Further Quantum Mechanical Evidence that Difluorotoluene Does Not Hydrogen Bond

Edward C. Sherer,[†] Sam J. Bono,[‡] and George C. Shields*[,‡]

Department of Chemistry, Hamilton College, 198 College Hill Rd., Clinton, New York 13323, and Department of Chemistry, University of Minnesota, 207 Pleasant St. SE, Minneapolis, Minnesota 55455

Received: January 17, 2001; In Final Form: April 12, 2001

Calculations were run on the methylated DNA base pairs adenine:thymine and adenine:difluorotoluene to further investigate the hydrogen-bonding properties of difluorotoluene (F). Geometries were optimized using hybrid density functional theory. Single-point calculations at the MP2(full) level were performed to obtain more rigorous energies. The functional counterpoise method was used to correct for the basis set superposition error (BSSE), and the interaction energies were also corrected for fragment relaxation. These corrections brought the B3LYP and MP2 interaction energies into excellent agreement. In the gas phase, the Gibbs free energies calculated at the B3LYP and MP2 levels of theory predict that A and T will spontaneously form an A:T pair while A:F spontaneously dissociates into A and F. Solvation effects on the pairing of the bases were explored using implicit solvent models for water and chloroform. In aqueous solution, both A:T and A:F are predicted to dissociate into their component monomers. Semiempirical calculations were performed on small sections of B-form DNA containing the two pairs, and the results provide support for the concept that base stacking is more important than hydrogen bonding for the stability of the A:F pair within a DNA helix.

1. Introduction

Specificity in the replication of DNA has often been attributed to Watson−Crick[1] hydrogen bonding between complementary nucleic acid bases. There exists a debate, however, over the relative importance of shape complementarity versus hydrogen-bonding interactions in the fidelity of DNA polymerases (refs 2−7 and references therein).

The hypothesis that shape mimicry is more important has been investigated by Kool and co-workers through an elegant set of experiments based on the exploration of replication fidelity using nonpolar homologues of the four canonical nucleotides.[8−16] Difluorotoluene (F), a nonpolar shape mimic of thymine (T), was designed to test shape complementarity.[10] Kool and co-workers solved the crystal structure of the F nucleoside, and no intermolecular hydrogen bonds were observed between the functional groups of F.[10] Instead, the bases were stacked together in hydrophobic regions, while the sugar rings were bound together by hydrogen bonds. In addition, pairing studies in chloroform did not show evidence for hydrogen bonding between adenine (A) and F.[9,17] Difluorotoluene is a good probe of hydrogen-bonding interactions in DNA replication since unwanted steric hindrance issues do not arise.[7] Difluorotoluene is more hydrophobic than thymine, which may influence its ability to stack better than thymine,[11] and stacking is one of the main factors in DNA helix stability. The roles of shape complementarity and hydrogen bonding as the determining factors in DNA replication have been studied using other nonpolar shape analogues.[10,14,18−22]

Replication studies using the Klenow fragment of DNA polymerase I show that F is successfully incorporated across from adenine, and vice versa, in a precise fashion.[8,11] Difluorotoluene has been incorporated into a DNA dodecamer as the complementary base to adenine, and melting temperature studies determined that the decrease in stability of a T to F exchange is 3.6 kcal/mol.[8] The free energy change upon substitution of T with F within a DNA helix was calculated to be 5.1 kcal/mol using free-energy perturbation methods.[23]

The difluorotoluene-based studies suggest that shape mimicry is a much stronger determinant of replication fidelity than hydrogen bonding. Difluorotoluene, however, does not function exactly as thymine. When two adenines exist adjacent to each other in the template strand, the polymerase stalls when trying to add two difluorotoluenes.[15] In addition, incorporation of F across from itself has nearly the same efficiency as formation of an A:F base pair.[8] Difluorotoluene is edited (proofread) by the 3′ exonuclease domain of the Klenow fragment of *E. coli* DNA polymerase I.[24]

A solution structure of a DNA duplex containing F has been determined using NMR, and the duplex was found to adopt a B-like conformation.[13] Molecular dynamics simulations have shown that A:F pairs exhibit higher base pair opening rates than canonical A:T pairs.[25,26] The decreased strength of the hydrogen bonds in the A:F pair was evident from these long time scale simulations. The base pair opening of A:F was reversible, though it was not determined whether the driving force for this reversibility was hydrogen bonding or base stacking. If difluorotoluene has any ability to hydrogen bond, then studies performed with this base analogue are inconclusive as to whether shape is more important than hydrogen bonding.

The strength of hydrogen bonding (interaction energy) between nucleic acid bases can be calculated using theoretical methods. A large body of work covering the calculation of interaction energies of nucleic acid base pairs exists, and we refer the reader to a recent review of nucleic acid calculations.[27] Previous theoretical calculations performed on the A:F base pair (Figure 1) have shown that hydrogen bonding between A and F is not as strong as between A and T and have provided further support for the theory that shape mimicry is likely very important for the high fidelity seen in DNA replication.[18,28−30]

* To whom correspondence should be sent: gshields@hamilton.edu.
[†] University of Minnesota.
[‡] Hamilton College.

10.1021/jp010197x CCC: $20.00
Published on Web 08/10/2001

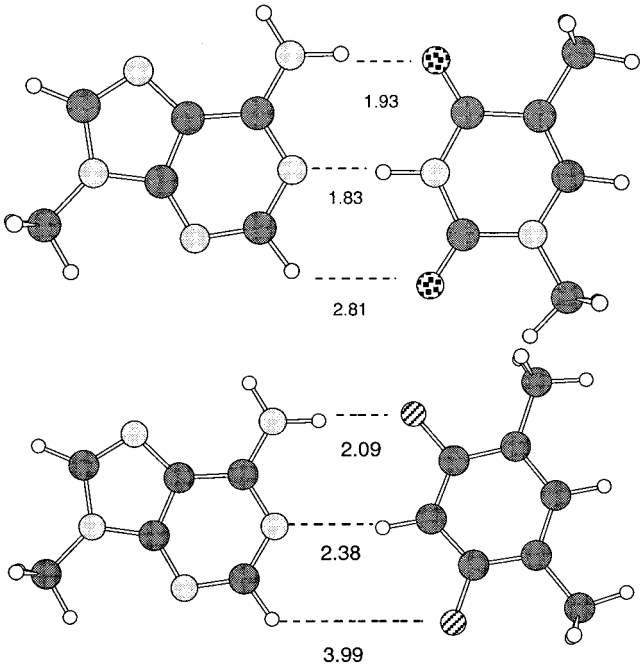


Figure 1. Gas-phase optimized structures of the methylated A:T (top) and A:F base pairs at the B3LYP/6-31G* level from this study. Bond distances are in angstroms. Structures are near planar. Atom spheres are black for carbon, white for hydrogen, gray for nitrogen, squared for oxygen, and lined for fluorine.

