level of computation provides good agreement with other high level calculations, as has been shown in previous papers.[3a,b,g]

Equilibrium structures were optimized using analytical gradient techniques.[10] As a first step, all investigated systems were optimized with 10^{-3} accuracy for the maximal gradient force and for the energy change without any restriction. Next, the C_s symmetry constraint was switched on in rest of the cases and the final geometry was determined with 10^{-5} accuracy. Nevertheless for those systems whose buckle angle was significant or steric repulsion occurs between the two fragments, a fully unconstrained optimization was also performed with the same criteria as in the C_s case. For the optimized structures, the Voronoi deformation density (VDD) atomic charges[11] were calculated.

Investigated systems

The two forms of the substituted aminouracils are shown in Fig. 1, where the intended connection points to the sugar-phosphate or peptide nucleic acid chain are closed by methyl groups. The primary consequence is that the 5-substituted molecule provides additional bonding capacities by the new N–H group in position 3. This way, it is richer in possible duplex arrangements, as one can notice in Fig. 2 and 3 in which we present all the calculated complexes of the modified aminouracils with natural DNA bases. Although there are in principle several other possibilities, we focused on those systems whose interaction energy is considerable (final geometries of the optimized systems can be found in the ESI†).

Energy level and VDD charge investigation of the two nucleobases

The energy level structure of the nucleobases **1a** and **1b** is analyzed. The order, differences or shifting of the levels show how the reactivity of a molecule is modified due to the change in structure or environment.[3j] In Table 1 we present the energy values (in eV) of the two highest occupied σ-orbitals (σ-HOMO−1), σ-HOMO) and the two lowest unoccupied σ-orbitals (σ-LUMO, σ-LUMO+1) for three cases, namely the two methyl substituted 6-aminouracils (left and right panels) and a purely hydrogen-closed form (in the middle). Furthermore, in Fig. 4, the results of the VDD charge analyzes are illustrated for these systems. Comparing the VDD charges, the systems are very similar: they are almost not affected by the substitution of a hydrogen atom by a methyl group.

This is in accord with the fact that the electronegativity of a hydrogen atom and a methyl group are not very different as compared to that of the heteroatoms. Moreover, as the oxygen atoms have negative electrostatic potential as well as a lone-pair-type orbital amplitude and the hydrogen atoms of the aromatic and amino N–H bonds have positive electrostatic potential as well as N–H antibonding σ* acceptor-orbital amplitude, the artificial bases posses the right electronic structure for achieving favorable electrostatic and donor–acceptor orbital interactions when forming a pair with a natural base.

In Table 1 the numbers were taken from the C_s-symmetric cases because in most of the calculations we applied this constraint. It has been previously shown[3b] that the effect of constraining the nucleobase geometry to C_s symmetry on the bond energy of natural Watson–Crick pairs (which are close to C_s symmetric) is in most cases very small or zero (see also later). In this way, the separation of the σ and π MOs is clear which is important in our analysis since the σ-orbitals have a

