Theoretical Study of the Hydrogen Bond Energy of Base Pairs Formed between Substituted 1-Methylcytosine Derivatives and 9-Methylguanine

Shun-ichi Kawahara,[†] Akio Kobori,[‡] Mitsuo Sekine,[‡] Kazunari Taira,[†] and Tadafumi Uchimaru*,[†]

National Institute of Advanced Industrial Science and Technology (AIST), AIST Tsukuba Central, Tsukuba, 305-8565, Japan, and Department of Life Science, Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, Nagatsuta, Midori-ku, Yokohama 226-8501, Japan

Received: June 27, 2001; In Final Form: September 25, 2001

The substitution effect on hydrogen bond energy of the Watson−Crick type base pair between 9-methylguanine and chemically modified 1-methylcytosine derivatives was evaluated by an ab initio molecular orbital theory. A remarkable tendency was observed, namely, that cytosine derivatives possessing an electron-donating group form a stable base pair with guanine. Nevertheless, neither the hydrogen bond distance nor the charge distribution was a valid index for the hydrogen bond status in C^X−G base pairing. An intramolecular hydrogen bond between the substituent and the *exo*-cyclic amino moiety also had an important role in the substitution effect of the base pair stability.

Introduction

The Watson−Crick type base pair formation, guanine (G)−cytosine (C) and adenine (A)−uracil (U)/thymine (T) (Figure 1) is fundamental for molecular recognition in the duplex formation of nucleic acids.[1] The processes of transcription from DNA to mRNA,[2] and of translation from mRNA to protein via tRNA,[3] are also based on the formation of the Watson−Crick type base pairs.

The molecular recognition via highly selective Watson−Crick base pairing has attracted widespread attention; for example, it has been applied to construction of artificial supermolecular systems,[4] to template synthesis,[5] and also especially to antisense technology,[6] which are topics of interest from the standpoint of control of expression of genetic information. These applications are based on the selective hydrogen bond formation of nucleic acid bases, so a molecule which is able to selectively form a stable complex is needed. However, there are no systematic studies targeting the improvement of the base pair stability. Thus, to improve the base pair stability, computer-aided molecular design of nucleic acid base analogues is highly demanded.

We have already reported ab initio molecular orbital study of the substitution effect on hydrogen bond energy in the base pairs between 9-methyl adenine (A) and modified 1-methyl uracil derivatives (U^X)[7] and between modified 9-methyl adenine (A^X) and 1-methyl uracil (U).[8] In the case of the substitution effect on the hydrgen bond energies in the U^X−A base pairs, a remarkable trend was observed: U^X possessing a stronger electron-withdrawing group (EWG) forms a more stable base pair. On the other hand, no remarkable trend was observed in the relation between the substituent on adenine derivatives and the hydrogen bond energies, in the case of the A^X−U base pairs.

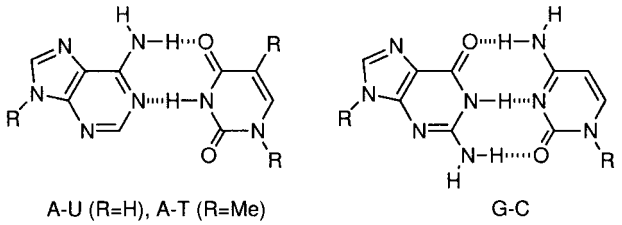

Figure 1. Watson−Crick base pairs.

We also reported that the substitution effect on hydrogen bond energy of the A^X−U base pairs, calculated by an ab initio method, was in good agreement with the substitution effect on experimentally observed binding properties.

Although there are many theoretical studies on the hydrogen bond energy of the Watson−Crick type base pairs between natural nucleic acid bases,[9] no systematic ab initio molecular orbital studies on modified base pairs have been reported except for our studies.[7,8] Theoretical studies are important for understanding the nature of the hydrogen bond in the base pair and are useful for applications such as those described above. We report herein an ab initio study regarding the substitution effect on hydrogen bond energy in the base pair between modified 1-methylcytosine derivatives (C^X, Figure 2) and 9-methylguanine (G).

Computational Methods

In most theoretical studies, the hydrogen bond energies of the Watson−Crick type base pairs were evaluated at the second-

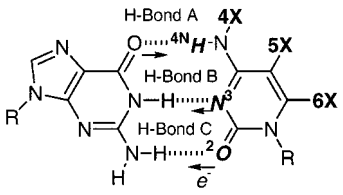

Figure 2. Hydrogen bond between C^X and G.

* To whom correspondence should be addressed. Phone: +81-298-61-4800. Fax: +81-298-61-4487. E-mail: t-uchimaru@aist.go.jp.
[†] National Institute of Advanced Industrial Science and Technology.
[‡] Tokyo Institute of Technology.