Since the A:F pair remains paired after geometry optimization at high levels of theory, and since the calculated interaction electronic energies and enthalpies are indicative of weak intermolecular forces, it cannot be ruled out that F hydrogen bonds, albeit weakly.

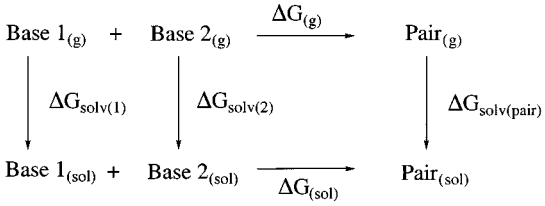
Most of the pairing interaction energies reported to date for the relative differences between A:T and A:F pairs have been reported in electronic energy or enthalpy.[18,28−31] The effect of adding in entropy to arrive at free energies has also been investigated.[23] Early calculations on A:F base pairing concluded that aqueous solvation would not prevent hydrogen bonds from forming between F and A.[31] In water, the canonical Watson−Crick base pair A:T does not form, rather the bases stack (ref 32 and references therein). In the gas phase and in nonpolar solvents, A:T hydrogen-bonded base pairs are believed to be the stable minima, and this point will be discussed further.

It has been thoroughly shown that the interaction energy between an A:F pair is decreased compared to an A:T pair. However, quantum mechanical calculations referenced above predict the A:F pair to maintain a roughly Watson−Crick geometry. We show here that this geometry does not form spontaneously from a free energy perspective, while the A:T pair does. Furthermore, the calculated results predict that both pairs will dissociate in water. Solvation in chloroform decreases the magnitude of the interaction energies for both pairs (in relation to gas-phase interaction energies), and comparison is made to experimental pairing energies when possible. Semiempirical geometry optimizations of a small section of B-form DNA pulled the A:F pair apart lending further support to its inability to form hydrogen bonds. The impact of these new insights on the experimental behavior of difluorotoluene is discussed.

2. Methods

The structures of the different DNA structures were built using Spartan 5.0.[33] Full geometry optimizations were performed on methylated A, T, F, A:F, and A:T species at the B3LYP level of theory using a 6-31G(d) basis set[34−40] as implemented in the Gaussian 98 program.[41] Single-point calculations were conducted at the B3LYP level using 6-311+G(2d,p) and 6-311+G(2df,2p) basis sets, and with correlation of core and valence electrons through second-order Møller−Plesset perturbation theory (MP2(full)) using the 6-311G(d,p) and 6-311++G-(d,p) basis sets.[34] The nature of each stationary point was confirmed with a frequency calculation.

SCHEME 1



The initial interaction energy was calculated by subtracting the electronic energy of the two monomers from the dimer (visualized in the free energy cycle found in Scheme 1), using eq 1. This ΔE must be converted to ΔG to determine whether pairing will occur spontaneously. The thermodynamic information and zero-point vibrational energy (ZPE) determined at the B3LYP/6-31G(d) level (scaled by 0.96)[42] were used to calculate enthalpies and free energies at 298 K (see eqs 2−4) using standard ideal-gas statistical mechanics and the rigid-rotor harmonic oscillator approximation.[34] After calculating the base pair interaction free energy, $\Delta G_{(g)}$ was converted from a reference state of 1 atm in the gas phase to a 1 M reference state at 298 K.

Since a limited basis set was used in the calculations, the electronic interaction energies may be overestimated due to BSSE. The BSSE corrections were calculated using eq 5, which is the functional counterpoise estimate (fCP).[43] In addition, the interaction energies were corrected for fragment relaxation (FR) using eqs 6−8, as suggested by Xantheas.[43] The notation $E_{AB}^{\alpha\cup\beta}(A)$ indicates that the energy of the structure in parentheses (A) from the optimized geometry of the subscript species AB was calculated using a basis set indicated by the superscript $\alpha \cup \beta$, where α and β are the basis functions of the independent monomers.[43] The fragment relaxation energies calculated with eqs 7 and 8 are the energy penalties induced by distorting the isolated monomer geometries to those found within the dimer. Further discussion of BSSE in regards to DFT and correlated levels of theory can be found in refs 44 and 45.

$$E = E_{AB}^{\alpha\cup\beta}(AB) - E_{A}^{\alpha}(A) - E_{B}^{\beta}(B) \quad (1)$$

$$H^\circ = E + H_{vib} + H_{rot} + H_{trans} \quad (H_{vib} \text{ includes ZPE}) \quad (2)$$

$$S^\circ = S_{vib} + S_{rot} + S_{trans} + S_{el} \quad (3)$$

$$G^\circ = H^\circ - TS^\circ \quad (4)$$

$$E(\text{fCP}) = E_{AB}^{\alpha\cup\beta}(AB) - E_{AB}^{\alpha\cup\beta}(A) - E_{AB}^{\alpha\cup\beta}(B) \quad (5)$$

$$E(\text{FR}) = E_{AB}^{\alpha\cup\beta}(AB) - E_{AB}^{\alpha\cup\beta}(A) - E_{AB}^{\alpha\cup\beta}(B) + E_{rel}^{\alpha}(A) + E_{rel}^{\beta}(B) \quad (6)$$

$$E_{rel}^{\alpha}(A) = E_{AB}^{\alpha}(A) - E_{A}^{\alpha}(A) \quad (7)$$

$$E_{rel}^{\beta}(B) = E_{AB}^{\beta}(B) - E_{B}^{\beta}(B) \quad (8)$$

$$\Delta\Delta G_{solv} = \Delta G_{solv(pair)} - \Delta G_{solv(1)} - \Delta G_{solv(2)} \quad (9)$$

