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

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

National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, 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: October 25, 2001; In Final Form: December 21, 2001

The substitution effect on hydrogen-bond energy of the Watson−Crick type base pair formation between 1-methylcytosine and chemically modified 9-methylguanine derivatives was evaluated by an ab initio molecular orbital theory. Introduction of an electron-withdrawing group on the 8-position or on the exo-cyclic amino moiety enforced the hydrogen bond. Neither the charge distribution nor the separation between the hydrogen bonding sites is found to be directly correlated with the strength of the hydrogen bonds.

Introduction

The Watson−Crick type base pair formation is essential for molecular recognition in the duplex formation of nucleic acid,[1] i.e., 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.

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, especially, to antisense technology,[6] which is a topic 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 that is able to selectively form a stable complex is needed. However, there are no systematic studies planning the improvement of base pair stability, because it is difficult and time-consuming to prepare a wide variety of nucleic acid base analogues and to experimentally measure their base pair formation ability. Thus, to improve the base pair stability, computer-aided molecular design of nucleic acid base analogues is highly demanded.

We have already reported an ab initio molecular orbital study on the substitution effect on hydrogen-bond energies for base pair formation between nucleic acid base analogues, i.e., base pairs between 9-methyladenine (A) and modified 1-methyluracil derivatives (U^X),[7] base pairs between modified 9-methyladenine derivatives (A^X) and 1-methyluracil (U),[8] and base pairs between modified 1-methylcytosine derivatives (C^X) and 9-methylguanine (G).[9] In the case of the substituent effect on uracil in the A−U^X base pair, a remarkable tendency was observed for U^X: U^X possessing a stronger electron-withdrawing group (EWG) forms a more stable base pair. Contrary to the substituent effect on uracil, C^X possessing an electron-donating group (EDG) forms a more stable base pair with G. On the other hand, no remarkable trend was observed in the relation between the substituent in adenine derivatives and the hydrogen-bond energies in the A^X−U base pair formation.

* Corresponding Author: Tadafumi Uchimaru, Tel.: +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.

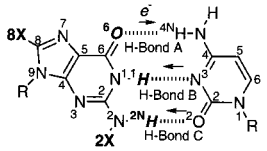

Figure 1. Watson−Crick base pairs between G^X and C.

Although there are many theoretical studies on the hydrogen-bond energy of the Watson−Crick type base pair between natural nucleic acid bases,[10] no systematic ab initio molecular orbital studies on modified base pairs have been reported except for our studies.[7−9] We report herein an ab initio study regarding the substitution effect on hydrogen-bond energy in the base pair between modified 9-methylguanine derivatives (G^X) and C (Figure 1).

The atom numbering used here is given in Figure 1. The superscript on the right side on atom symbol represents the position of ring atom, e.g. N^1 on guanine indicates the ring nitrogen atom at the 1 position of guanine. The superscript on the left side on atom symbol represents the position of exo-cyclic atom, e.g. ^6O on guanine indicates the exo-cyclic oxygen atom at the 6 position of guanine. In addition, ^{2N}H represents the hydrogen atom in the exo-cyclic amino moiety at the 2 position.

Computational Methods

In most theoretical studies, the hydrogen-bond energies of the Watson−Crick type base pairs were evaluated at the second-order Møller−Plesset (MP2) level of theory using double-ζ basis sets with polarization.[10] Rablen et al. showed[11] 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)[12] were in good agreement with the results of the complete basis set (CBS-Q[13]). Sponer et al. reported[10m] that the hydrogen-bond energies of some model compounds in MP2/6-31G*(0.25)//HF/6-31G** [14] reproduced relatively well the result obtained using much larger basis sets. They also found[15] that the contribution of higher level electron correlation to hydrogen-bond energy was minimal and that MP2 interaction energies were close to the results of coupled cluster electron correlation (CCSD(T)[16]) data. The electrostatic contribution to hydrogen-bond energy is

