

Effective Fragment Potential Study of the Interaction of DNA Bases

Quentin A. Smith and Mark S. Gordon*

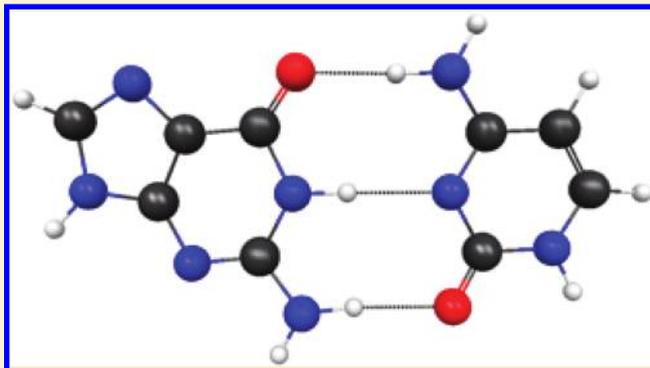
Department of Chemistry and Ames Laboratory, Iowa State University, Ames, Iowa 50011, United States

Lyudmila V. Slipchenko

Department of Chemistry, Purdue University, West Lafayette, Indiana 47907, United States

Supporting Information

ABSTRACT: Hydrogen-bonded and stacked structures of adenine–thymine and guanine–cytosine nucleotide base pairs, along with their methylated analogues, are examined with the ab initio based general effective fragment potential (EFP2) method. A comparison of coupled cluster with single, double, and perturbative triple (CCSD(T)) energies is presented, along with an EFP2 energy decomposition to illustrate the components of the interaction energy.



INTRODUCTION

Accurately modeling complexes as large as paired DNA nucleotide bases with ab initio methods remains difficult and computationally expensive. The most complete set of high-level ab initio calculations on these systems is that of Hobza et al.,¹ performed at the estimated coupled cluster with singles, doubles, and perturbative triples (CCSD(T))² level of theory. Recently, the CCSD(T) energies of the adenine–thymine complexes found in Hobza’s study were calculated with an improved basis set.³ Refs 1 and 3 provide benchmarks with which to test the accuracy of other methods.

The general effective fragment potential (EFP2) method⁴ is a fast, ab initio based method that has shown success in modeling dimers of benzene,⁵ benzene–water,⁶ substituted benzenes,⁷ and pyridine⁸ systems with noncovalent interactions similar to DNA. Previously, EFP2 has been used to examine interactions in stacked and H-bonded adenine and thymine dimers and oligomers.⁹ This work expands on the previous study by using the EFP2 method to model both the hydrogen-bonded and stacked DNA nucleotide base pairs of guanine–cytosine and methylated adenine–thymine and guanine–cytosine.

The adenine–thymine (A···T) and guanine–cytosine (G···C) Watson–Crick (WC) or hydrogen-bonded structures are the canonical nucleotide base pairs. Their methylated analogues are denoted mA···mT (9-methyladenine–1-methylthymine) and mG···mC (9-methylguanine–1-methylcytosine). Multiple hydrogen bonds give these complexes large interaction energies that are dominated by Coulomb forces.¹ G···C WC

and mG···mC WC are stabilized by three hydrogen bonds, denoted R_1 , R_2 , and R_3 in Figures 1 and 2. A···T WC and mA···mT WC have two hydrogen bonds, denoted R_1 and R_2 in Figures 3 and 4. While the intermonomer separation labeled R_3 in Figures 3 and 4 is not a true hydrogen bond, this distance is included for illustrative purposes.

The stacked nucleotide base structures are more difficult to describe than the hydrogen-bonded species without resorting to very high levels of theory. The difficulty in modeling these structures arises because of the inability of most methods, even the simpler electronic structure methods (e.g., Hartree–Fock (HF) and the most commonly used density functional theory (DFT) functionals), to accurately portray the dispersion energy.¹⁰ Second order perturbation theory¹¹ (MP2) is, in addition, known to overestimate the binding energy of stacked nucleotide bases by about 20%, relative to the more reliable (and more costly) coupled cluster method with single, double, and perturbative triple (CCSD(T)) excitations.^{1,12} Indeed, the Δ CCSD(T) correction term (the difference between CCSD(T) and MP2 energies) is positive (decreased binding) for the stacked nucleotide bases, in contrast to the hydrogen-bonded structures for which the correction is negative.¹ DFT often fails to find any bound nucleotide base pairs with a stacked motif,¹³

Special Issue: Pavel Hobza Festschrift

Received: May 23, 2011

Revised: August 2, 2011

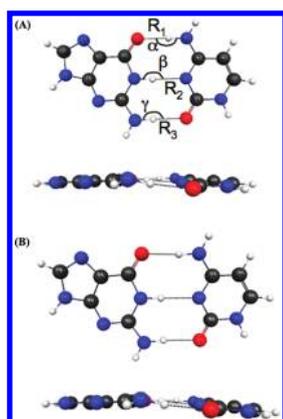


Figure 1. Top and side views of the guanine–cytosine hydrogen-bonded complex: (A) depicts the RI-MP2/cc-pVTZ optimized geometry of ref 1; (B) is the complex obtained by EFP2 optimization with the nucleotide base geometries of ref 1. Hydrogen bond lengths R_1 , R_2 , and R_3 and angles α , β , and γ (chosen to show the linearity of each hydrogen bond) are given in Table 1 for each complex.