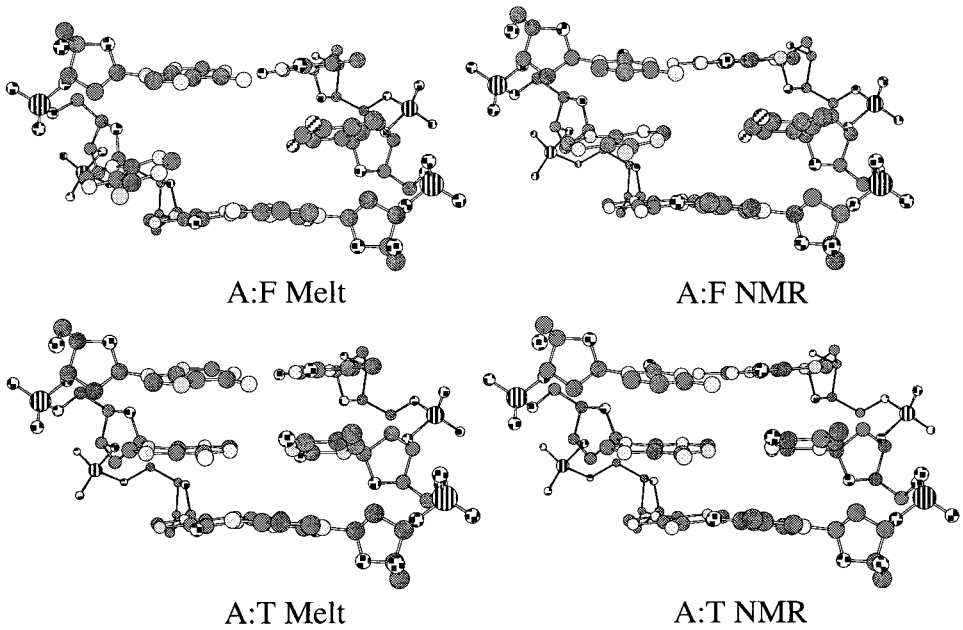


Figure 2. PM3-optimized structures of the Melt (T−T−G) and NMR (C−T−T) helix sections. Difluorotoluene, when present, replaces the central thymine. Hydrogen atoms have been removed for clarity and the view is into the minor groove. Atom representations are identical to Figure 1, with phosphorus represented by vertical lines.

Solvation values were obtained using AMSOL.[46] This program is capable of determining free energies of solvation of a solute molecule in numerous solvents with highly accurate results, including solvation free energies and water/chloroform partition coefficients for nucleic acid bases.[47−49] Solvation calculations in water and chloroform were performed using SM5.4/PM3. The geometries calculated using density functional theory were either restrained (SM5.4/PM3//B3LYP/6-31G*) or allowed to optimize at the PM3 level (SM5.4/PM3) to calculate the solvation free energies.[50] Addition of the differential solvation free energy ($\Delta\Delta G_{solv}$) calculated using eq 9, to the free energy of interaction in the gas phase allowed for the calculation of the free energy of solvation in a given solvent, $\Delta G_{(sol)}$ (reference state of 1 M, 298 K).

Investigation of the helical contribution to the stability of the various base pairs was done using sections of B-form DNA. Two different sequences shown in Figure 2 were built by flanking the central A:F or A:T base pair by two other base pairs. The two investigated sequences were the flanking base pairs that exist in the dodecamer used in the experimental melting temperature study (labeled melt) and the NMR structural study (labeled NMR) of the A:F pair.[8,13] Each helix section contained three DNA base pairs (189 atoms) with a total charge of −4. Since the phosphates were not neutralized, part of the helical backbone was restrained for gas-phase optimizations to prevent unwinding. For the restrained optimizations, the atoms comprising O3′ to O5′ of each central nucleotide unit were allowed to optimize, while the base pairs above and below the central pairs were frozen in position. Each of the central base pairs (starting from a canonical Watson−Crick geometry) were optimized at the PM3 level.[50] Extension of this type of calculation to release restraints by either moving to correlated levels of theory (to account for base stacking) or including solvation (even in a continuum approach) are computationally demanding.

All of the above methodology has been recently applied, with good success, to investigate the base pairing properties of 2-aminopurine with cytosine.[51]

TABLE 1: Gas-Phase Interaction Energies (kcal/mol)

	I[a]	II[b]	III[c]	IV[d]	V[e]	VI[f]
			A:T (Methylated)			
ΔE	−11.6	−16.2	−12.0	−12.2	−17.0	−15.9
ΔE(fCP)[g]			−12.9	−13.0	−13.5	−13.8
ΔE(FR)[h]			−11.6	−11.7	−12.0	−12.4
ΔH[i]			−10.1	−10.2	−10.6	−11.0
ΔG(1 atm)[i]			0.8	0.7	0.3	−0.2
ΔG(1 M, 298 K)[i]			−1.1	−1.2	−1.6	−2.0
			A:F (Methylated)			
ΔE	−3.9	−5.9	−3.1	−3.2	−6.9	−6.7
ΔE(fCP)			−3.0	−2.9	−4.2	−4.4
ΔE(FR)			−2.8	−2.7	−4.0	−4.3
ΔH			−1.4	−1.3	−2.6	−2.9
ΔG(1 atm)			7.0	7.1	5.8	5.5
ΔG(1 M,298 K)			5.1	5.2	3.9	3.7

[a] I, HF/6-31G*. [b] II, B3LYP/6-31G*. [c] III, B3LYP/6-311+G(2d,p)//B3LYP/6-31G*. [d] IV, B3LYP/6-311+G(2df,2p)//B3LYP/6-31G*. [e] V, MP2(full)/6-311G(d,p)//B3LYP/6-31G*. [f] VI, MP2(full)/6-311++G(d,p)//B3LYP/6-31G*. [g] fCP, functional counterpoise correction for BSSE. [h] FR, energies corrected for fragment relaxation energy + fCP. [i] Corrected for fCP and FR.