10.1021/jp0139538 CCC: $22.00
Published on Web 03/03/2002

large,[17] so the effect of electron correlation should be relatively small. Thus, the conclusion of Sponer et al. would be quite reasonable and also be generally applicable to various types of hydrogen-bonding systems. We already reported an MP2 level ab initio study regarding the basis set effect on the calculated hydrogen-bond energies of Watson−Crick type base pairs.[18] The computational levels of MP2/6-31+G(2d′,p′)[13]//HF/6-31G(d,p) provide quite reasonable values for the hydrogen-bond energies of A−U and G−C base pairs, which were in excellent agreement not only with the values calculated at the MP2/6-311++G(3df,p)//HF/6-311++G(3d,p) level but also with the values reported by Rablen et al.[11] Thus, the 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 a base pair, and triple-, quadruple- and quintuple-ζ basis sets were applied to model complexes of the base pair, for discussion of the basis set effect on the hydrogen-bond energy.[10m] From the results of the model compounds, Sponer et al. pointed out that double-ζ basis sets were apparently underestimated in comparison with quintuple-ζ basis sets. However, we consider that the margin of error, which originates from the incompleteness of the basis set, should be comparable for all G^{X}−C base pairs. On the other hand, the hydrogen-bond energies of the A−T and G−C base pairs, calculated in the Slater-type orbital triple-ζ basis set (TZ2P) using DFT (BP86, PW91 and BLYP),[10n] were in good agreement with our results for the 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) level calculations.

The structures of G^{X}−C, as well as those of nucleic acid bases G^{X} and C, were optimized in the 6-31G(d,p) basis set at the HF level of theory. In all cases, C_s symmetry was assumed: all atoms, except for hydrogen atoms in the methyl group(s), were placed on the plane of symmetry. A preliminary conformer search with HF/3-21G(d) calculations was carried out in some cases.

The energies of the optimized structures were evaluated with single-point calculations on the 6-31+G(2d′,p′) basis set at the MP2 level of theory.

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.[19] Hereafter, we refer to the molecular interaction energy without and with BSSE correction as δE and ΔE^{HB}, respectively (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).

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

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

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

Additionally, energy estimation of the two important conformers in MP2/6-31+G(2d′,p′)//HF/6-31G(d,p) was carried out in the case of derivatives possessing a modified exo-cyclic amino group: 2-*N*-methyl-1-methyl guanine (G^{2NMe}) and 4-*N*-formyl-1-methyl guanine (G^{2Nfo}). In such derivatives, the rotation of the substituted amino group gives rise to two conformers (Figure 2), one of which (conformer **I**) can form a base pair with C via three hydrogen bonds, while the other conformer (conformer **II**) cannot form a base pair, because of the inhibition of hydrogen-bond formation by the substituent on the amino group. For both G^{2NMe} and G^{2Nfo}, conformer **I** was found to be lower in energy than conformer **II**. Thus, $\Delta E^{HB}(G^{X}-C)$ of these derivatives was calculated on the basis of the hydrogen-bond forming conformer **I**.[20]

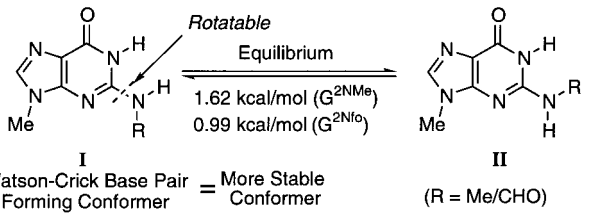

Figure 2. Rotatable exo-cyclic bonds in G^{2NMe} amd G^{2Nfor}.

A nonplanarity of the exo-cyclic amino moiety on the bases in higher level calculations, especially for G, was reported by Hobza et al.[10h,k] However, the energy difference between planar and nonplanar optimized structures is minimal, especially for higher level calculation (MP2/6-311G(2df,p)), except for G, and the structures of the bases in Watson−Crick type base pairs are planar.[10h,k] Thus, the energy derived from the nonplanarity of C is negligible. We considered two molecular interaction energies: ΔE^{HB}, as described above, the hydrogen-bond energy based on the planar G^{X} structure, and ΔE^{Total}, including the energy difference derived from the nonplanarity of the G^{X} (i.e., ΔE^{NP}). ΔE^{NP} values were calculated at the MP2/6-31+G(2d′,p′)//MP2/6-31G(d,p) level of calculations and are defined as

$$\Delta E^{NP} = E(G^{X}\text{Planar}) - E(G^{X}\text{Nonplanar}) \quad (4)$$

where $E(G^{X}\text{Planar})$ [$=E(G^{X})$ in eq 1−3] shows the energy of G^{X}, whose structure was optimized in C_s symmetry assumption, and $E(G^{X}\text{Nonplanar})$ shows the energy of G^{X}, which has a pyramidal exo-cyclic amino group. ΔE^{Total} is calculated as

$$\Delta E^{Total} = \Delta E^{HB}(G^{X}-C) + \Delta E^{NP} \quad (5)$$

Conformer search calculations of some derivatives 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

