Hydrogen-Bonding Interactions in Peptide Nucleic Acid and Deoxyribonucleic Acid: A Comparative Study

Holly Elizabeth Herbert, Mathew D. Halls, Hrant P. Hratchian, and Krishnan Raghavachari*

Department of Chemistry, Indiana University, Bloomington, Indiana 47405

Received: October 13, 2005; In Final Form: December 8, 2005

Peptide nucleic acid (PNA) is a synthetic analogue of deoxyribonucleic acid (DNA) capable of tightly binding to itself and DNA with high specificity. Using hybrid density functional methods, hydrogen-bond (H-bond) strengths have been evaluated for isolated Watson−Crick base pairs, PNA base pairs, and charged as well as neutral DNA base pairs. Heterogeneous base pairs of PNA with charged and neutral DNA have also been investigated. The competing effects of short-range H-bonding and long-range Coulombic repulsions in charged DNA base pairs have been analyzed. Polarizable continuum models have been employed to evaluate solvation effects on the binding energies.

Introduction

Molecular recognition by nucleic acids is a topic of marked interest due, in part, to ongoing developments of myriad innovative chemical and analytical processes.[1,2] At the heart of these approaches are the Watson−Crick bases, which associate preferentially into H-bonded purine−pyrimidine base pairs (Chart 1). The adenine−thymine (AT) base pair forms two H-bonds in its central region, while the guanine−cytosine (GC) base pair forms three (Chart 1). These H-bonds play an integral role in the association of bases in the double helix and help maintain planarity within the base pair. Predictable H-bond affinity and base pair structures allow for highly targeted binding at specific sites, which leads to applications such as amplification of polymerase chain reaction (PCR) and gel electrophoresis results,[1−4] effective blocking of transcription and/or translation of specific gene sequences in rational drug design,[1,2,5−7] and constructing periodic arrays of complementary nucleic acids for, among other things, hypothetical drug delivery vessels in nucleic acid nanotechnology.[1,2,8−12]

The DNA backbone is composed of a chiral phosphodiester moiety that is bound to a H-bonded purine−pyrimidine base pair (Chart 2).[13] The phosphate group, located on the exterior of the double helix, is negatively charged, giving each DNA base pair a total −2 charge.[14] The anionic nature of the DNA backbone suggests that H-bonding interactions of the bases may be weakened by Coulombic repulsion between backbone units. This repulsion may be negated by associating countercations, such as Na^+ or Mg^{2+}, with the DNA phosphate groups, and indeed, crystallographic data verify that countercations associate with the DNA backbone.[15−18]

Peptide nucleic acid (PNA), first described by Wittung and co-workers,[19,20] is a synthetic analogue of DNA composed of the Watson−Crick base pairs linked to a neutral, achiral 2-aminoethyl glycine backbone (Chart 3).[1,2,20−27] The neutral backbone assembles into a loose helix, making it structurally homomorphic to DNA.[1,2,19,26,28−30]

Interestingly, PNA appears to form a more stable double helix with DNA than DNA does with itself, making PNA a valuable analytical tool with exciting therapeutic potential.[1,2] Explorative work by Griffin and Smith indicates that the thermal stability of PNA−PNA and PNA−DNA base pairs exceeds that of DNA−DNA base pairs.[31] Gräslund and co-workers attributed the additional stability in PNA-containing base pairs to the neutral character of the PNA backbone,[32] suggesting that the total binding stability of DNA base pairs is dependent upon ion concentration or the total dielectric environment. Furthermore, their results imply that the stability of PNA base pairs is independent of such effects.

The goal of the present investigation is to perform a comparative evaluation of PNA and DNA bonding at the single base pair level using electronic structure methods. Specifically, we focus on H-bond energies of homogeneous base pairs formed from charged DNA, neutral DNA (containing sodium countercations), and PNA, as well as heterogeneous charged and neutral DNA−PNA base pairs. In addition, Coulombic interactions and

CHART 1

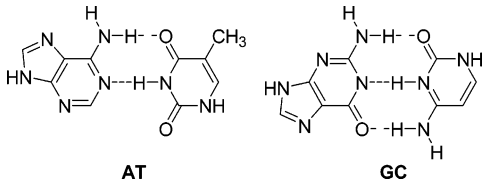


CHART 2

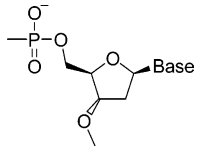
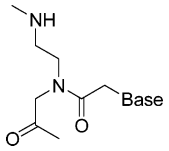


CHART 3



* To whom correspondence should be addressed. E-mail: kraghava@indiana.edu.

