

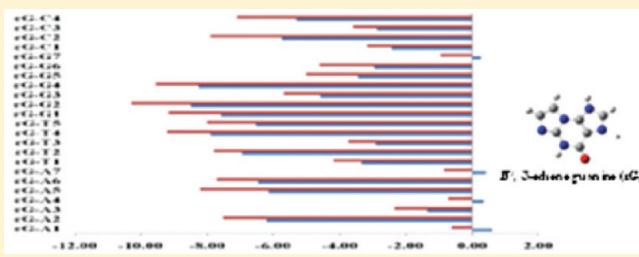
Model Calculations for the Misincorporation of Nucleotides Opposite Five-Membered Exocyclic DNA Adduct: $N^2,3$ -Ethenoguanine

Venkatesan Srinivasadesikan, Prabhat K. Sahu, and Shyi-Long Lee*

Department of Chemistry and Biochemistry, National Chung Cheng University, Chia-Yi, 621 Taiwan

S Supporting Information

ABSTRACT: Five-membered exocyclic DNA adducts are biologically very significant because of their potential to block DNA replication and transcription. $N^2,3$ -Ethenoguanine ($N^2,3\text{-}\varepsilon\text{G}$) has been identified in the liver DNA of vinyl chloride-exposed rats as a five-membered DNA adduct. Singer et al. (Carcinogenesis 1987, 8, 745–747) reported that the misincorporation of thymine (T), with two hydrogen bonds to $N^2,3\text{-}\varepsilon\text{G}$, represents the mutagenic event. Although the base-pairing specificity and mode of misincorporation have been studied experimentally for the $N^2,3$ -ethenoguanine adduct, molecular-level information is not yet clear. In this study, we have considered all four different DNA nucleotides paired with the $N^2,3$ -ethenoguanine adduct for model calculations toward the determination of base-pairing specificity. To provide insight into the mutagenic process of DNA damage based on geometric characteristics and electronic properties, the B3LYP and M06 methods were employed for these model calculations. Single-point energy calculations at the MP2/6-311++G** level on the corresponding optimized geometries were also carried out to better estimate the hydrogen-bonding strengths. The polarizable conductor calculation model (CPCM), which accounts for the overall polarizability of the solvent, was also employed. The computed reaction enthalpy values lie in the order $\varepsilon\text{G}-\text{G}(2)$ (10.3 kcal/mol) > $\varepsilon\text{G}-\text{G}(4)$ (9.6 kcal/mol) > $\varepsilon\text{G}-\text{T}(4)$ (9.2 kcal/mol) > $\varepsilon\text{G}-\text{G}(1)$ (9.1 kcal/mol) > $\varepsilon\text{G}-\text{A}(5)$ (8.2 kcal/mol) > $\varepsilon\text{G}-\text{C}(2)$ (7.9 kcal/mol) at the M06 level, which indicates that guanine and thymine are most favorable for mispairing with the $N^2,3$ -ethenoguanine adduct.



INTRODUCTION

Five-membered exocyclic DNA adducts are biologically very significant because of their potential to block DNA replication and transcription, induce DNA strand breaks, trigger apoptosis, and cause gene mutations and chromosomal aberrations. These effects could lead directly to carcinogenesis.¹ Common exocyclic DNA adducts, such as $1,N^6$ -ethenoadenine (εA), $3,N^4$ -ethenocytosine (εC), $N^2,3$ -ethenoguanine ($N^2,3\text{-}\varepsilon\text{G}$), and $1,N^2$ -ethenoguanine ($1,N^2\text{-}\varepsilon\text{G}$), are responsible for blocking Watson–Crick base pairing.² The annelation products of exocyclic five-membered rings in DNA bases arising from the reactions of electrophiles from vinyl halides; vinyl monomers, including chloroacetaldehyde; and endogenous lipid peroxidation of malondialdehyde, crotonaldehyde, 4-hydroxy-2-nonenal, 4,5-epoxy-2(E)-decanal, 4-hydroperoxynonenal, and trans-fatty acids and their metabolic products are also of significance in the formation of DNA adducts.³ Among the different five-membered DNA adducts, $N^2,3$ -ethenoguanine ($N^2,3\text{-}\varepsilon\text{G}$) has been identified in the liver DNA of vinyl chloride-exposed rats, the major route of human exposure.⁴ Singer et al.⁵ reported that misincorporation of thymine (T), having two hydrogen bonds with $N^2,3\text{-}\varepsilon\text{G}$, represents the mutagenic event. Again, $N^2,3\text{-}\varepsilon\text{G}$ is the only derivative formed in vivo by the human carcinogen vinyl chloride that can be shown to have a high probability of causing transitions that could initiate malignant transformation.⁵ Cheng et al.⁶ determined the base-pairing specificity of $N^2,3\text{-}\varepsilon\text{G}$ in

Escherichia coli through genetic reversion assay and concluded that $N^2,3\text{-}\varepsilon\text{G}$ specifically induces $\text{G} \rightarrow \text{A}$ transitions, with 13% mutagenic potential. Failure of the mismatch repair system can lead to genomic instability, whereas damage in genes responsible for maintaining DNA stability can lead to the risk of developing cancer and many other debilitating diseases.⁷ Understanding the complexity of the recognition and repair of erroneous base pairing has been a major area of research in biological and medical communities for many decades.⁸ Only limited computational studies of five-membered exocyclic DNA adducts for the determination of cancer etiology have been reported.⁹ In this article, we report model calculations on the misincorporation of nucleotides opposite the five-membered exocyclic $N^2,3$ -ethenoguanine adduct ($N^2,3\text{-}\varepsilon\text{G}$). The present study could provide useful information, such as geometric characteristics, electronic properties, and physical parameters, for the misincorporation of DNA nucleotides in the $N^2,3$ -ethenoguanine adduct, based on different conformations and unique mutagenic properties that are essential for the continuous effort to understand the base-pairing specificity toward the determination of cancer etiology.

Received: March 23, 2011

Revised: July 5, 2011

Published: July 21, 2011

■ MODEL SYSTEM AND COMPUTATIONAL DETAILS

The DNA double helix is stabilized by hydrogen bonds between the bases attached to the double strands. In our model systems, the pentose (five-carbon) sugar group and phosphate

group were capped with hydrogen atoms. We considered all possible different positions of interaction between the $N^{2,3}$ -ethenoguanine adduct ($N^{2,3}\text{-}\varepsilon\text{G}$) and the DNA nucleotides (i.e., $\varepsilon\text{G}\text{--A}$, $\varepsilon\text{G}\text{--T}$, $\varepsilon\text{G}\text{--G}$, $\varepsilon\text{G}\text{--C}$), keeping in mind the capped