10.1021/jp0124645 CCC: $20.00
Published on Web 10/27/2001

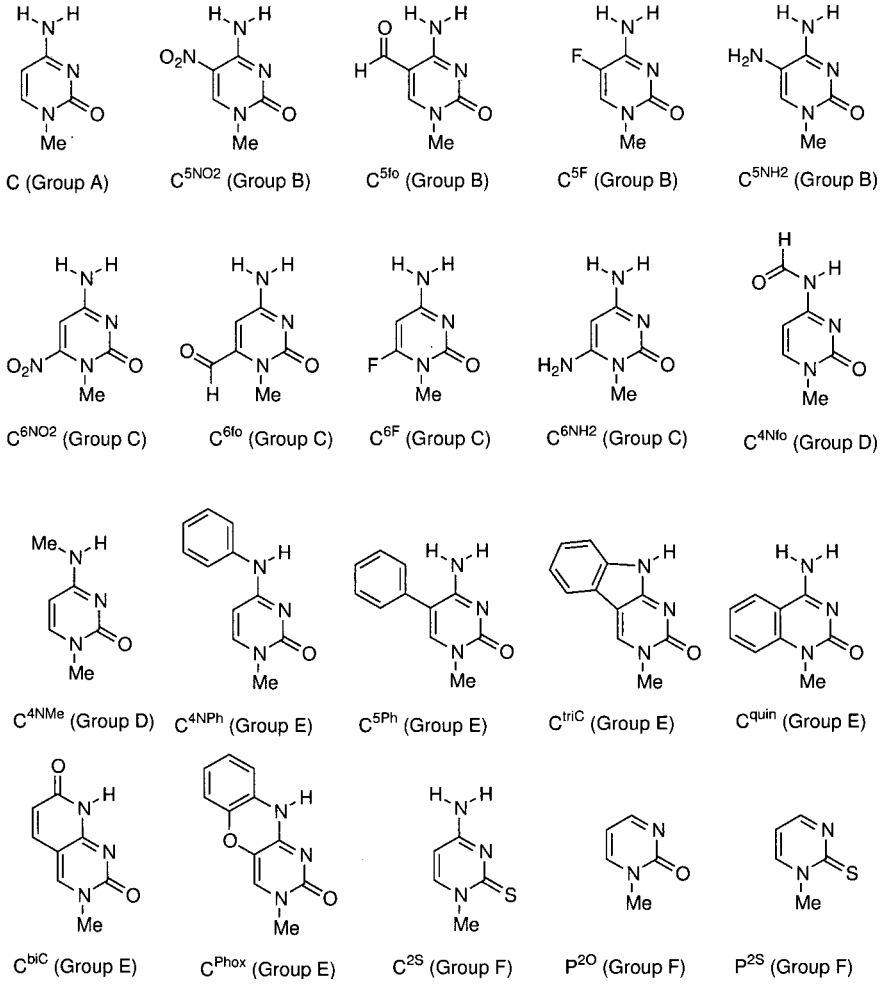

Figure 3. Substituent introduced 1-methylcytosine derivatives (C^X) in this study.

order Møller−Plesset (MP2) level of theory using double-ζ basis sets with polarization.[9] Rablen et al. showed[10] that hydrogen bond energies of small molecules calculated at the level of B3LYP/6-31++G(2d(X+),p)//B3LYP/6-31++G(d(X+),p)[11] were in good agreement with the results of the complete basis set approach (CBS-Q[12]). Sponer et al. reported[9m] that the hydrogen bond energies of some model compounds in MP2/6-31G*(0.25)//MP2/6-31G*(0.25)[13] reproduced relatively well the result of much larger basis sets. They also found[14] that the contribution of higher-level electron correlation was small on hydrogen bond energy and that MP2 interaction energies were close to the results of coupled cluster electron correlation (CCSD(T)[15]) data. Hydrogen bond energy is mainly characterized by electrostatic contribution,[16] and the contribution of electron correlation higher than MP2 level should be relatively small.[14] Considering these two results, the hydrogen bond energies of Watson−Crick type base pairs can be studied in MP2 level electron correlation. We already reported an ab initio study regarding the basis set effect on the calculated hydrogen bond energies of Watson−Crick type base pairs at the MP2 levels of theory.[17] The values of hydrogen bond energies of A−U and G−C base pairs, evaluated at the computational levels of MP2/6-31+G(2d′,p′)[12]//HF/6-31G(d,p), were in excellent agreement not only with the values calculated at MP2/6-311++G(3df,p)//HF/6-311++G(3d,p) but also with the values reported by Rablen et al.[13] Thus, MP2/6-31+G(2d′,p′)//HF/6-31G(d,p) level calculation was employed for estimation of the hydrogen bond energies of the Watson−Crick type base pairs in this report. Recently, Dunning's triple-ζ basis sets were applied to nucleic acid base pairs,[9n] and triple-, quadruple-, and quintuple-ζ basis sets were applied to a model complex of the base pair, for discussion of the basis set effect on the hydrogen bond energy.[9n] From the results of the model compounds, Sponer et al. pointed out that double-ζ basis sets should underestimate the hydrogen bond energies in the base pairs, in comparison with quintuple-ζ basis sets. However, we consider that the error, which originates from the basis set employed, should be comparable for all C^X−G base pairs. On the other hand, the hydrogen bond energies of A−T and G−C base pairs, calculated in the Slater-type orbital triple-ζ basis set (TZ2P) using DFT (BP86, PW91 and BLYP),[9o] were in good agreement with our results in MP2/6-31+G(2d′,p′)//HF/6-31G(d,p) level calculation. Thus, the substituent effects in nucleic acid bases on the hydrogen bond energy for base pair formation can be discussed, at least qualitatively, on the basis of the energy estimates derived from MP2/6-31+G(2d′,p′)//HF/6-31G(d,p) calculations.

The hydrogen bond energies of the Watson−Crick type base pairs were evaluated by a supermolecular method. The basis set super position error (BSSE) for hydrogen bond energies was corrected by using the counterpoise method.[18] Hereafter, we refer to the molecular interaction energy without BSSE correction as δE and the energy with BSSE correction as ΔE^{HB} (eqs 1 and 2). Thus, the more negative ΔE^{HB} means the more stable hydrogen bond. $\Delta\Delta E$ was defined as the substitution effect on ΔE^{HB} (eq 3):

$$\delta E(C^X-G) = E(C^X\text{-}G) - (E(C^X) + E(G)) \quad (1)$$

$$\Delta E^{HB}(C^X-G) = \delta E(C^X-G) + BSSE \quad (2)$$

$$\Delta\Delta E = \Delta E^{HB}(C^X-G) - \Delta E^{HB}(C-G) \quad (3)$$

As shown in eqs 1 and 2, $\Delta E^{HB}(C^X-G)$ includes the total interaction energy, and the deformation energy was not separated from $\Delta E^{HB}(C^X-G)$, because of our standpoint in this research: the substitution effect on the interaction energy, including the deformation energy, is an important aspect of this work.