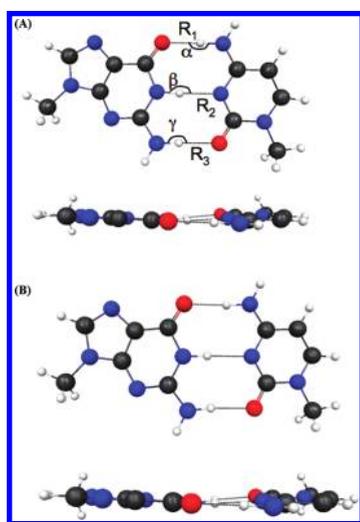


Figure 2. Top and side views of the methylated guanine–cytosine hydrogen-bonded complex: (A) depicts the RI-MP2/cc-pVTZ optimized geometry of ref 1; (B) is the complex obtained by EFP2 optimization with the nucleotide base geometries of ref 1. Hydrogen bond lengths R_1 , R_2 , and R_3 and angles α , β , and γ (chosen to show the linearity of each hydrogen bond) are given in Table 1 for each complex.

although DFT can be made to perform better when dispersion is explicitly introduced through parametrization or hybrid methods.^{14–16}

THEORETICAL METHODS

The general effective fragment potential (EFP2) method is coded in the GAMESS (General Atomic and Molecular Electronic Structure System)¹⁷ quantum chemistry software package, which was used for all calculations in this study. The EFP2 total interaction energy is decomposed into Coulomb, exchange-repulsion, polarization (induction), dispersion, and charge transfer terms. The Coulomb interaction is calculated using Stone's distributed multipolar expansion,¹⁸ carried out through octopole

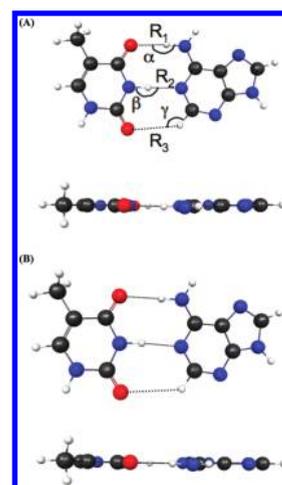


Figure 3. Top and side views of the adenine–thymine hydrogen-bonded complex: (A) depicts the RI-MP2/cc-pVTZ optimized geometry of ref 1; (B) is the complex obtained by EFP2 optimization with the nucleotide base geometries of ref 1. Intermolecular distances R_1 , R_2 , and R_3 and angles α , β , and γ are given in Table 1 for each complex.

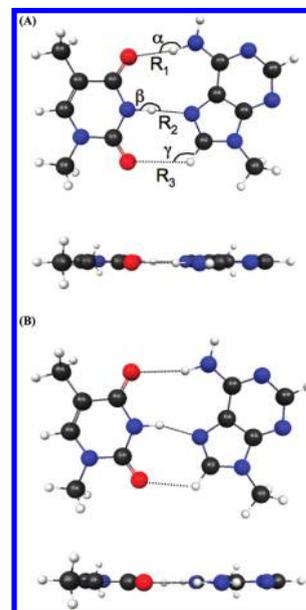


Figure 4. Top and side views of the methylated adenine–thymine hydrogen-bonded complex: (A) depicts the RI-MP2/cc-pVTZ optimized geometry of ref 1; (B) is the complex obtained by EFP2 optimization with the nucleotide base geometries of ref 1. Intermolecular distances R_1 , R_2 , and R_3 and angles α , β , and γ are given in Table 1 for each complex.

moments. The analytic distributed multipolar analysis (DMA) was used in this study, although a numerical DMA¹⁹ is also available. Charge penetration, required when fragments approach each other closely, is modeled by an exponential damping function that multiplies the distributed multipoles, up through the dipole–quadrupole term, but not the quadrupole–quadrupole or higher order terms.²⁰ Exchange-repulsion is derived as an expansion in the intermolecular overlap, truncated at the quadratic term.²¹ Polarization is represented by the sum of localized molecular orbital (LMO) polarizability tensors. Polarizable points are located at the LMO centroids, corresponding to the

bonds and lone pairs of the molecule. These LMO polarizabilities are determined using the coupled-perturbed Hartree–Fock equations.²² The dispersion interaction is expressed in terms of imaginary frequency-dependent polarizabilities, with an explicitly derived C_6 coefficient and an estimated C_8 coefficient.²² A Gaussian-type damping function was used with the polarization term and an overlap-based damping function with the dispersion term. Charge-transfer is computed from a perturbative treatment of the interaction between occupied orbitals on one fragment and virtual orbitals on a second fragment.²³ Charge transfer is not included in the present study because a previous study²³ demonstrated that this term is very small for most neutral molecules.

Geometries. Two sets (referred to as sets 1 and 2) of adenine–thymine (A···T) and guanine–cytosine (G···C) stacked and hydrogen-bonded geometries were examined with the EFP2 method, along with two sets of the corresponding methylated structures: 9-methyladenine–1-methylthymine (mA···mT) and 9-methylguanine–1-methylcytosine (mG···mC). The geometries of the individual nucleotide bases were obtained from RI-MP2 geometry optimizations¹ performed on the paired nucleotide bases. All EFP2 potentials were generated²⁴ with the analytic DMA²⁵ and the 6-311++G(3df,2p) basis set. EFP2 fragments are rigid, having fixed internal geometries.

Set 1. The first set of structures was taken directly from the benchmark data set referred to as JSCH-2005 in ref 1. The stacked structures chosen for the present study are denoted A···T S, G···C S, mA···mT S, and mG···mC S in JSCH-2005; the hydrogen-bonded structures are denoted A···T WC (Watson–Crick), G···C WC, mA···mT WC, and mG···mC WC. The geometries for all of these structures were obtained through optimization with the RI-MP2/cc-pVTZ method.¹ To analyze these structures with the EFP2 method, EFP potentials were generated separately for each nucleotide base. These EFP2 nucleotide bases were then recombined to produce paired geometries identical to the JSCH-2005 structures. Set 1 tests the ability of the EFP2 method to reproduce estimated CCSD(T) interaction energies.¹

Set 2. The second set of nucleotide base geometries began with the same nucleotide bases and EFP2 potentials described in Set 1. The internal coordinates of the individual nucleotide bases remain fixed. However, in this case, instead of constraining the nucleotide base pairs to the paired geometries found with RI-MP2 in ref 1, an EFP2 geometry optimization was performed. Set 2 tests the ability of the EFP2 method to reproduce paired geometries.

