8516 *J. Phys. Chem. A* **1999,** *103,* 8516−8523

Ab Initio and Density Functional Studies of Substituent Effects of an A−U Base Pair on the Stability of Hydrogen Bonding

Shun-ichi Kawahara,[†,‡,§] Takeshi Wada,[†] Susumu Kawauchi,[||] Tadafumi Uchimaru,[‡] and Mitsuo Sekine*,[†]

Department of Life Science, Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, Nagatsuta, Midori-ku, Yokohama 226, Japan, Department of Physical Chemistry, National Institute of Materials and Chemical Research, Agency of Industrial Science and Technology, MITI, Tsukuba Science City 305, Japan, and Department of Polymer Science, Tokyo Institute of Technology, 2−12−1 O-okayama, Meguro-ku, Tokyo 152, Japan

Received: May 4, 1999

Systematic substituent effects on the stability of the hydrogen bonding between 9-methyladenine (A) and 1-methyluracil derivatives (U^X) having various substituent groups were studied. The hydrogen bond stability of mismatched base pairs between 9-methylguanine (G) and the tautomeric enol structures ($U^{X'}$) of U^X was also estimated. Geometry optimization of these bases and their base pairs using the 6-31G* basis set was carried out by Hartree−Fock (HF) self-consistent field (SCF) and the density functional theory (DFT) using Becke's three-parameter hybrid method with Predew/Wang 91 expression (B3PW91). The strength of hydrogen bonding was evaluated at the HF level and the second-order Møller−Plesset correlation energy correction (MP2) level in some cases. The calculated values for the hydrogen bonding strength using DFT were situated between those obtained by using HF and MP2 but closer to the latter than the former. Introduction of an electron-withdrawing group (EWG) into the uracil ring resulted in the formation of more stable hydrogen bonding. These results indicate that the hydrogen bond between ^{3}H of U and N^1 of A is crucial for the A−U base pairing.

Introduction

There are several basic principles for construction of the unique structure of a nucleic acid duplex. Nucleic acid is composed of a base, sugar, and phosphodiester backbone, and their components exhibit complex intra- and/or intermolecular interactions. The hydrogen bonding of base pairs is a fundamental force for molecular recognition in the duplex formation of DNA−DNA, DNA−RNA, and RNA−RNA[1] and plays an important role in heredity and mutation.[2] Moreover, its importance is now stressed because of recent medicinal applications of base-modified artificial oligonucleotides to gene therapy[3] and antisense/antigene[4] strategies. Therefore, many studies directed toward the chemical synthesis of functional base-modified nucleic acid derivatives have been reported.[5] However, there are few theoretical studies of the substituent effect of base pairs on the stability of the hydrogen bond.[6] Particularly, no systematic studies of ab initio molecular orbital calculations of modified base pairs have appeared to date. Only natural bases, adenine, guanine, uracil, thymine, and cytosine, and their base pairs have been chosen for the theoretical calculation during the past two decades.[7] Although there are several types of base pairs,[1b,7e] the Watson−Crick-type base pairs are the most important. The Watson−Crick A−U base pair has the simplest hydrogen bonding mode among the hitherto known nucleic acid base pairs. It is, however, difficult to predict the substituent effect on the whole hydrogen bond energy, because the directions of the charge localization in the two hydrogen bonds are opposite to each other. Namely, U^X has an electron-acceptor site at the exocyclic hydrogen (^{3}H(U)) of the uracil 3-position and an electron-donor site at the exocyclic oxygen (O^4(U)) of the uracil 4-position when it forms hydrogen bonds with the endocyclic nitrogen (N^1(A)) of the adenine 1-position and the exocyclic hydrogen ($^6N^6$H(A)) of the adenine 6-position, respectively (Figure 1). The former hydrogen bond of ^{3}H(U)-N^1(A) seems to be reinforced by introduction of an EWG on U^X, while the latter is weakened.

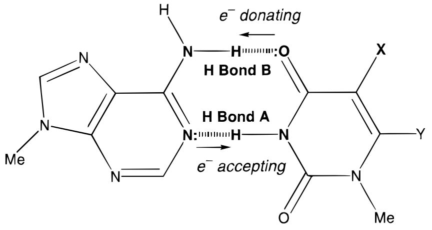

Figure 1. Electron transfer mode in hydrogen bonds between 9-methylladenine (A) and 1-methyluracil derivative (U^X).

In this paper, we report systematic, theoretical studies of substituent effects on the hydrogen bond stability of A−U^X base pairs, where thymine is considered as a 5-methyl derivative of U. In addition, the results obtained by using HF, MP2, and DFT calculations were compared. During the past decade, studies for the prediction of chemical structures using DFT[8] have been increasingly reported.

Computational Methods

The geometries of all of the bases and base pairs studied were optimized[9] by using the 6-31G* and 6-311++G** basis sets at the HF and B3PW91[10] levels. Single-point energy calculations

---

[†] Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology.

[‡] National Institute of Materials and Chemical Research.

[§] Present address: National Institute of Materials and Chemical Research.

[||] Department of Polymer Science, Tokyo Institute of Technology.

10.1021/jp9914525 CCC: $18.00 © 1999 American Chemical Society
Published on Web 09/30/1999

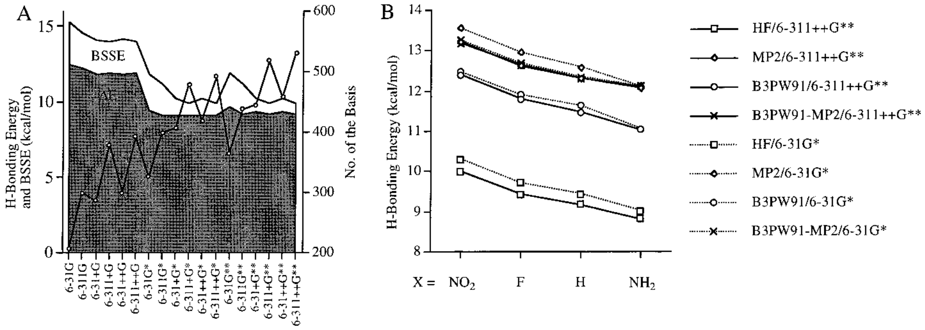

Figure 2. (A) Substitution effects of base pairs between 9-methyladenine (A) and substituted 1-methyluracil derivatives (U^X) on hydrogen bonding energies (Δ*E*, kcal/mol) at various levels using the 6-31G* and 6-311++G** basis sets. (B) Substitution effects of base pairs between A and U^X on hydrogen bonding energies (Δ*E*, kcal/mol) at the HF or HF-MP level and B3PW91 or B3PW91-MP2 using the 6-31G* basis set.