3. Results and Discussion

Interaction Energies in the Gas Phase. Interaction energies are calculated, in the most reduced case, by taking differences of electronic energies. Two methods for calculating a more rigorous interaction energy were undertaken. The fCP method provides a correction for BSSE that does not incorporate fragment relaxation energies. Upon examination of Table 1, correction for BSSE using the fCP method (eq 5) decreases ΔE by 0.9 and 0.8 kcal/mol for the B3LYP/6-311+G(2d,p) and B3LYP/6-311+G(2df,2p) calculations of methylated A:T and increases ΔE by 3.5 and 2.1 kcal/mol for the MP2(full)/6-311G-(d,p) and MP2/6-311++G(d,p) calculations, respectively. Incorporation of fragment relaxation (eqs 6−8), which is a measure of the penalty to distort the isolated A and T monomer geometries to those found within the dimer, increases ΔE by 0.7 kcal/mol for B3LYP and 1.5 kcal/mol for MP2. For the A:F system, fragment relaxation makes a negligible contribution,

and the magnitude of the BSSE is lower as well. The BSSE corrections for the MP2 calculations are not trivial. An interesting result of the BSSE and fragment relaxation corrections is that they bring the DFT and MP2 interaction energies for formation of A:T into better agreement with each other. A spread of 5.0 kcal/mol in the interaction electronic energies drops to 0.9 kcal/mol after correcting for the BSSE and fragment relaxation. In addition, the MP2 energies are affected more by the size of the basis set than the B3LYP energies, as has been shown in previous theoretical studies.[52]

Correction for BSSE using the fCP method should make ΔE more positive, as using a larger basis set to calculate the energy of the monomer in the dimer geometry lowers the energy for rigid subsystems. Correction for fragment relaxation then accounts for the inaccuracy that results from assuming that the geometry of the monomer in the dimer structure is unperturbed from its isolated geometry, and this correction makes ΔE more positive (eqs 4−8). All of the fragment relaxation corrections are positive in Table 1, as are most of the fCP corrections, with the exception of the fCP corrections for the A:T system calculated with B3LYP. Since DFT methods are relatively insensitive to the basis set size, the net effect of BSSE and fragment relaxation corrections on the B3LYP calculations yields corrected interaction electronic energies that are within half a kcal/mol of the uncorrected values. The A:T geometry has more fragment distortion from the individual monomers than the A:F geometry. The magnitude of BSSE corrections are dependent upon the intermolecular forces whose magnitude varies with intermolecular separation. In the A:F pair, the hydrogen bonds are weaker, and in turn, the base pairs are further apart in space.

The gas-phase enthalpies listed in Table 1 show that when A and T pair, 10.1−11.0 kcal/mol is released. These energies are in good agreement with benchmark calculations performed by Hobza and Sponer.[52] When difluorotoluene and adenine pair, 1.3−2.9 kcal/mol is released. The decreased interaction energy for A:F relative to A:T has been reported previously.[18,28−30] Experimentally, the gas-phase interaction enthalpy of formation of an A:T pair is −12.1 kcal/mol.[53,54] Our MP2 numbers, corrected for BSSE including fragment relaxation, reproduce this value best, an observation consistent with previous high-level calculations.[27] Our values for ΔH are −10.2 and −11.0 kcal/mol for the B3LYP/6-311+G(2df,2p) and MP2/6-311++G-(d,p) single-point calculations. The underestimation of the interaction energy caused by using basis sets that do not reach the basis set limit has been approximated to be between 2.0 and 2.5 kcal/mol.[52] The addition of this "correction" term brings the calculated A:T interaction energy into fair agreement with the experimental value of −12.1 kcal/mol.

When the enthalpies are converted to free energies at a standard state of 1 M, 298 K, the entropic effects are significant. The addition of the $T\Delta S$ terms to the interaction energies will make the process of pair formation less favorable.[55] Prior calculations of the magnitude of the entropy correction for nucleic acid base pairs have shown that the magnitude of the entropy term differs by less than 40% for all the base pairs studied.[55] Specifically, Hobza and Sponer calculated $T\Delta S$ for A:T (Watson−Crick) to be −11.1 kcal/mol. The highest level calculations used in this study provide a similar value of −10.8 kcal/mol. When the entropy term is calculated for A:F, it would be expected to be slightly decreased since the pair is more loosely bound, and the calculated value drops to −8.4 kcal/mol. Florián et al. have estimated the entropy correction for the A:T pair to be −4.4 kcal/mol using a linear relationship between experimental base stacking enthalpies and entropies, which, as stated by the authors, may not be directly applicable to hydrogen-bonded pairs.[56]

TABLE 2: Differential Solvation Free Energies, $\Delta\Delta G_{solv}$ (kcal/mol)

	water[a]	chloroform[a]	water[b]	chloroform[b]
A:T methylated	10.5	5.2	10.7	5.5
A:F methylated	3.0	0.3	4.6	1.6

[a] SM5.4/PM3. [b] SM5.4/PM3//B3LYP/6-31G*.

TABLE 3: Solution-Phase Interaction Energies (kcal/mol)[a]

	III[b]		IV		V		VI	
	ΔH	ΔG[c]	ΔH	ΔG	ΔH	ΔG	ΔH	ΔG
A:T Methylated								
water	−1.3	9.6	−1.4	9.5	−1.8	9.1	−2.2	8.7
chloroform	−6.5	4.4	−6.6	4.3	−7.0	3.9	−7.4	3.5
A:F Methylated								
water	1.3	9.7	1.4	9.8	0.1	8.5	−0.2	8.3
chloroform	−1.7	6.7	−1.6	6.8	−2.9	5.5	−3.2	5.3

[a] Calculated using SM5.4/PM3//B3LYP/6-31G* values. [b] Levels as indicated in Table 1. [c] 1 M, 298 K.

The addition of the $T\Delta S$ terms to the interaction enthalpies to arrive at free energies predicts that A and T will spontaneously form a dimer in the gas phase ($\Delta G = -1.1$ to -2.0 kcal/mol), but A and F will not ($\Delta G = 5.1-3.7$ kcal/mol). Reducing the enthalpies by an additional 2.0−2.5 kcal/mol to account for the medium-sized basis sets employed here will still not be enough to make the formation of A:F in the gas phase a spontaneous event. Thus we conclude that difluorotoluene and adenine will not spontaneously form a hydrogen-bonded complex in the gas phase.