Interaction Energies. The estimated CCSD(T) interaction energies are taken from ref 1. In that work, RI-MP2 energies were counterpoise-corrected for basis set superposition error (BSSE)²⁵ and extrapolated to the complete basis set (CBS) limit by the two-point extrapolation scheme of Helgaker et al.²⁶ Small-basis CCSD(T) single point energy calculations were performed at the RI-MP2 optimized geometries; the difference between the small-basis CCSD(T) energy and the small-basis MP2 energy is the CCSD(T) energy correction term, $\Delta\text{CCSD}(T)$. The final estimated CCSD(T)/CBS interaction energy is given by $\Delta E_{\text{CBS}}^{\text{CCSD}(T)} = \Delta E_{\text{CBS}}^{\text{MP2}} + \Delta\text{CCSD}(T)$, where $\Delta\text{CCSD}(T) = (\Delta E_{\text{CBS}}^{\text{CCSD}(T)} - \Delta E_{\text{small-basis}}^{\text{MP2}})$.

Components of the Interaction Energy. The EFP2 method has previously been shown to agree closely with symmetry adapted perturbation theory (SAPT) predictions in systems similar to the paired nucleotide bases (e.g., dimers of benzene,⁵ substituted benzenes,⁷ and pyridine⁸). A recent SAPT study²⁷

Table 1. Geometries of Hydrogen Bonded Complexes^a

	opt type	R_1	R_2	R_3	α	β	γ
G···C WC	MP2	1.76	1.91	1.92	178.5	175.1	175.6
	EFP2	1.87	1.98	1.95	175.3	175.3	172.4
mG···mC WC	MP2	1.88	1.87	1.73	175.3	176.0	178.7
	EFP2	1.83	1.97	1.95	176.0	175.3	171.6
A···T WC	MP2	1.93	1.82	2.83	173.6	179.1	132.7
	EFP2	2.01	1.93	2.94	178.5	173.5	127.5
mA···mT WC	MP2	1.93	1.75	2.72	169.3	175.7	120.8
	EFP2	2.23	1.78	2.39	166.2	159.9	119.1

^a Lengths (in Å) and angles (in degrees) of the hydrogen bonds in the Watson–Crick (WC) complexes of Figures 1–4. “Opt type” refers to the level of theory used for the geometry optimization, either RI-MP2 (ref 1) or EFP2. Lengths are indicated by R_1 , R_2 , and R_3 in Figures 1 and 2 for guanine–cytosine (G···C) and methylated guanine–cytosine (mG···mC) and by R_1 and R_2 in Figures 3 and 4 for adenine–thymine (A···T) and methylated adenine–thymine (mA···mT). For A···T and mA···mT, intermonomer separation R_3 is also given, although it is not a true hydrogen bond.

examined stacked and hydrogen-bonded adenine–thymine; a comparison of EFP2 interaction energy components with those of ref 27 is presented here. Previous studies examining the paired nucleotide bases using a combined SAPT/density functional theory (DFT) approach may also be found in the literature.^{15,16}

RESULTS AND DISCUSSION

Hydrogen-Bonded Structures. EFP2 geometry optimizations (Set 2) accurately predict the structures of the hydrogen-bonded complexes. The lengths of the hydrogen bonds corresponding to R_1 , R_2 , and R_3 (Figures 1 and 2) or to R_1 and R_2 (Figures 3 and 4) are given in Table 1. The intermonomer separation R_3 in Figures 3 and 4 are also given in Table 1, although R_3 is not a true hydrogen bond. The angles denoted by Greek letters in Figures 1–4 are chosen to show the linearity of each hydrogen bond. These values are reported in Table 1.

The EFP2 G···C WC hydrogen bond lengths differ from the RI-MP2/cc-pVTZ¹ lengths by 0.03–0.11 Å (Table 1). The EFP2 A···T WC hydrogen bond lengths are 0.08–0.11 Å longer than those found with RI-MP2¹. The nonmethylated EFP2 structures depicted in Figures 1B and 3B are very similar to the RI-MP2¹ structures of Figures 1A and 3A, respectively. Structural differences are greater between the EFP2-optimized methylated WC geometries compared to the corresponding RI-MP2¹ geometries. In EFP2-optimized mG···mC WC, the hydrogen bond R_1 in Figure 2 is 0.05 Å shorter than it is in RI-MP2 optimized¹ mG···mC WC, while R_3 is 0.22 Å longer; thus, the order of bond lengths changes from $R_1 > R_2 > R_3$ with RI-MP2 to $R_1 < R_2 < R_3$ with EFP2. In EFP2-optimized mA···mT WC, R_3 in Figure 4 is 0.33 Å shorter than in RI-MP2 optimized¹ mA···mT WC, while R_1 is 0.30 Å longer. This indicates that the methylated nucleotide bases optimized with EFP2 are slightly tilted relative to their orientation in the RI-MP2 optimized¹ configurations.