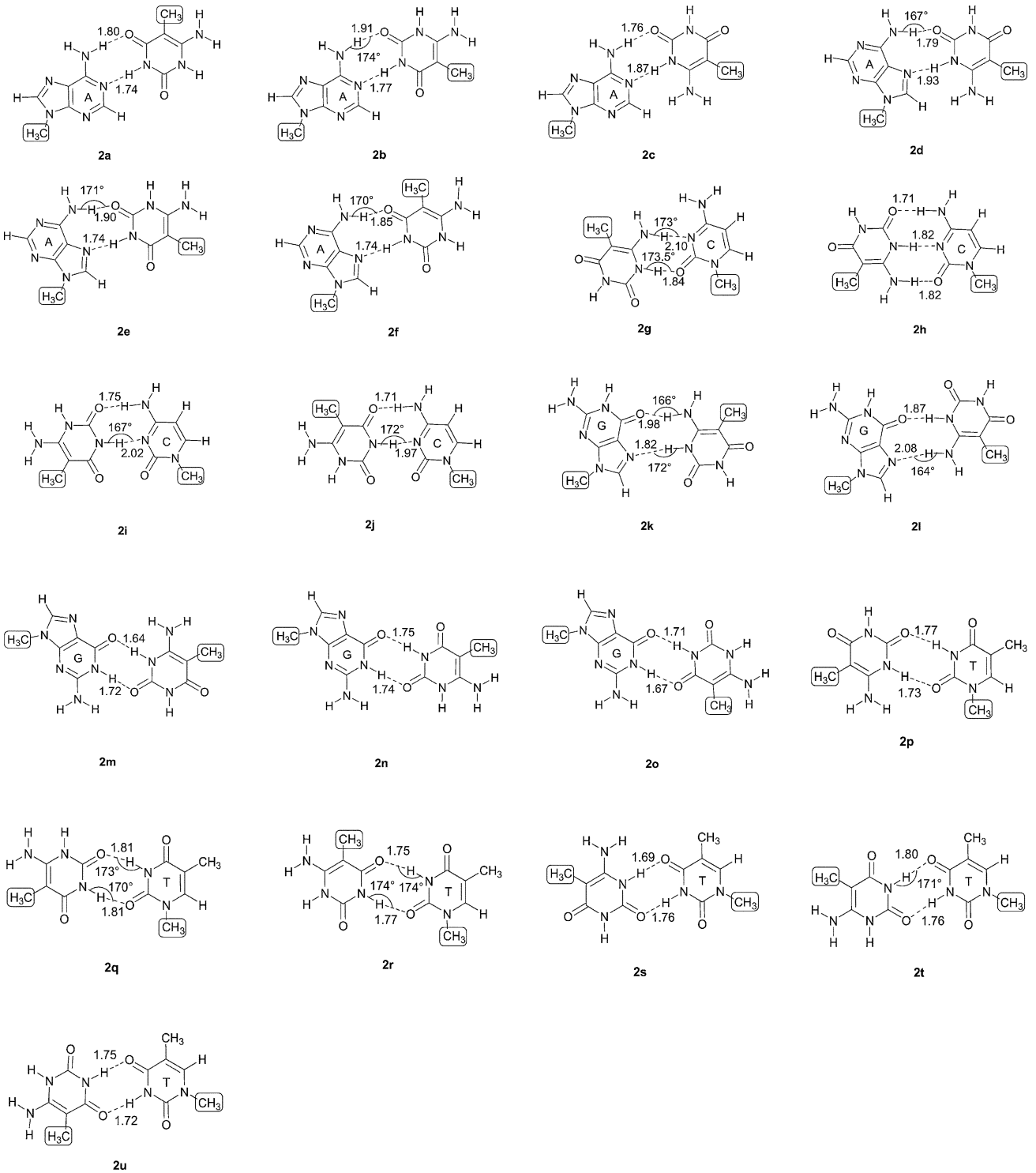


Fig. 2 The calculated duplex systems of 5-methyl-6-aminouracil with methylated adenine (**2a–2f**), cytosine (**2g–2j**), guanine (**2k–2o**) and thymine (**2p–2u**). The connections to the backbone chain are closed by methyl groups (encircled in the structures). The numbers refer to the optimized distances (in Å) and angles (in degrees) at the symmetric restricted case. The H-bond angle is shown only for instances where the deviation from 180° is larger than 5°.

primary role in H-bonding: the σ-type lone-pair orbital of the heteroatom donates charge into the unoccupied σ* N–H antibonding orbitals of the other fragment.[3f] Furthermore, for the natural DNA bases the sum of the σ-type orbital interaction energies is usually one order of magnitude larger than the total π-type orbital interaction energy.[3c,f] Finally, since the interaction is proportional to the energy level gap between the interacting σ and σ* orbitals, the shift or the order of levels provides some insight into the bonding capabilities of nucleobase candidates.[3j]