In the present work, we studied twenty C analogues (C^X),[19] whose structures and abbreviations are shown in Figure 3. The structures of C^X-G, as well as those of nucleic acid bases C^X and G, were optimized in the 6-31G(d,p) basis set at the HF level of theory. C_s symmetry was assumed: all atoms, except for hydrogen atoms in the methyl group(s) and the phenyl group of C^{4NPh} and C^{5Ph},[20] were placed on the plane of symmetry. A nonplanarity of the bases in higher level calculations, especially for G, was reported by Hobza et al; however, the structures of the bases in Watson−Crick type base pairs were planar.[9k,l] Thus, the errors derived from the assumption of the planarity of the bases should be comparable for all of the C^X-G base pairs. Therefore, the relative substituent effects in nucleic acid bases on the hydrogen bond energy for base pair formation can be discussed, on the basis of the energy estimates based on the planar structures. The energies of the optimized structures were evaluated with single-point calculations using the 6-31+G-(2d′,p′) basis set at the MP2 level of theory.

A preliminary conformer search with HF/3-21G calculations was carried out in some cases. Additionally, energy estimation of the two important conformers was carried out at the level of MP2/6-31+G(2d′,p′)//HF/6-31G(d,p) in the case of derivatives possessing a modified *exo*-cyclic amino group: 4-*N*-methyl-1-methylcytosine (C^{4NMe}), for 4-*N*-formyl-1-methylcytosine (C^{4Nfo}), and 4-*N*-phenyl-1-methylcytosine (C^{4NPh}). In such derivatives, there are conformational isomers because of the rotation of the amino group and the substituent (Figure 4). $\Delta E^{HB}(C^X-G)$ of these derivatives were calculated based on the hydrogen bond forming conformer (**I**). For C^{4NMe}, the conformer (**I**) was found to be higher in energy than the conformer (**II**). We refer to the molecular interaction energies calculated based on the conformer (**II**) as $\Delta E^{total}(C^{4NMe}-G)$. On the other hand, for C^{4Nfo} and C^{4NPh}, the conformer (**I**) was found to be lower in energy than the conformer (**II**), so only $\Delta E^{HB}(C^{4Nfo}-G)$ and $\Delta E^{HB}(C^{4NPh}-G)$ were considered for these derivatives. However, in the case of C^{4NPh}, the difference in the energy was small; thus, base pair formation between such cytosine analogues and guanine should be entropically unfavorable. There are also two conformers derived from the rotation of the formyl group in the cases of $C^{5fo}$and C^{6fo}. ΔE^{HB} of these derivatives were calclated based on the most stable conformers except where noted otherwise. In the cases of C^{5NO_2}, C^{5NH_2}, and C^{quin}, planar structures of the *exo*-cyclic amino moieties were not energetically minimum, because of steric hindrance. Thus, we calculated ΔE^{HB} based on two structures: (a) based on the planar structures (referred to as $C^{5NO_2\ddagger}$, $C^{5NH_2\ddagger}$, and $C^{quin\ddagger}$) to estimate typical substitution effect and (b) based on the most stable structures to estimate real substitution effect.

In the case of 5/6-formyl substituted derivatives (C^{5fo} and C^{6fo}), there are conformers/conformations, as shown in Figure 5. A remarkable intramolecular hydrogen bond is observed only in C^{5fo}(a). Conformation b was the transition state (TS) in the formyl group rotation. To estimate the effect of the intramolecular hydrogen bond, the hydrogen bond energies based on the rotamers were also calculated.

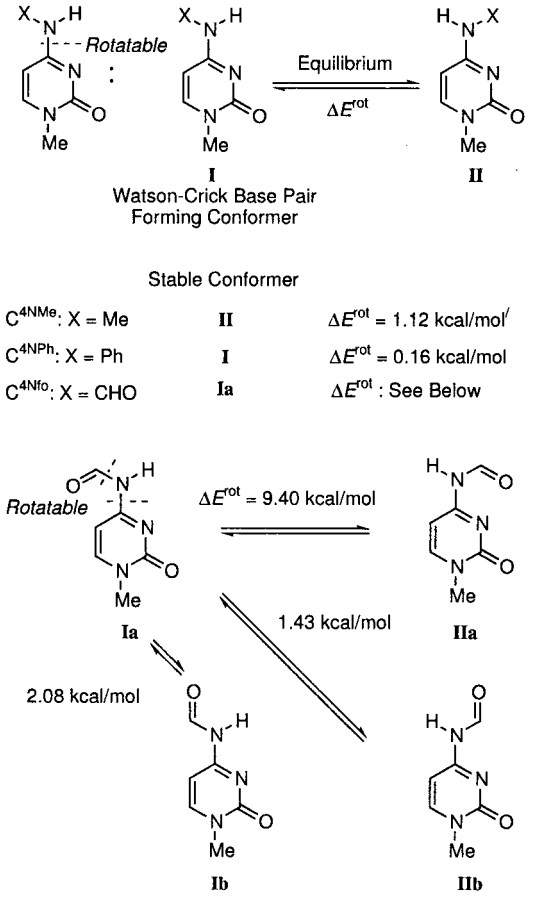

Figure 4. Rotatable *exo*-cyclic bonds in C^{4NMe}, C^{4NPh}, and C^{4Nfo}.

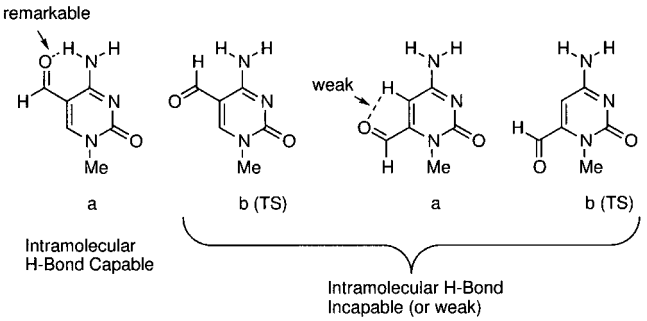

Figure 5. Rotation of formyl group and intramolecular hydrogen bond in C^{5fo} and C^{6fo}.