In the present work, we studied 14 guanine analogues (G^{X}), whose structures and abbreviations are shown in Figure 3.[23] G^{X} shown in Figure 3 was classified into groups A−D. Group A is unmodified guanine (G). For group B, a substitution group was introduced at the 8-position on G[24] or at the exo-cyclic amino moiety of G, an EWG was introduced on the 8-position of G (G^{8F}, G^{8oxo}, and G^{8NO_2}), an EDG was introduced on the 8-position of G (G^{8NH_2}), and a formyl group was introduced as an EWG (G^{2Nfo}), or a methyl group was introduced as an EDG (G^{2NMe}) on the exo-cyclic amino moiety on the 2-position. For group C, the number of the hydrogen bonds was changed (H and H^{6S}) or one of the hydrogen bonds was weakened by the replacement of the exo-cyclic oxygen atom by a sulfur atom (G^{6S}). For group D, a carbon atom in the purine ring was replaced with a nitrogen atom (G^{8N}) or a nitrogen atom in the purine ring was replaced with a carbon atom (G^{3C}, G^{7C}, and G^{9C}).

Table 1 shows the results of theoretically estimated ΔE^{HB}, $\Delta\Delta E$, BSSE, ΔE^{NP}, and ΔE^{Total} (kcal/mol) of each G^{X}. The substitution effects of each group are discussed as follows. To discuss a typical effect of the substituent on hydrogen-bond

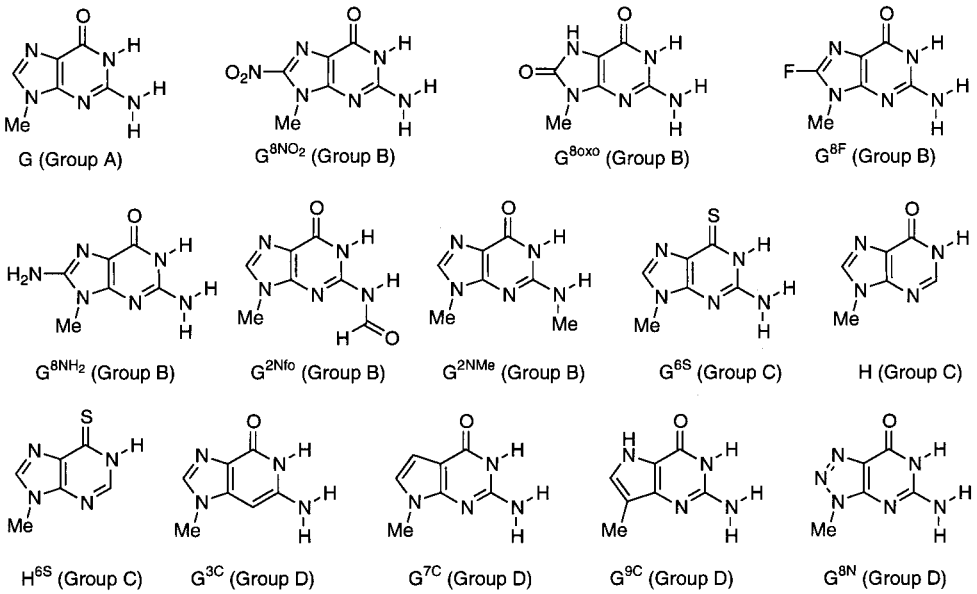

Figure 3. Substituents introduced to the 9-methyladenine derivatives (G^X) in this study.

TABLE 1: ΔE^{HB}, BSSE, $\Delta\Delta E$, ΔE^{NP}, and ΔE^{Total} (kcal/mol) of Each G^X Calculated at the MP2/6-31+G(2d′,p′)//HF/6-31G(d,p) Level

G^X	ΔE^{HB}	BSSE	$\Delta\Delta E$	ΔE^{NP}	ΔE^{Total}
G	−26.08	2.83		1.47	−24.61
G^{8NO_2}	−27.49	2.91	−1.41	2.88	−24.62
G^{8oxo}	−26.90	2.89	−0.82	1.37	−25.53
G^{8F}	−26.54	2.88	−0.46	1.38	−25.16
G^{8NH_2}	−25.79	2.86	0.29	5.78	−20.00
G^{2Nfo}	−26.68	3.18	−0.60	0.00	−26.68
G^{2NMe}	−26.06	3.01	0.02	0.00	−26.05
G^{6S}	−23.52	2.66	2.56	1.36	−22.16
H	−20.25	2.40	5.83		−20.25
H^{6S}	−19.24	2.45	6.84		−19.24
G^{3C}	−25.75	2.82	0.33	2.55	−23.21
G^{7C}	−25.19	2.83	0.89	1.77	−23.43
G^{9C}	−24.19	2.79	1.89	2.08	−22.11
G^{8N}	−26.73	2.84	−0.65	1.11	−25.62