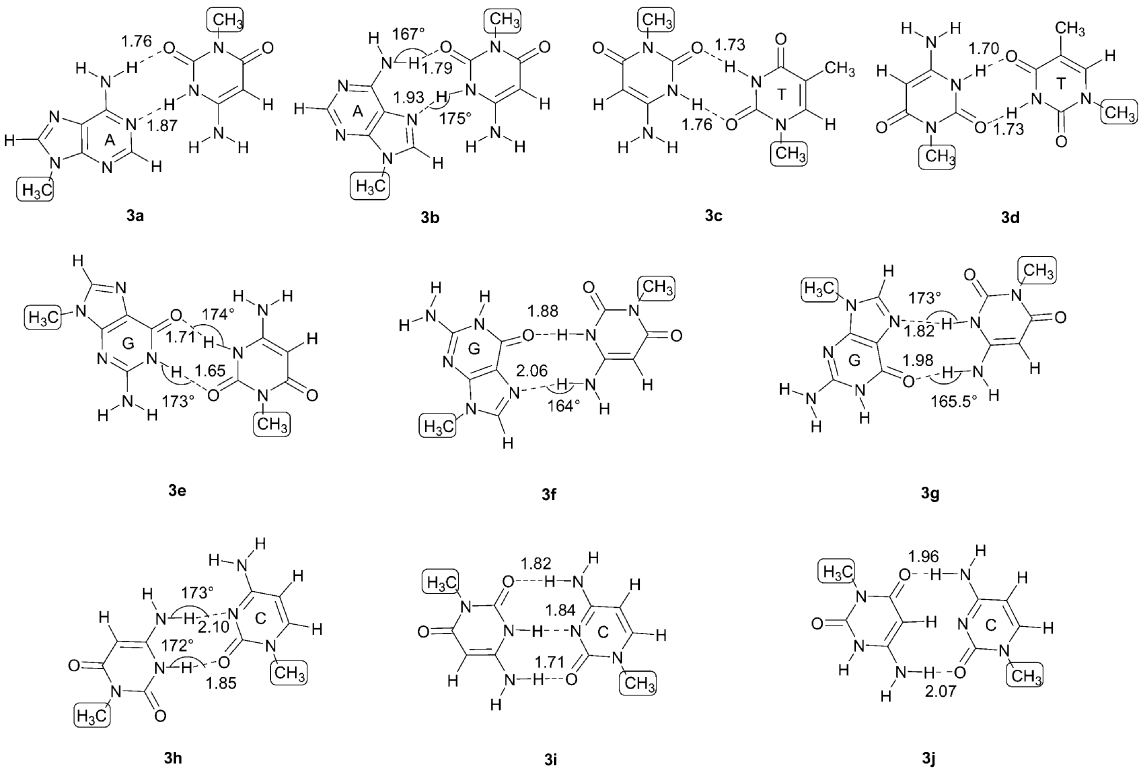


Fig. 3 The calculated duplex systems of 3-methyl-6-aminouracil with methylated adenine (**3a, 3b**), thymine (**3c, 3d**), guanine (**3e–3g**) and cytosine (**3h–3j**). The connections to the backbone chain are closed by methyl groups (encircled in the structures). The numbers refer to the optimized distances (in Å) and angles (in degrees) at the symmetric restricted case. The H-bond angle is shown only for instances where the deviation from 180° is larger than 5°.

Table 1 The σ-HOMO−1, σ-HOMO, σ-LUMO and σ-LUMO+1 energy values (in eV) of 3-methyl-6-aminouracil, 6-aminouracil and 5-methyl-6-aminouracil molecules

Orbital	3-Methyl-6-aminouracil	6-Aminouracil	5-Methyl-6-aminouracil
σ-LUMO+1	0.146	0.053	−0.055
σ-LUMO	−0.951	−1.044	−0.988
σ-HOMO	−5.895	−6.003	−5.996
σ-HOMO−1	−6.826	−7.026	−6.939

We have found that the energy levels shift very slightly due to the exchange of a hydrogen atom by a methyl group (see Table 1). This is in line with previous work[3b] in which it was shown that substituting a hydrogen atom by a methyl group at the nitrogen atom which is linked in DNA to the phosphate-sugar backbone, leaves the hydrogen bonds in the natural DNA base pairs AT and GC essentially unaffected. For both artificial bases the σ-HOMO has the largest lobe on the O4 atom. The absolute HOMOs are in the π-electron system and they have higher orbital energy in the methyl substituted cases. The absolute LUMOs also have π-symmetry and their energies are also raised by the methyl substitution. The energy level of other σ orbitals are shifted up slightly but the changes are very small and the 3- and 5-substituted isomers are accordingly expected to provide very similar interaction energies under analogous circumstances.

The intrinsic differences in the chemical behaviour of O2 and O4 atoms of **1a** and **1b** are expected to be larger than the effect caused by methyl substitution. As mentioned before, the largest lobe of the σ-HOMO orbital is centered at the O4 atom while the σ-HOMO−1 has the largest distribution on the O2 atom. This asymmetry stems from the pyrimidine ring and not from the methyl substituent. When the hydrogen bond changes from the O2 to the O4 atom, *e.g.* **2q** *vs.* **2r**, the latter provides a stronger interaction energy (see next section). The σ-HOMO−1 and the σ-HOMO functions of the 5-substituted molecule **1b** are shown in Fig. 5 (**5a** and **5b**, respectively). The unoccupied orbitals σ*-LUMO, σ*-LUMO+1 and σ*-LUMO+2 are also given in Fig. 5 (**5c**, **5d** and **5e**, respectively). These orbitals have clearly antibonding character in the arylic as well as the amino N–H bonds.

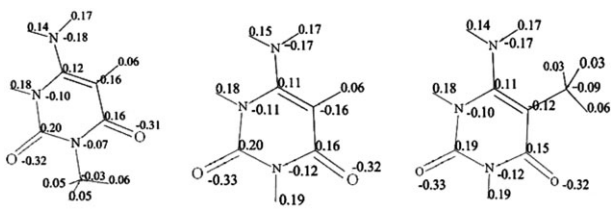


Fig. 4 VDD charges (in a.u.) of the 3- and 5-methyl substituted (left and right panel, respectively) and the purely hydrogen closed form (middle panel) of 6-aminouracil.