Conformer search calculations were carried out using the SPARTAN program.[21] Structure optimization and energy estimation calculations were both carried out using the Gaussian 94 program.[22]

Results and Discussion

C^X shown in Figure 3 was classified into the following six groups: group A: unmodified 1-methylcytosine (C); group B: an EWG or electron-donating group (EDG) was introduced at the 5 position on C (C^{5NO_2}, C,[5fo] C^{5F}, and C^{5NH_2}); group C: an EWG or an EDG was introduced at the 6 position on C (C^{6NO_2}, C,[6fo] C^{6F}, and C^{6NH_2});[23] group D: an EWG or EDG was introduced on the *exo*-cyclic amino moiety at the 4 position (^4N) of C (C^{4Nfo} and C^{4NMe}); group E: bases having larger conjugate systems (C^{5Ph}, C^{4NPh}, C^{triC}, C^{quin}, C^{biC}. and C^{Phox});

TABLE 1: Counterpoise Corrected Hydrogen Bond Energies (ΔE^{HB}), the Substitution Effects ($\Delta\Delta E$), and the Basis Set Super Position Errors (BSSE) (kcal/mol) of C^X-G Base Pairs Calculated at MP2/6-31+G(2d′,p′)//HF/6-31G(d,p) Level

C^X	ΔE^{HB}	BSSE	$\Delta\Delta E$
C	−26.08	2.83	
C^{5NO_2}	−23.49	2.91	+2.59
$C^{5NO_2‡}$	−23.41	2.91	+2.67
C^{5For}(a)	−23.64	2.88	+2.44
C^{5For}(b)‡	−24.85	2.89	+1.23
C^{5F}	−25.82	2.90	+0.26
$C^{5NH_2‡}$	−27.51	2.94	−1.43
C^{5NH_2}	−25.84	2.92	+0.24
C^{6NO_2}	−24.84	2.91	+1.24
C^{6For}(a)	−25.36	2.87	+0.72
C^{6For}(b)‡	−25.73	2.88	+0.35
C^{6F}	−25.79	2.43	+0.29
C^{6NH_2}	−26.21	2.89	−0.13
C^{4Nfo}	−26.05	3.12	+0.03
C^{4NMe}	−27.09	3.04	−1.02
ΔE^{total} =	−25.97		
C^{4NPh}	−27.23	3.37	−1.15
C^{5Ph}	−25.75	3.07	+0.33
C^{triC}	−27.23	3.18	−1.15
$C^{quin‡}$	−26.31	2.99	−0.23
C^{quin}	−26.30	2.99	−0.22
C^{biC}	−21.51	3.06	+4.57
C^{Phox}	−28.35	3.42	−2.27
C^{2S}	−23.58	2.57	+2.50
P^{2O}	−17.77	2.26	+8.31
P^{2S}	−18.10	2.34	+7.98

and group F: bases resulting from the replacement of the *exo*-cyclic oxygen atom by a sulfur atom (C^{2S}), and/or deletion of the *exo*-cyclic amino moiety (P^{2S} and P^{2O}).

Table 1 shows the results of theoretically estimated ΔE^{HB}, $\Delta\Delta E$, and BSSE (kcal/mol) of each C^X. Figure 6 shows the substitution effect on ΔE^{HB} in groups B and C of C^X-G and the corresponding those of U^X-A. Opposite to the substitution effect in uracil, a remarkable tendency was observed, namely, that C^X possessing stronger EWG forms a less stable base pair with G. The cytosine derivatives act as an electron-acceptor in H−bond A and act as an electron-donor in H−bonds B and C, as shown in Figure 2. So, it is considered that introduction of an EWG on C^X makes H−bond A stronger and the H−Bonds B and C weaker. Conversely, an EDG on C^X makes H−bond A weaker and H−bonds B and C stronger. Thus, the sum of the substitution effects on the H−bonds B and C overcome the substitution effect on the H−bond A. The substitution effect of each substituent on the 5 position in ΔE^{HB} of C^X was larger than that of U^X: ΔE^{HB} of the $C^{5NO_2}-G$ base pair was 2.59 kcal/mol less negative than that of the C−G base pair; on the other hand, ΔE^{HB} of the $U^{5NO_2}-A$ base pair was 0.92 kcal/mol more negative than that of the U−A base pair.[7a] Also, ΔE^{HB} of the $C^{5NH_2‡}-G$ base pair was 1.43 kcal/mol more negative than that of the C−G base pair; on the other hand, ΔE^{HB} of the $U^{5NH_2}-A$ base pair was 0.28 kcal/mol less negative than that of the U−A base pair.[7a] The substitution effect on the 5 position was larger than that on the 6 position. For example, ΔE^{HB} of the $C^{5NO_2}-G$ base pair was 1.35 kcal/mol less negative than $C^{6NO_2}-G$ base pair, and ΔE^{HB} of the $C^{5NH_2‡}-G$ base pair was 1.30 kcal/mol more negative than $C^{6NH_2}-G$ base pair. Nevertheless, ΔE^{HB} of $C^{5NH_2}-G$ (based on the most stable structure, not C_s symmetry) base pair was 0.25 kcal/mol less negative than that of the C−G base pair. The dihedral angle between the *exo*-cyclic amino moiety on the 5 position and the purine ring was about 120°, because of steric hindrance. So the electron-donating property of the amino moiety was lost in this case. ΔE^{HB} of the $C^{quin}-G$ and $C^{quin‡}-G$ base pairs were almost the same, because the most stable structure of C^{quin} was nearly planar.