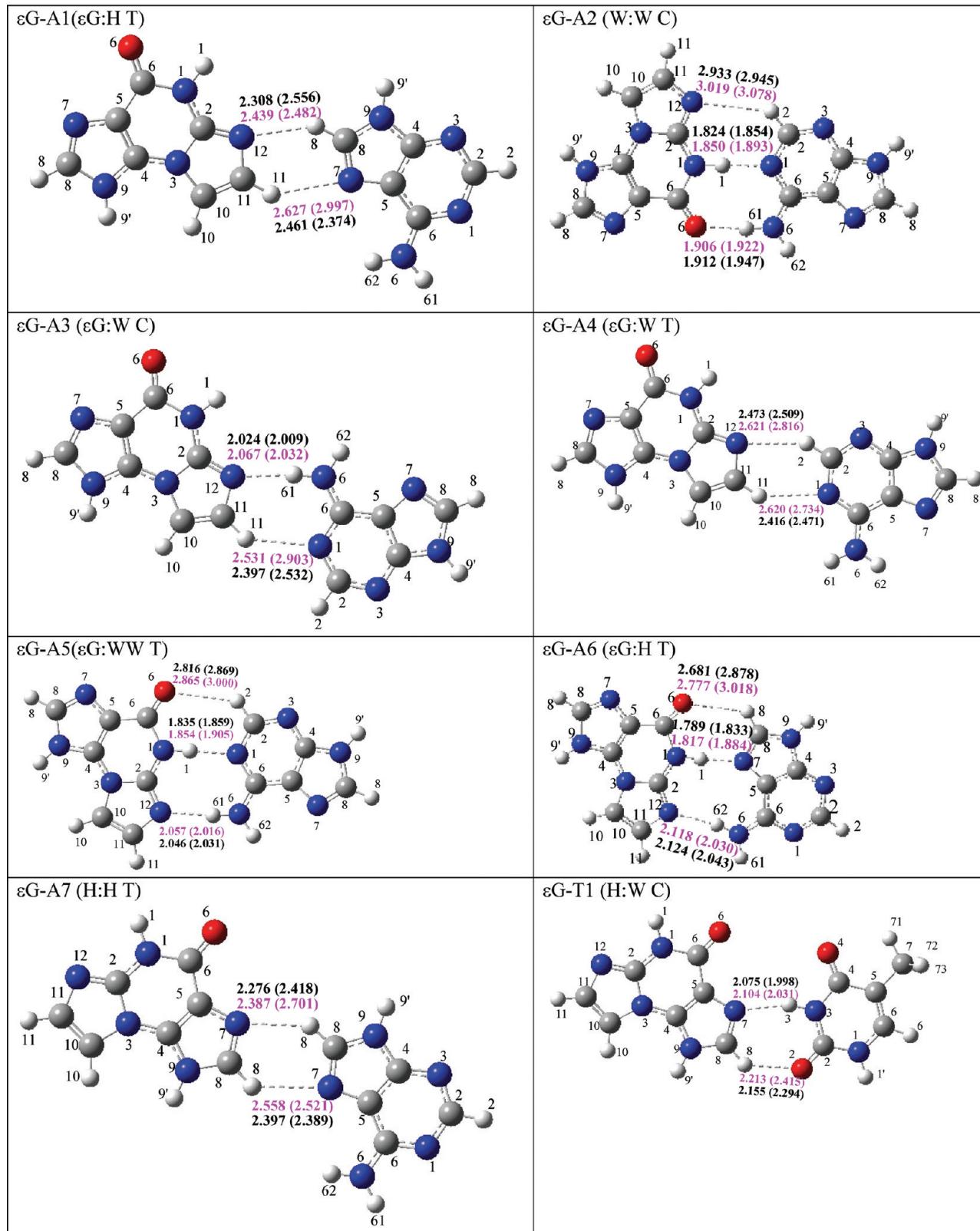


Figure 1. Continued

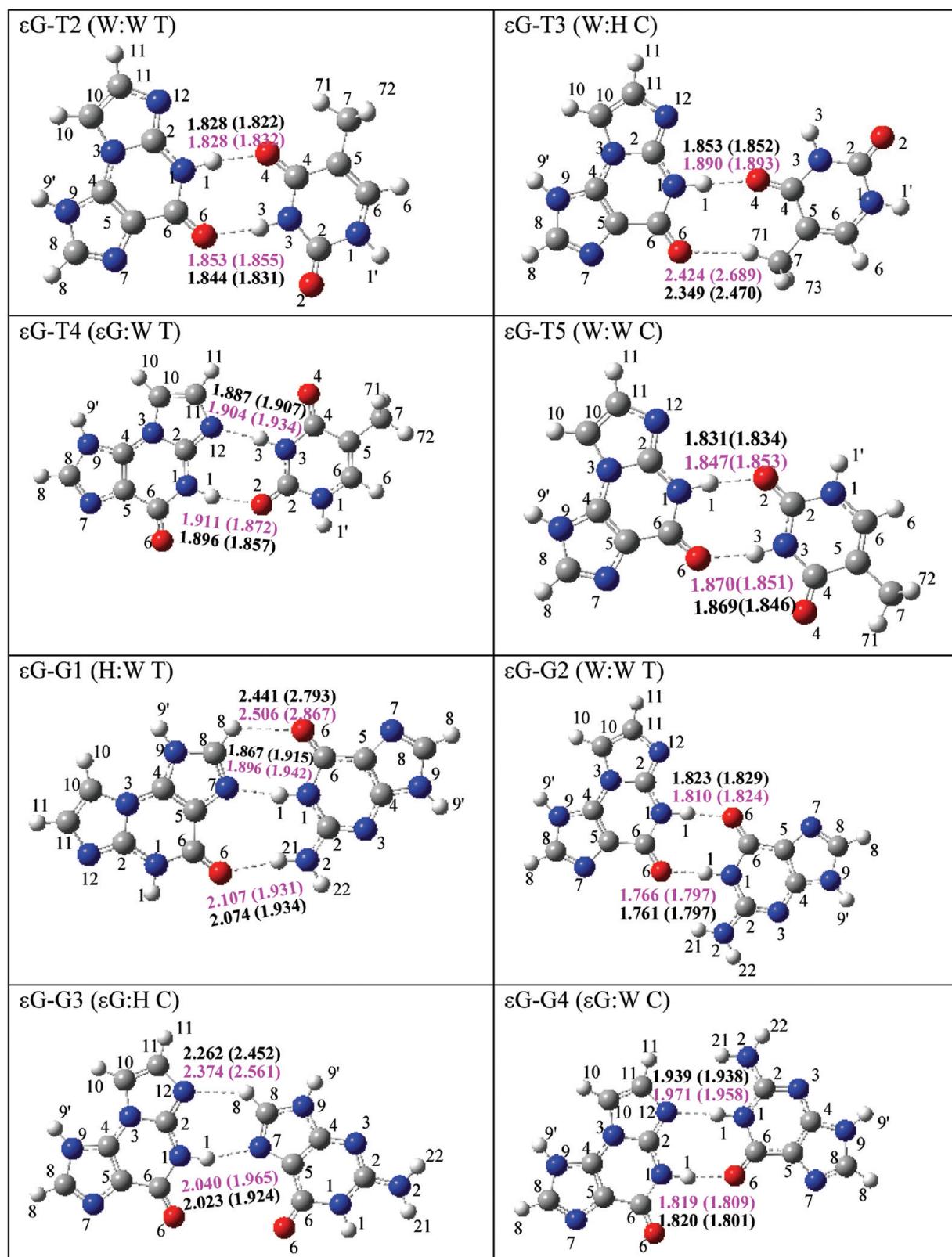


Figure 1. Continued

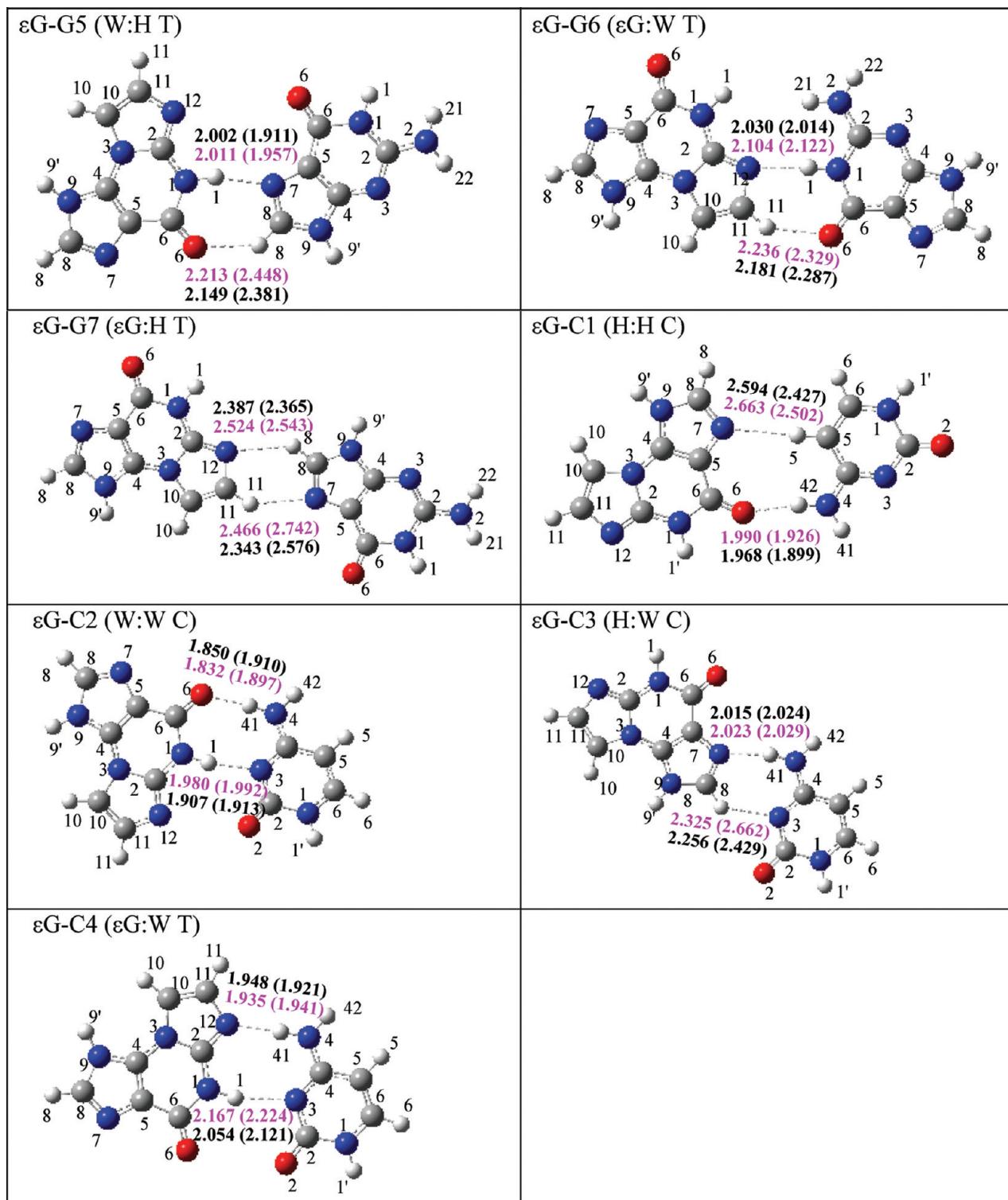


Figure 1. Optimized structures for $N^2,3$ -ethenoguanine adduct and DNA base complexes in the gas phase (aqueous phase) using different DFT levels with the 6-31+G* basis set. Computed values at different DFT levels: B3LYP, magenta; M06, black.

hydrogen atoms for the excluded sugar and phosphate groups. The geometries and harmonic vibrational frequencies of εG , adenine, cytosine, guanine, thymine, and all of the different resultant possible geometries for εG with these DNA nucleotide complexes were calculated using density functional theory (DFT) methods

with the Pople-type split valence basis set (6-31+G*). DFT combines accuracy with low computational cost in such a way that it constitutes an attractive alternative to other post-Hartree–Fock (HF) procedures.^{10–13} B3LYP,^{14–18} a hybrid density functional, was considered because of several successful results on hydrogen

Table 1. Hydrogen-Bonding Strengths (kcal/mol) for N^2 ,3-Ethenoguanine Adduct and DNA Base Complexes in the Gas Phase (ΔE_g) and Aqueous Phase (ΔE_{aq}) Calculated Using DFT Levels/6-311++G**//DFT Levels/6-31+G* and MP2 Levels/6-311++G**//DFT Levels/6-31+G*^a

complex	B3LYP//B3LYP		M06//M06		MP2//B3LYP		MP2//M06	
	ΔE_g	ΔE_{aq}						
$\varepsilon G-A1$	-3.23 (-2.96)	-0.64	-4.80 (-4.40)	-2.20	-5.82 (-4.54)	-2.90	-5.91 (-4.46)	-3.07
$\varepsilon G-A2$	-13.65 (-12.94)	-8.10	-15.85 (-15.15)	-10.54	-16.59 (-13.65)	-11.44	-16.67 (-13.68)	-11.53
$\varepsilon G-A3$	-6.16 (-5.78)	-3.25	-7.90 (-7.43)	-4.93	-8.71 (-7.02)	-5.75	-8.74 (-6.93)	-5.78
$\varepsilon G-A4$	-1.46 (-1.20)	-0.26	-3.19 (-2.83)	-1.74	-3.48 (-2.23)	-2.52	-4.10 (-2.77)	-2.59
$\varepsilon G-A5$	-12.73 (-12.06)	-8.30	-15.00 (-14.31)	-10.74	-16.09 (-15.42)	-11.83	-16.14 (-13.39)	-11.88