Interaction energy and optimized structure of the duplex systems

In Table 2 we present the energy decomposition of the total interaction energies according to the calculation method

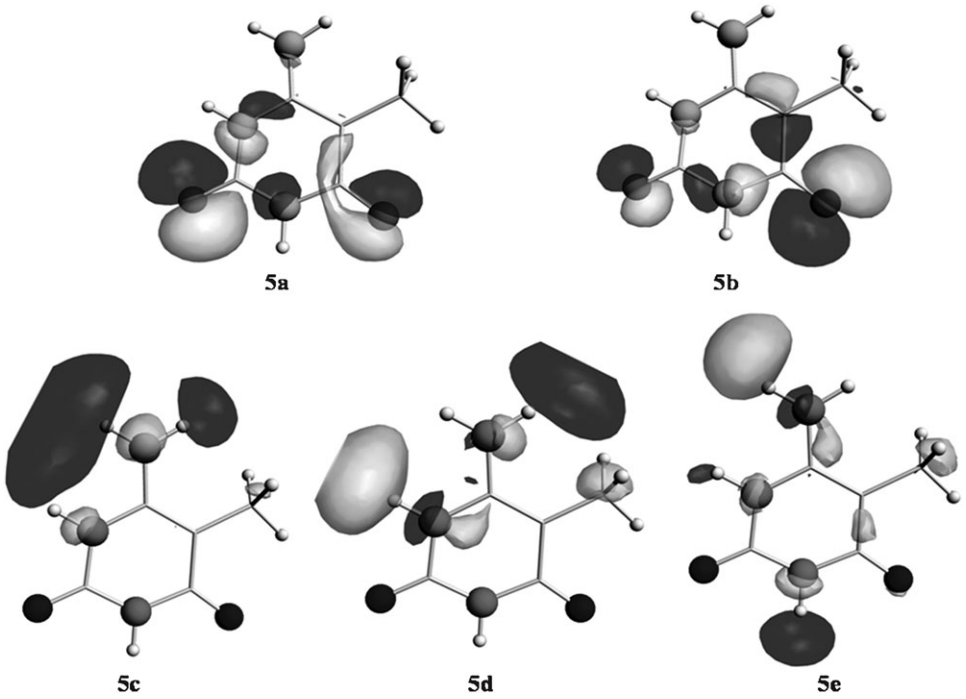


Fig. 5 The σ-HOMO−1 (**5a**), σ-HOMO (**5b**), σ*-LUMO (**5c**), σ*-LUMO+1 (**5d**) and σ*-LUMO+2 (**5e**) orbitals of 5-methyl-6-aminouracil. White atoms: hydrogen; light grey: nitrogen; dark grey: oxygen. The dark and light regions of the orbitals correspond to positive and negative signs of the wave function value, respectively.

described in the theoretical background section. One can see that the orbital interactions ΔE_{oi} play an important role in the hydrogen bonds and constitute a sizeable component of 36–43% of the total bonding forces $\Delta V_{elst} + \Delta E_{oi}$. This is in good agreement with the results of previous studies on natural and modified base-pair systems.[3] Furthermore, Table 2 also shows that the σ-type orbital interactions are one order of magnitude larger than the π-type ones. For comparison we mention that the interaction energy of the natural base pairs are −13.0 kcal mol^{-1} for the adenine-thymine (A–T) and −26.1 kcal mol^{-1} for the guanine–cytosine (G–C) base-pair (values taken form ref. 3*b*) according to the same type of calculation. Taking into account the basis set superposition error (BSSE) an BSSE-corrected B3LYP/6–31G** calculation yields −12.3 and −25.5 kcal mol^{-1} interaction energies for A–T and the (slightly nonplanar) G–C base pairs, respectively.[12] Furthermore, MP2/6-31G** yields −11.8 and −23.4 kcal mol^{-1}, respectively, using HF/6-31G** optimized geometry.[12]

Returning to our investigated systems, Table 2 shows that the interaction energies of the artificial base pairs are in most of the cases closer to the adenine–thymine level, but in two arrangements they approach the guanine–cytosine bonding energy. In the case of the **2h** and **3i**, the numbers of the H-bonds reveal the similarity with the G–C pair, but for **3e** and **2m**, we must take into account the bifurcations. As it is well known, the proton acceptor heteroatom can interact with more than one hydrogen atom in a bifurcated H-bond. In Table 3, we present those arrangements where bifurcation can occur.