The EFP2 total interaction energies (Table 2) of the hydrogen-bonded nucleotide bases are generally in good agreement with the estimated CCSD(T) interaction energies.^{1,3} EFP2 underbinds the G···C WC 1 complex by 1.7 kcal/mol (~5% of the total binding energy) compared to the estimated CCSD(T) value.¹ When allowed to relax to the slightly (0.03–0.11 Å)

Table 2. Energies of Hydrogen-Bonded Complexes^a

	opt type	EFP2 Coulomb	EFP2 ex-rep	EFP2 pol	EFP2 disp	EFP2 total	CCSD(T)
G···C WC	MP2	−41.4	36.4	−14.7	−10.7	−30.4	−32.1
	EFP2	−37.6	28.1	−12.3	−9.4	−31.2	
mG···mC WC	MP2	−44.4	41.4	−16.6	−11.7	−31.3	−31.6
	EFP2	−39.4	30.6	−13.5	−10.0	−32.3	
A···T WC	MP2	−25.3	26.7	−7.2	−8.2	−14.0	−16.7*
	EFP2	−22.0	19.0	−5.4	−7.0	−15.3	
mA···mT WC	MP2	−28.8	29.8	−8.1	−9.1	−16.1	−18.2
	EFP2	−27.5	26.3	−6.9	−9.2	−17.3	

^a EFP2 energy components (Coulomb, exchange-repulsion, polarization, dispersion) and total interaction energy for the hydrogen-bonded nucleotide base complexes, shown in Figures 1–4. “Opt type” refers to the level of theory used for the geometry optimization, either RI-MP2 (ref 1) or EFP2. EFP2 total energies are compared with the estimated CCSD(T)/CBS interaction energies shown for complexes with RI-MP2 optimized geometries taken from ref 1 (Set 1). These CCSD(T)/CBS energies are from ref 1, except the A···T WC energy (*), which is from ref 3. Energies in kcal/mol.

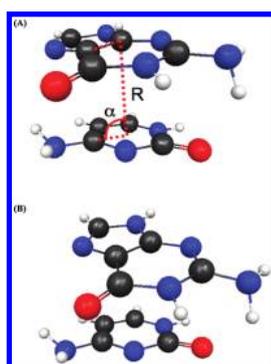


Figure 5. Guanine–cytosine stacked complex: (A) depicts the RI-MP2/cc-pVTZ optimized geometry of ref 1; (B) is the complex obtained by EFP2 optimization with the nucleobase geometries of ref 1. Intermonomer distance R , angle α (chosen to demonstrate the relative displacement of the nucleotide bases with respect to one another), and the dihedral angle between the planes of the nucleotide bases (dashed red line) are given in Table 3 for each complex.

more widely separated EFP2 optimized geometry, the EFP2 G···C WC total interaction energy differs from the estimated CCSD(T) value¹ by less than 1 kcal/mol. The EFP2 interaction energy of the RI-MP2 optimized¹ methylated guanine–cytosine complex differs from the estimated CCSD(T) value¹ by only 0.3 kcal/mol. The greatest discrepancy between the EFP2 and estimated CCSD(T)³ interaction energies is in the adenine–thymine complex. EFP2 under-binds the structure at both geometries examined, by 2.7 kcal/mol when constrained to the RI-MP2¹ geometry and by 1.4 kcal/mol when an EFP2 geometry optimization is used. EFP2 also under-binds the RI-MP2 optimized¹ methylated adenine–thymine structure compared to the estimated CCSD(T) value,¹ by 2.1 kcal/mol. Although this discrepancy decreases to less than 1 kcal/mol with the EFP2-optimized structure, the EFP2-optimized structure also differs the most from its corresponding RI-MP2¹ geometry, as discussed above.

For the RI-MP2 optimized¹ hydrogen-bonded structures, the root-mean-square deviation (rmsd) between the EFP2 and est. CCSD(T)¹ total interaction energies is 1.9 kcal/mol, with a maximum unsigned difference of 2.7 kcal/mol (corresponding to A···T WC). The rmsd between the EFP2 and est. CCSD(T)¹ total interaction energies when the hydrogen-bonded structures are optimized with EFP2 (whereas the est. CCSD(T)¹ values

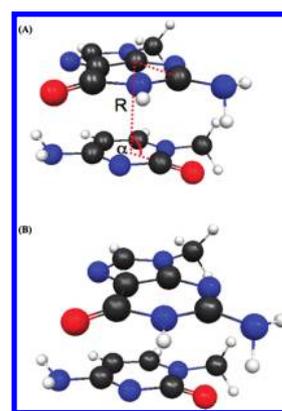


Figure 6. Methylated guanine–cytosine stacked complex: (A) depicts the RI-MP2/cc-pVTZ optimized geometry of ref 1; (B) is the complex obtained by EFP2 optimization with the nucleotide base geometries of ref 1. Intermonomer distance R , angle α (chosen to demonstrate the relative displacement of the nucleotide bases with respect to one another), and the dihedral angle between the planes of the nucleobases (dashed red line) are given in Table 3 for each complex.

correspond to the RI-MP2 optimized¹ structures) is 1.0 kcal/mol. The maximum unsigned difference, still corresponding to the A···T WC structure, is 1.4 kcal/mol.

The Coulomb interaction is the predominant attractive force in the hydrogen-bonded complexes (Table 2), although, especially in the case of adenine–thymine and their methylated analogues, the exchange-repulsion is similar in size and opposite in sign. The EFP2 exchange-repulsion term exceeds the Coulomb attraction in the RI-MP2¹ geometries of A···T WC and mA···mT WC. At the EFP2-optimized geometries, the EFP2 Coulomb magnitude exceeds the exchange-repulsion magnitude by only 3.0 kcal/mol in A···T WC and by 1.2 kcal/mol in mA···mT WC. While the Coulomb term is ~ 5 –10 kcal/mol larger than the exchange-repulsion in all G···C WC and mG···mC WC structures, the total EFP2 interaction energies are -30.4 kcal/mol (RI-MP2 optimized¹ G···C WC) to -32.3 kcal/mol (EFP2-optimized mG···mC WC). The Coulomb + exchange-repulsion accounts for only $\sim 10\%$ (RI-MP2 optimized¹ mG···mC WC) to 30% (EFP2-optimized G···C WC) of the total interaction energy. Thus, polarization and dispersion make significant contributions to the binding of the hydrogen-bonded nucleotide base pairs.