$\varepsilon G-A6$	-13.78 (-13.10)	-8.43	-15.76 (-15.01)	-10.56	-17.46 (-14.52)	-12.21	-17.52 (-14.52)	-12.28
$\varepsilon G-A7$	-4.99 (-4.70)	-0.80	-6.62 (-6.19)	-2.51	-7.66 (-6.32)	-3.17	-7.72 (-6.23)	-3.22
$\varepsilon G-T1$	-6.99 (-6.47)	-3.98	-8.75 (-8.13)	-6.05	-9.94 (-7.83)	-7.01	-9.94 (-7.73)	-6.99
$\varepsilon G-T2$	-11.44 (-10.71)	-8.19	-13.02 (-12.30)	-9.91	-13.54 (-10.47)	-10.96	-13.53 (-10.48)	-10.95
$\varepsilon G-T3$	-6.38 (-4.83)	-3.52	-7.82 (-7.26)	-5.16	-8.63 (7.16)	-6.17	-8.65 (-6.42)	-6.24
$\varepsilon G-T4$	-13.34 (-12.58)	-9.75	-15.40 (-14.59)	-12.03	-16.70 (-13.69)	-13.37	-16.71 (-13.66)	-13.40
$\varepsilon G-T5$	-10.97 (-10.26)	-7.82	-12.59 (-11.87)	-9.59	-13.15 (-10.18)	-10.62	-13.07 (-10.13)	-10.62
$\varepsilon G-G1$	-17.28 (-16.59)	-8.22	-19.58 (-18.79)	-10.65	-19.05 (-16.24)	-10.74	-19.19 (-16.27)	-10.90
$\varepsilon G-G2$	-17.01 (-16.27)	-9.06	-18.94 (-18.21)	-11.23	-17.90 (-14.53)	-11.28	-18.01 (-14.61)	-11.38
$\varepsilon G-G3$	-6.29 (-5.74)	-4.70	-8.46 (-7.79)	-7.07	-8.93 (-6.61)	-7.33	-8.93 (-6.54)	-7.51
$\varepsilon G-G4$	-12.81 (-12.07)	-9.07	-15.22 (-14.43)	-11.47	-15.24 (-12.11)	-11.95	-15.38 (-12.19)	-12.09
$\varepsilon G-G5$	-8.45 (-7.92)	-4.84	-10.31 (-9.70)	-6.88	-11.31 (-9.13)	-8.18	-11.28 (-9.05)	-8.24
$\varepsilon G-G6$	-8.27 (-7.78)	-3.70	-10.64 (-10.04)	-6.04	-10.06 (-7.91)	-5.95	-10.24 (-7.92)	-6.24
$\varepsilon G-G7$	-3.86 (-3.57)	-0.44	-5.64 (-5.22)	-2.13	-5.14 (-3.75)	-1.92	-5.28 (-3.76)	-2.11
$\varepsilon G-C1$	-8.01 (-7.62)	-3.90	-9.15 (-8.65)	-5.32	-10.07 (-8.38)	-6.33	-10.14 (-8.36)	-6.29
$\varepsilon G-C2$	-12.34 (-11.52)	-7.24	-14.99 (-14.07)	-10.19	-16.03 (-12.46)	-11.69	-16.37 (-12.48)	-12.06
$\varepsilon G-C3$	-12.60 (-12.09)	-3.51	-14.77 (-14.16)	-5.71	-15.33 (-13.25)	-6.21	-15.40 (-13.19)	-6.30
$\varepsilon G-C4$	-9.77 (-9.03)	-6.96	-12.39 (-11.63)	-9.53	-13.65 (-10.79)	-10.43	-13.60 (-10.58)	-10.41

^a BSSE-corrected binding energies (kcal/mol) in parentheses.