energy, we mainly discuss ΔE^{HB}. For group B, in contrast to the substitution effect on adenine,[8] a remarkable trend in ΔE^{HB} was observed: the substitution effect in introducing an EWG on the 8-position and on the exo-cyclic amino moiety made the G^X−C base pair energetically more favorable. The guanine derivatives act as an electron donor in H-bond A; on the other hand, they act as an electron-acceptor in H-bonds B and C, as shown in Figure 1. So, it is considered that introduction of an EDG on G^X makes H-bond A stronger and H-bonds B and C weaker. Conversely, an EWG on G^X should weaken H-bond A and strengthen H-bonds B and C. Thus, the sum of the substitution effect on the H-bonds B and C overcomes the substitution effect on the H-bond A. Consequently, ΔE^{HB} of the G^{8NO_2}−C base pair (−27.49 kcal/mol) was the most negative in the present study, and ΔE^{HB} of the G^{8NH_2}−C base pair (−25.79 kcal/mol) was the least negative in group B. This trend was opposite to the substitution effect in cytosine[9] (see Figure 4), as expected. The fluctuation in ΔE^{HB} resulting from introduction of the substituent in G^{8X} was smaller than that in C^{5X}. As already reported, an intramolecular hydrogen bond between the substituent on the 5-position and the exo-cyclic amino moiety has an important effect on ΔE^{HB}, in the case of C^{5X}; on the other hand, the substituent in G^{8X} is unable to form such an intramolecular hydrogen bond.[9] The substitution effect on the exo-cyclic amino moiety of the guanine derivatives was also opposite to that of the corresponding cytosine derivatives. ΔE^{NP}

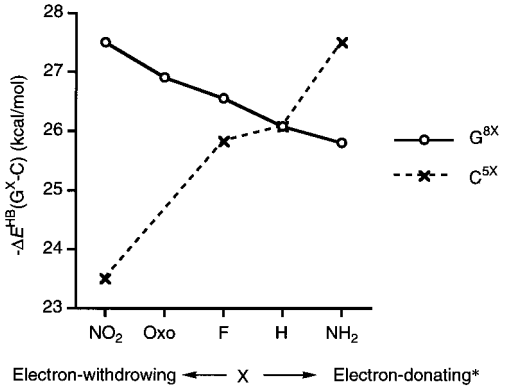

Figure 4. Substitution effect in the base pair hydrogen-bond energy of C^{5X} (--×--) and G^{8X} (—O—) in group B. *See ref 25.

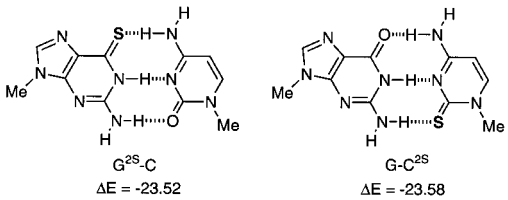

Figure 5. Hydrogen bond and ΔE^{HB} in G^{6S}−C and G−C^{6S} base pair.

of the G^{8NO_2}−C (2.88 kcal/mol) and G^{8NH_2}−C base pairs (5.78 kcal/mol) was larger than that of the G−C base pair (1.47 kcal/mol); thus, the total hydrogen-bond stability of the G^{8NO_2}–C base pair (−24.62 kcal/mol) was almost the same as that of the G−C base pair (−24.61 kcal/mol), and base pair stability should be largely decreased by introduction of the amino moiety on the 8-position. The exo-cyclic amino moiety on G^{2Nfo} and G^{2NMe} was planar. For group C, as expected, the deletion of the exo-cyclic amino moiety or replacing the exo-cyclic oxygen by sulfur causes a decrease of the base pair stability. It is quite reasonable that the substitution effect of this group was larger than that of the other groups. The effect of the deletion of the exo-cyclic amino moiety was larger than the effect of replacing the exo-cyclic oxygen by sulfur: ΔE^{HB} of the H−C base pair (−20.25 kcal/mol) was 5.83 kcal/mol less negative and ΔE^{HB} of the G^{6S}−C base pair (−23.52 kcal/mol) was 2.56 kcal/mol less negative than that of the G−C base pair. ΔE^{HB} of the H^{6S}−C base pair (−19.24 kcal/mol) was 1.01 kcal/mol less negative