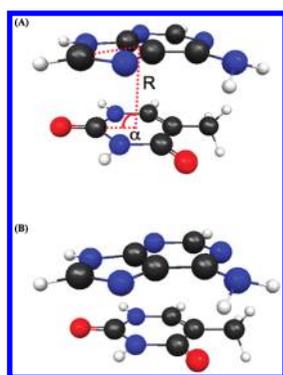


Figure 7. Adenine–thymine stacked complex: (A) depicts the RI-MP2/cc-pVTZ optimized geometry of ref 1; (B) is the complex obtained by EFP2 optimization with the nucleobase geometries of ref 1. Intermonomer distance R , angle α (chosen to demonstrate the relative displacement of the nucleotide bases with respect to one another), and the dihedral angle between the planes of the nucleotide bases (dashed red line) are given in Table 3 for each complex.

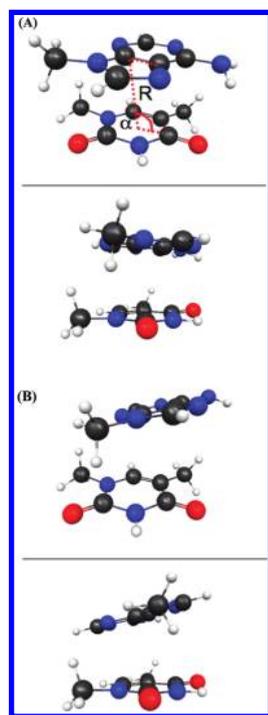


Figure 8. Methylated adenine–thymine stacked complex: (A) depicts the RI-MP2/cc-pVTZ optimized geometry of ref 1; (B) is the complex obtained by EFP2 optimization with the nucleotide base geometries of ref 1. Two views are shown of each complex to illustrate the relative orientations of the monomers. Intermonomer distance R , angle α (chosen to demonstrate the relative displacement of the nucleotide bases with respect to one another), and the dihedral angle between the planes of the nucleotide bases (dashed red line) are given in Table 3 for each complex.

Stacked Structures. EFP2 geometry optimizations produce stacked structures for guanine–cytosine (Figure 5B), methylated guanine–cytosine (Figure 6B), and adenine–thymine (Figure 7B) that are similar to their RI-MP2 optimized¹ counterparts (Figures 5A–7A). The EFP2-optimized structure for methylated adenine–thymine (Figure 8B) is, however,

Table 3. Geometries of Stacked Complexes^a

	opt type	R	α	dihedral
$G \cdots C$ S	MP2	3.47	85.5	10.3
	EFP2	3.54	95.6	4.0
$mG \cdots mC$ S	MP2	3.21	91.6	−11.4
	EFP2	3.47	78.1	−15.2
$A \cdots T$ S	MP2	3.18	87.6	29.8
	EFP2	3.53	83.2	27.8
$mA \cdots mT$ S	MP2	3.14	100.6	−35.8
	EFP2	3.32	91.1	30.7

^a Lengths (in Å) and angles (in degrees) of the hydrogen bonds in the Watson–Crick (WC) complexes of Figures 5–8. “Opt type” refers to the level of theory used for the geometry optimization, either RI-MP2 (ref 1) or EFP2. Angle α (defined in Figures 5A–8A) indicates the relative displacement of the nucleotide bases. The dihedral angle (also defined in Figures 5A–8A) goes between ring planes of the nucleotide bases and indicates how parallel these ring planes are.

significantly different from that obtained with RI-MP2¹ (Figure 8A); between these two structures, the dihedral angle between nucleotide base ring planes changes by 65°.

The EFP2-optimized stacked guanine–cytosine complex ($G \cdots C$ S) is similar to the RI-MP2 optimized¹ $G \cdots C$ S. The nucleotide bases in EFP2-optimized $G \cdots C$ S are just 0.07 Å more widely separated than in the RI-MP2 optimized¹ $G \cdots C$ S structure (Table 3), and they are oriented only slightly differently, with the dihedral angle between the ring planes (Figure 5A) differing by 6.3°. The greatest difference is their relative displacement, indicated by the angle α , which differs by ~10°. EFP2 interaction energies for RI-MP2 optimized¹ $G \cdots C$ S and EFP2-optimized $G \cdots C$ S differ by about 1 kcal/mol (Table 4). The EFP2 method overbinds the $G \cdots C$ S complex more than any other stacked complex except EFP2-optimized $mA \cdots mT$ S. The EFP2 interaction energy of RI-MP2 optimized¹ $G \cdots C$ S is 1.6 kcal/mol lower than the estimated CCSD(T)¹ energy¹ (Table 4).

In the EFP2-optimized methylated complex $mG \cdots mC$ S (Figure 6B), the guanine nucleotide base is displaced over the cytosine relative to its orientation in RI-MP2 optimized¹ $mG \cdots mC$ S (Figure 6A). The angle α , a measure of relative displacement of the nucleotide bases, is 91.6° in RI-MP2 optimized¹ $mG \cdots mC$ S and 78.1° in EFP2-optimized $mG \cdots mC$ S (Table 3). However, the nucleotide base ring planes are similarly oriented in these two structures, with dihedral angles differing by less than 4° (Table 3). While the EFP2 interaction energy of the RI-MP2 optimized¹ $mG \cdots mC$ S complex is 2.3 kcal/mol lower in magnitude than the estimated CCSD(T)¹ energy, the EFP2 interaction energy of the EFP2-optimized $mG \cdots mC$ S complex is just 0.3 kcal/mol higher than the estimated CCSD(T)¹ energy (Table 4).

The EFP2-optimized $A \cdots T$ S structure (Figure 7B) is the most similar among the stacked complexes to its RI-MP2¹ optimized counterpart (Figure 7A). The distance between nucleotide bases in EFP2-optimized $A \cdots T$ S is 0.35 Å greater than in RI-MP2 optimized¹ $A \cdots T$ S, though the relative displacement (given by α) and dihedral angle between nucleotide base planes differ by just 4.4 and 2° (Table 3). The EFP2 interaction energy of EFP2-optimized $A \cdots T$ S is only 0.3 kcal/mol larger in magnitude than the estimated CCSD(T)¹ interaction energy, while for $A \cdots T$ S optimized with RI-MP2,¹ it is 1.6 kcal/mol lower.