Interaction Energies in Solution. Conflicting reports on the comparative effect of solvation for the A:F and A:T pairs range from solvent having no effect[31] to solvent disfavoring only A:F formation,[29] to solvent disfavoring both A:F and A:T pair formations.[30] There is no question that solvent will have an effect on base pairing, and it is proposed to play a large role in the determination of replication fidelity.[7,18]

The dielectric constants in the gas and aqueous phases are 1.0 and 78.3, respectively. The corresponding variation in electrostatic fields has a pronounced effect on the energetics of pair formation. Pairing of two monomers in solution partially involves desolvation of the monomers, formation of the dimer, solvation of the newly formed dimer, and solvent/solute reorganization. A measure of the favorability of this process is found in the differential free energy of solvation ($\Delta\Delta G_{solv}$). Two sets of solvation energies are provided in Table 2. The first and second columns contain structures fully optimized with PM3 using a continuum dielectric. The third and fourth columns contain solvation single-point calculations (SPC's) on DFT geometries. In Table 2, the penalty to form an A:F pair is much lower than the solvation penalty to form an A:T pair in water. There is little energetic cost to desolvate the nonpolar base F. In the less polar solvent chloroform (dielectric of 4.71), the favorability of forming both the A:F and A:T pairs increases (though remains unfavorable for both) compared to their formation in water.

Combining the ΔG values for formation of the base pairs in the gas phase (1 M, 298 K) with the differential free energies of solvation yields the overall $\Delta G_{(sol)}$ value for the thermodynamic cycle diagrammed in Scheme 1. These values for $\Delta G_{(sol)}$ are listed in Table 3, where we have used the four highest level

gas-phase SPC's from Table 1 (III−VI) and the solvation SPC's on the B3LYP geometries reported in Table 2. For A:T, we calculate a differential free energy of solvation of 10.5−10.7 kcal/mol. Using a Langevin dipoles solvation model, Florián et al. have estimated this differential solvation energy to be less costly at only 4.0 kcal/mol.[56] We turn now to an examination of the differential solution interaction energy: $\Delta\Delta G_{(sol)} = \Delta G_{(sol)}$ A:T $- \Delta G_{(sol)}$ A:F, starting with water as the solvent.

Using the SCI-PCM model in conjunction with RHF/6-31G* geometries, Wang and Houk found that pairing of A:T and A:F in water was disfavored by 5.2 and 1.4 kcal/mol, respectively,[30] a differential free energy of 3.8 kcal/mol. In contrast to this finding, we see that an aqueous environment nearly cancels out any energetic advantage of the A:T pair over the A:F pair. The differential interaction free energy in water (1 M, 298 K) is at most 0.6 kcal/mol (see Table 3), while the average signed difference of methods III−VI is −0.2 kcal/mol, a difference that is probably too small to be meaningful.

Formation of both pairs in water was calculated to be unfavorable. This is in contrast to Florián et al. calculating the free energy of forming an A:T hydrogen-bonded base pair to be slightly favorable (−0.8 kcal/mol) at 1 atm and 298 K.[56] This is in part due to the decreased entropic contribution and differential free energy of solvation values calculated in that study.

In contrast to aqueous solution, formation of A:T is clearly favored over A:F in the gas phase, as the average $\Delta\Delta G_{(g)}$ is 6.0 kcal/mol (columns III−VI, Table 1).

Nonpolar Solvent/Chloroform. Since the relative pairing advantage of A:T over A:F is dependent upon the dielectric constant, it is interesting to speculate which calculation best represents the effects of the helix environment. Beveridge et al. have summarized recent work on the calculation of the dielectric field of DNA and other macromolecules.[57] The dielectric constant of DNA has been estimated to be between 30 and 60, depending upon which subsection of the DNA is of interest, and the water around DNA had a decreased dielectric constant of 41.3, as compared to bulk water. Using AMBER, CHARMM, GROMOS, and OPLS, Arora and Jayaram found that no single dielectric constant reproduced the electrostatic environment of DNA well.[58] Fitting a function to various dielectric constants, they find that approximately 2.0 kcal/mol will be released for each hydrogen bond formed between base pairs in the helix environment. Since the dielectric of DNA is decreased compared to water but is larger than the gas-phase constant, it is thought that a nonpolar solvent might reproduce the dielectric of DNA well.[59−61]

The preference for stacked pairs versus hydrogen-bonded pairs is dependent upon the dielectric of the medium. Experimental studies show that in the gas phase, bases hydrogen bond and preferentially form A:T and G:C pairs.[54] When the N9 position of adenine and the N1 position of thymine are unmethylated, the Watson−Crick conformation is not the global minimum in the gas phase.[62] However, as stated in that report, the global minimum would not exist in the helix environment, where the Watson−Crick geometry is more populated.

In water, bases are observed to stack rather than form hydrogen bonds.[63] However, Florián et al. calculate the energetic preference of stacked pairs versus hydrogen-bonded dimers to be rather low in aqueous solution, where all the calculated interaction energies for many base pairs fell into a small range between +0.3 and −1.9 kcal/mol.[56] When the bases are solvated in chloroform, the preference of G to pair with C, and A with T, is still observed, strongly supporting hydrogen-bonded pairs in chloroform.[59,60,64,65] In chloroform, the geometry of the adenine:uracil (A:U) pair is sometimes difficult to assign as strictly hydrogen bonded or stacked.[61,66] Investigating the conformational preference of an adenine dinucleotide in several solvents by molecular dynamics showed that the stacking preferences in water were lost to a perpendicular configuration in chloroform.[67] Recent investigations of the pairing ability of A:F in chloroform show no evidence for pair formation, while A:T remains paired.[9,17]

TABLE 4: Hydrogen Bond Lengths from the PM3 Helix Optimizations (Å)

	H2- - -O(F)2[a]	N1- - -H3	H61- - -O(F)4
A:T NMR[b]	1.86	1.88	3.29
A:T melt	2.26	1.85	2.88
A:F NMR	2.79		5.99
A:F melt	5.28		9.24

[a] Atom label on the left applies to A and atom label on right to the pyrimidine. [b] Base pair is propeller twisted.