TABLE 2: Charge Distribution (*e*) and Sum of the Charge Distributions[a] of the Guanine Derivatives in Mulliken, CHelpG and NPA Method

	Mulliken				CHelpG				NPA			
G^X	^1H	^{2N}H	^6O	sum	^1H	^{2N}H	^6O	sum	^1H	^{2N}H	^6O	sum
G	0.220	0.230	−0.421	0.871	0.409	0.436	−0.559	1.405	0.441	0.428	−0.566	1.435
G^{8NO_2}	0.225	0.234	−0.398	0.857	0.412	0.446	−0.536	1.394	0.444	0.431	−0.544	1.420
G^{8oxo}	0.225	0.231	−0.474	0.930	0.417	0.451	−0.605	1.473	0.444	0.429	−0.602	1.475
G^{8F}	0.222	0.231	−0.422	0.875	0.416	0.446	−0.571	1.433	0.442	0.429	−0.567	1.438
G^{8NH_2}	0.220	0.228	−0.441	0.889	0.418	0.448	−0.602	1.468	0.440	0.427	−0.581	1.448
G^{2Nfo}	0.221	0.221	−0.378	0.820	0.393	0.361	−0.539	1.293	0.446	0.440	−0.552	1.438
G^{2NMe}	0.212	0.204	−0.413	0.829	0.382	0.334	−0.561	1.276	0.438	0.425	−0.568	1.431
G^{6S}	0.221	0.233	−0.444	0.898	0.262	0.468	−0.358	1.088	0.448	0.431	−0.105	0.985
H	0.237		−0.422	0.660	0.382		−0.563	0.945	0.451		−0.561	1.012
H^{6S}	0.238		−0.429	0.667	0.256		−0.350	0.606	0.459		−0.087	0.546
G^{3C}	0.211	0.227	−0.434	0.872	0.399	0.433	−0.581	1.412	0.436	0.427	−0.584	1.447
G^{7C}	0.221	0.227	−0.461	0.909	0.397	0.448	−0.604	1.445	0.439	0.427	−0.591	1.457
G^{9C}	0.220	0.225	−0.480	0.925	0.393	0.446	−0.613	1.451	0.440	0.424	−0.604	1.468
G^{8N}	0.221	0.233	−0.395	0.849	0.403	0.439	−0.533	1.375	0.443	0.430	−0.547	1.420

[a] Sum of the charge distribution = |the charge distribution of ^1H| + |the charge distribution of ^{2N}H| + |the charge distribution of ^6O|.

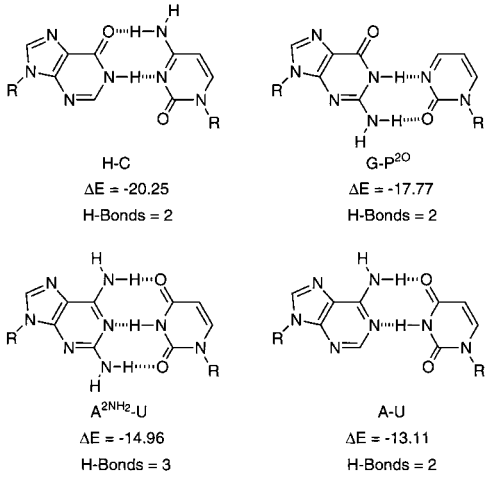

Figure 6. Hydrogen bond and ΔE^{HB} in H−C, G−P^{2O}, and A^{2NH_2}–U base pair.