10.1021/jp055865j CCC: $33.50
Published on Web 01/27/2006

TABLE 1: Basis Set Effects on Calculated Binding Energies of Isolated AT and GC Base Pairs[a]

H-bond energy (kcal/mol)	experiment	6-31G(d,p)	6-31+G(d,p)	6-311+G(d,p)	6-311+G(2df,p)	6-311+G(3df,2p)
AT	−13.0	−15.2	−11.7	−11.7	−11.2	−11.1
GC	−21.0	−28.8	−25.0	−24.5	−23.9	−23.5

[a] All energies were calculated at the B3LYP level using optimized B3LYP/6-31G(d,p) geometries. Zero-point effects computed at the B3LYP/6-31G(d,p) level are included.

solvation effects have been investigated with appropriate theoretical models.

Theoretical Models and Calibration

Studies of H-bond interactions in DNA base pairs are prevalent in the literature, and a number of theoretical investigations have previously been carried out[33−47] using a spectrum of methods ranging from hybrid density functional models to sophisticated electron correlation techniques including second-order perturbation theory (MP2) and augmented coupled cluster theory [CCSD(T)]. These studies provide us with benchmark values of the interaction energies to calibrate our results.

We have performed preliminary calculations on isolated AT and GC base pairs to assess the performance of theoretical models in predicting accurate H-bond interaction energies. Our goal is to derive a practical, yet accurate, theoretical model not only applicable to the current systems of interest (isolated base pairs) but also useful in future investigations of H-bond interactions involving multiple base pairs. To explore these systems, we selected the widely used B3LYP density functional method (Becke's three parameter exchange functional[48] along with the Lee−Yang−Parr correlation functional[49]). Previous studies with B3LYP have suggested that the method yields accurate H-bond energies when used with large basis sets. While B3LYP is appropriate for H-bond calculations, we note that an augmented method will be necessary when studying extended helical systems due to the importance of base pair stacking interactions. We performed geometry optimizations for AT and GC base pairs as well as the isolated bases with the standard polarized 6-31G(d,p) basis set.[50,51] Single-point calculations were then carried out with a series of larger basis sets containing additional diffuse and polarization functions, and the resulting H-bond energies are listed in Table 1. Zero-point energy corrections (ZPE) evaluated at the B3LYP/6-31G(d,p) level have been included in the interaction energies listed in Table 1.

At the B3LYP/6-31G(d,p) level of theory, where optimizations were performed, the calculated interaction energies including ZPE are −15.2 and −28.8 kcal/mol for AT and GC pairs, respectively. These values are significantly larger than the corresponding experimentally determined values of −13.0 and −21.0 kcal/mol for AT and GC pairs, respectively.[52] It is well-known that larger basis sets containing additional diffuse and/or polarization functions may have an influence on such computed interaction energies. Indeed, larger basis sets decrease the computed interaction energies and lead to improved agreement with experiment (Table 1). Interestingly, most of this effect comes from the addition of one set of diffuse functions to the basis set [6-31+G(d,p)[50,51]] that decreases the binding energies of the AT and GC base pairs by 3.5 and 3.8 kcal/mol, respectively (Table 1). Optimizations of the isolated base pairs using 6-31+G(d,p) did not give geometries of notable difference from those determined using 6-31G(d,p). Subsequent additions to the basis sets (Table 1) yielded relatively small improvements. We have thus used this model chemistry, B3LYP/6-31+G(d,p)//B3LYP/6-31G(d,p), for all interaction energies reported in this work (here, the symbol // separates the level of theory used for single-point calculations from that used in the optimizations). We note that corrections due to basis set superposition error (BSSE) have not been included. However, the data in Table 1 suggest that BSSE for the chosen model chemistry is likely to be small since the 6-31+G(d,p) is near the basis set limit for the systems of interest.[53]