Assuming a correctly formed Watson−Crick pair, the $\Delta H°$ and $\Delta G°$ for formation of a 9-ethyladenine:1-cyclohexyluracil base pair in chloroform was experimentally determined to be −6.2 and −2.7 kcal/mol, respectively.[61] We have calculated the interaction free energy of methylated A:T to be unfavorable by roughly 4 kcal/mol at a reference state of 1 M, 298 K. This does not agree with the related experimental value of −2.7 kcal/mol, even if one was to account for the underestimated interaction energy. Yet our calculated $\Delta H°$ for pair formation of A:T in chloroform has an average value of −6.9 kcal/mol, a value that agrees very well with experiment. The potential energy surface near the minimum of the hydrogen-bonded base pairs is fairly flat, and low modes contribute substantially to the vibrational contribution to the calculated free energies.[55] Even though the frequencies have been scaled to make them more accurate, the free energies appear to be less accurate than the enthalpies. The difficulty in calculating accurate free energies is further supported by the large variation in calculated $T\Delta S$ terms for A:T pairs, as opposed to fair agreement between calculated interaction enthalpies for A:T pairs using various methods found in the works cited previously.

Helical Optimizations. Even though we have shown that the free energy of pairing A:F in the gas phase is unfavorable, it is certain that the helix environment influences this interaction. For this reason, small sections of B-form DNA were built so that optimizations could be run to observe the effect of helix constraints on the pairing behavior. When attempting to optimize a system containing nearly 190 atoms, relatively low levels of theory must be used. The advantages of using PM3 over AM1 when dealing with hydrogen bonding, including nucleotide bases, has been discussed previously.[68,69] Figure 2 reveals that the structures of the helices incorporating A:F show evidence of base pair opening. The bond distances found in Table 4 clearly show both A:F pairs to be in an open conformation, though the NMR structure has A:F paired in a closed conformation. In contrast to A:F, the A:T base pairs maintain a correct Watson−Crick pair after optimization (slightly propeller twisted in the case of the NMR structure), revealing the influence of hydrogen bonding on this pair. The hydrogen bond lengths for the A:T pair found in Table 4 are within the normal range for a Watson−Crick pair.

The opening of the A:F pair is not surprising but lends further support to the decreased hydrogen bond strength of F.[13,25] If hydrogen bonding can exist between two bases in a dimer, by virtue of a negative value for ΔH, PM3 will likely exhibit the

correct bonding behavior.[68,69] It is significant that the helix optimizations maintain a correct A:T Watson−Crick pair, while unpairing A:F. Though an A:F pair would be expected to form in the gas-phase based on ΔH alone, the constraints of the helix pulls these bases apart.

In a comparison of semiempirical, ab initio, density functional, and molecular dynamics theories, it was determined that AMBER reproduced the stacking and hydrogen-bonding interaction energies best relative to MP2 theory.[70] The inability of PM3 to replicate the MP2 values was stressed. In AMBER molecular dynamics studies, the A:F pair exhibited high rates of base pair opening, and though the A:F Watson−Crick hydrogen bonds were formed, broken, and re-formed numerous times, the driving force for this process was not identified as either base stacking or hydrogen bonding.[25,26]

It seems clear from the outcome of the PM3 helix optimizations of the A:F base pair that the attractive forces between A and F are not strong enough to maintain an A:F Watson−Crick pair within the helix. While the driving force for the opening of adenine into solution in the molecular dynamics studies is better solvation of adenine, we can now identify the driving force for the re-formation of the A:F pair as base stacking and not hydrogen bonding, as this study makes clear that A and F will not spontaneously form A:F Watson−Crick base pairs under any of the dielectric conditions studied. Though the semiempirical structures must be interpreted with caution, because of the failure of these methods to model base stacking, the poor hydrogen-bonding ability of F is certainly evident. It must be stressed that we are not attempting to calculate stacking energies with a level of theory that completely lacks this ability. For this reason, recent advances in augmented density functional theory (some of which account for weak van der Waals forces, and thus stacking interactions) may be helpful in looking further into the issues discussed.[71−73]

4. Replication Fidelity

The base pairs A:T and A:F are replicated with high fidelity. If A:F does not hydrogen bond to A, then there must be some explanation for why these two pairs are formed by DNA polymerase I. We have shown that any energetic advantage of T to pair with A, over F with A, is canceled out by aqueous solvation (Table 3). In concert with the excellent steric mimicry of A:T by A:F, the equal energetic penalty to form the two pairs helps to explain why both pairs are replicated. Nonstandard base pairs with more positive free energies of interaction will not be as successfully replicated.

If the dielectric of DNA can be assumed to average near that of chloroform, a nonpolar solvent, our results suggest that A:T pair formation does have a slight advantage over A:F pair formation. This is consistent with the decreased fidelity in the replication of A:F as compared to A:T. The above assumption is reasonable since formation of A:U in chloroform has a free energy of roughly −3.0 kcal/mol experimentally, and this is on the order of −4.0 kcal/mol, given the estimation of −2.0 kcal/mol per hydrogen bond formed within the helix environment, by Arora et al.[58] Arora et al. feel that a good dielectric constant to model DNA is 4.0, a value close to that of chloroform.

The unfavorable pairing free energies in every dielectric studied here help to elucidate the behavior of F during replication extension. DNA translocation within *Bacillus stearothermophilus* DNA polymerase I has been coupled with the minor-groove region (MGR) of the polymerase.[74] Interactions between the polymerase and the DNA helix, as well as the proposed modes of translation of the DNA, are similar in other DNA polymerases such as rat DNA polymerase β and *Taq* DNA polymerase.[75,76] Specifically, in DNA polymerase I, the MGR makes three sets of hydrogen bonds to the DNA minor groove,[74] in a base pair independent fashion.[77] Mismatched pairs in the MGR can stall the action of a given DNA polymerase.[78,79] Replication studies show that when two adenines are adjacent in the template strand, the Klenow fragment of *E. coli* DNA polymerase I stalls when trying to incorporate two difluorotoluenes, even though incorporation of one F does not induce a stall.[15] This stall likely results from the inability of the polymerase to translocate the DNA since the ability of the amino acids in the MGR to form hydrogen bonds with difluorotoluene is negligible.[7,80,81] If two A:F pairs are present in the MGR, there are not nearly enough hydrogen-bonding interactions between the amino acid residues in the MGR and the minor groove of the DNA to allow for translocation, assuming the MGR interactions somehow ratchet the DNA through the polymerase. This in turn leads to a stall during replication. Stalls caused by other nonpolar analogues[20] are likely caused by the same decrease in hydrogen-bonding ability. Though the absolute magnitudes of the calculated free energies from this study seem to be less accurate than the calculated enthalpies, the results lend further support to the inability of difluorotoluene to form hydrogen bonds.