than that of the H−C base pair, as expected. The substitution effect of the exo-cyclic oxygen by sulfur in G (G−C → G^{6S}−C; see Figure 5; $\Delta\Delta E$ = 2.56 kcal/mol) was almost the same as that of C (G−C → G−C^{6S}; $\Delta\Delta E$ = 2.50 kcal/mol).[8] The effect of the deletion of the exo-cyclic amino moiety of G (G−C → H−C; see Figure 6; $\Delta\Delta E$ = 5.83 kcal/mol) was smaller than that in cytosine[9] (G−C → G−P^{2O}; $\Delta\Delta E$ = 8.31 kcal/mol). It is considered that the deletion of the exo-cyclic amino moiety of G (H-bond C is deleted) makes H-bond A weaker and H-bond B stronger, because the electron density of the purine ring is decreased by removing the electron-donating amino moiety. On the other hand, the deletion of the exo-cyclic amino moiety of C (H-bond A is deleted) makes the pyrimidine ring electron-deficient; then it is considered to make both H-bonds B and C weaker.[26] ΔE^{HB} of the H−C base pair (Figure 6, −20.25 kcal/mol), which forms only two hydrogen bonds, was more negative not only than that of the A−U base pair (−13.11 kcal/mol) but also than that of the A^{2NH_2}−U base pair (Figure 6, −14.96 kcal/mol),[8] which forms three hydrogen bonds. Jorgensen et al. proposed secondary interactions in multiple hydrogen-bonded complexes.[27] On the basis of their proposal, the most stable complex should be obtained when the orientation of all the hydrogen bonds is the same, and on the other hand, the stability of the hydrogen-bond complex should be least negative when the hydrogen bonds formed in alternate orientation. According to their proposal, ΔE^{HB} of the G−P^{2O} base pair (−17.77 kcal/mol) was more negative than that of

TABLE 3: Hydrogen-Bond Length (Å)

	hydrogen-bond length			
G^X	H-bond A	H-bond B	H-bond C	average
G	1.9115	2.0455	2.0021	1.9864
G^{8NO_2}	1.9457	2.0292	1.9627	1.9792
G^{8oxo}	1.9254	2.0242	1.9872	1.9789
G^{8F}	1.9144	2.0390	1.9906	1.9813
G^{8NH_2}	1.8971	2.0480	2.0148	1.9866
G^{2Nfo}	1.9306	2.0289	1.9322	1.9639
G^{2NMe}	1.9036	2.0620	2.0183	1.9946
G^{6S}	2.4523	2.3166	1.8883	2.2190
H	1.9919	1.9351		1.9635
H^{6S}	2.5585	2.0247		2.2916
G^{3C}	1.8810	2.0793	2.0121	1.9908
G^{7C}	1.9085	2.0464	2.0217	1.9922
G^{9C}	1.9234	2.0507	2.0430	2.0057
G^{8N}	1.9395	2.0395	1.9764	1.9851

the A−U base pair (−13.11 kcal/mol). Therefore, the secondary interactions in the former are considered to affect favorably stable base pair formations comparing with those in the latter. However, ΔE^{HB} of the G−P^{2O} base pair was less negative than that of the H−C base pair (−20.25 kcal/mol). Orientation of the hydrogen bonds in the H−C and A−U base pairs is alternated; conversely, that in G−P^{2O} is in the same direction. Obviously, the hydrogen-bond energy per hydrogen bond is different for each base pair, and the base pair stability is determined not only by the number and the direction of the hydrogen bonds in the base pair but also the hydrogen-bond capability of the hydrogen bond sites. For group D, only ΔE^{HB} of the G^{8N}−C base pair (−26.73 kcal/mol) was more negative than that of the G−C. G^{8N} is the only derivative whose aromatic carbon is substituted to nitrogen. This type of substitution causes the aromatic ring to become electron-deficient, so it is equivalent to introducing an EWG on the aromatic ring. On the other hand, substitution from nitrogen to carbon is considered as making the aromatic ring electron-rich; thus, ΔE^{HB} became less negative. This was the same trend as the substitution effect in group B. ΔE^{NP} of the G^{3C}−C (2.55 kcal/mol) and G^{9C}−C base pairs (2.08 kcal/mol) was larger than that of the G−C base pair.

We also studied the charge distribution of the atoms that participate in the hydrogen bonds and the hydrogen-bond lengths (see Tables 2 and 3). Nevertheless, no remarkable trends were observed in the relationship between the hydrogen-bond energy and the atomic charge[28] nor between the hydrogen-bond energy and hydrogen-bond length, except for the relationship between the hydrogen-bond length and charge distribution of ^1H and ^{2N}H. Figure 7A−C shows the relationship between the charge