bonding, even for the weakly bonded systems. However, to better describe the dispersion contribution to the hydrogen-bonding strength, the M06¹⁹ method was also considered for these model calculations. Single-point energy calculations at the MP2²⁰/6-311++G** level on B3LYP- and M06-optimized geometries were also carried out to better estimate the hydrogen-bonding strengths. The hydrogen-bonding strengths were calculated according to the equation

$$\Delta E = E_{\varepsilon G-\text{DNA nucleotide}} - E_{\varepsilon G} - E_{\text{DNA nucleotide}} \quad (1)$$

where $E_{\varepsilon G-\text{DNA nucleotide}}$ corresponds to the total energy of the complex and $E_{\varepsilon G}$ and $E_{\text{DNA nucleotide}}$ are the energies of isolated monomers. The calculated hydrogen-bonding energies were corrected for basis set superposition error (BSSE)²¹ using the counterpoise method.²² The continuum model CPCM (polarizable conductor calculation model along with united-atom topological model)²³ ($\varepsilon = 78.39$) that accounts for the overall polarizability of the solvent was employed for the optimization and enthalpy computations at various DFT levels with the 6-31+G* basis set. Single-point energy calculations using the CPCM model were also carried out with the 6-311++G** basis set. Relative binding energies (RBEs) for both the gas and solvent phases were determined by considering the most tightly bound complex as the reference value. All calculations were performed using the Gaussian 09 suite of programs.²⁴

RESULTS AND DISCUSSION

Optimized Geometric Parameters for N^2 ,3-Ethenoguanine Adduct and DNA Nucleotide Complexes. Although the base-pairing specificity and mode of misincorporation have been studied experimentally^{5,6} for the N^2 ,3-ethenoguanine adduct, molecular-level information is not yet clear. Therefore, as mentioned earlier, we considered all four different DNA nucleotides paired with the N^2 ,3-ethenoguanine adduct for the model calculations toward determining the base-pairing specificity.

The optimized structures for the N^2 ,3-ethenoguanine–DNA nucleotide complexes ($\varepsilon G-A$, $\varepsilon G-T$, $\varepsilon G-G$, $\varepsilon G-C$) are shown in Figure 1. For all of these different conformations, we also provide the accepted nomenclature for the εG –DNA nucleotide pairs (see Figure 1) according to Leontis et al. and others,²⁵ with some modifications for the interacting edges of the DNA bases with the guanine adduct (εG) to those compared with normal bases. The capped hydrogen for the sugar and phosphate group remains far from the hydrogen-bond interaction, justifying its exclusion from our model calculations, which is also supported by earlier calculations.²⁶ The fundamental importance of the hydrogen bond lies in its role in molecular association and is considered as an important factor in stabilizing biomolecular structures. Because biochemical processes are predominantly solution-based, with water serving as the ubiquitous solvent, our discussion is based on aqueous-phase computational results. The optimized hydrogen-bond distances and bond

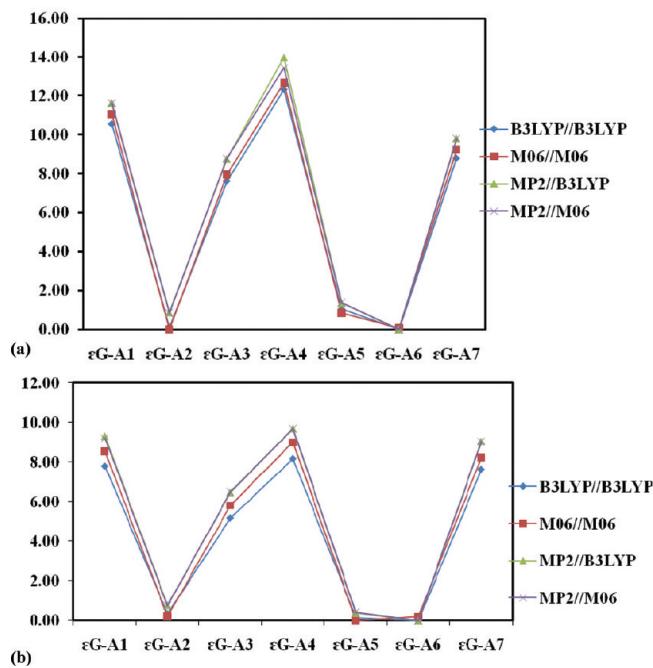


Figure 2. Relative binding energies (kcal/mol) for $N^2,3$ -ethenoguanine adduct and adenine ($\varepsilon G-A$) complexes in the (a) gas and (b) aqueous phases using DFT and MP2 levels/6-311++G**//DFT levels/6-31+G*.

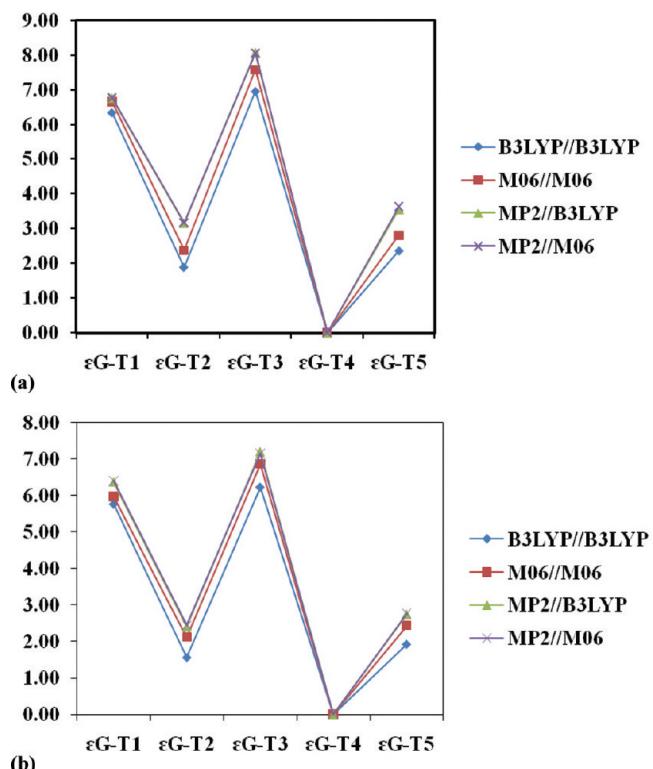


Figure 3. Relative binding energies (kcal/mol) for $N^2,3$ -ethenoguanine adduct and thymine ($\varepsilon G-T$) complexes in the (a) gas and (b) aqueous phases using DFT and MP2 levels/6-311++G**//DFT levels/6-31+G*.

angles for the several different conformations of $N^2,3$ -ethenoguanine adduct and DNA nucleotide complexes are listed in

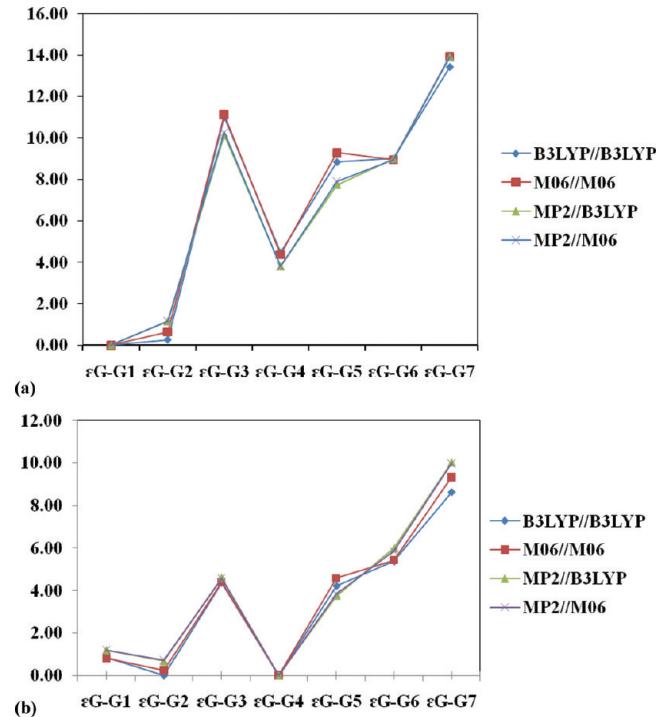


Figure 4. Relative binding energies (kcal/mol) for $N^2,3$ -ethenoguanine adduct and guanine ($\varepsilon G-G$) complexes in the (a) gas and (b) aqueous phases using DFT and MP2 levels/6-311++G**//DFT levels/6-31+G*.