The computed H-bond interaction energies using our model chemistry are −11.7 and −25.0 kcal/mol for the AT and GC pairs, respectively. These values are about 2 kcal/mol lower than the best values (−14.0 kcal/mol for AT and −27.0 kcal/mol for GC) obtained by Hobza and co-workers[54] using MP2 results extrapolated to the complete basis set limit and corrected for higher-order correlation [CCSD(T)] and zero-point corrections. Hobza and co-workers rationalized the somewhat large discrepancy between their calculations and the experimental findings by concluding that experimental results were essentially an average of several different base-pairing configurations, such as Hoogsteen pairing,[54] but that the Watson−Crick pairs should exhibit binding energies close to their theoretical predictions. Overall, our model chemistry is reasonably close to the best results from previous studies and should be reliable in predicting changes in H-bond strengths when applied to systems with similar base pair couplings.

Starting geometries for isolated, PNA, and charged DNA base pairs were taken from available crystal structures.[15,30,55] Complete geometry optimizations were then performed for each of the structures at the B3LYP/6-31G(d,p) level to reach minimum energy configurations. Our target in this study was to identify minima that most closely resemble the geometries of the base pairs in their native helical conformations. Although some of these structures may not be global minima, the structures resemble the native conformation which facilitates a helical stacking of the base pairs. The nature of each minimum was verified by analytic harmonic frequency calculations that also yielded zero-point energy corrections. Single-point calculations with the addition of diffuse functions, as discussed previously, gave the interaction energies.

The DNA−DNA base pairs, as previously mentioned, are doubly anionic and have repulsive Coulomb interactions. We have also optimized the geometries of analogous neutral forms of the DNA base pairs by associating a countercation with the anionic phosphate groups. Each phosphate group is associated with a sodium ion in such neutral DNA studies. The optimized structures contain sodium ions in bridging positions between two phosphate oxygens (vide infra). Finally, we have also optimized hetero base pairs (both anionic and neutral forms containing sodium) to investigate DNA−(base pair)−PNA interactions.

Solvation effects were modeled using the conductor-like polariazable continuum model (CPCM)[56,57] and the integral-equation formalism polarizable continuum model (IEF-PCM).[58−61] The cavity was defined by interlocking spheres with radii determined by united atom Kohn−Sham approximations. The convergence criterion was defined to be 10^{-8} in CPCM[62], and the area of the tesserae in IEF-PCM was defined to be 0.4 Å^2.[63] All calculations reported in this study were carried out using the Gaussian 03 suite of electronic structure programs.[64]

TABLE 2: Optimized H-Bond Lengths and Angles for AT Base Pairs[a]

	bond length (Å)		bond angle (deg)	
	NH−O	N−HN	NH−O	N−HN
isolated AT base pair	2.97	2.84	173.7	179.4
PNA_AT_PNA	2.93	2.85	174.6	179.6
DNA_AT_DNA	2.94	2.90	175.5	178.8
Na_DNA_AT_DNA_Na	2.93	2.85	174.9	178.4
PNA_AT_DNA	2.86	2.91	177.8	178.3
DNA_AT_PNA	2.98	2.80	172.2	177.8
PNA_AT_DNA_Na	2.92	2.86	175.3	179.4
Na_DNA_AT_PNA	2.93	2.85	174.4	179.6

[a] Bonds and bond angles are read from adenine to thymine.

Results and Discussion

H−Bond Distances and Angles. There are two regions of interest in geometry exploration for the base pairs and their various configurations. The first is the H-bonding region between the Watson−Crick bases, while the second is the geometry of the backbone. The H-bonding region can be described by bond lengths and angles between the H-bond donors and the acceptors, while the backbone geometry can be defined by a series of dihedral angles. We discuss the H-bonding region in this section and the backbone geometry in a later section.