Table 3 shows that single (*e.g.* **2s**) or double (**2m**, **3e**) bifurcation can be distinguished in these systems. As the interaction energy is roughly proportional with $(1/r)^{12}$, where r is the bond length, it is clear that the second interaction is very weak in those situations where only one bifurcation is presented.

Finally, we should note that it is difficult to separate the difference in the total energy caused by the presence of the bifurcation or the changes of the participating heteroatoms (*e.g.*: **2n** *vs.* **2o**). We can interpret clearly the total energy difference as the natural consequence of the appearance of the bifurcations only in the double bifurcations. These systems are examples requiring a detailed study on bifurcated hydrogen bonds, which will be published elsewhere.

Now, we focus on the differences between systems optimized under the constraint of C_s symmetry and fully optimized systems (*i.e.*, in C_1 symmetry). We see that energy differences can amount to about 1 kcal mol^{-1}. This is mainly due to the repulsion in the flat geometry, which can occur between two hydrogen atoms or two oxygen atoms coming too close together. In the fully optimized geometry the system can relax (see, for instance **2d** and **2i** in Fig. 2).

If we compare systems **2p** and **2s** with bonding energies of −13.9 and −15.6 kcal mol^{-1}, respectively, systems **2q** and **2t** with bonding energies of −9.0 and −9.5 kcal mol^{-1}, respectively, and systems **2r** and **2u** with bonding energies of −13.9 and −15.6 kcal mol^{-1}, respectively, we see that the most strongly bound system of each pair of systems is that with a hydrogen bond formed with the O4 oxygen of thymine (**2s**, **2t** and **2u**). It has been shown previously,[3c,f] that the σ-HOMO of thymine is mainly a lone pair on the O4 and higher in energy than the σ-HOMO−1, which is a lone pair on the O2. Therefore, a hydrogen bond formed with the O4 atom which has the strongest amplitude from the higher energy σ-HOMO results in a slightly stronger bond than a hydrogen bond formed with the O2 atom.

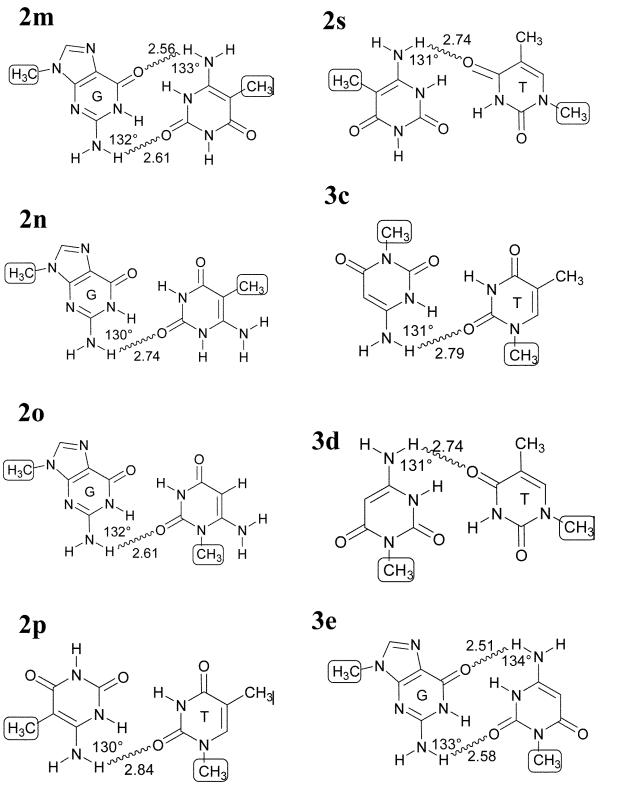
Table 2 Analysis of the total bonding energy ΔE between the base pairs (in kcal mol^{-1}). All calculations were performed in C_s symmetry, except for "NS" which are unconstrained optimizations in C_1 symmetry