Table S1a–d (cf. Supporting Information for detail descriptions). From Table S1a–d (Supporting Information) and Figure 1, it can be observed that four different types of hydrogen bonds ($N-H \cdots N$, $N-H \cdots O$, $C-H \cdots O$, and $C-H \cdots N$) play vital roles in stabilizing different εG -DNA nucleotide complexes. For all these of different complexes, based on both B3LYP and M06 computations, the $N-H \cdots N$ and $N-H \cdots O$ hydrogen bonds are shorter and the corresponding bond angles are more linear than those for the $C-H \cdots N$ and $C-H \cdots O$ hydrogen bonds, indicating a stronger interaction for the former pair. As it was already reported by Czyzniowska et al.²⁷ regarding the weak contribution from $C-H \cdots O$ - and $C-H \cdots N$ -type hydrogen bonds and also according to our previous computations⁹ on such types of weak hydrogen bonds, the present investigation also reveals the similar nature of longer bond lengths associated with bond angles that deviate from linearity. On comparing with normal G-C WC base pair²⁸ and $\varepsilon G-C(2)$, it can be noticed that the bond angles $\angle OHN$ and $\angle NHN$ decrease by around 4° and 8° , respectively. Again, the $\angle NHO$ bond is missing because the free amine group of guanine is covalently linked with the etheno group in $\varepsilon G-C(2)$. It can also be observed that the exocyclic double bond in etheno group maintains planarity for some stable conformations, even though two hydrogen-bond lengths and hydrogen-bond angles cannot completely describe the planarity.

Energetics and Hydrogen-Bonding Strength for $N^2,3$ -Ethenoguanine Adduct and DNA Nucleotide Complexes. The computed total energies and the BSSE-corrected hydrogen-bonding strengths for several different conformations of the $N^2,3$ -ethenoguanine adduct and DNA bases are listed in Table S2a–d (cf. Supporting Information). Table 1 summarizes the hydrogen-bonding strengths in the gas phase (along with

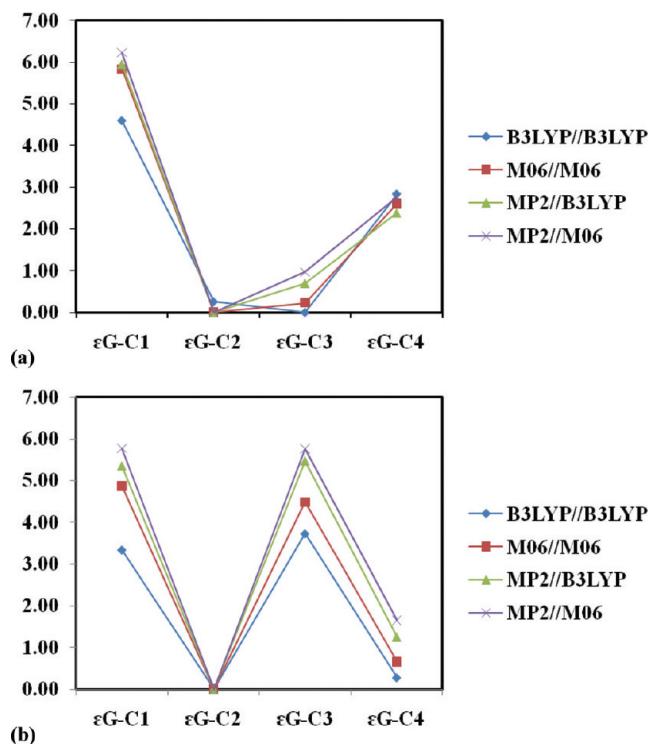


Figure 5. Relative binding energies (kcal/mol) for $N^2,3$ -ethenoguanine adduct and cytosine ($\varepsilon G-C$) complexes in the (a) gas and (b) aqueous phases using DFT and MP2 levels/6-311++G**//DFT levels/6-31+G*.

the BSSE corrections) and in the aqueous phase computed using DFT and MP2 levels/6-311++G**//DFT levels/6-31+G*.

$N^2,3$ -Ethenoguanine Adduct and Adenine ($\varepsilon G-A$) Complexes. Parts a and b of Figure 2 show the computed relative binding energies for the $N^2,3$ -ethenoguanine adduct and adenine complexes $\varepsilon G-A(1)$, $\varepsilon G-A(2)$, $\varepsilon G-A(3)$, $\varepsilon G-A(4)$, $\varepsilon G-A(5)$, $\varepsilon G-A(6)$, and $\varepsilon G-A(7)$ at the DFT and MP2 levels with the 6-311++G** basis set for the gas and aqueous phases, respectively (cf. Supporting Information, Table S3). From Table 1, one can see that the computed hydrogen-bonding strengths for $\varepsilon G-A(2)$, $\varepsilon G-A(5)$, and $\varepsilon G-A(6)$ complexes are higher (differing within 0.4 kcal/mol) in the aqueous phase (within 1.0 kcal/mol in the gas phase) among the $\varepsilon G-A$ complexes. The reason for the higher binding strengths is the strong interaction through $N-H \cdots N$ and $N-H \cdots O$ hydrogen bonds. On shifting to the aqueous phase, a decrease in the hydrogen-bonding strength is observed because of solvent polarization effects.

$N^2,3$ -Ethenoguanine Adduct and Thymine ($\varepsilon G-T$) Complexes. Parts a and b of Figure 3 show the computed relative binding energies for the $N^2,3$ -ethenoguanine adduct and thymine complexes $\varepsilon G-T(1)$, $\varepsilon G-T(2)$, $\varepsilon G-T(3)$, $\varepsilon G-T(4)$, and $\varepsilon G-T(5)$ at the DFT and MP2 levels with the 6-311++G** basis set for the gas and aqueous phases, respectively (cf. Supporting Information, Table S4). Table 1 also indicates that the computed hydrogen-bonding strength for the $\varepsilon G-T(4)$ complex is highest among the $\varepsilon G-T$ complexes, which is due to the strong interaction associated with $N-H \cdots N$ and $N-H \cdots O$ hydrogen bonds.

$N^2,3$ -Ethenoguanine Adduct and Guanine ($\varepsilon G-G$) Complexes. Parts a and b of Figure 4 show the computed relative binding energies for the $N^2,3$ -ethenoguanine adduct and guanine

complexes $\varepsilon G-G(1)$, $\varepsilon G-G(2)$, $\varepsilon G-G(3)$, $\varepsilon G-G(4)$, $\varepsilon G-G(5)$, $\varepsilon G-G(6)$, and $\varepsilon G-G(7)$ as obtained at the DFT and MP2 levels with the 6-311++G** basis set in the gas and aqueous phases, respectively (cf. Supporting Information, Table S5). It was found that the $\varepsilon G-G(1)$ and $\varepsilon G-G(2)$ complexes have higher computed hydrogen-bonding strengths (lying within 1.0 kcal/mol) (see Table 1). However, on shifting to the aqueous phase, as a result of solvent polarization, the hydrogen-bonding strengths for the $\varepsilon G-G(4)$ complex (strong interactions through $N-H \cdots N$ and $N-H \cdots O$ hydrogen bonds) and the $\varepsilon G-G(2)$ complex (strong interactions through two $N-H \cdots O$ hydrogen bonds) are very similar (within 1.0 kcal/mol). As described earlier [see Figure 1 and Table S1c (Supporting Information)], the hydrogen-bond lengths are 1.93 Å for $N12-H1$ and 1.80 Å for $O6-H1$ at the M06 level and 1.95 and 1.80 Å, respectively, at the B3LYP level for the $\varepsilon G-G(4)$ complex, and two strong $N-H \cdots O$ hydrogen bonds with $O6-H1$ lengths of 1.82 and 1.79 Å at the M06 and B3LYP levels, respectively, are obtained for $\varepsilon G-G(2)$ complex. It is also noted that both the $\varepsilon G-G(2)$ and $\varepsilon G-G(4)$ complexes are planar and that the linearity in their bond angles also increases due to hydration (see Table S1c, Supporting Information).