Table 4. Energies of Stacked Complexes^a

	opt type	EFP2 Coulomb	EFP2 ex-rep	EFP2 pol	EFP2 disp	EFP2 total	CCSD(T)
G···C S	MP2	−20.6	20.6	−3.1	−17.5	−20.6	−19.0
	EFP2	−22.1	22.2	−3.7	−18.1	−21.7	
mG···mC S	MP2	−21.9	30.3	−3.1	−23.4	−18.1	−20.4
	EFP2	−19.7	22.3	−3.4	−20.0	−20.7	
A···T S	MP2	−8.6	16.1	−0.7	−17.6	−10.7	−11.7*
	EFP2	−10.9	13.9	−0.6	−14.9	−12.6	
mA···mT S	MP2	−11.3	26.3	−1.3	−25.1	−11.4	−14.6
	EFP2	−15.9	18.3	−1.1	−18.2	−16.8	

^aEFP2 energy components (Coulomb, exchange-repulsion, polarization, dispersion) and total interaction energy for the stacked nucleotide base complexes shown in Figures 5–8. “Opt type” refers to the level of theory used for the geometry optimization, either RI-MP2 (ref 1) or EFP2. The estimated CCSD(T)/CBS interaction energies are shown for complexes with geometries taken from ref 1 (Set 1). These CCSD(T)/CBS energies are from ref 1, except the A···T WC energy (*), which is from ref 3. Energies in kcal/mol.

Among the stacked structures, as with the hydrogen-bonded structures, the greatest discrepancy between RI-MP2¹ and EFP2-optimized geometries occurs for the methylated adenine–thymine complex. The difference between the estimated CCSD(T)¹ and EFP2 interaction energies of RI-MP2 optimized¹ mA···mT S is the greatest of all of the stacked structures. Two views of mA···mT S are shown in Figure 8, where it is evident that the nucleotide bases in mA···mT S optimized with EFP2 (Figure 8B) are rotated relative to their orientation in mA···mT S optimized with RI-MP2¹ (Figure 8A). The dihedral angle between the nucleotide base planes differs greatly between the two structures; this angle is -35.8° in RI-MP2¹ optimized mA···mT S and is $+30.7^\circ$ in EFP2-optimized mA···mT S (Table 3). Compared to the estimated CCSD(T) interaction energies,¹ EFP2 underbinds the RI-MP2 optimized¹ mA···mT S complex by 3.2 kcal/mol. The EFP2-optimized mA···mT S structure is 2.2 kcal/mol more strongly bound compared to the estimated CCSD(T)¹ value.

While the hydrogen-bonded structures are more strongly bound than their respective stacked counterparts, the stacked guanine–cytosine complexes have a greater interaction energy than the hydrogen-bonded adenine–thymine complexes. The EFP2 interaction energy for the RI-MP2 optimized¹ G···C S structure is 6.6 kcal/mol more strongly bound than that of the RI-MP2 optimized¹ A···T WC structure (2.3 kcal/mol more strongly bound with estimated CCSD(T)^{1,3}) and 4.5 kcal/mol more strongly bound than that of RI-MP2 optimized¹ mA···mT WC (0.8 kcal/mol with estimated CCSD(T)¹; Tables 2 and 4). The RI-MP2 optimized¹ stacked methylated structure mG···mC S has an EFP2 interaction energy 4.1 kcal/mol greater (in magnitude) than RI-MP2 optimized¹ A···T WC (3.7 kcal/mol with estimated CCSD(T)^{1,3}) and 2.0 kcal/mol greater than RI-MP2 optimized¹ mA···mT WC (2.2 kcal/mol with estimated CCSD(T)¹) (Tables 2 and 4).

The rmsd between the EFP2 and est. CCSD(T)^{1,3} energies for the RI-MP2 geometry¹ is 2.2 kcal/mol. The maximum unsigned difference, corresponding to mA···mT S, is 3.2 kcal/mol. Comparing EFP2 interaction energies for the EFP2-optimized stacked structures with est. CCSD(T)¹ interaction energies for the RI-MP2 optimized¹ structures gives an rmsd of 1.8 kcal/mol and a maximum unsigned difference of 2.7 kcal/mol (for G···C S).

Dispersion is the single greatest attractive contribution to the EFP2 total interaction energy for the stacked adenine–thymine pairs, but not for guanine–cytosine (Table 4).

In mA···mT S optimized with RI-MP2,¹ the dispersion energy is -25.1 kcal/mol, the highest for any structure examined. This is more than twice the binding contribution of the Coulomb term (-11.3 kcal/mol). The magnitude of the EFP2 dispersion energy for RI-MP2 optimized¹ A···T S exceeds the magnitude of the EFP2 exchange-repulsion by 1.5 kcal/mol (Table 4). In the guanine–cytosine pairs, the Coulomb term predominates, although the dispersion term is also large (within 4 kcal/mol of the Coulomb contribution for both EFP2-optimized and RI-MP2 optimized¹ G···C S structures). The exchange-repulsion term equals or approximately equals the magnitude of the Coulomb attraction in both G···C S structures, so dispersion remains extremely important for binding in stacked guanine–cytosine pairs. Among the methylated stacked pairs, the magnitude of the dispersion term is greater than the magnitude of the Coulomb term in all four structures examined. The exchange-repulsion term in all four methylated stacked pairs is greater in magnitude than either the Coulomb or dispersion terms. So, in fact, all of the interaction types make significant contributions to the binding of the stacked pairs, except for the polarization, which remains small in all species.