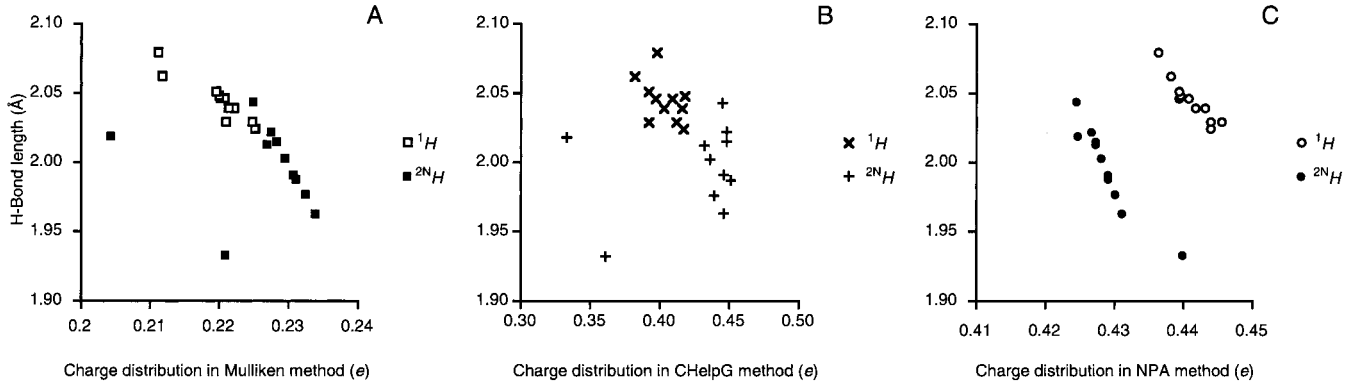

Figure 7. Relationship between the charge distribution on ^{1}H (*e*) and H-bond B length (Å) and between the charge distribution on ^{2N}H (*e*) and H-bond C length (Å): A, Mulliken; B, ChelpG; and C, NPA.

distribution on ^{1}H and H-bond B length and between the charge distribution on ^{2N}H and H-bond C length, in the Mulliken, ChelpG,[29] and NPA[30] methods. Remarkable trends were observed in the Mulliken (Figure 7A) and NPA (Figure 7C) methods relevant to the relationships between the charge distribution on ^{1}H and H-bond B length and between the charge distribution on ^{2N}H and H-bond C length: the more positively charged proton forms the shorter hydrogen bond. On the other hand, no remarkable trend was observed in the relationship between the charge distribution on ^{6}O and H-bond A length in all three charge distribution analysis methods.[28] Thus, the hydrogen-bond properties in the G^X−C base pairs could not be comprehensively interpreted by the charge distribution analysis nor by the length of the hydrogen bonds. Only the charge distribution of the proton in the hydrogen bonds was correlated with the hydrogen-bond length. Platts also reported that charge distribution was not a valid predictor for the hydrogen-bond stability.[31] Guerra et al. reported that charge-transfer interaction has an important role in base pair hydrogen bonding.[32] Our results in the charge distribution analysis were in good agreement with these results and suggested that the hydrogen-bond property in a base pair should not be characterized only by the electrostatic contribution.

Conclusion

The substitution effect on hydrogen-bond energy of the Watson−Crick type base pair between 1-methylcytosine and chemically modified 9-methylguanine derivatives was evaluated by ab initio molecular orbital theory. Opposite to the substitution effect in cytosine, and in contrast to the substitution effect in adenine, introducing an electron-withdrawing group on the 8-position or on the exo-cyclic amino moiety enforced the base pair stability. The guanine derivatives, which have a nitro group on the 8-position, could form the most stable hydrogen bonds with cytosine in the present study. It was difficult to comprehensively interpret the hydrogen-bond properties of the base pairs between 1-methylcytosine and chemically modified 9-methylguanine derivatives by use of charge distribution or the length of the hydrogen bonds. Thus, the hydrogen-bond property in a base pair should not be characterized only by the electrostatic contribution.

Acknowledgment. The services and computational time made available by the 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: Discussion of charge distribution analysis and figures illustrating the relation between charge distribution and hydrogen-bond length, between the sum of charge distribution and hydrogen-bond energy, and between average hydrogen-bond length and ΔE^{HB}. 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) Murray, J. A. H., Ed. *Antisense RNA and DNA*; 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) Kawahara, S.-I.; Wada, T.; Kawauchi, S.; Uchimaru, T.; Sekine, M. *J. Phys Chem. A* **1999**, *103*, 8516.

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

(9) Kawahara, S.-I.; Kobori, A.; Sekine, M.; Taira, K.; Uchimaru, T. *J. Phys. Chem. A* **2001**, *105*, 10596.