$N^2,3$ -Ethenoguanine Adduct and Cytosine ($\varepsilon G-C$) Complexes. Parts a and b of Figure 5 show the computed relative binding energies for the $N^2,3$ -ethenoguanine adduct and cytosine complexes $\varepsilon G-C(1)$, $\varepsilon G-C(2)$, $\varepsilon G-C(3)$, and $\varepsilon G-C(4)$ at the DFT and MP2 levels with the 6-311++G** basis set for the gas and aqueous phases, respectively (cf. Supporting Information, Table S6). It can be seen that the computed hydrogen-bonding strengths for the $\varepsilon G-C(2)$ and $\varepsilon G-C(3)$ complexes (within 0.5 kcal/mol) are higher in the gas phase (see Table 1). On shifting to the aqueous phase, the highest hydrogen-bonding strength is obtained for the $\varepsilon G-C(2)$ complex because of a less pronounced solvent polarization effect on the strong interaction associated with the $N-H \cdots N$ and $N-H \cdots O$ hydrogen bonds.

In general, Table 1 shows that the hydrogen-bonding strengths as calculated by M06 are about 1.5–2 kcal/mol more attractive than those calculated by B3LYP, because the M06 functional considers the dispersion interaction in addition to the electrostatic and exchange-correlation components of the Hamiltonian. As no interatomic contact is smaller than the corresponding van der Waals radii, the constant shift due to the dispersion attraction is reflected in the higher values reported in Table 1. Even after the BSSE-corrected hydrogen-bonding strengths were computed at the MP2 level using a larger basis set on the corresponding optimized geometries obtained at different DFT levels, it was noticed that the computed values were similar for each of the different conformations (see Table 1). Among different conformations of the $N^2,3$ -ethenoguanine adduct and DNA bases, the highest hydrogen-bonding strengths at the MP2//DFT levels (in the gas phase) were found to be both qualitatively and quantitatively similar and were in the order $\varepsilon G-G(1)$ (19 kcal/mol) > $\varepsilon G-A(6)$ (17.5 kcal/mol) > $\varepsilon G-T(4)$ (16.7 kcal/mol) > $\varepsilon G-C(2)$ (16 kcal/mol). The qualitative trend (in the gas phase) was also found to be the same: $\varepsilon G-G(1)$ (17.2 kcal/mol) ≈ $\varepsilon G-G(2)$ (17.0 kcal/mol) > $\varepsilon G-A(6)$ (13.7 kcal/mol) ≈ $\varepsilon G-A(2)$ (13.6 kcal/mol) > $\varepsilon G-T(4)$ (13.3 kcal/mol) > $\varepsilon G-C(3)$ (12.6 kcal/mol) ≈ $\varepsilon G-C(2)$ (12.3 kcal/mol) at the B3LYP level and $\varepsilon G-G(1)$ (19.6 kcal/mol) ≈ $\varepsilon G-G(2)$ (18.9 kcal/mol) > $\varepsilon G-A(2)$ (15.9 kcal/mol) ≈ $\varepsilon G-A(6)$ (15.8 kcal/mol) > $\varepsilon G-T(4)$ (15.4 kcal/mol) > $\varepsilon G-C(2)$ (14.9 kcal/mol) ≈ $\varepsilon G-C(3)$ (14.7 kcal/mol) at the M06 level. The important conclusion that can be

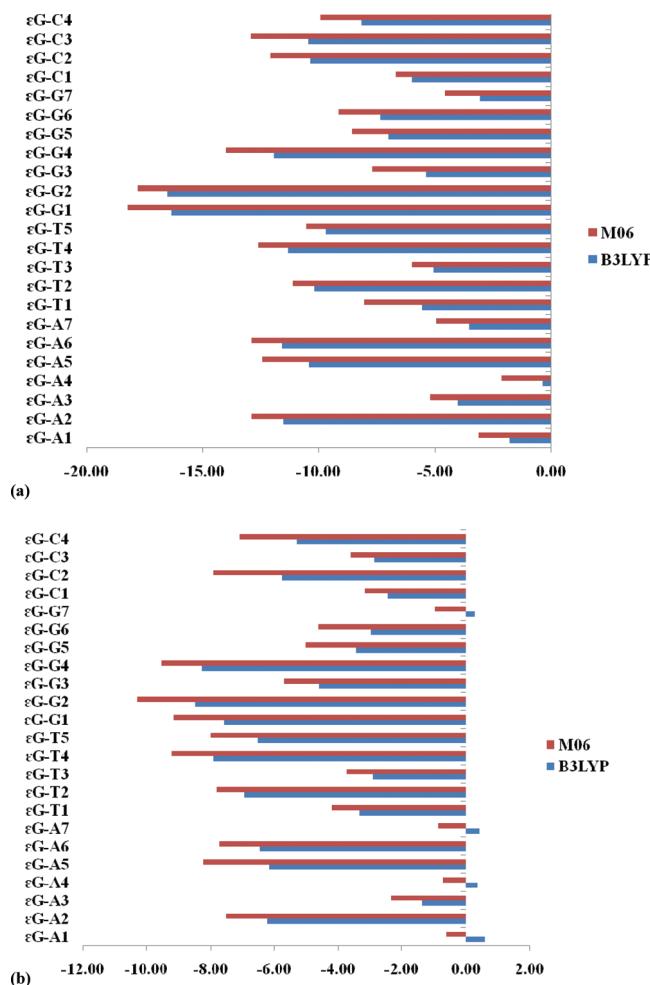


Figure 6. Computed reaction enthalpy values (kcal/mol) for the $N^2,3$ -ethenoguanine adduct and DNA bases at different DFT levels in the (a) gas and (b) solvent phases at 298.15 K.

drawn for the gas-phase computations is that $\varepsilon G-G(1)$ is the most favorable and potential candidate for the mispairing; however, the other pairs $\varepsilon G-A(6)$, $\varepsilon G-T(4)$, and $\varepsilon G-C(2)$ are also close in binding energy strength (within 1.0 kcal/mol). Compared with earlier reported interaction energies of the G-C WC base pair by Sponer et al.²⁹ (27.5 kcal/mol), Mo³⁰ (25.2 kcal/mol), and Roy et al.³¹ (26.5 kcal/mol), the corresponding adduct base pair $\varepsilon G-C(2)$ has an energy of 12.3 kcal/mol, because of the missing N-H \cdots O interaction (free amine group of guanine is covalently linked with the etheno group). As compared to the methyl-group-capped G-G base-pair interaction energy (17.58 kcal/mol),³¹ the corresponding $\varepsilon G-G(1)$ energy in our study is also similar (17.28 kcal/mol), which reveals that the etheno group does not play any role and, rather, the interbase interaction prevails. Again, earlier MP2-level computations³² reported that the stabilization energies for G-CWC, GA1, GT1, and GT2 are 25.4, 15.7, 14.7, and 14.3 kcal/mol, respectively. In our present investigation, the hydrogen-bonding strengths for the corresponding adduct base pairs $\varepsilon G-C(2)$, $\varepsilon G-A(2)$, $\varepsilon G-T(2)$, and $\varepsilon G-T(5)$ as obtained at the MP2/6-311++G(d,p)//B3LYP/6-31+G(d) level were found to be 16.0, 16.6, 13.5, and 13.2 kcal/mol, respectively. The discrepancy found in our study is obvious and was mentioned earlier for $\varepsilon G-C(2)$; however, for $\varepsilon G-A(2)$,

$\varepsilon G-T(2)$, and $\varepsilon G-T(5)$, we believe that the strong interbase hydrogen-bond interaction prevails over the direct etheno group interaction.