	ΔE_{Pauli}	ΔV_{elstat}	ΔE_{oi}^{σ}	ΔE_{oi}^{π}	ΔE_{oi}	ΔE_{int}	ΔE_{prep}	ΔE
2a	40.3	−32.9	−21.0	−1.7	−22.7	−15.3	2.3	−13.0
2b	37.0	−30.4	−19.0	−1.5	−20.5	−13.9	2.0	−11.9
2c NS	36.7	−32.2			−21.7	−17.2	2.4	−14.8
2c	34.3	−30.1	−18.8	−2.3	−21.1	−16.8	2.2	−14.7
2d NS	32.4	−29.3			−19.4	−16.2	2.4	−13.8
2d	27.9	−25.2	−15.3	−2.2	−17.4	−14.7	2.1	−12.6
2e	33.5	−29.3	−17.2	−1.5	−18.7	−14.5	1.8	−12.7
2f	35.1	−30.7	−18.2	−1.6	−19.9	−15.4	1.9	−13.5
2g NS	19.4	−21.5			−12.3	−14.4	1.2	−13.1
2g	19.0	−21.1	−10.6	−1.5	−12.1	−14.3	0.4	−13.9
2h	50.4	−47.6	−28.0	−4.4	−32.4	−29.7	3.6	−26.0
2i NS	29.7	−24.1			−17.6	−12.0	2.1	−9.9
2i	26.7	−21.2	−14.2	−1.8	−16.0	−10.5	1.7	−8.7
2j NS	34.0	−27.3			−20.5	−13.8	2.3	−11.5
2j	30.9	−24.9	−16.8	−2.0	−18.8	−12.8	2.1	−10.7
2k NS	26.4	−28.9			−16.3	−18.8	1.6	−17.3
2k	26.7	−29.3	−14.3	−2.2	−16.5	−19.1	1.3	−17.8
2l NS	17.7	−20.4			−11.2	−14.0	0.7	−13.3
2l	17.3	−20.7	−9.7	−1.5	−11.1	−14.5	0.7	−13.7
2m	46.5	−44.0	−26.3	−4.4	−30.7	−28.2	4.4	−23.8
2n	34.4	−29.1	−18.8	−2.5	−21.3	−16.0	2.5	−13.5
2o	42.3	−35.0	−23.5	−3.1	−26.6	−19.3	3.4	−15.9
2p	32.7	−28.6	−17.8	−2.4	−20.2	−16.1	2.2	−13.9
2q	25.6	−21.5	−13.1	−1.3	−14.4	−10.4	1.3	−9.0
2r	30.4	−24.7	−16.1	−1.6	−17.7	−12.0	1.7	−10.3
2s	37.1	−32.1	−20.5	−2.8	−23.3	−18.3	2.7	−15.6
2t	28.6	−23.2	−15.1	−1.5	−16.5	−11.1	1.6	−9.5
2u	33.6	−26.6	−18.1	−1.8	−19.9	−13.0	2.0	−10.9
3a NS	37.0	−32.3			−21.8	−17.1	2.3	−14.7
3a	34.5	−30.0	−18.9	−2.3	−21.2	−16.7	2.1	−14.6
3b NS	32.0	−29.0			−19.0	−16.0	2.3	−13.7
3b	27.8	−25.0	−15.2	−2.2	−17.4	−14.5	1.9	−12.6
3c	33.4	−29.0	−18.2	−2.4	−20.5	−16.2	2.2	−13.9
3d	37.7	−32.4	−20.8	−2.8	−23.6	−18.2	3.8	−14.5
3e	47.0	−44.1	−26.6	−4.4	−31.0	−28.1	4.4	−23.7
3f NS	17.8	−20.4			−11.3	−13.9	0.7	−13.2
3f	17.4	−20.5	−9.7	−1.4	−11.2	−14.2	0.7	−13.5
3g NS	26.1	−28.6			−16.1	−18.6	1.7	−16.9
3g	26.6	−29.0	−14.2	−2.2	−16.5	−18.9	1.2	−17.7
3h NS	18.6	−20.6			−11.8	−13.8	0.9	−12.9
3h	18.8	−20.9	−10.5	−1.5	−11.9	−13.9	0.8	−13.1
3i	49.5	−46.9	−27.6	−4.4	−32.0	−29.4	3.4	−26.0
3j	31.2	−30.6	−15.4	−2.6	−18.0	−17.4	2.4	−15.0

Incorporation into natural DNA

Table 2 reveals that the bonding energy of the two artificial nucleobases with one of the natural DNA bases amounts mostly to values around the bonding energy of A–T with a few exceptions which are more strongly bound. For a possible incorporation into natural DNA, a pair of an artificial base with a natural DNA base must also fit into the Watson–Crick geometry.[3h] This means that the methyl groups, which are mimicking the sugar-phosphate backbones, must be on the same side of the Watson–Crick base pairs (*i.e.*, on the minor-groove side) and that the distance between the two base atoms (carbon or nitrogen) that are connected to the sugar–phosphate backbone (modelled here by a methyl group), must be around 8.9 Å.[3i] This demand is satisfied, when **1a** pairs with adenine to form the complex **2b** (distance between N9 and C5 is 9.0 Å) or with guanine to form **2o** (distance between N9 and C5 is 8.9 Å) and when **1a** pairs with guanine to form **3e** (distance between N9 and N3 is 8.8 Å). The pairs between the artificial bases and thymine or cytosine are too small to fit into DNA. There are of course also other factors (*e.g.* solvent effects[13] or pH dependence[3j]) which will determine if the artificial bases can be incorporated. This is however beyond the scope of this work.