TABLE 1: Substitution Effects of Base Pairs between 9-Methyladenine (A) and Substituted 1-Methyluracil Derivatives (U^X) on Hydrogen Bond Energies (Δ*E*, kcal/mol) at Various Levels Using the 6-31G* and 6-311++G Basis Sets**

	HF		HF-MP2		B3PW91		B3PW91-MP2	
substituent	6-31G*	6-311++G**	6-31G*	6-311++G**	6-31G*	6-311++G**	6-31G*	6-311++G**
5-NO_2	−10.31	−9.99	−13.55	−13.22	−12.48	−12.41	−13.25	−13.19
5-F	−9.72	−9.43	−12.96	−12.63	−11.90	−11.80	−12.69	−12.64
5-H	−9.43	−9.18	−12.59	−12.31	−11.63	−11.47	−12.34	−12.31
5-NH_2	−9.01	−8.82	−12.12	−12.08	−11.06	−11.03	−12.01	−12.12

of the optimized structures using second-order Møller−Plesset perturbation (MP2) for electron correlation were carried out at the same basis sets. All of the MP2 calculations were performed with the frozen core approximation. It was impossible to optimize base pair structures at the MP2/6-31G* level because such calculations require extremely long periods of time.

An initial geometry of the two bases of a base pair was set on a plane[11] as Saenger's structure.[1b] The most stable conformation was found by preliminary calculations changing the dihedral angle of all the substituents at every 30° step at the HF/3-21G level.[12]

Hydrogen bond energy was evaluated using the supermolecular method.[13] It is important to consider the basis set superposition error (BSSE)[13] when intermolecular interaction is evaluated in this method. BSSE was corrected by the counterpoise method introduced by Boys and Bernardi.[14] The hydrogen bond energy ($\Delta E(\mathrm{A{-}U^X})$) of $\mathrm{A{-}U^X}$ was evaluated according to eqs 1−3. $E(\mathrm{A{-}U^X})$, $E(\mathrm{A})$, and $E(\mathrm{U^X})$ refer to the total energy of a fully geometry-optimized base pair, the energy of A with the optimized geometry, and the energy of U^X with the optimized geometry, respectively. $E(\mathrm{A{-}})$ and $E(\mathrm{U^X{-}})$ refer to the energies of A and U^X, respectively, in which the geometry of the optimized base pair $\mathrm{A{-}U^X}$ is preserved. $E(\mathrm{A{-}u^X})$ and $E(\mathrm{a{-}U^X})$ refer to those of A and U^X where calculation was done when the "ghost" molecule, shown by a small letter "u^X" or "a" was put in place of the optimized structure $\mathrm{A{-}U^X}$.

$$\delta E(\mathrm{A{-}U^X}) = E(\mathrm{A{-}U^X}) - (E(\mathrm{A}) + E(\mathrm{U^X})) \qquad (1)$$

$$\Delta E(\mathrm{A{-}U^X}) = \delta E(\mathrm{A{-}U^X}) + \mathrm{BSSE} \qquad (2)$$

$$\mathrm{BSSE} = E(\mathrm{A{-}}) - E(\mathrm{A{-}u^X}) + E(\mathrm{U^X{-}}) - E(\mathrm{a{-}U^X}) \qquad (3)$$

Kollmann et al.[7] described that the hydrogen bond energies of both Watson−Crick-type and Hoogsteen-type A−T base pairs were evaluated satisfactorily by ab initio calculations of the optimized structure at the HF/6-31G* level followed by a single-point energy calculation with the MP2/DZP method. They did not compare the calculated value with its experimental value for the Watson−Crick-type base pair although they compared the calculated value with its experimental data for the more stable Hoogsteen-type A−T base pair. Actually, this calculated value was not fitted to the experimental data.[15]

Therefore, the effect of basis function was evaluated at the HF level. The structures of A, U, and A−U base pairs were optimized with some basis functions from 6-31G to 6-311G++G**. Dependence of the basis set on hydrogen bond energy and BSSE was studied. Figure 2A shows the result that the hydrogen bond energy (Δ*E*) calculated at the 6-31G* basis set with BSSE correction is very close to that obtained by using the 6-311++G** basis set.[16] Next, dependence of the basis set on the substitution effect of the hydrogen bond energies of $\mathrm{A{-}U^X}$ base pairs was studied by using four typical substituents (X = NO_2, F, H, and NH_2), as shown in Figure 2B. The geometries of these modified base pairs were optimized at the HF and B3PW91 levels using 6-31G* and 6-311++G** and then single-point calculations at the MP2 level were done. Table 1 shows the summary of these results. Paizs and Suhai suggestion that the valence triple-ζ quality basis set improved by diffuse functions in the DFT studies is useful for hydrogen bonding complexes.[17] However, our results showed that the substitution effect on the hydrogen bond energies of $\mathrm{A{-}U^X}$ had a similar tendency in each of the four different levels using 6-31G* or 6-311++G**, as shown in Figure 2B. Particularly, the results of HF-MP2/6-311++G**, B3PW91-MP2/6-311++G**, and B3PW91-MP2/6-31G* are very close to each other. The results obtained by using MP2/6−31* gave a little higher but essentially parallel Δ*E* value for the substituents NO_2, F, and H and almost the same Δ*E* value as those from HF-MP2/6-311++G**, B3PW91-MP2/6-311++G**, and B3PW91-MP2/6-31G* for the substituent of NH_2. Considering the results described above and the number of the basis functions, the 6-31G* basis set was selected for evaluation of relative hydrogen bond energies of other substituents. Sponer and Hobza reported in their work[18] that use of electric correlation levels higher than MP2 does not influence the energy in hydrogen bonding and π-stacking by comparison of MP2 with CCSD(T) using AUG-ccpVDZ in smaller molecules. Therefore, higher correlations were not examined.

Considering the fact that the total energies obtained by using B3PW91-MP2 are lower than those derived from HF-MP2, the B3PW91-MP2 calculations seem more suitable.[19] However, it

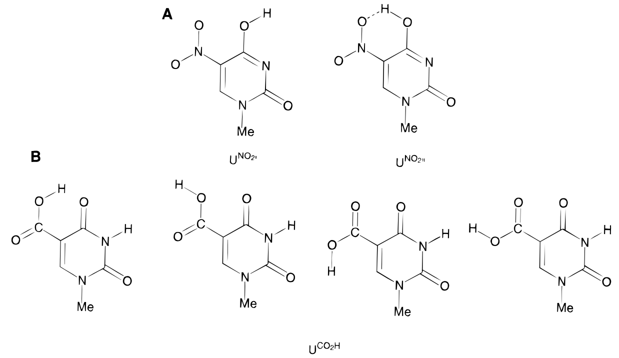

Figure 3. Structures of two tautomers of U^{NO_2}.

was found that the $\Delta\Delta E$ values calculated by the two methods are very small. Generally, it is recognized that the MO method lacks accuracy to absolutely evaluate a small difference less than 1 kcal/mol in energy between two substrates. Nonetheless, the comparative substituent effect of the hydrogen bond energy can be predicted as far as a series of compounds having a common fundamental structure are concerned. Therefore, relative values among various hydrogen bond energies of base pairs calculated were used for discussion in this paper.

The solvation effect was considered by changing the dielectric constant in the Onsager reaction field calculation[20] ("SCRF= dipole" in Gaussian 94). The structures of A−U^X optimized in gas phase were used to evaluate the molecular volumes (a_0) and then optimized in the reaction field at the HF level. Successively, single-point calculations of the structures thus optimized were performed at the MP2 level. The procedure of the BSSE correction in the solvent is the same as that used for nonsolvated calculations. The energy of each species (U^X−, A−, A−u^X, a−U^X) in solvated structures was used. The following dielectric constants were used for base pair formation: $\epsilon = 40$ (when DNA forms duplexes, the surrounding circumstance around the DNA duplexes corresponds to $\epsilon = 40$[21] This value is almost the same as the dielectric constant in DMF); and $\epsilon = 78.5$ (this value corresponds to aqueous solution). In the case of $U^{NO2'}$, the tautomer protonated ^{4}O from the opposite side of the hydrogen bond ($U^{NO2''}$) is more stable than the tautomer protonated in the same direction as the hydrogen bond ($U^{NO2'}$). However, $U^{NO2'}$ was used to evaluate the tautomerism because $U^{NO2''}$ cannot act as a C analogue (Figure 2A).