On shifting to the aqueous phase, the stability trend at the MP2//DFT levels was found to be in the order $\varepsilon G-T(4)$ (13.4 kcal/mol) $>$ $\varepsilon G-A(6)$ (12.2 kcal/mol) \approx $\varepsilon G-G(4)$ (12.0 kcal/mol) \approx $\varepsilon G-C(2)$ (12.0 kcal/mol). The qualitative trend (in the aqueous phase) was also observed to be the same: $\varepsilon G-T(4)$ (9.75 kcal/mol) $>$ $\varepsilon G-G(4)$ (9.07 kcal/mol) \approx $\varepsilon G-G(2)$ (9.06 kcal/mol) $>$ $\varepsilon G-A(6)$ (8.43 kcal/mol) \approx $\varepsilon G-A(5)$ (8.3 kcal/mol) \approx $\varepsilon G-A(2)$ (8.1 kcal/mol) $>$ $\varepsilon G-C(2)$ (7.2 kcal/mol) at the B3LYP level and $\varepsilon G-T(4)$ (12.03 kcal/mol) $>$ $\varepsilon G-G(4)$ (11.47 kcal/mol) \approx $\varepsilon G-G(2)$ (11.23 kcal/mol) $>$ $\varepsilon G-A(5)$ (10.74 kcal/mol) \approx $\varepsilon G-A(6)$ (10.56 kcal/mol) \approx $\varepsilon G-A(2)$ (10.54 kcal/mol) $>$ $\varepsilon G-C(2)$ (10.2 kcal/mol) at the M06 level. Therefore, qualitatively, there is no significant difference with respect to the dispersion contribution to the hydrogen-bonding strength to obtain the most stable conformer or the trend based on DFT levels. However, in a quantitative sense, B3LYP underestimates the hydrogen-bonding strength, whereas M06 provides better performance. From the aqueous-phase computations, it was concluded that the $N^2,3$ -ethenoguanine adduct is most favorable for forming a pair with thymine as the nucleotide [$\varepsilon G-T(4)$ conformation], which is also in line with the experimental observation^{5,6} of a high level of $\varepsilon G \rightarrow A$ transition resulting from $\varepsilon G-T$ pairing upon replication of εG -containing DNA. In addition, this pairing resembles that of the wobble pair G:T and would not be expected to distort the helix significantly. However, according to our computations, the other pairs with different conformations $\varepsilon G-G(4)$, $\varepsilon G-G(2)$, $\varepsilon G-A(6)$, $\varepsilon G-A(5)$, and $\varepsilon G-A(2)$ are very close in binding energy strength and cannot be ignored for the mispairing.

Table S7a,b (Supporting Information) reports the computed reaction enthalpies in the gas and aqueous phases for the $N^2,3$ -ethenoguanine and DNA bases at the B3LYP and M06 levels at 298.15 K. Figure 6a,b presents comparative plots of the computed reaction enthalpies. From Figure 6a, it can be observed that the reaction enthalpy values in the gas phase lie in the order $\varepsilon G-G(2)$ (16.5 kcal/mol) \approx $\varepsilon G-G(1)$ (16.3 kcal/mol) $>$ $\varepsilon G-A(6)$ (11.5 kcal/mol) \approx $\varepsilon G-A(2)$ (11.5 kcal/mol) $>$ $\varepsilon G-T(4)$ (11.3 kcal/mol) $>$ $\varepsilon G-C(3)$ (10.4 kcal/mol) \approx $\varepsilon G-C(2)$ (10.4 kcal/mol) at the B3LYP level and $\varepsilon G-G(1)$ (18.2 kcal/mol) \approx $\varepsilon G-G(2)$ (17.8 kcal/mol) $>$ $\varepsilon G-A(2)$ (12.9 kcal/mol) \approx $\varepsilon G-A(6)$ (12.9 kcal/mol) \approx $\varepsilon G-C(3)$ (12.9 kcal/mol) $>$ $\varepsilon G-T(4)$ (12.6 kcal/mol) $>$ $\varepsilon G-C(2)$ (12.1 kcal/mol) at the M06 level. These results are similar to the trend obtained during hydrogen-bonding strength computations. The qualitative trend is the same for the B3LYP (although the values are underestimated) and M06 levels. On shifting to the aqueous phase, the reaction enthalpy values lie in the order $\varepsilon G-G(2)$ (8.5 kcal/mol) \approx $\varepsilon G-G(4)$ (8.3 kcal/mol) $>$ $\varepsilon G-T(4)$ (7.9 kcal/mol) $>$ $\varepsilon G-G(1)$ (7.6 kcal/mol) $>$ $\varepsilon G-A(6)$ (6.5 kcal/mol) \approx $\varepsilon G-A(2)$ (6.2 kcal/mol) $>$ $\varepsilon G-C(2)$ (5.7 kcal/mol) at the B3LYP level and $\varepsilon G-G(2)$ (10.3 kcal/mol) $>$ $\varepsilon G-G(4)$ (9.6 kcal/mol) $>$ $\varepsilon G-T(4)$ (9.2 kcal/mol) $>$ $\varepsilon G-G(1)$ (9.1 kcal/mol) $>$ $\varepsilon G-A(5)$ (8.2 kcal/mol) $>$ $\varepsilon G-C(2)$ (7.9 kcal/mol) at the M06 level. Therefore, it was found that $\varepsilon G-G(2)$, $\varepsilon G-G(4)$, $\varepsilon G-G(1)$, and $\varepsilon G-T(4)$ have similar reaction enthalpy values (within 0.6 kcal/mol) and, hence, these are most favorable mispairs for understanding the base-pairing specificity with the $N^2,3$ -ethenoguanine adduct.

SUMMARY AND CONCLUSIONS

The present computations based on qualitative trends as well as the quantitative values provide a good comparison between the performances of the B3LYP and M06 level of theories. Our aim was to perform model calculations of DNA base misincorporation with the N^2 ,3-ethenoguanine adduct, which are essential for the continuous effort to understand base-pairing specificity and the mode of misincorporation. It is known that DNA damage is an alteration of the structure of the DNA molecule that prevents the molecule from being replicated and, hence, from being inherited. Double-stranded DNA damage, if not repaired, results in the loss of genetic information, which can be harmful or lethal to an individual's gametes and offspring.³³

Several different conformations of the complexes $\varepsilon G-A$, $\varepsilon G-T$, $\varepsilon G-G$, and $\varepsilon G-C$ with regard to different types of hydrogen-bond interactions were investigated in terms of base-pairing specificity and the mode of misincorporation. The computed reaction enthalpy values lie in the order $\varepsilon G-G(2)$ (10.3 kcal/mol) > $\varepsilon G-G(4)$ (9.6 kcal/mol) > $\varepsilon G-T(4)$ (9.2 kcal/mol) > $\varepsilon G-G(1)$ (9.1 kcal/mol) > $\varepsilon G-A(S)$ (8.2 kcal/mol) > $\varepsilon G-C(2)$ (7.9 kcal/mol) at the M06 level, which indicates that both guanine and thymine are most favorable to be mispaired with the N^2 ,3-ethenoguanine adduct. Specifically, we believe that the strong interbase hydrogen-bond interaction prevails over the weak etheno group interaction for the $\varepsilon G-G(2)$, $\varepsilon G-G(4)$, $\varepsilon G-G(1)$, and $\varepsilon G-T(4)$ conformations. Although the $\varepsilon G-T(4)$ conformation was found in line with experimental observation^{5,6} of a high level of $\varepsilon G \rightarrow A$ transition resulting from $\varepsilon G-T$ pairing upon replication of εG -containing DNA, our investigations also suggest that the $\varepsilon G-G(2)$, $\varepsilon G-G(4)$, and $\varepsilon G-G(1)$ conformations are also potent and favorable mispairs for understanding the base-pairing specificity with the N^2 ,3-ethenoguanine adduct.

ASSOCIATED CONTENT

Supporting Information. Optimized hydrogen-bond distances and bond angles for the several different conformations of N^2 ,3-ethenoguanine adduct and DNA nucleotide complexes (Table S1a–d), computed total energy and the BSSE-corrected hydrogen-bonding strength for the several different conformations of N^2 ,3-ethenoguanine adduct and DNA bases (Table S2a–d), computed relative binding energies for the N^2 ,3-ethenoguanine adduct and DNA nucleotide complexes (Tables S3–S6), computed reaction enthalpy values for the N^2 ,3-ethenoguanine adduct and DNA nucleotide complexes (Table S7a,b). This information is available free of charge via the Internet at <http://pubs.acs.org/>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: chesll@ccu.edu.tw.