Table 3 Duplex structures containing bifurcated H-bonds with the values (distances in Å and angles in degrees) of the supposed secondary weaker bond (wavy line) at the optimized structure



Conclusions

Two methyl substituted forms of 6-aminouracil were examined in this paper. In the MO analyses of the monomers, we found that the orbital levels are only slightly shifted by a substitution of a hydrogen atom for a methyl group. Also, the charge analyses show that the molecules are very similar and capable of forming hydrogen bonds with the natural DNA bases. However, a difference of about 1 kcal mol^{-1} in hydrogen bond energy was shown when a hydrogen bond was formed to the O4 of one of the artificial bases or thymine or to the O2 of these same three monomers. This is not explained by the atomic charge of the O2 and O4 atoms, because they are almost equal, but by the higher lying MO energy level of the lone pair on O4, which leads to a stronger interaction with the N–H antibonding acceptor orbital on the other base.

Furthermore, it was shown that for the artificial pairs between the methyl-substituted 6-aminouracil and a natural DNA base the orbital interactions play an important role in the hydrogen bonds and constitute a sizeable component of 36–43% of the bonding forces, $\Delta V_{elst} + \Delta E_{oi}$.

The bonding energy of the two artificial nucleobases with one of the natural DNA bases amounts mostly to values around the bonding energy of A–T with a few exceptions which are more strongly bound. For a possible incorporation into natural DNA, a pair of an artificial base with a natural DNA base must also fit into the Watson–Crick geometry. Based on our theoretical considerations, possible candidates are **2b**, **2o** and **3e**. There are of course also other factors (*e.g.* solvent or pH effects) which will determine if the artificial bases can be incorporated.

Acknowledgements

This work was done in the framework of project HPC-EUROPA (RII3-CT-2003-506079)) and the EU6 project (Grant TRIoH, LSHB-CT-2003-503480). It was also supported by the grant OTKA No. K61577. We also thank the National Research School Combination Catalysis (NRSC-C) and the Netherlands Organization for Scientific Research (NWO-CW and NWO-NCF) for financial support. Excellent service by the Stichting Academisch Rekencentrum Amsterdam (SARA) is gratefully acknowledged.

References

1 (*a*) A. M. Leconte, G. T. Hwang, S. Matsuda, P. Capek, Y. Hari and F. E. Romesberg, *J. Am. Chem. Soc.*, 2008, **130**, 2336; (*b*) I. Hirao, M. Kimoto, T. Mitsui, T. Fujiwara, R. Kawai, A. Sato, Y. Harada and S. Yokoyama, *Nat. Methods*, 2006, **3**, 729; (*c*) A. K. Ogawa, Y. Wu, D. L. McMinn, J. Liu, P. G. Schultz and F. E. Romesberg, *J. Am. Chem. Soc.*, 2000, **122**, 3274.

2 (*a*) P. Hobza and J. Sponer, *Chem. Rev.*, 1999, **99**, 3247; (*b*) J. Sponer, J. Leszczynski and P. Hobza, *Biopolymers*, 2001, **61**, 3; (*c*) J. Sponer and P. Hobza, *J. Phys. Chem. A*, 2000, **104**, 4592; (*d*) P. Hobza, J. Sponer, E. Cubero, M. Orozco and F. J. Luque, *J. Phys. Chem. B*, 2000, **104**, 6286; (*e*) S. M. LaPointe, C. A. Wheaton and S. D. Wetmore, *Chem. Phys. Lett.*, 2004, **400**, 487.