TABLE 2: Substitution Effects of Base Pairs between 9-Methyladenine (A) and Substituted 1-Methyluracil Derivatives (U^X) on Hydrogen Bonding Energies (ΔE, kcal/mol) at the HF or HF-MP Level and B3PW91 or B3PW91-MP2 Using the 6-31G* Basis Set

substituent	HF/6-31G*[a]	HF MP2/6-31G*[b]	B3PW91/ 6-31G*[c]	B3PW91- MP2/6-31G*[d]
5-NO_2	−10.31	−13.55	−12.48	−13.25
5-CN	−10.25	−13.58	−12.57	−13.29
5-CF_3	−9.96	−13.25	−12.24	−13.01
5-F	−9.72	−12.96	−11.90	−12.69
5-$COCH_3$	−9.46	−12.70	−11.59	−12.43
5-CO_2H	−9.93	−13.33	−12.16	−13.16
5-CO_2Me	−9.59	−12.79	−11.75	−12.50
5-$CONH_2$	−9.38	−12.61	−11.50	−12.37
5-H	−9.43	−12.59	−11.63	−12.34
5-CH_3	−9.51	−12.50	−11.42	−12.26
5-CCH	−9.67	−12.91	−11.88	−12.63
5-NH_2	−9.01	−12.12	−11.06	−12.01
5-NMe_2	−8.65	−11.83	−10.53	−11.58
quin	−9.18	−12.39	N/A[e]	N/A
5-Ph	−9.51	−12.89	N/A	N/A
5-Pf	−10.02	−13.48	N/A	N/A
6-aza	−9.79	−13.10	−12.17	−12.87
6-NO_2	−10.06	−13.46	−12.39	−13.21
6-F	−9.85	−13.04	−12.05	−12.72
6-CO_2H	−9.56	−12.86	−11.83	−12.62
6-NH_2	−9.69	−12.75	−11.83	−12.41
2-S	−8.66	−11.99	−10.83	−11.82
4-S	−4.12	−7.83	−6.49	−8.12
dft	−2.40	−4.13	−2.29	−4.26

[a] Hydrogen bond energy of the geometry of the base pair optimized at the HF/6-31G* level. [b] Hydrogen bond energy of the base pair obtained after a single-point MP2 calculation using the geometry of the base pair optimized at the HF/6-31G* level. [c] Hydrogen bond energy of the geometry of the base pair optimized at the B3PW91/6−31* level. [d] Hydrogen bond energy of the base pair obtained after a single-point MP2 calculation using the geometry of the base pair optimized at the B3PW91/6−31* level. [e] The base pairs having these structures were not optimized because the molecules are too big to be calculated at these levels.

All MO calculations were carried out using the Gaussian 94[22] program on a Cray C-916/12256 supercomputer.

Results and Discussions

Hydrogen Bond Energies in Gas Phase. Figure 4 shows the uracil analogues and abbreviations for the hydrogen bond

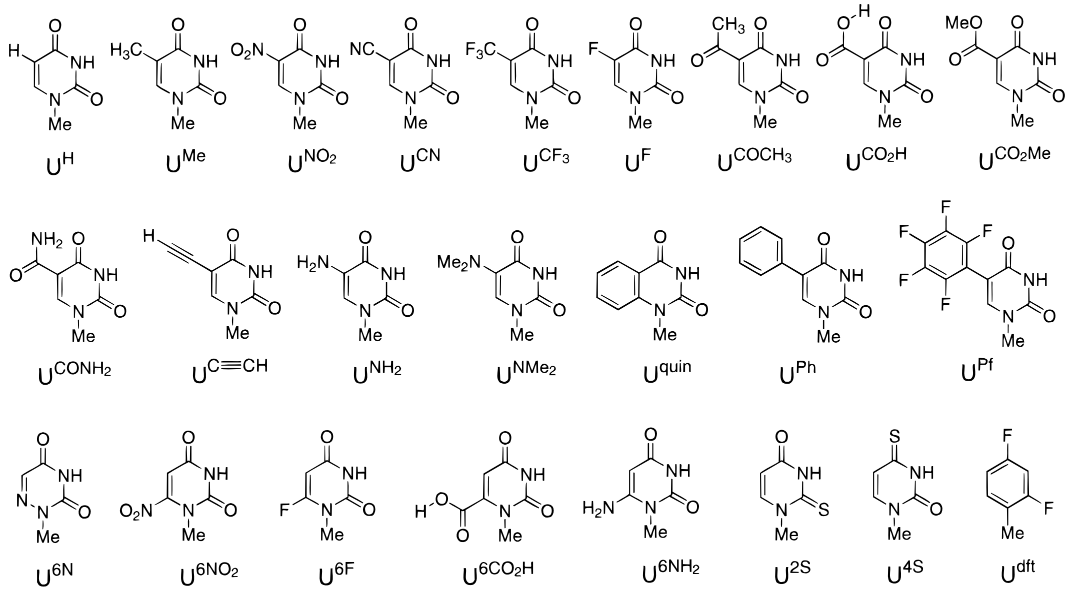

Figure 4. Structures and abbreviations of substitued uracil derivatives (U^X) used for ab initio calculation in this study.

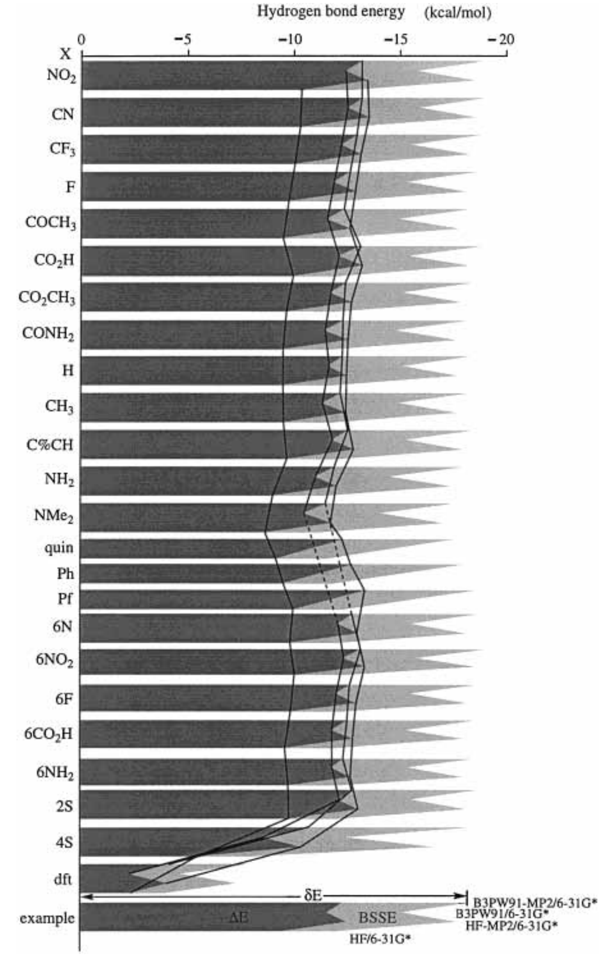

Figure 5. Hydrogen bond energies and BSSEs of A−U^X calculated by various basis functions.