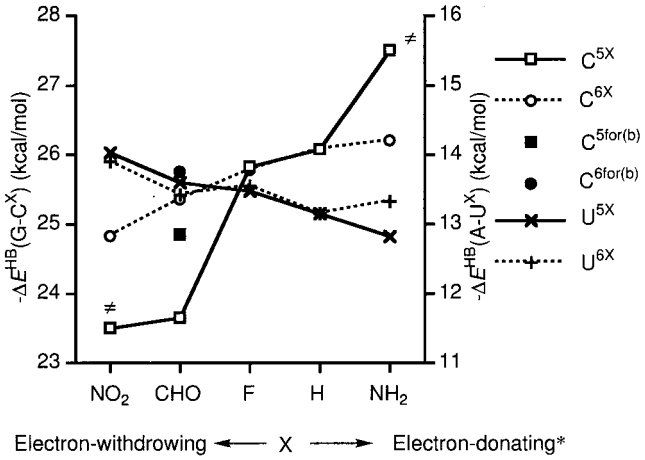

Figure 6. Substitution effect in the base pair hydrogen bond energy of 5 or 6 position on cytisine (C^{5X}, −□− and C^{6X}, - -○- -) and uracil (U^{5X}, −X− and U^{6X}, - -+- -) derivatives. For the asterisk, see ref 24. For ≠, ΔE^{HB} of the $C^{5NH_2}-G$ base pair in this figure shows the result based on the C_s structure (= $\Delta E^{HB}(C^{5NH_2‡}-G)$ in Table 1). See also Figure 7 and footnote 20.

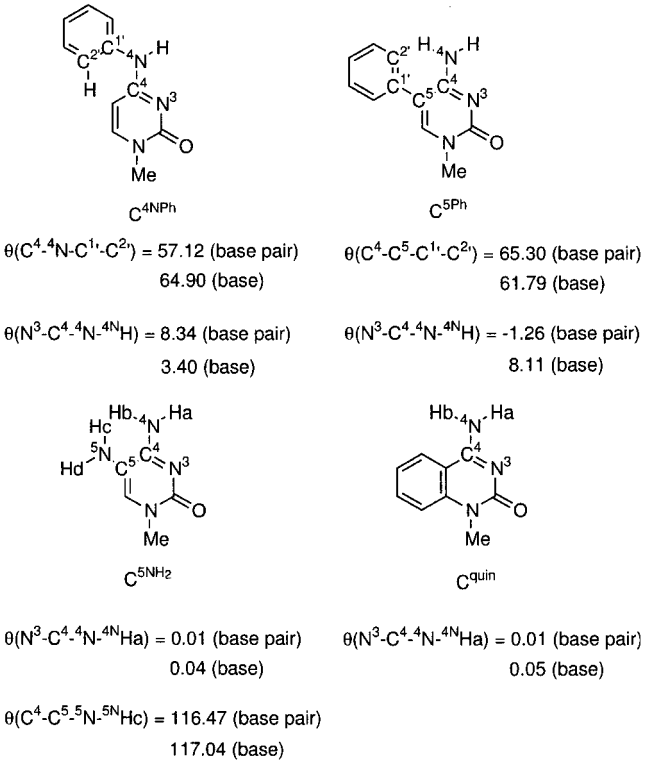

Figure 7. Important dihedral angles (θ, degree) between the purine ring and the substituent in C^{4NPh}, C^{5Ph}, C^{5NH_2}, and C^{quin}.

An intramolecular hydrogen bond, as shown in Figure 5, should be considered as one reason, why the substitution effect on the 5 position and the 6 position was different. As shown in Figure 6, the substitution effect of the formyl group in C^{5fo}(b), which is unable to form such hydrogen bond, was much smaller and ΔE^{HB} of C^{5fo}(b)−G was closer to that of C^{6fo}(a)−G and C^{6fo}(b)−G. In both cases of C^{5fo} and C^{6fo}, the substitution effect became smaller by removing the intramolecular hydrogen bond. The ΔE^{HB} difference of C^{6fo}(a)−G and C^{6fo}(b)−G was smaller than that of C^{5fo}(a)−G and C^{5fo}(b)−G. Table 2 shows the sum of the charge distributions on the substituent of C^{5NO_2}, C^{6NO_2},

TABLE 2: Sum of the Charge Distributions (*e*) of the Substituents

C^X	Mulliken	CHelpG	NPA
C^{5NO_2}	−0.418	−0.077	−0.258
C^{6NO_2}	−0.279	−0.054	−0.173
C^{5For}(a)	−0.281	0.008	−0.015
C^{5For}(b)‡	−0.189	0.038	−0.001
C^{6For}(a)	0.007	0.057	0.073
C^{6For}(b)‡	−0.007	0.040	0.060

C^{5fo}(a), C^{5fo}(b), C^{6fo}(a), and C^{6fo}(b). In all calculation methods, the electron negativity of the nitro group in C^{5NO_2} was higher than that of C^{6NO_2} and the electron negativity of the formyl group in C^{5fo}(a) was higher than that of C^{5fo}(b), C^{6fo}(a), and C^{6fo}(b). So, it is considered that the formation of the intramolecular hydrogen bond enhances the electron-withdrawing ability of the substituent by accepting the hydrogen bond, and the substitution effect on ΔE^{HB} was enlarged for the 5 position substituted derivatives. However, ΔE^{HB} of C^{5fo}(b)−G and C^{6fo}(b)−G was more negative than that of C−G; thus, the conclusion that C^X possessing stronger EWG forms a less stable base pair with G is not required to be changed even after removing the intramolecular hydrogen bond.