3 (*a*) C. Fonseca Guerra and F. M. Bickelhaupt, *Angew. Chem.*, 1999, **111**, 3120; C. Fonseca Guerra and F. M. Bickelhaupt, *Angew. Chem., Int. Ed.*, 1999, **38**, 2942; (*b*) C. Fonseca Guerra, F. M. Bickelhaupt, J. G. Snijders and E. J. Baerends, *J. Am. Chem. Soc.*, 2000, **122**, 4117; (*c*) C. Fonseca Guerra, F. M. Bickelhaupt and E. J. Baerends, *Cryst. Growth Des.*, 2002, **2**, 239; (*d*) C. Fonseca Guerra and F. M. Bickelhaupt, in *Modern Methods for Theoretical Physical Chemistry of Biopolymers*, ed. E. B. Starikov, J. P. Lewis and S. Tanaka, Elsevier, Amsterdam, 2006, pp. 79–97; (*e*) C. Fonseca Guerra and F. M. Bickelhaupt, in *Computational Studies of RNA and DNA*, ed. J. Sponer and F. Lankas, Springer, Berlin, 2006, vol. **2**, pp. 463–484; (*f*) C. Fonseca Guerra, F. M. Bickelhaupt, J. G. Snijders and E. J. Baerends, *Chem.–Eur. J.*, 1999, **5**, 3581; (*g*) T. van der Wijst, C. Fonseca Guerra, M. Swart and F. M. Bickelhaupt, *Chem. Phys. Lett.*, 2006, **426**, 415; (*h*) C. Fonseca Guerra and F. M. Bickelhaupt, *Angew. Chem.*, 2002, **114**, 2194; C. Fonseca Guerra and F. M. Bickelhaupt, *Angew. Chem., Int. Ed.*, 2002, **41**, 2092; (*i*) C. Fonseca Guerra, F. M. Bickelhaupt, S. Saha and F. Wang, *J. Phys. Chem. A*, 2006, **110**, 4012; (*j*) C. Fonseca Guerra, T. van der Wijst and F. M. Bickelhaupt, *Chem.–Eur. J.*, 2006, **12**, 3032; (*k*) C. Fonseca Guerra, F. M. Bickelhaupt and E. J. Baerends, *ChemPhysChem*, 2004, **5**, 481; (*l*) M. Swart, C. Fonseca Guerra and F. M. Bickelhaupt, *J. Am. Chem. Soc.*, 2004, **126**, 16718.

4 E. Szájli, G. Paragi and L. Kovács, *Nucleosides, Nucleotides Nucleic Acids*, 2005, **24**, 907.

5 ADF 2006: E. J. Baerends, J. Autschbach, A. Bérces, F. M. Bickelhaupt, C. Bo, P. L. de Boeij, P. M. Boerrigter, L. Cavallo, D. P. Chong, L. Deng, R. M. Dickson, D. E. Ellis, L. Fan, T. H. Fischer, C. Fonseca Guerra, S. J. A. van Gisbergen, J. A. Groeneveld, O. V. Gritsenko, M. Grüning, F. E. Harris, P. van den Hoek, C. R. Jacob, H. Jacobsen, L. Jensen, G. van Kessel, F. Kootstra, E. van Lenthe, D. A. McCormack, A. Michalak, J. Neugebauer, V. P. Osinga, S. Patchkovskii, P. H. T. Philipsen, D. Post, C. C. Pye, W. Ravenek, P. Ros, P. R. T. Schipper, G. Schreckenbach, J. G. Snijders, M. Solà, M. Swart, D. Swerhone, G. te Velde, P. Vernooijs, L. Versluis, L. Visscher, O. Visser, F. Wang, T. A. Wesolowski, E. M. van Wezenbeek, G. Wiesenekker, S. K. Wolff, T. K. Woo, A. L. Yakovlev and T. Ziegler, *SCM, Amsterdam*, The Netherlands.

6 A. D. Becke, *J. Chem. Phys.*, 1986, **84**, 4524; A. Becke, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**, 3098; J. P. Perdew, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1986, **33**, 8822 (*Phys. Rev. B: Condens. Matter Mater. Phys.*, 1986, **34**, 7406) (Erratum).

7 (*a*) T. Ziegler and A. Rauk, *Inorg. Chem.*, 1979, **18**, 1755; (*b*) T. Ziegler and A. Rauk, *Inorg. Chem.*, 1979, **18**, 1558; (*c*) T. Ziegler and A. Rauk, *Theor. Chim. Acta*, 1977, **46**, 1.

8 F. M. Bickelhaupt and E. J. Baerends, in *Rev. Comput. Chem.*, ed. K. B. Lipkowitz and D. B. Boyd, Wiley-VCH, New York, 2000, vol. 15, pp. 1–86.

9 J. G. Snijders, E. J. Baerends and P. Vernooijs, *At. Data Nucl. Data Tables*, 1982, **26**, 483.

10 L. Versluis and T. Ziegler, *J. Chem. Phys.*, 1988, **88**, 322.

11 C. Fonseca Guerra, J.-W. Handgraaf, E. J. Baerends and F. M. Bickelhaupt, *J. Comput. Chem.*, 2004, **25**, 189.

12 J. Sponer, P. Jurecka and P. Hobza, *J. Am. Chem. Soc.*, 2004, **126**, 10142.

13 (*a*) D. H. Turner, Naoki Sugimoto, R. Kierzek and S. D. Dreiker, *J. Am. Chem. Soc.*, 1987, **109**, 3783; (*b*) E. Stofer, C. Chipot and R. Lavery, *J. Am. Chem. Soc.*, 1999, **121**, 9503; (*c*) J. Florián, J. Sponer and A. Warshel, *J. Phys. Chem. B*, 1999, **103**, 884.