energies evaluated in this study. For the following reasons, most of the compounds have a substituent group at the 5-position: (1) The 5-substituent group can be located toward the major groove of double-stranded nucleic acids so that it does not obstruct the duplex formation. (2) It can be easily introduced by several chemical methods. Some kinds of 6-substituted compounds (6-NO_2, 6-F, etc.), which are not suitable for double-strand formation because of the unfavorable syn conformation of the base moiety, were also evaluated to compare them with the 5-substituted ones. These substituted derivatives can be classified as follows: (a) unsubstituted natural bases (EWG: X = H and 5-Me); (b) bases having electron-withdrawing groups (X = 5-NO_2, 5-CN, etc.); (c) bases having electron-donating groups (EDG: X = 5-NH_2 and 5-NMe_2); (d) bases having longer conjugate systems (X = 5-Ph, 5-Pf, and quin); (e) difluorotoluene (dft) as the nonpolar hydrogen bonding substrate.[5z]

Table 2 and Figure 5 show the hydrogen bond energies δE(A−U^X), ΔE(A−U^X), and BSSE, which were calculated at the HF/6-31G*, HF-MP2/6-31G*, B3PW91/6-31G*, and B3PW91-MP2/6-31G* levels. Because of their larger structures, some structures (U^{quin}, U^{Ph}, and U^{Pf}) could not be optimized with B3PW91/6-31G*. As mentioned above, the total hydrogen bond energy of the base pair between A and U^X is enforced by introduction of an EWG. Moreover, a stronger EWG is more effective for enforcement of the hydrogen bond in all of the calculation methods tested. This tendency found in 5-substituted A−U base pairs was also seen in 6-substituted A−U base pairs. From these results, it is deduced that the hydrogen bond between ^{3}H(U)-N^1(A) (H-bond A in Figure 1) plays a more important role in these systems.

Some inherent differences in these regioisomers were found: (1) the 5-NO_2 group has a stronger effect on the hydrogen bond energy than the 6-NO_2 group and the 5-F group has a smaller effect than the 6-F group; (2) the 5-CO_2H group has a greater effect than the 5-F group, while reversibly the 6-CO_2H group has a lesser effect than the 6-F group. The marked effect of the 5-CO_2H group might be explained in terms of the intramoleculer hydrogen bonding between H of the 5-CO_2H group and ^{4}O as shown in Figure 3B, which leads to significant localization of $\delta+$ on the ^{4}O atom. The geometry optimization suggests that the structure having this cyclic hydrogen bond is the most stable.[23]

From the results for smaller systems such as $(H_2O)_2$,[24] electrostatic energy contributes largely to the hydrogen bond energy. In other words, the local charge distributions $X^{\delta-}-H^{\delta+}$ are important.[25] Therefore, the atomic charges of the associated atoms (N^3(U), ^{3}H(U), and O^4(U)) were calculated by Mulliken's and natural population methods. These results suggest that the atomic charges considerably affect the hydrogen bonding, as shown in Table 3. There is a linear relationship between the atomic charge of ^{3}H(U) of U^X and the hydrogen bond energy as shown in Figure 6.[26] It is reasonable that the data plots of U^{4S} and dft are located far from those of a series of 5-substituted uracil derivatives because of the destruction and/or decrease of the hydrogen bond ability that were caused by replacement of ^{4}O by S in U^{4S} or both N^3 by C and ^{4}O by F in dft. These substitutions are essentially different from the other cases, since these substituted atoms directly participate in hydrogen bonding.

Additivity of the Substituent Effect. Next, the additivity of substitution effects was investigated at the HF/6-31G* and B3PW91/6-31G* levels. Figure 7 shows di- or trisubstituted U derivatives at the 2- (O or S), 4- (O or S), 5- (NO_2, F, CO_2H, H or NH_2), and 6- (NO_2, F, CO_2H, H, NH_2, or endocyclic substitution from C to N) positions.

Substitution parameters ($\Delta\Delta E$) prescribed for calculation of additivity of substitution effects are defined according to eq 4 and shown in Table 4 from the results of Table 2.

$$\Delta\Delta E(U^X) = \Delta E(U^X) - \Delta E(U^H) \qquad (4)$$

Table 5 shows the hydrogen bond energies of di- or trisubstituted uracil derivatives, and Figure 8 shows the relationship between the energy difference ΔE of di- or trisubstituted uracil derivatives (vertical axis) and the additivity on the basis of estimated values (horizontal axis). Except for the ^{4}O-substituted derivatives (for both methods) and U^{2S5NH_2} (for HF), the additivity of substituent effects holds good agreement with both methods (<0.5 kcal/mol). The reason the ^{4}O-substituted derivatives do not hold the additivity is apparently because this substitution loses one of the hydrogen bonds. As described in Figure 1, the substituent effect on hydrogen bond A is opposite to that of hydrogen bond B, and the hydrogen bond energies in this paper are evaluated as the sum of all of the interactions. Therefore, the loss of one hydrogen bond causes a serious change in the substituent effect so that the additivity fails.

H-bonding Energies in Solvents. All of the discussions above are based on the results in gas phase. However, nucleic acids form duplexes in water. Therefore, the effect of substituent groups on hydrogen bonding at $\epsilon = 40$ and 78.5 was studied. Table 6 shows the energies of hydrogen bonding and the effects of the substituent. The $\Delta\Delta E$ values were calculated according to eq 4. The hydrogen bond energies are weakened by

TABLE 3: Atomic Charges of 5- and 6-Substituted Uracil Derivatives (U^X) Associated with Hydrogen Bonding with 9-Methyladenine (A)