■ EFFECTS OF METHYLATION

In general, methylation increases the magnitude of all energy components among both the hydrogen bonded and the stacked base pairs, regardless of whether the pairs are optimized with EFP2 or held fixed at the RI-MP2 optimized¹ geometries. The stacked G···C structure is a notable exception. The EFP2-optimized mG···mC S structure has an EFP2 Coulomb energy term 2.5 kcal/mol *lower* in magnitude than the nonmethylated analogue, while the EFP2 Coulomb energy term for the RI-MP2 optimized¹ mG···mC S follows the aforementioned trend; it is 1.3 kcal/mol higher in magnitude than that of the nonmethylated structure. Additionally, the EFP2 polarization term in the EFP2-optimized mG···mC S structure is 0.3 kcal/mol lower in magnitude than the nonmethylated structure; with the RI-MP2 optimized¹ structure, this difference is zero. Without exception, for all structures examined, methylation results in a more attractive EFP2 dispersion energy term. This increase in magnitude is 0.6–3.3 kcal/mol for EFP2-optimized structures and 0.9–7.5 kcal/mol for the RI-MP2 optimized¹ structures.

For both of the hydrogen-bonded pairs as well as for mA···mT S, the increase in the magnitudes of the attractive energy

Table 5. EFP2 and SAPT Interaction Energy Components for Adenine–Thymine Pairs^a

	Coulomb	exch-rep	pol	disp	total	CCSD(T)
A···T WC						−16.7
SAPT	−26.6	31.7	−11.9	−10.6	−17.4	
EFP2	−25.3	26.7	−7.2	−8.2	−14.0	
A···T S						−11.7
SAPT	−10.6	18.2	−2.5	−18.2	−13.1	
EFP2	−8.6	16.1	−0.7	−17.6	−10.7	

^aSAPT interaction energy values are from ref 27. Est. CCSD(T)/CBS energies are from ref 3. Energies in kcal/mol.

components due to methylation are somewhat greater, cumulatively, than the increase in exchange-repulsion. This results in methylated structures that are 0.7–2.1 kcal/mol (for RI-MP2 optimized¹ structures) or 1.0–4.2 kcal/mol (for EFP2-optimized structures) more strongly bound than their nonmethylated analogues. On the other hand, the EFP2 total interaction energies for both the RI-MP2 optimized and EFP2 optimized mG···mC S structures are lower than those for the nonmethylated G···C S. For EFP2-optimized mG···mC S, this energy lowering is due to the decrease in attractiveness of the EFP2 Coulomb energy term, which, as noted above, is an exception to the general trend (this term tends to become more attractive with methylation). For RI-MP2 optimized¹ mG···mC S, the EFP2 exchange-repulsion term grows by nearly 10 kcal/mol, offsetting the more modest (7.2 kcal/mol) increases in the magnitudes of the attractive energy components.

■ COMPARISON WITH SAPT

Interaction energy components found with density fitting SAPT²⁷ are available only for the stacked and hydrogen-bonded adenine–thymine base pairs (Table 5). EFP2 at RI-MP2 optimized¹ geometries underestimate the magnitude of all interaction energy components compared to SAPT,²⁷ but SAPT overestimates the total interaction energies of these base pairs compared to the CCSD(T) values.³ For the hydrogen-bonded complex, compared to the SAPT values,²⁷ EFP2 underestimates the magnitude of the Coulomb term by 1.3 kcal/mol, the polarization term by 4.7 kcal/mol, and the dispersion term by 2.4 kcal/mol, although EFP2 also predicts a 5.0 kcal/mol less repulsive exchange-repulsion term. The total EFP2 interaction energy is 3.4 kcal/mol less strongly bound than the total SAPT energy, or 2.7 kcal/mol less strongly bound than the est. CCSD(T)/CBS energy. For the stacked adenine–thymine complex, the EFP2 Coulomb, polarization, and dispersion terms are 2.0, 1.8, and 0.6 kcal/mol lower in magnitude than the SAPT values.²⁷ The EFP2 exchange-repulsion is 2.1 kcal/mol lower than that of SAPT.²⁷ The EFP2 total interaction energy is about 2.4 kcal/mol lower in energy than the SAPT total,²⁷ but only 1 kcal/mol lower in energy than the est. CCSD(T)/CBS total interaction energy.³

Regardless of the differences between the EFP2 and SAPT values, similar conclusions about the relative importance of each interaction energy component may be drawn from either set of data. As discussed above, the relatively large exchange-repulsion term exceeds the magnitude of the Coulomb term for both the stacked and the hydrogen-bonded adenine–thymine pairs; these complexes would not be bound without polarization

and dispersion. Compared to the polarization term, dispersion is of significantly greater importance in the stacked complex, whereas polarization is ~1 kcal/mol larger in magnitude than dispersion in the hydrogen-bonded complex.

■ CONCLUSIONS

The EFP2-predicted structures of the hydrogen-bonded and stacked guanine–cytosine, methylated guanine–cytosine, and adenine–thymine pairs are in good agreement with RI-MP2.¹ Greater discrepancy is found between EFP2 and RI-MP2 optimized geometries of methylated adenine–thymine complexes.

The EFP2 total interaction energies are in fair to excellent agreement with the corresponding estimated CCSD(T)¹ results. The root-mean-square deviation (rmsd) between the EFP2 and est. CCSD(T)¹ total interaction energies is 1.9 kcal/mol for the RI-MP2 optimized¹ hydrogen-bonded (Watson–Crick or WC) structures, with a maximum unsigned difference of 2.7 kcal/mol (corresponding to A···T WC). The rmsd between the EFP2 and est. CCSD(T)¹ total interaction energies when the hydrogen-bonded structures are relaxed to their EFP2-preferred geometries (while the est. CCSD(T)¹ values still correspond to the RI-MP2 optimized¹ structures) is 1.0 kcal/mol. The maximum unsigned difference is 1.4 kcal/mol (A···T WC). Among the stacked structures, the rmsd between EFP2 and est. CCSD(T)¹ energies for structures optimized with RI-MP2¹ is 2.2 kcal/mol. The maximum unsigned difference is 3.2 kcal/mol (mA···mT S). Comparing EFP2 interaction energies for the EFP2-optimized stacked structures with est. CCSD(T)¹ interaction energies for the RI-MP2 optimized¹ structures gives an rmsd of 1.8 kcal/mol and a maximum unsigned difference of 2.7 kcal/mol (G···C S).