5. Conclusions

Gas-phase optimizations of A:T and A:F show that both structures adopt canonical Watson−Crick base pairs, though the A:F hydrogen bonds are longer. At the DFT and MP2 levels of theory, the gas-phase interaction free energies predict that A and T will spontaneously from an A:T pair while A:F spontaneously dissociates into A and F. Addition of the BSSE and fragment relaxation energies brought DFT and MP2 interaction energies into excellent agreement with each other. The BSSE corrections had a large effect on the interaction energies calculated with the MP2 method, and the fragment relaxation correction was much larger for A:T as compared to A:F, a consequence of the weaker forces between A and F. The calculated free energies were not as accurate as the calculated enthalpies of pair formation, as judged by pairing thermodynamics of A:T in chloroform.

Solvation has a pronounced effect on the energetics. The small cost to desolvate F causes the energy difference between pairing A:T and A:F in water to become negligible. Both A:T and A:F are predicted to dissociate into their monomer components when immersed in water. The monomer components most likely form stacked pairs in aqueous solution rather than hydrogen-bonded structures. When the solvent becomes less polar, as in chloroform, the favorability of the A:T pair over the A:F pair increases by 2 kcal/mol.

Low-level helix optimizations provided support for the notion that base stacking is more important than hydrogen bonding for the stability of the A:F pair within a DNA helix. Replication studies[15] show that when two adenines are next to each other in the template strand, the polymerase stalls when trying to incorporate two difluorotoluenes. This stall likely results from the inability of the polymerase to translocate the DNA, since the ability of the MGR of the protein to form hydrogen bonds with the minor groove of the DNA when two difluorotoluenes are adjacent is negligible. This study, while building off previous theoretical studies of A:F and A:T pairs, shows that F is unable to hydrogen bond with A.

Acknowledgment. We thank Chris Cramer, Karl Kirschner, Charlie Laughton, Modesto Orozco, and anonymous reviewers

for helpful insight. E.C.S. acknowledges Chris Cramer for support. Acknowledgment is made to the donors of the Petroleum Research Fund, administered by the American Chemical Society, and to the NIH, and to Hamilton College for support of this work.