ACKNOWLEDGMENT

This research was supported by the National Science Council (NSC) of Taiwan and the computational resource is partially supported by National Center for High-Performance Computing (NCHC), Hsin-Chu, Taiwan.

REFERENCES

- (1) (a) Blair, I. A. *Exp. Gerontol.* **2001**, *36*, 1473–1481. (b) Burcham, P. C. *Mutagenesis* **1998**, *13*, 287–305.
- (2) (a) Choi, J. H.; Pfeifer, G. P. *Mutat. Res.* **2004**, *568*, 245–256. (b) Gros, L.; Ishchenko, A. A.; Saparbaev, M. *Mutat. Res.* **2003**, *531*, 219–229. (c) Douki, T.; Odin, F.; Caillat, S.; Favier, A.; Cadet, J. *Free Radical Biol. Med.* **2004**, *37*, 62–70. (d) Speina, E.; Kierzek, A. M.; Tudek, B. *Mutat. Res.* **2003**, *531*, 205–217. (e) Mishina, Y.; Yang, C.-G.; He, C. *J. Am. Chem. Soc.* **2005**, *127*, 14594–14595.
- (3) (a) Green, T.; Hathaway, D. E. *Chem.-Biol. Interact.* **1978**, *22*, 211–224. (b) Eberle, G.; Barbin, A.; Laib, R. J.; Ciroussel, F.; Thomale, J.; Bartsch, H.; Rajewsky, M. F. *Carcinogenesis* **1989**, *10*, 209–212. (c) Fedtke, N.; Boucheron, J. A.; Walker, V. E.; Swenberg, J. A. *Carcinogenesis* **1990**, *11*, 1287–1292. (d) Guengerich, F. P.; Mason, P. S.; Stott, W. T.; Fox, T. R.; Watanabe, P. G. *Cancer Res.* **1981**, *41*, 4391–4398. (e) Guengerich, F. P.; Kim, D. H. *Chem. Res. Toxicol.* **1991**, *4*, 413–421. (f) Guengerich, F. P. *Chem. Res. Toxicol.* **1992**, *5*, 2–5. (g) Bartsch, H.; Barbin, A.; Marion, M. J.; Nair, J.; Guichard, Y. *Drug Metab. Rev.* **1994**, *26*, 349–371. (h) Nair, U.; Bartsch, H.; Nair, J. *Free Radical Biol. Med.* **2007**, *43*, 1109–1120. (i) Lee, S. H.; Oe, T.; Blair, I. A. *Chem. Res. Toxicol.* **2002**, *15*, 300–304. (j) Petrova, K. V.; Jalluri, R. S.; Kozekov, I. D.; Rizzo, C. J. *Chem. Res. Toxicol.* **2007**, *20*, 1685–1692. (k) Lee, S. H.; Arora, J. A.; Oe, T.; Blair, I. A. *Chem. Res. Toxicol.* **2005**, *18*, 780–786. (l) Lee, S. H.; Blair, I. A. *Chem. Res. Toxicol.* **2000**, *13*, 698–702. (m) Kawai, Y.; Uchida, K.; Osawa, T. *Free Radical Biol. Med.* **2004**, *36*, 529–541. (n) Lee, S. H.; Silva Elipe, M. V.; Arora, J. S.; Blair, I. A. *Chem. Res. Toxicol.* **2005**, *18*, 566–578.
- (4) Creech, J. L.; Johnson, M. N. *J. Occup. Med.* **1974**, *16*, 150–151.
- (5) Singer, B.; Spengler, S. J.; Chavez, F.; Kusmierenk, J. T. *Carcinogenesis* **1987**, *8*, 745–747.
- (6) Cheng, K. C.; Preston, B. D.; Cahill, D. S.; Dosanjh, M. K.; Singer, B.; Loeb, L. A. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 9974–9978.
- (7) Loft, S.; Poulsen, H. E. *J. Mol. Med.* **1996**, *74*, 297–312.
- (8) Schofield, M. J.; Hsieh, P. *Annu. Rev. Microbiol.* **2003**, *57*, 579.
- (b) Berashevich, J.; Chakraborty, T. *J. Chem. Phys.* **2009**, *130*, 015101.
- (9) (a) Sahu, P. K.; Kuo, C.-W.; Lee, S.-L. *J. Phys. Chem. B* **2007**, *111*, 2991–2998. (b) Sahu, P. K.; Wang, C.-H.; Lee, S.-L. *J. Phys. Chem. B* **2009**, *113*, 14529–14535. (c) Sahu, P. K.; Jhong, M.-L.; Srinivasade-sikan, V.; Lee, S.-L. *J. Chem. Inf. Model.* **2011** manuscript submitted.
- (10) Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford Science: Oxford, U.K., 1989.
- (11) Magalhaes, A. L.; Madail, S. R. R. S.; Ramos, M. J. *Theor. Chem. Acc.* **2000**, *105*, 68–76.
- (12) Fuqiang, B.; Kathryn, N. R.; James, W. G.; Russell, J. B. *Theor. Chem. Acc.* **2002**, *108*, 1–11.
- (13) Loos, P. F.; Assfeld, X.; Rivail, J. L. *Theor. Chem. Acc.* **2007**, *118*, 165–171.
- (14) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (15) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (16) Becke, A. D. *J. Chem. Phys.* **1997**, *107*, 8554–8560.
- (17) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- (18) Schmid, H. L.; Becke, A. D. *J. Chem. Phys.* **1998**, *108*, 9624–9631.
- (19) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
- (20) (a) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.* **1988**, *153*, 503–506. (b) Saebo, S.; Almloef, J. *Chem. Phys. Lett.* **1989**, *154*, 83–89. (c) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett.* **1990**, *166*, 275–280. (d) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. *Chem. Phys. Lett.* **1990**, *166*, 281–289.
- (21) van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. C. M.; van Lenthe, J. H. *Chem. Rev.* **1994**, *94*, 1873–1885.
- (22) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.
- (23) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.
- (24) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N. ;

Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R. E.; Stratmann, O.; Yazyev, A. J.; Austin, R.; Cammi, C.; Pomelli, J. W.; Ochterski, R.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision A.1; Gaussian, Inc.: Wallingford, CT, 2009.

(25) (a) Leontis, N. B.; Stombaugh, J.; Westhof, E. *Nucleic Acids Res.* **2002**, *30*, 3497–3531. (b) Lemieux, S.; Major, F. *Nucleic Acids Res.* **2002**, *30*, 4250–4263. (c) Das, J.; Mukherjee, S.; Mitra, A.; Bhattacharyya, D. *J. Biomol. Struct. Dyn.* **2006**, *24*, 149–161. (d) Lu, X.-J.; Olson, W. K. *Nat. Protocols* **2008**, *3*, 1213–1227.

(26) Volk, D. E.; Thiviyananthan, V.; Somasunderam, A.; Gorenstein, D. G. *Org. Biomol. Chem.* **2006**, *4*, 1741–1745.

(27) Czyżnikowska, Z.; Góra, R. W.; Zaleśny, R.; Lipkowski, P.; Jarzembska, K. N.; Dominiak, P. M.; Leszczynski, J. *J. Phys. Chem. B* **2010**, *114*, 9629–9644.

(28) Mallajosyula, S. S.; Datta, A.; Pati, S. K. *Synth. Met.* **2005**, *155*, 398–401.

(29) Sponer, J.; Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142–10151.

(30) Mo, Y. *J. Mol. Model.* **2006**, *12*, 665–672.

(31) Roy, A.; Panigrahi, S.; Bhattacharyya, M.; Bhattacharyya, D. *J. Phys. Chem. B* **2008**, *112*, 3786–3796.

(32) Hobza, P.; Kabelac, M.; Sponer, J.; Mezlik, P.; Vondrasek, J. *J. Comput. Chem.* **1997**, *18*, 1136–1150.

(33) Bernstein, H.; Byerly, H.; Hopf, F.; Michod, R. *Science* **1985**, *229*, 1277–1281.