An accurate description of the dispersion energy is essential to determine the binding energies of the nucleotide base pairs, even for the hydrogen-bonded structures. While the Coulomb interaction is the predominant attractive energy term for hydrogen-bonded structures, its magnitude is only ~1–3 kcal/mol larger than the exchange-repulsion for EFP2-optimized A···T WC and EFP2-optimized mA···mT WC. The EFP2 exchange-repulsion term slightly (~1 kcal/mol) exceeds the magnitude of the EFP2 Coulomb term in the RI-MP2 optimized¹ A···T WC and mA···mT WC structures. In the guanine–cytosine and methylated guanine–cytosine hydrogen-bonded structures, the magnitude of the Coulomb term is ~9 kcal/mol larger than the exchange-repulsion term among the EFP2-optimized structures; however, this ~9 kcal/mol accounts for only ~30% of the total binding energy, the remainder being comprised of the dispersion and polarization energies. Thus, a computational method that fails to describe the dispersion and polarization terms accurately will substantially underestimate the binding energies of these complexes, including those of the Coulomb-dominated hydrogen bonded pairs. That the EFP2 method captures these interactions accurately and with a low computational cost demonstrates its utility in modeling DNA base pairs and similar biologically important systems.

Due to its low computational cost, especially compared to high-level ab initio methods such as MP2, SAPT, and CCSD(T), yet reasonable agreement with these methods, EFP2 may be a useful tool for modeling larger complexes of nucleotide bases, including strands of DNA.

■ ASSOCIATED CONTENT

S Supporting Information. Cartesian coordinates of paired nucleotide base structures found with EFP2. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ ACKNOWLEDGMENT

This work was supported by a grant from the Air Force Office of Scientific Research. L.V.S. acknowledges support by NSF Career (Grant CHE-0955419), ACS PRF (Grant 49271-DNI6), and Purdue University. The authors gratefully acknowledge Dr. Lori Burns, Ed Hohenstein, and Professor David Sherrill for providing SAPT data and for helpful discussions.

■ REFERENCES

- (1) Jurecka, P.; Spöner, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.
- (2) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.
- (3) Takatani, T.; Hohenstein, E. G.; Malagoli, M.; Marshall, M. S.; Sherrill, C. D. *J. Chem. Phys.* **2010**, *132*, 144104.
- (4) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevents, W. J. *J. Phys. Chem. A* **2001**, *105*, 293. Jensen, J. H.; Gordon, M. S. *Mol. Phys.* **1996**, *89*, 1313.
- (5) Slipchenko, L. V.; Gordon, M. S. *J. Comput. Chem.* **2007**, *28*, 276.
- (6) Gordon, M. S.; Slipchenko, L. V.; Li, H.; Jensen, J. H. *Annu. Rep. Comp. Chem.* **2007**, *3*, 177.
- (7) Smith, T.; Slipchenko, L. V.; Gordon, M. S. *J. Phys. Chem. A* **2008**, *112*, 5286.
- (8) Smith, Q. A.; Slipchenko, L. V.; Gordon, M. S. *J. Phys. Chem. A* **2011**, *115*, 4598.
- (9) Ghosh, D.; Kosenkov, D.; Vanovschi, V.; Williams, C. F.; Herbert, J. M.; Gordon, M. S.; Schmidt, M. W.; Slipchenko, L. V.; Krylov, A. I. *J. Phys. Chem. A* **2010**, *114*, 12739.
- (10) Langner, K. M.; Sokalski, W. A.; Leszczynski, J. *J. Chem. Phys.* **2007**, *127*, 111102.
- (11) Möller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
- (12) Cybulski, S. M.; Lytle, M. L. *J. Chem. Phys.* **2007**, *127*, 141102.
- (13) Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608.
- (14) Johnson, E. R.; Becke, A. D. *J. Chem. Phys.* **2006**, *124*, 174104.
- (15) Hesselmann, A.; Jansen, G.; Schütz, M. *J. Am. Chem. Soc.* **2006**, *128*, 11730.
- (16) Sedláč, R.; Jurečka, P.; Hobza, P. *J. Chem. Phys.* **2007**, *127*, 075104.
- (17) Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.
- (18) Stone, A. J. *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, 1996.
- (19) Stone, A. J. *J. Chem. Theory Comput.* **2005**, *1*, 1128.
- (20) Slipchenko, L. V.; Gordon, M. S. *Mol. Phys.* **2009**, *107*, 999.
- (21) Jensen, J. H.; Gordon, M. S. *Mol. Phys.* **1996**, *89*, 1313.
- (22) Adamovic, I.; Gordon, M. S. *Mol. Phys.* **2005**, *103*, 379. Amos, R. D.; Handy, N. C.; Knowles, P. J.; Rice, J. E.; Stone, A. J. *J. Phys. Chem.* **1985**, *89*, 2186.
- (23) Li, H.; Gordon, M. S.; Jensen, J. H. *J. Chem. Phys.* **2006**, *124*, 214108.
- (24) EFP2 geometry optimizations failed to converge on planar hydrogen bonded structures when the numerical DMA was used.
- (25) Boys, S. F.; Bernardi, F. *Mol. Phys.* **2002**, *100*, 65.
- (26) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Olsen, J. *Chem. Phys. Lett.* **1999**, *302*, 437. Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243.
- (27) Hohenstein, E. G.; Sherrill, C. D. *J. Chem. Phys.* **2010**, *133*, 014101.