Alkylation of the *exo*-cyclic amino group on the 4 position (^{4}N) enforced the hydrogen bond ($\Delta\Delta E = -1.02$ kcal/mol) of the base pairs. The acylation on the same position had almost no effect ($\Delta\Delta E = +0.03$ kcal/mol). The reason the substitution effect of the formyl group on the 4 position was much smaller than that of the 5 position is considered to be that the formyl group on the 4 position reduces the electron density on the *exo*-cyclic amino moiety more effectively; on the other hand, the formyl group on the 5 position reduces the electron density on ^{2}O and/or the pyrimidine ring mainly. C^{4Nfo} formed shorter H−bonds A and C than C^{5for} but formed longer H−bonds B than C^{5for}. Then C^{4Nfo} formed a more stable base pair with G than C^{5for}. On the other hand, C^{4NMe} formed slightly longer H−bond A and shorter H−bond C as compared with cytosine, but it formed slightly longer H−bond B. From the result of the hydrogen bond length of C^{4NMe}−G base pair, the methyl group of C^{4NMe} has almost no effect on the electron density of ^{4}N but enriches the electron density of ^{2}O effectively. Thus, it was difficult to describe the substitution effect on ^{4}N.

Although the substitution effect in group E was relatively large, there was no remarkable trend in the substitution effect on ΔE^{HB} in group E, in contrast to groups B and C. Introduction of the phenyl group on ^{4}N also enforced the hydrogen bond of the base pairs (1.15 kcal/mol); on the other hand, introduction of the phenyl group on the 5 position slightly weakened the hydrogen bond of the base pairs (0.33 kcal/mol). C^{triC}−G formed a more stable base pair than C−G (1.50 kcal/mol). ΔE^{HB} of the C^{quin}−G base pair was 0.23 kcal/mol negative than that of the C−G base pair. C^{biC} is known as a fluorescence C analogue, but ΔE^{HB} of the C^{biC}−G base pair was much less negative than that of the C−G base pair (4.57 kcal/mol). ΔE^{HB} of C^{Phox}−G was the most negative (= the most stable base pair) in this study ($\Delta\Delta E = -2.27$ kcal/mol). Lin et al. reported a DNA oligomer, which contains C^{Phox} instead of C and showed a higher melting temperature (*T*m) than a "normal" DNA oligomer; that is, the duplex stability of the DNA oligomer was increased by the base substitution from C to C^{Phox}.[19p,25] The duplex stability is dependent not only on the base pair hydrogen bond stability, but also other various factors, e.g., stacking of the bases, conformation of the sugars, etc. However, the base pair hydrogen bond stability should be one of the factors, and this result suggests that the C^{Phox} can form stable hydrogen bonds with G.

As expected, the substitution of *exo*-cyclic oxygen to sulfur (C^{2S}) weakened the hydrogen bond energy of the base pair ($\Delta\Delta E = +2.50$ kcal/mol). This substitution effect was also observed in the case of U^{4S} ($\Delta\Delta E = +1.52$ kcal/mol). Removing the *exo*-cyclic amino moiety or replacing the *exo*-cyclic oxygen by sulfur also causes a large decrease in the base pair stability (8.31 kcal/mol). However, ΔE^{HB} of the P^{2O}−G base pair (−17.77 kcal/mol), which contains two hydrogen bonds, was more negative than ΔE^{HB} of the A^{2NH_2}−U base pair (9-methyl 2-aminoadenine − 1-methyl uracil base pair, −14.96 kcal/mol).[8] Surprisingly, ΔE^{HB} of the P^{2S}−G base pair (−18.10 kcal/mol) was almost the same as that of the P^{2O}−G base pair. Although an *exo*-cyclic sulfur atom in heterocycles can act as a hydrogen acceptor (=electron donor) in a hydrogen bond,[26] hydrogen bond capability of the *exo*-cyclic sulfur atom is considered much weaker than that of an oxygen atom. It is considered that H−bond B of the P^{2S}−G base pair is enforced by the substitution of *exo*-cyclic oxygen by sulfur, which is less electron-negative; however, the H−bond B length of P^{2S}−G (2.095 Å) is longer than that of P^{2O}−G (1.961 Å). We are now investigating this unexpected result in more detail.

Considering the fact that hydrogen bond energy is mainly characterized by electrostatic contribution,[16] the charge distribution analysis is considered a good method to study the substitution effect on the strength of each hydrogen bond in the base pairs. Nevertheless, from the results of the charge distribution analysis and the relationship between hydrogen bond length and the substituent, expected trends of the relationship between the hydrogen bond energy, the hydrogen bond distance, and the charge distribution were not observed.[27] Neither the hydrogen bond distance nor the charge distribution was a valid index for the hydrogen bond status in C^X−G base pairing. Platts also reported that charge distribution was not a valid indicator for the hydrogen bond stability.[28] Guerra et al. reported that charge-transfer interaction should have an important role in base pair hydrogen bonding.[29] Our results in the charge distribution analysis suggested that not only electrostatic but also some other contribution should be considered for characterizing hydrogen bonding property in the base pairs, which appears to be consistent with the results of Platts and Guerra et al.

Conclusion

The substitution effect on hydrogen bond energy of the Watson−Crick type base pair between G and C^X was estimated by an ab initio molecular orbital theory. Opposite to the substitution effect in uracil, there was a remarkable tendency for C^X possessing stronger EWG to form a less stable base pair with G. The substitution effect on the 5 position in C^X was greater than that on the 6 position. Intramolecular hydrogen bond formation between the substituent and the *exo*-cyclic amino group plays an important role in the difference of the substitution effect in the 5 position and the 6 position. Neither the hydrogen bond distance nor the charge distribution was a valid index for the hydrogen bond status in C^X−G base pairing; thus, considering electrostatic contribution alone is not enough for characterizing a hydrogen bond property in the base pairs.

Acknowledgment. The services and computational time made available by National Institute of Materials and Chemical Research and Tsukuba Advanced Computing Center (TACC) have been essential to this study and are gratefully acknowledged. We also thank the Science and Technology Agency for financial support.

Supporting Information Available: Figures (S1−S6) and Tables (T1 and T2) showing the hydrogen bond energy, the

hydrogen bond distance, and the charge distribution. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Watson, J. D.; Crick, H. C. *Nature* **1953**, *171*, 737.