X	N^3	3-H	4-O	X	N^3	3-H	4-O	method
5-NO_2	−0.9469	0.4360	−0.5278	5-NH_2	−0.9356	0.4250	−0.5989	Mull/HF[a]
	−0.7644	0.4634	−0.6358		−0.7530	0.4551	−0.6978	NPA/HF[b]
	−0.7622	0.3909	−0.4523		−0.7545	0.3784	−0.5269	Mull/B3PW91
	−0.6761	0.4645	−0.5374		−0.6645	0.4561	−0.6127	NPA/B3PW91
5-CN	−0.9374	0.4342	−0.5509	5-NMe_2	−0.9330	0.4253	−0.5982	Mull/HF
	−0.7611	0.4619	−0.6585		−0.7537	0.4556	−0.6990	NPA/HF
	−0.7527	0.3897	−0.4720		−0.7533	0.3790	−0.5227	Mull/B3PW91
	−0.6740	0.4638	−0.5589		−0.6650	0.4567	−0.6122	NPA/B3PW91
5-CF_3	−0.9383	0.4310	−0.5589	6-aza	−0.9228	0.4304	−0.5631	Mull/HF
	−0.7612	0.4597	−0.6649		−0.7576	0.4582	−0.6597	NPA/HF
	−0.7574	0.3855	−0.4828		−0.7389	0.3842	−0.4910	Mull/B3PW91
	−0.6722	0.4610	−0.5695		−0.6686	0.4586	−0.5660	NPA/B3PW91
5-F	−0.9430	0.4309	−0.5600	6-NO_2	−0.9295	0.4342	−0.5669	Mull/HF
	−0.7625	0.4588	−0.6591		−0.7570	0.4621	−0.6565	NPA/HF
	−0.7577	0.3854	−0.4876		−0.7445	0.3886	−0.4117	Mull/B3PW91
	−0.6754	0.4598	−0.5678		−0.6682	0.4631	−0.5101	NPA/B3PW91
5-$COCH_3$	−0.9350	0.4288	−0.5915	6-F	−1.0646	0.4286	−0.2585	Mull/HF
	−0.7596	0.4581	−0.6965		−0.7659	0.4582	−0.6834	NPA/HF
	−0.7513	0.3836	−0.5031		−0.7508	0.3828	−0.5118	Mull/B3PW91
	−0.6713	0.4596	−0.5929		−0.6796	0.4593	−0.5849	NPA/B3PW91
5-CO_2H	−0.9356	0.4346	−0.6356	6-CO_2H	−0.9295	0.4342	−0.5669	Mull/HF
	−0.7503	0.4622	−0.7243		−0.7565	0.4574	−0.6727	NPA/HF
	−0.7536	0.3902	−0.5488		−0.7416	0.3824	−0.5057	Mull/B3PW91
	−0.6574	0.4640	−0.6216		−0.6681	0.4585	−0.5758	NPA/B3PW91
5-CO_2Me	−0.9379	0.4277	−0.5586	6-NH_2	−0.9308	0.4217	−0.6221	Mull/HF
	−0.7638	0.4574	−0.6652		−0.7622	0.4535	−0.7125	NPA/HF
	−0.7533	0.3824	−0.4800		−0.7483	0.3746	−0.5399	Mull/B3PW91
	−0.6756	0.4588	−0.5652		−0.6780	0.4540	−0.6116	NPA/B3PW91
5-$CONH_2$	−0.9356	0.4298	−0.6207	2-S	−0.8506	0.4446	−0.5776	Mull/HF
	−0.7547	0.4587	−0.7168		−0.7173	0.4667	−0.6717	NPA/HF
	−0.7527	0.3842	−0.5373		−0.6829	0.3940	−0.5010	Mull/B3PW91
	−0.6641	0.4600	−0.6194		−0.6273	0.4650	−0.5765	NPA/B3PW91
5-H	−0.9290	0.4244	−0.5897	4-S	−0.8500	0.4411	−0.2587	Mull/HF
	−0.7637	0.4546	−0.6847		−0.7084	0.4644	−0.1945	NPA/HF
	−0.7443	0.3788	−0.5096		−0.6871	0.3915	−0.2179	Mull/B3PW91
	−0.6764	0.4558	−0.5865		−0.6222	0.4639	−0.1207	NPA/B3PW91
5-CH_3	−0.9333	0.4234	−0.5866	dft	−0.3661	0.2420	−0.3857	Mull/HF
	−0.7582	0.4542	−0.6899		−0.3811	0.2584	−0.3870	NPA/HF
	−0.7443	0.3788	−0.5096		−0.3157	0.1971	−0.2958	Mull/B3PW91
	−0.6764	0.4558	−0.5865		−0.3769	0.2721	−0.3279	NPA/B3PW91
5-CCH	−0.9345	0.4279	−0.5622	quin	−0.9392	0.4251	−0.5868	Mull/HF
	−0.7612	0.4573	−0.6686		−0.7581	0.4556	−0.6905	NPA/HF
	−0.7485	0.3824	−0.4847					Mull/B3PW91
	−0.6735	0.4589	−0.5709					NPA/B3PW91
5-Ph	−0.9346	0.4250	−0.5814	5-Pf	−0.9376	0.4296	−0.5772	Mull/HF
	−0.7586	0.4553	−0.6864		−0.7577	0.4588	−0.6804	NPA/HF

[a] Net atomic charge calculated by the Mulliken method. [b] Net atomic charge calculated by the natural population analysis method.

solvatation as expected. The difference between the results obtained by using $\epsilon = 40$ and $\epsilon = 78.5$ is very small. Although the tendency of substituent effects in solvents is almost the same as that in gas phase, two remarkable differences were found. First, hydrogen bonding energy is slightly enforced by introducing the NH_2 (EDG) group into the uracil residue at the 5-position. Second, the substitution effects ($\Delta\Delta E$) by the introduction of EWG are much greater in solvents than in gas phase.

Tautomerism of U^X. The existence of nucleobase tautomers has a pronounced effect on molecular recognition in the duplex formation. As shown in Figure 9, the enol tautomer ($U^{X\prime}$) of U^X is a cytosine (C) equivalent with respect to hydrogen bond formation. Therefore, if such a tautomer can be stabilized more sufficiently than the parent structure U^X by introducing a 5-substituent, $U^{X\prime}$ can form hydrogen bonding with G so that accurate molecular recognition becomes impossible.

Both the experimental results[27] and theoretical studies[7b−d,f−m,p−r,t−v] suggest that this type of unusual hydrogen bonding is not serious because there is a big difference in energy between the imide and enol tautomers in the case of the natural base of U or T. Recently, Roberts[28] et al. reported theoretical and experimental consideration of the tautomerism of isocytosine (*iso*-C) and isoguanine (*iso*-G), base pair formation between them, and unmodified bases under solvated conditions. Taking into account their results, the NO_2, F, H and NH_2 groups were selected as the representative substituents for this purpose. From the results of the calculation of the A−U^X base pairs, the order of the electronic effect of these substituents was clear. In addition, the 4-SH tautomer of U^{4S} was calculated as an example of an obviously undesirable tautomer. The stability of A′−$U^{X\prime}$, i.e., the fully tautomerized form of A−U^X, was not studied because it is so unstable that there is no local minimum. Table 7 shows the energy difference between the imide and enol tautomers of each substituted uracil derivative in gas phase, in duplex ($\epsilon = 40$), and in water ($\epsilon = 78.5$). In all cases, the imide tautomers are much more stable than the enol tautomers. The energy differences between the tautomers are smaller in the Onsager model than in gas phase in all of the substituents tested. There is also a tendency that the energy gap between the imide