(10) (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. Phys. Chem.* **1996**, *100*, 1965. (m) Sponer, J.; Hobza, P. *J. Phys. Chem. A* **2000**, *104*, 4592. (n) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *J. Am. Chem. Soc.* **2000**, *122*, 4117. In fact, there are some theoretical studies for individual modified base pair. For example, the hydrogen-bonding and stacking properties of thioguanine and thiouracils were studied: (o) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. A* **1997**, *101*, 9489.

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

(12) 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.

(13) (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.

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

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

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

(17) (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.

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

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

(20) However, the difference in energy between the conformers was not so large, so base pair formation between such guanine analogues and cytosine should be entropically unfavorable.

(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.; 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) For G^{2NMe}: (a) Rife, J. P.; Cheng, C. S.; Moore, P. B.; Strobel, S. A. *Nucleosides Nucleotides* **1998**, *17*, 2281. For G^{6S} and H^{6S}: (b) Clivio, P.; Fourrey, J. L.; Favre, A. *J. Chem. Soc., Perkin Trans. 1* **1993**, 2585. (c) Chae, W.-G.; Chan, T. C. K.; Chang, C.-J. *Tetrahedron* **1998**, *54*, 8661. For H: (d) Jiricny, J.; Wood, S. G.; Martin, D.; Ubasawa, A. *Nucleic Acids Res*. **1986**, *14*, 6579. For G^{3C}: (e) Liu, M.-C.; Luo, M.-Z.; Mozdziesz, D. E.; Lin, T.-S.; Dutschman, G. E.; Cheng, Y.-C.; Sartorell, A. C. *Nucleosides Nucleotides* **1999**, *18*, 55. For G^{7C}: (f) Ramasamy, K.; Imamura, N.; Robins, R. K.; Revankar, G. R. *Tetrahedron Lett*. **1987**, *28*, 5107. (g) Seela, F.; Ramzaeva, N.; Chen, Y. *Bioorg. Med. Chem. Lett.* **1995**, *5*, 3049. For G^{9C}: (h) Rao, T. S.; Lewis, A. F.; Durland, R. H.; Revankar, G. R. *Tetrahedron Lett*. **1993**, *34*, 6709. (i) Lim, M. I.; Ren, W. Y.; Otter, B. A.; Klein, R. S. *J. Org. Chem*. **1983**, *48*, 780. For G^{8N}: (j) Shealy, Y. F.; Clayton, J. D.; Arnett, G.; Shannon, W. M. *J. Med. Chem*. **1984**, *27*, 670.

(24) The nucleic acid base orientation of most of the 8-substituted purine is syn in nucleotides,[33] so these derivatives are not suitable for the expected stable base pair formation in nucleotides such as an antisense application. However, if the structure of the backbone is substantially altered, e.g., in PNA,[34] the substituent at the 8-position will not be unfavorable for base pair formation via hydrogen bonds.

(25) Hammet parameter: X = NO_2, $\sigma_m = 0.71$, $\sigma_p = 0.81$; 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. *Correlation Analysis in Chemisry*; Chapman, N. B., Shorter, J., Eds.; Plenum Press: New York, 1978.

(26) The effect of the deletion of one exo-cyclic amino moiety in the G−C base pair was unrelated to the secondary interaction rule.[27] The secondary interactions between the 2-position exo-cyclic amino moiety of G and the 3-position endo-cyclic nitrogen atom of C act attractively, while the secondary interactions between the 4-position exo-cyclic amino moiety of C and the 1-position endo-cyclic nitrogen atom of G act repulsively. Thus, the effect of the deletion of one 2-position exo-cyclic amino moiety of G should be larger than that of C, from the standpoint of the secondary interaction.

(27) (a) Jorgensen, W. L.; Pranata, J. *J. Am. Chem. Soc*. **1990**, *112*, 2008−2010. (b) Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. *J. Am. Chem. Soc*. **1991**, *113*, 2810−2819.

(28) See Supporting Information. Additionally, Mulliken charge distribution analysis is highly basis set dependent. For example, Sponer et al. reported that the Mulliken charge analysis obtained from MP2 calculations using 6-31G*(0.25) and 6-31G** basis sets suggested quite different atomic charge for the sulfur atom in 6-thioguanine and for the corresponding oxygen atom on guanine.[10o]

(29) Breneman, C. M.; Wiberg, K. B. *J. Comp. Chem.* **1990**, *11*, 361.

(30) NBO ver. 3.1, Glendening, E. D.; Reed, A. E.; Carpenter, J. E.; Weinhold, F.; Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev*. **1988**, *88*, 899.

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

(32) (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.

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

(34) (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.