References and Notes

(1) Watson, J.; Crick, H. C. *Nature* **1953**, *171*, 737−738.
(2) Kutcha, R. D.; Mizrahi, V.; Benkovic, P. A.; Johnson, K. A.; Benkovic, S. J. *Biochemistry* **1987**, *26*, 6,.
(3) Sloane, D. L.; Goodman, M. F. *Nucl. Acids Res.* **1988**, *16*, 6465−6475.
(4) Echols, H.; Goodman, M. F. *Annu. Rev. Biochem.* **1991**, *60*, 477−511.
(5) Diederichsen, U. *Angew. Chem., Int. Ed. Engl.* **1998**, *37*, 1655−1657.
(6) Goodman, M. F. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10493−10495.
(7) Kool, E. *Biopolymers* **1998**, *48*, 3−17.
(8) Moran, S.; Ren, R. X.-F.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 10506−10511.
(9) Moran, S.; Ren, R. X.-F.; Rumney IV, S.; Kool, E. T. *J. Am. Chem. Soc.* **1997**, *119*, 2056−2057.
(10) Guckian, K. M.; Kool, E. T. *Angew. Chem., Int. Ed. Engl.* **1997**, *36*, 2825−2828.
(11) Guckian, K. M.; Schweitzer, B. A.; Ren, R. X.-F.; Sheils, C. J.; Paris, P. L.; Tahmassebi, D. C.; Kool, E. T. *J. Am. Chem. Soc.* **1996**, *118*, 8182−8183.
(12) Guckian, K. M.; Morales, J. C.; Kool, E. T. *J. Org. Chem.* **1998**, *63*, 3, 9652−9656.
(13) Guckian, K. M.; Krugh, T. R.; Kool, E. T. *Nat. Struct. Biol.* **1998**, *5*, 954−959.
(14) Matray, T. J.; Kool, E. T. *J. Am. Chem. Soc.* **1998**, *120*, 6191−6192.
(15) Liu, D.; Moran, S.; Kool, E. T. *Chem. Biol.* **1997**, *4*, 919−926.
(16) Kool, E. T.; Morales, J. C.; Guckian, K. M. *Angew. Chem., Int. Ed. Engl.* **2000**, *39*,.
(17) Schmidt, K. S.; Sigel, R. K.; Filippov, D. V.; van der Marel, G. A.; Lippert, B.; Reedijk, J. *New J. Chem.* **2000**, *24*, 195−197.
(18) Barsky, D.; Kool, E. T.; Colvin, M. E. *J. Biomol. Struct. Dyn.* **1999**, *16*, 1119−1134.
(19) Berger, M.; Ogawa, A. K.; McMinn, D. L.; Wu, Y.; Schultz, P. G.; Romesberg, F. E. *Angew. Chem., Int. Ed. Engl.* **2000**, *39*, 2940−2942.
(20) Ogawa, A. K.; Wu, Y.; Berger, M.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2000**, *122*, 8803−8804.
(21) Morales, J. C.; Kool, E. T. *Nat. Struct. Biol.* **1998**, *5*, 950−954.
(22) Guckian, K. M.; Krugh, T. R.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122*, 6841−6847.
(23) Florian, J.; Goodman, B. F.; Warshel, A. *J. Phys. Chem. B* **2000**, *104*, 10092−10099.
(24) Morales, J. C.; Kool, E. T. *Biochemistry* **2000**, *39*, 2626−2632.
(25) Cubero, E.; Sherer, E. C.; Luque, F. J.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **1999**, *121*, 8653−8654.
(26) Cubero, E.; Laughton, C. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2000**, *122*, 6891−6899.
(27) Hobza, P.; Sponer, J. *Chem. Rev.* **1999**, *99*, 3247−3276.
(28) Meyer, M.; Suhnel, J. *J. Biomol. Struct. Dyn.* **1997**, *15*, 619−624.
(29) Santhosh, C.; Mishra, P. C. *Int. J. Quantum Chem.* **1998**, *68*, 351−355.
(30) Wang, X.; Houk, K. N. *Chem. Commun.* **1998**, 2631−2632.
(31) Evans, T. A.; Seddon, K. R. *Chem. Commun.* **1997**, 2023−2024.
(32) Cieplak, P.; Kollman, P. A. *J. Am. Chem. Soc.* **1988**, *110*, 3734−3739.
(33) Spartan5.0. Wavefunction, Inc., 18401 Von Karman Ave., Ste. 370, Irvine, CA 92612, U.S.A.
(34) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.
(35) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257−2261.
(36) Hariharan, P. C.; Pople, J. A. *Chem. Phys. Lett.* **1972**, *66*, 217.
(37) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724−728.
(38) Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040−1046.
(39) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(40) Miehlich, A.; Savin, A.; Stoll, H.; Pruess, H. *Chem. Phys. Lett.* **1989**, *157*, 200.
(41) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; 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. A.; 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 98*; Gaussian, Inc.: Pittsburgh, PA, 1998.
(42) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502−16513.
(43) Xantheas, S. S. *J. Chem. Phys.* **1996**, *104*, 8821−8824.
(44) Paizs, B.; Suhai, S. *J. Comput. Chem.* **1998**, *19*, 575−584.
(45) Halász, G. J.; Vibók, Á.; Mayer, I. *J. Comput. Chem.* **1999**, *20*, 274−283.
(46) Hawkins, G. D.; Giesen, D. J.; Lynch, G. C.; Chambers, C. C.; Rossi, I.; Storer, J. W.; Li, J.; Winget, P.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. Based in part on AMPAC-version 2.1 by D. A. Liotard, E. F. Healy, J. M. Ruiz, and M. J. S. Dewar. 1999. AMSOL-version 6.6. University of Minnesota, Minneapolis.
(47) Giesen, D. J.; Chambers, C. C.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1997**, *101*, 5084−5088.
(48) Eksterowitcz, J. E.; Miller, J. L.; Kollman, P. A. *J. Phys. Chem. B* **1997**, *101*, 10971−10975.
(49) Li, J.; Cramer, C. J.; Truhlar, D. G. *Biophys. Chem.* **1999**, *78*, 147−155.
(50) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 221−264.
(51) Sherer, E. C.; Cramer, C. J. *J. Comput. Chem.* **2001**, *22*, 1167−1179.
(52) Sponer, J.; Hobza, P. *J. Phys. Chem.* **2000**, *104*, 4592−4597.
(53) Brameld, K.; Dasgupta, S.; Goddard, W. A., III. *J. Phys. Chem. B* **1997**, *101*, 4851−4859.
(54) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149−1170.
(55) Hobza, P.; Sponer, J. *Chem. Phys. Lett.* **1996**, *261*, 379−384.
(56) Florian, J.; Sponer, J.; Warshel, A. *J. Phys. Chem. B* **1999**, *103*, 884−892.
(57) Young, M.; Jayaram, B.; Beveridge, D. *J. Phys. Chem. B* **1998**, *102*, 7666−7669.
(58) Arora, N.; Jayaram, B. *J. Phys. Chem. B* **1998**, *102*, 6139−6144.
(59) Kyogoku, K.; Lord, R. C.; Rich, A. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *57*, 250−257.
(60) Kyogoku, Y.; Lord, R. C.; Rich, A. *Science* **1966**, *154*, 518−520.
(61) Binford Jr., J. S.; Holloway, D. M. *J. Mol. Biol.* **1968**, *31*, 91−99.
(62) Kratochvil, M.; Sponer, J.; Hobza, P. *J. Am. Chem. Soc.* **2000**, *122*, 3495−3499.
(63) Solie, T. N.; Schellman, J. A. *J. Mol. Biol.* **1968**, *33*, 61−77.
(64) Katz, L.; Penman, S. *J. Mol. Biol.* **1966**, *15*, 220−231.
(65) Hamlin, R. M.; Lord, R. C.; Rich, A. *Science* **1965**, *148*, 1734.
(66) Miller, J. H.; Sobell, H. M. *J. Mol. Biol.* **1967**, *24*, 345−350.
(67) Norberg, J.; Nilsson, L. *Biophys. J.* **1998**, *74*, 394−402.
(68) Jurema, M. W.; Shields, G. C. *J. Comput. Chem.* **1993**, *14*, 89−104.
(69) Lively, T. N.; Jurema, M. W.; Shields, G. C. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1994**, *21*, 95−107.
(70) Hobza, P.; Kabelác, M.; Sponer, J.; Mejzlík, P.; Vondrásek, J. *J. Comput. Chem.* **1997**, *18*, 1136−1150.
(71) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149−5155.
(72) Kurita, N.; Araki, M.; Nakao, K.; Kobayashi, K. *Int. J. Quantum Chem.* **2000**, *76*, 677−685.
(73) Kurita, N.; Kobayashi, K. *Comput. Chem.* **2000**, *24*, 351−357.
(74) Kiefer, J. R.; Mao, C.; Braman, J. C.; Beese, L. S. *Nature* **1998**, *391*, 304−307.
(75) Pelletier, H.; Sawaya, M. R.; Kumar, A.; Wilson, S. H.; Kraut, J. *Science* **1994**, *264*, 1891−1903.
(76) Eom, S. H.; Wang, J.; Steitz, T. A. *Nature* **1996**, *382*, 278−281.
(77) Beard, W. A.; Wilson, S. H. *Chem. Biol.* **1998**, *5*, R7−R13.
(78) Doublie, S.; Tabor, S.; Long, A. M.; Richardson, C. C.; Ellenberger, T. *Nature* **1998**, *391*, 251−258.
(79) Johnson, K. A. *Annu. Rev. Biochem.* **1993**, *62*, 685−713.
(80) Morales, J. C.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122*, 1001−1007.
(81) Morales, J. C.; Kool, E. T. *J. Am. Chem. Soc.* **1999**, *121*, 2323−2324.