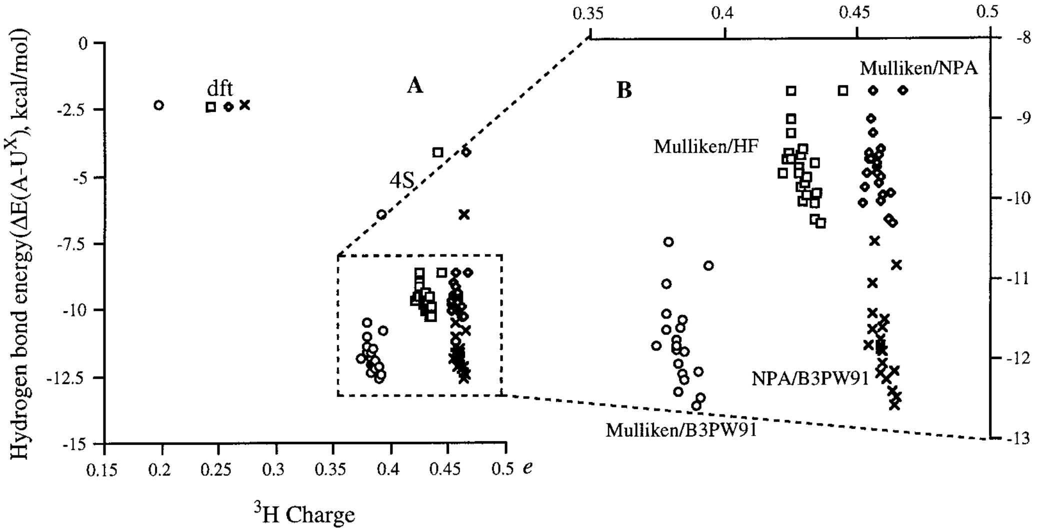

Figure 6. Relationship between the atomic charge of ^{3}H of U^X and the hydrogen bond energy of A—U^X.

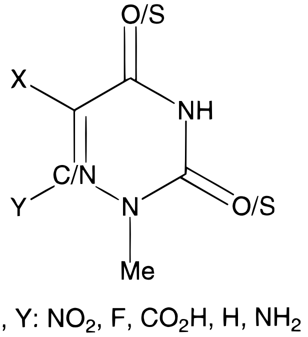

Figure 7. Di- or trisubstituted 1-methyluracil derivatives (U^X).

TABLE 4: Substitution Parameters (ΔΔ*E*, kcal/mol) for Calculation of the Additivity of Substitution Effects (kcal/mol)

X	HF/6-31G*	B3PW91/6-31G*
5-NO_2	−0.87	−0.85
5-F	−0.28	−0.27
5-CO_2H	−0.50	−0.53
5-H	0.00	0.00
5-NH_2	0.42	0.57
6-N	−0.35	−0.53
6-NO_2	−0.63	−0.76
6-F	−0.41	−0.42
6-CO_2H	−0.12	−0.20
6-NH_2	−0.25	−0.20
2-S	0.78	0.80
4-S	5.31	5.15

and enol forms of U^X becomes smaller than that of the unmodified U by introducing an EWG in solvents. Unexpectedly, the enol tautomer (U$^{4S\prime}$) of U^{4S} is less stable than in the tautomer (U$^{NH_2\prime}$) of U^{NH_2}. Because these results suggest that these 5-substituted uracil derivatives can exist as the normal imino form, the accuracy of molecular recognition might maintain without interaction with the mismatched base of guanine.

Conclusion

TABLE 5: Hydrogen Bond Energies (Δ*E*, kcal/mol) of Base Pairs between 9-Methyladenine (A) and Di- or Trisubstituted Uracil Derivatives (U^X) Based on the Actual Calculation and the Additivity Rule (kcal/mol)

	actual calcd		additivity calcd	
X	HF	B3PW91	HF	B3PW91
2,4-diS	−5.10	−7.26	−3.34	−5.68
5-F,6-CO_2H	−9.87	−12.11	−9.84	−12.10
5-F,6-NO_2	−10.42	−12.89	−10.34	−12.66
5-F,6-NH_2	−9.94	−11.94	−9.97	−12.10
5,6-di-F	−10.16	−12.30	−10.13	−12.32
5-F,6-aza	−10.10	−12.45	−10.07	−12.44
2-S,5-F	−8.88	−11.03	−8.94	−11.10
4-S,5-F	−7.96	−10.25	−4.40	−6.76
2,4-di-S,5-F	−5.02	−6.95	−3.63	−5.95
2-S,5-F,6-NO_2	−9.15	−11.55	−9.57	−11.86
2-S,5-F,6-NH_2	−9.15	−11.25	−9.20	−11.30
2-S,5-NO_2	−9.26	−11.27	−9.53	−11.68
4-S,5-NO_2	−7.61	−9.66	−4.99	−7.33
2,4-di-S,5-NO_2	−4.43	−6.22	−4.22	−6.53
2-S,6-aza	−9.14	−11.49	−9.01	−11.36
4-S,6-aza	−8.51	−10.54	−4.47	−7.02
2,4-di-S,6-aza	−4.82	−7.97	−3.70	−6.22
5-$NO_{2,6}$-aza	−10.75	−13.16	−10.66	−13.01
5-$NH_{2,6}$-aza	−9.26	−11.37	−9.37	−11.59
2-S,5-NH_2	−7.07	−10.29	−8.24	−10.25
4-S,5-NH_2	−7.49	−9.30	−3.70	−5.91
2,4-di-S,5-NH_2	−4.52	−6.27	−2.92	−5.11

The hydrogen bond energies between A and substituted U derivatives are studied by MO (HF/6-31G* and MP2−HF/6-31G*) and DFT (B3PW91/6-31G* and B3PW91-MP2/6-31G*) methods. In the A—U^X base pair, the hydrogen bond between N^3(U)-^{3}H(U)-N^1(A) is more important than the hydrogen bond between O^4(U)-^{6}H(A)-^{6}N(A). The hydrogen bond energies in A—U^X are reinforced by introduction of an EWG group, and electrostatic energy contributes significantly to stabilization of the base pair. The additivity of substituent effects is valid except for ^{4}O-substituted derivatives. Although the tautomerization of U^X leads to an enol tautomer by introduction of an EWG, the imide tautomer is much more stable than the enol tautomer.

Tautomerism of U^X does not seem to play an important role from the results of the calculation both in the gas phase and in the solvent because of its big energy differences. The substituent effect on the hydrogen bond energies of A—U^X base pair in the solvents is similar to that in the gas phase.

The difference in the hydrogen bond energy depends on the calculation methods more greatly than on the substituent effect. The absolute values of Δ*E* are B3PW91-MP2 ≅ HF-MP2 > B3PW91 > HF, and the results of B3PW91 are closer to those of B3PW91-MP2 and HF-MP2 than HF. The tendency of the substitution effect is almost the same in all methods.

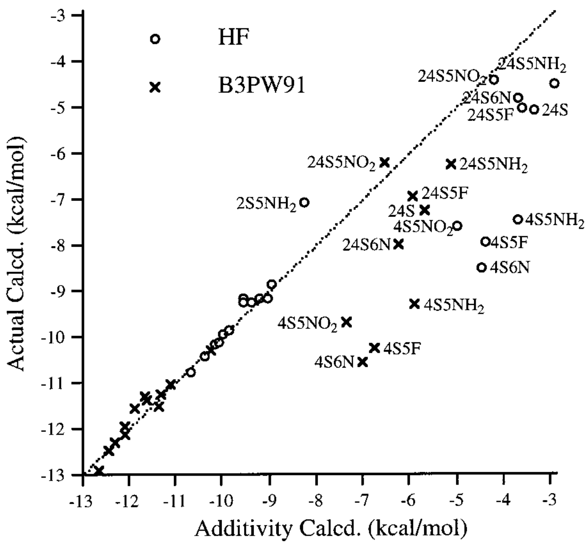

Figure 8. Additivity of substitution effects.