(2) Brenner, S.; Jacob, F.; Meselson, M. *Nature* **1961**, *190*, 576.

(3) Hopfield, J. J. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 4135.

(4) (a) Sessler, J. L.; Wang, R. *Angew. Chem., Int. Ed. Engl*. **1998**, *37*, 1726. (b) González, J. J.; Prados, P.; de Mendoza, J. *Angew. Chem., Int. Ed. Engl*. **1999**, *38*, 525.

(5) (a) Tjivikua, T.; Ballester, P.; Rebek, J., Jr. *J. Am. Chem. Soc*. **1990**, *112*, 1249. (b) Nowick, J. S.; Feng, Q.; Tjivikua, T.; Ballester, P.; Rebek, J., Jr. *J. Am. Chem. Soc*. **1991**, *113*, 8831. (c) Menger, F. M.; Eliseev, A. V.; Khanjin, N. A.; Sherrod, M. J. *J. Org. Chem*. **1995**, *60*, 2870.

(6) (a) *Antisense RNA and DNA*; Murray, J. A. H., Ed.; Wiley-Liss: New York, 1992. (b) Uhlmann, E.; Peyman, A. *Chem. Rev*. **1990**, *90*, 543. (c) Englisch, U.; Gauss, D. H. *Angew. Chem., Int. Ed. Engl*. **1991**, *30*, 613. (c) Stein, C. A.; Cheng, Y.-C. *Science* **1993**, *261*, 1004−1012. (d) Freier, S. M.; Altmann, K. H. *Nucleic Acid Res*. **1997**, *25*, 4429−4443. (e) Agrawal, S.; Zhao, Q. Y. *Curr. Opin. Chem. Biol*. **1998**, *2*, 519−528.

(7) (a) Kawahara, S.-I.; Wada, T.; Kawauchi, S.; Uchimaru, T.; Sekine, M. *J. Phys Chem. A* **1999**, *103*, 8516. (b) Kawahara, S.-I.; Uchimaru, T.; Sekine, M. *J. Mol. Struct. (THEOCHEM)* **2000**, *530*, 109.

(8) Kawahara, S.-I.; Uchimaru, T.; Taira, K.; Sekine, M. *J. Phys Chem. A* **2001**, *105*, 3894.

(9) (a) Scheiner, S.; Kern, C. W. *Chem. Phys. Lett*. **1978**, *57*, 331. (b) Scheiner, S.; Kern, C. W. *J. Am. Chem. Soc*. **1979**, *101*, 4081. (c) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc*. **1987**, *109*, 1302. (d) Trollope, K. I.; Gould, I. R.; Hiller, I. H. *Chem. Phys. Lett*. **1993**, *209*, 113. (e) Destexhe, A.; Smets, J.; Adamowicz, L.; Maes, G. *J. Phys. Chem*. **1994**, *98*, 1506. (f) Gould, I. R.; Kollman, P. A. *J. Am. Chem. Soc*. **1994**, *116*, 2493. (g) Hobza, P.; Sponer, J.; Polásek, M. *J. Am. Chem. Soc.* **1995**, *117*, 792. (h) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Biomol. Strct. Dyn.* **1996**, *14*, 117. (i) Sponer, J.; Hobza, P. *Chem. Phys.* **1996**, *204*, 365. (j) Sponer, J.; Hobza, P. *Chem. Phys. Lett.* **1996**, *261*, 379. (k) Sponer, J.; Hobza, P. *Chem. Rev*. **1999**, *99*, 3247. (l) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn*. **1996**, *14*, 117. (m) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem*. **1996**, *100*, 1965. (n) Sponer, J.; Hobza, P. *J. Phys. Chem*. *A* **2000**, *104*, 4592. (o) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *J. Am. Chem. Soc*. **2000**, *122*, 4117.

(10) Rablen, P. R.; Lockman, J. W.; Jorgensen, W. L. *J. Phys. Chem. A* **1998**, *102*, 3782.

(11) X/A//Y/B shows energy evaluation at the level of theory X with basis set A at the geometry optimized at the level of theory Y with basis set B.

(12) (a) Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Al-Laham, M. A.; Sheirley, W. A.; Mantzaris, J. *J. Chem. Phys*. **1988**, *89*, 2193. (b) Petersson, G. A.; Al-Laham, M. A. *J. Chem. Phys*. **1991**, *94*, 6081.

(13) Kroon-Batenburg, L. M. J.; van Duijneveldt, F. B. *J. Mol. Struct*. **1985**, *121*, 185.

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

(15) Pople, J. A., Head-Gordon, M.; Raghavachari, K. *J. Phys. Chem.* **1987**, *87*, 5968.

(16) (a) Morokuma, K. *Acc. Chem. Res.* **1977**, *10*, 294. (b) Rendell, A. P. L.; Bacskay, G. B.; Hush, N. S. *Chem. Phys. Lett.* **1985**, *117*, 400. (c) Chen, W.; Gordon, M. S. *J. Phys. Chem.* **1996**, *100*, 14316.

(17) Kawahara, S.-I.; Uchimaru, T. *Phys. Chem. Chem. Phys*. **2000**, *2*, 2869.