TABLE 6: Hydrogen Bond Energies (ΔE, kcal/mol) of Base Pairs between 9-Methyladenine (A) and Substituted 1-Methyluracil Derivatives (U^X) in Solvents ($\epsilon = 40$ and 78.5) and Gas Phase and Substitution Parameters (ΔΔE, kcal/mol)

	HF/6-31G*				HF-MP2/6-31G*			
X	ϵ=40	ϵ=78.5	gas	ΔΔ*E*	ϵ=78.5	ΔΔ*E*	gas	ΔΔ*E*
5-NO_2	−7.48	−7.41	−11.38	−2.47	−11.34	−2.42	−13.55	−0.96
5-F	−6.35	−6.27	−10.07	−1.16	−10.01	−1.21	−12.96	−0.47
5-H	−5.02	−4.92	−8.91		−8.82		−12.59	
5-NH_2	−5.41	−5.33	−9.09	−0.18	−9.02	−0.20	−12.12	+0.47

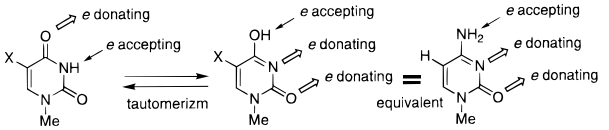

Figure 9. Hydrogen bonding modes of U^X, $U^{X'}$, and 1-methylcytosine (C).

TABLE 7: Energy Differences (kcal/mol) in Tautomerism between 5-Substituted Unidine Derivatives (U^X) and Their Tautomers ($U^{X'}$)

X	gas phase	$\epsilon = 40$	$\epsilon = 78.5$
5-NO_2	13.32	15.36	15.40
5-F	14.24	16.48	16.44
5-H	13.43	17.52	17.46
5-NH_2	16.84	24.51	24.40
4S	15.34	17.33	15.93

These results would provide important information for the molecular design of functional nucleic acid derivatives as antisense molecules and hydrogen bond character including tautomerization of nucleobases.

Acknowledgment. The services and computational time made available by Computer Center of Tokyo Institute of Technology and Research Information Processing Center (RIPS) have been essential to this study and are gratefully acknowledged. This work was supported by a grant from "Research for the Future" Program of the Japan Society for the Promotion of Science (JSPS-RFTF97I00301) and a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture, Japan.

Supporting Information Available: Full tables of the structural parameters (bond lengths, bond angles and dihedral angles of the optimized structure), total energies of component molecules of base pairs and supermolecules. This information is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) (a) Watson, J.; Crick, H. C. *Nature* **1953**, *171*, 757−738. (b) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, Berlin, Heidelberg, Tokyo, 1984.

(2) Watson, J. D.; Hopkins, N. H.; Roberts, J. W.; Steitz, J. A.; Weiner, A. M. *Molecular Biology of the Gene*; Benjamin/Cummings Publishing Company, Inc.: New York, 1987.

(3) (a) Thompson, L. *Correcting the Code−Inventing the Genetic Cure for the Human Body*; Simon & Schuster: New York, 1994. (b) Lyon, J.; Gorner, P. *Altered Fates−Gene Therapy and the Retooling of Human Life*, W. W. Norton & Company: New York, 1995, and references therein.

(4) Murray, J. A. H., Ed. *Antisense RNA and DNA*; Wiley-Liss, Inc.: New York, 1992, and references therein.

(5) Base Modification References. For review: (a) Limbach, P. A.; Crain, P. F.; AcCloskey, J. A. *Nucleic Acids Res.* **1994**, *22*, 2183−2196. For 5-NO_2: (b) Wempen, I.; Doerr, I. L.; Kaplan, L.; Fox. J. J. *J. Am. Chem. Soc.* **1960**, *82*, 1624−1629. (c) Niedballa, U.; Vorbrüggen, H. *Nucleic Acid Chem.* **1978**, *1*, 431−433. For 5-CN: (d) Dornaw, A.; Petsch, G. *Ann.* **1954**, *588*, 45−61. (e) Bleakley, R. C.; Jones, A. S.; Walker, R. T. *Nucleic Acids Res.* **1975**, *2*, 683−690. For 5-CF_3: (f) Heidelberger, C.; Parsons, D. G.; Remy, D. G. *J. Am. Chem. Soc.* **1962**, *84*, 3597−3598. For 5-F: (g) Meinert, H.; Cech, D. *Z. Chem.* **1972**, *12*, 292−293. (h) Meinert, H.; Cech, D. *Z. Chem.* **1972**, *12*, 335−336. (i) Earl, R. A.; Townsend, L. B. *J. Heterocycl. Chem.* **1972**, *9*, 1141−1143. For C%CH: (j) Barr, P. J.; Jones, A. S.; Serafinowsky, P.; Walker, R. T. *J. Chem. Soc., Perkin Trans. 1* **1978**, 1263−1267. For 5-$COCH_3$: (k) Evans, C. H.; Jones, A. S.; Walker, R. T. *Tetrahedron* **1973**, *29*, 1611−1514. For 5-CO_2H: (l) Agathocleous, D. C.; Shaw, G. *J. Chem. Soc., Perkin Trans. 1* **1993**, 2555−2559. (m) Maurier, N.; Trabaud, C.; Niddam, V.; Graciet, J.-C.; Camplo, M.; Chermann, J.-C.; Kraus, J.-L. *Nucleosides Nucleotides* **1996**, *15*, 1397−1409. For 5-NH_2: (n) Ferrer, E.; Neubauer, G.; Mann, M.; Erutja, R. *J. Chem. Soc., Perkin Trans. 1* **1997**, 2051−2057. For 5-NMe_2: (o) Senda, S.; Hirota, K.; Asao, T. *Chem. Pharm. Bull.* **1974**, *22*, 189−195. For quin: (p) Butt, A.; Parveen, R. R. P. *J. Sci. Ind. Res.* **1972**, *15*, 234−236. For 5-Ph: (q) Hirai, K.; Matsuda, H.; Kishida, Y. *Chem. Pharm. Bull.* **1973**, *21*, 1090−1095. For 6N: (r) Saenger, W.; Suck, D. *Nature* **1973**, *242*, 610−612. For 6-F: (s) Korsakov, M. V.; Ushmorov, A. G.; Yakovleva, V. D.; Studentsov, E. P.; Ivin, B. A. *Khim. −Farm. Zh.* **1984**, *18*, 1041−1044. For 6-NH_2: (t) Shim, J. L.; Niess, R.; Broom, A. D. *J. Org. Chem.* **1972**, *37*, 578−581. (u) Ueda, T.; Ohtsuka, H. *Chem. Pharm. Bull.* **1973**, *21*, 1530−1534. For 2S: (v) Brown, D. M.; Parihar, D. B.; Todd, A.; Varadarajan, S. *J. Chem. Soc.* **1958**, 3028. (w) Ueda, T.; Nishino, H. *Chem. Pharm. Bull.* **1969**, *17*, 920−926. For 4S: (x) Fox, J. J.; Van Praag, D.; Wempen, I.; Doerr, I. L.; Cheong, L.; Knoll, J. E.; Eidinoff, M. L.; Bendich, A.; Brown, G. B. *J. Am. Chem. Soc.* **1959**, *81*, 178−187. (y) Lipsett, M. N. *J. Biol. Chem.* **1965**, *240*, 3975−3978. For dft: (z) Schweitzer, B. A.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, *117*, 1863−1872.

(6) Kyogoku, Y.; Lord, R. C.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *57*, 250−257.

(7) (a) Scheiner, S.; Kern, C. W. *Chem. Phys. Lett.* **1978**, *57*, 331−333. (b) Czerminski, R.; Lesyng, B.; Pohorille, A. *Int. J. Quantum Chem.* **1979**, *16*, 605−613. (c) Czerminski, R.; Lesyng, B.; Pohorille, A. *Int. J. Quantum Chem.* **1979**, *16*, 1141−1148. (d) Scheiner, S.; Kern, C. W. *J. Am. Chem. Soc.* **1979**, *101*, 4081−4085. (e) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc.* **1987**, *109*, 1302−1307. (f) Person, W. B.; Szczepaniak, K.; Szczesniak, M.; Kwiatkowski, J. S.; Czerminski, L. H.; Czerminski, R. *J. Mol. Struct.* **1989**, *194*, 239−258. (g) Le's, A.; Adamowicz, L. *J. Phys. Chem.* **1989**, *93*, 7078−7081. (h) Czerminski, R.; Szczepaniak, K.; Person, W. B.; Kwiatkowski, J. S. *J. Mol. Struct.* **1990**, *237*, 151−163. (i) Le's, A.; Adamowicz, L. *J. Am. Chem. Soc.* **1990**, *112*, 1504−1509. (j) Sabio, M.; Topiol, S.; Lumma, W. C., Jr. *J. Phys. Chem.* **1990**, *94*, 1366−1372. (k) Katritzky, A. R.; Szafran, M. *J. Chem. Soc., Perkin Trans. 2* **1990**, 871−876. (l) Katritzky, A. R.; Karelson, M. *J. Am. Chem. Soc.* **1991**, *113*, 1561−1566. (m) Leszczy'nski, J.; Lammertsma, K. *J. Phys. Chem.* **1991**, *95*, 3128−3132. (n) Trollope, K. I.; Gould, I. R.; Hiller, I. H. *Chem. Phys. Lett.* **1993**, *209*, 113−116. (o) Gould, I. R.; Kollman P. A. *J. Am. Chem. Soc.* **1994**, *116*, 2493−2499. (p) Florián, J.; Hrouda, V.; Hobza, P. *J. Am. Chem. Soc.* **1994**, *116*, 1457−1460. (q) Destexhe, A.; Smets, J.; Adamowicz, L.; Maes, G. *J. Phys. Chem.* **1994**, *98*, 1506−1514. (r) Jiang, S.-P.; Raghunathan, G.; Ting, K.-L.; Xuan, J. C.; Jernigen, R. L. *J. Biomol. Struct. Dyn.* **1994**, *12*, 367−382. (s) Hobza, P.; Sponer, J.; Polásek, M. *J. Am. Chem. Soc.* **1995**, *117*, 792−798. (t) Gould, I. R.; Burton, N. A.; Hall, R. J.; Hillier, I. H. *THEOCHEM.* **1995**, *331*, 147−154. (u) Paglieri, L.; Corongiu, G.; Estrin, D. A. *Int. J. Quantum Chem.* **1995**, *56*, 615−625. (v) Florián, J.; Leszczynski, J. *J. Am. Chem. Soc.* **1996**, *118*, 3010−3017. (w) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1996**, *14*, 117−135. (x) Sponer, J.; Hobza, P. *Chem. Phys.* **1996**, *204*, 365−372.

(8) (a) Zeigler, T. *Chem. Rev.* **1991**, *91*, 651−667. (b) Johnson, B. G.; Gill, P. M. W.; Pople, J. A. *J. Chem. Phys.* **1993**, *98*, 5612−5626.

(9) See Supporting Information (Tables 2 and 4).

(10) (a) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652. (b) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. *J. Chem. Phys.* **1989**, *90*, 5622−5629. (c) Curtiss, L. A.; Jones, C.; Trucks, G. W.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1990**, *93*, 2537−2545.

(11) (a) Some theoretical studies[11b] suggested the nucleic acid bases form most stable H-bonding at nonplanner structures. But energy improvement by nonplanarity is very small. So we adopt the planar structure as simplified models. (b) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965−1974, and references therein.

(12) The dihedral angles of these substituents are 0° (on plane) except for Ph (ca. 45°) and Pf (ca. 60°). See Supporting Information (Tables 2 and 4) for the details.

(13) (a) Hobza, P.; Zahradník, R. *Chem. Rev.* **1988**, *88*, 871−897. (b) Chalasinski, G.; Gutowski, M. *Chem. Rev.* **1988**, *88*, 943−962. (c) Buckingham, A. D.; Fowler, P. W.; Hutson, J. M. *Chem. Rev.* **1988**, *88*, 963−988.

(14) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553−566.

(15) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149−1170. This paper deals with a Hoogsteen-type A−T pair.

(16) See Supporting Information (Table 1) for the details.

(17) Paizs, B.; Suhai, S. *J. Comput. Chem.* **1998**, *19*, 575−584.

(18) Sponer, J.; Hobza, P. *Chem. Phys. Lett.* **1997**, *267*, 263−270.

(19) See Supporting Information (Table 5) for the details.

(20) Onsager, L. *J. Am. Chem. Soc.* **1938**, *58*, 1486−1493.

(21) Florian, J.; Leszcynski, J. *J. Am. Chem. Soc.* **1996**, *118*, 3010−3017.

(22) Frisch, M. J.; Trucks, G. W.; Schlegel. H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzales, C.; Pople, J. A. Gaussian, Inc.: Pittsburgh, PA, 1995.

(23) See Supporting Information (Table 2) for the details.

(24) (a) Morokuma, K. *Acc. Chem. Res.* **1977**, *10*, 294−300. (b) Frisch, M. J.; Del Bene, J. E.; Binkley, J. S.; Schaefer, H. F., III *J. Chem. Phys.* **1986**, *84*, 2279−2289. (c) Loushin, S. K.; Liu, S.; Dykstra, C. E. *J. Chem. Phys.* **1986**, *84*, 2720−2725.

(25) Considerations based on frontier molecular orbital theory are not so good to explain the substitution effect on hydrogen bonding. See Supporting Information.

(26) The relationship between the atomic charges of ^{3}H(U) in A−U^{X} and the hydrogen bond energies was worse than that between the atomic charges of ^{3}H(U) in U^{X} and the hydrogen bond energies of A−U^{X}.

(27) (a) Kroon, J.; Kanters, J. A.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Duijneveldt, F. B.; Vliegenthart, J. A. *J. Mol. Struct.* **1975**, *24*, 109−129. (b) Kaun, E.; Rüterjans, H.; Hull, W. E. *FEBS Lett.* **1982**, *141*, 217−221.

(28) Roberts, C.; Bandaru, R.; Switzer, C. *J. Am. Chem. Soc.* **1997**, *119*, 4640−4649.

(29) (a) Piebar, M.; Kroon, P. A.; Prestegrad, J. H.; Chan, S. I. *J. Am. Chem. Soc.* **1973**, *95*, 3408. (b) Wong, Y. P.; Wong, K. L.; Kearns, D. R. *Biochem. Biophys. Res. Commun.* **1972**, *6*, 1580−1587.