(18) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(19) For C^{5NH_2} and C^{5NO_2}: (a) Fox, J. J.; van Praag, D. *J. Org. Chem.* **1961**, *26*, 526. For C^{5for}: (b) Moriya, J.; Yokogawa, T.; Wakita, K.; Ueda, T.; Nishikawa, K.; Crain, P. F.; Hashizume, T.; Pomerants, S. C.; McCloskey, J. A.; Kawai, G.; Hayashi, N.; Yokoyama, S.; Watanabe, K. *Biochemistry* **1994**, *33*, 2234. For C^{5F}: (c) Marasco, C. J., Jr.; Sufrin, J. R. *J. Org. Chem*. **1992**, *57*, 6363. (d) Osterman, D. G.; DePillis, G. D.; Wu, J. C.; Matsuda, A.; Santi, D. V. *Biochemistry* **1988**, *27*, 5204. (e) Chen, L.; MacMillan, A. M.; Chang, W.; Ezaz-Nikpay, K.; Lane, W. S.; Verdine, G. L. *Biochemistry* **1991**, *30*, 11018. (f) Smith, S. S.; Kaplan, B. E.; Sowers, L. C.; Newman, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4744. For C^{4Nacyl}: (g) Edmonds, C. C.; Crain, P. F.; Hashizume, T.; Gupta, R.; Stetter, K. O.; McCloskey, J. A. *J. Chem. Soc., Chem. Commun.* **1987**, 909. (h) Limbach, P. A.; Crain, P. F.; McCloskey, J. A. *Nucleic Acid Res*. **1994**, *22*, 2183. (i) Wada, T.; Kobori, A.; Kawahara, S.-I.; Sekine, M. *Tetrahedron Lett.* **1998**, *39*, 6907. For $C^{4Nalkyl}$: (j) Fox, J. J.; Van Praag, D.; Wempon, I.; Doerr, I. L.; Cheong, L.; Knol, J. E.; Eidinoff, M. L.; Bendich, A.; Brown, G. B. *J. Am. Chem. Soc*. **1959**, *81*, 178. (k) Dubin, D. T.; Taylor, R. H.; Davenport, L. W. *Nucleic Acid Res*. **1978**, *5*, 4385. (l) Schott, H.; Häussler, M. P.; Schwendener, R. *Liebigs Ann. Chem.* **1994**, 465. For C^{4NAr}: (m) Huang, C.-Y.; Bi, G.; Miller, P. S. *Nucleic Acid Res*. **1996**, *24*, 2606. For C^{biC}: (n) Bergstrom, D. E.; Inoue, H.; Reddy, P. A. *J. Org. Chem.* **1982**, *47*, 2174. (o) Staubli, A. B.; Dervan, P. *Nucleic Acid Res*. **1994**, *22*, 2637. For C^{Phox}: (p) Lin, K.-Y.; Jones, R. J.; Matteucci, M. *J. Am. Chem. Soc.* **1995**, *117*, 3873. For C^{2S}: (q) Carbon, J.; David, H.; Studier, M. H. *Science* **1968**, *161*, 1146. (r) Ueda, T.; Nishino, H. *Chem. Pharm. Bull.* **1969**, *17*, 920.

(20) The important diheadral angles (θ, degree) between the rings in C^{4NPh}, C^{5Ph}, C^{5NH_2}, and C^{quin} are shown in Figure 7.

(21) Deppmeir, D. J.; Driessen, A. J.; Hehere, T. S.; Hehere, W. J.; Johnson, J. A.; Klunzinger, P. E.; Lou, L.; Yu, J.; Baker, J.; Carpenter, J. E.; Dixon, R. W.; Fielder, S. S.; Johnson, H. C.; Kahn, S. D.; Leonard, J. M.; Piero, W. J. *SPARTAN*; Wavefunction, Inc.: Irvine, CA, 1998.

(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.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; 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.; Gonzalez, C.; Pople, J. A. *Gaussian 94*; Gaussian, Inc.: Pittsburgh, PA, 1995.

(23) Looking at the DNA/RNA-duplex standard geometry structure, close contacts between the substituent at the 8 position and the backbone are obvious. Because of steric hindrance, the nucleic acid base orientation of most of the 6-substituted C^X is syn in nucleotides.[27] Thus, the modified base possessing a substituent at the 6 position may not form a favorable hydrogen bonding base pair in the standard DNA/RNA−duplex structure and these derivatives would not be suitable for antisense application. We calculated ΔE^{HB} of these derivatives to compare the substitution effect of 5 substituted ones. However, if the structure of the backbone is substantially altered, e.g., in PNA,[30] the substituent at the 6 position will not be unfavorable for base pair formation via hydrogen bonds.

(24) Hammet parameter: $X = NO_2$; $\sigma_m = 0.71$, $\sigma_p = 0.81$, $X = CHO$; $\sigma_m = 0.41$, $\sigma_p = 0.47$, $X = F$; $\sigma_m = 0.34$, $\sigma_p = 0.06$, $X = H$; $\sigma_m = \sigma_p = 0.00$, $X = NH_2$; $\sigma_m = -0.09$, $\sigma_p = -0.57$. (a) Hammet, L. P. *Physical Organic Chemistry*, 2nd ed.; McGraw-Hill: New York, 1970. (b) Exner, O. In *Correlation Analysis in Chemisry*; Chapman, N. B., Shorter, J., Eds.; Plenum Press: New York, 1978.

(25) Lin et al. concluded that the reason for the rise in *T*m of their DNA oligomer is enhancement of the stacking effect (aromatic $\pi-\pi$ overlap between heterocycles of DNA helices).[19p]

(26) Saenger, W. *Principals of Nucleic Acid Structure*; Springer-Verlag: New York, 1984; Chapter 4.

(27) See the Supplementary Material.

(28) (a) Platts, J. A. *Phys. Chem. Chem. Phys*. **2000**, *2*, 973. (b) Platts, J. A. *Phys. Chem. Chem. Phys*. **2000**, *2*, 3115.

(29) (a) Guerra, C. F.; Bickelhaupt, F. M. *Angew. Chem., Int. Ed. Engl*. **1999**, *38*, 2942. (b) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *Chem. Eur. J*. **1999**, *12*, 3581.

(30) (a) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Science* **1991**, *254*, 1497. (b) Hyrup, B.; Nielsen, P. E. *Bioorg. & Med. Chem*. **1996**, *4*, 5. (c) Uhlmann, E.; Peyman, A.; Breipohl, G.; Will, D. W. *Angew. Chem., Int. Ed. Engl*. **1998**, *37*, 2796.