AT base pair H-bond lengths and angles are presented in Table 2. The corresponding values for the GC base pair are presented in Table 3. Our computed geometrical parameters for the isolated AT and GC base pairs agree with previously published results.[65] The AT base pair contains two H-bonds. In the isolated AT base pair, the exterior N−H···O bond has a calculated length of 2.97 Å and an angle of 173.7°, and the interior N···H−N bond has a bond length of 2.84 Å and an angle of 179.4°. The GC base pair contains three H-bonds. In the isolated GC base pair, the exterior N−H···O bond has a length of 2.92 Å and an angle of 180.0°. The interior N−H···N bond length is 2.93 Å, the angle is 178.1°, and the other exterior O···H−N bond has a length of 2.79 Å and an angle of 179.5°. These bond lengths remain roughly constant in most AT and GC base pairs, particularly for homogeneous base pairs. Surprisingly, the addition of countercations to neutralize the charge in DNA base pairs has only a minor effect on the H-bond lengths. For example, the three H-bonds in the doubly anionic DNA−GC−DNA have values of 2.82, 2.96, and 2.90 Å, while the corresponding values in neutral Na−DNA−AT−DNA−Na are 2.81, 2.93, and 2.91 Å, respectively. However, significant differences are observed in some cases, particularly in the charged heterogeneous base pairs. For example, the ordering of the short and long H-bond distances is reversed in PNA−AT−DNA relative to isolated AT. The NH···O bond has a length of 2.80 Å in PNA−GC−DNA and 3.00 Å in DNA−GC−PNA. In both cases, the bond deviates from the average calculated base pair value by ∼0.1 Å. As discussed in the next section, there is a corresponding change in the strength of the interaction.

Hydrogen Bond Strengths. H-bond strengths are evaluated according to

$$E(\text{H-bond}) = E(\text{base pair}) - E(\text{base 1}) - E(\text{base 2}) \quad (1)$$

where E(base pair) is the energy of the optimized base pair and E(base 1) and E(base 2) are the energies of the isolated Watson−Crick bases contained within the pair.[34,39,41,44,65−69] The computed interaction energy values are presented in Table 4 for all base pairs. From these data, a number of important observations may be drawn. First, the addition of a PNA backbone to the base pairs does not provide a significant change in H-bond strengths (−11.8 kcal/mol for PNA−AT−PNA vs −11.7 for isolated AT and −24.7 kcal/mol for PNA−GC−PNA vs −25.0 for isolated GC). In contrast, the addition of the anionic DNA backbone has a dramatic effect on the H-bond strengths, yielding interaction energies of +5.7 and −11.0 kcal/mol for the DNA−AT−DNA and DNA−GC−DNA base pairs, respectively. Notice that the AT base pair is not bound with respect to separated bases in this case (i.e., with respect to DNA−A and DNA−T). At first sight, this appears somewhat surprising since the corresponding H-bond lengths did not change significantly with the introduction of the charged DNA backbone. The decrease in the binding is evidently due to the repulsion between the two negative charges in the phosphate backbones. Upon neutralization with sodium ions (or with point charges, vide infra), the H-bond strengths are similar to those observed for the isolated Watson−Crick base pairs. Thus, the binding energies in Na−DNA−AT−DNA−Na (−10.9 kcal/mol) and Na−DNA−GC−DNA−Na (−24.5 kcal/mol) are very close to those of the isolated base pairs. Clearly, charge−charge repulsion is a key interaction when calculating DNA H-bond strengths, and this aspect is explored in the following section.

Heterogeneous base pairs comprised of both a DNA and a PNA backbone also demonstrate similar H-bond strengths to those of the isolated Watson−Crick base pairs. In this case, the addition of the countercation has a much smaller effect on the binding energies. This is not surprising since the presence of only one negative charge does not cause any unfavorable interactions. There are two different structural arrangements in the heterogeneous base pairs. PNA−AT−DNA and DNA−AT−PNA have very similar binding energies close to that of the isolated base pairs, with or without the presence of countercations. However, the charged heterogeneous GC base pairs (DNA−GC−PNA and PNA−GC−DNA) show a deviation from the isolated Watson−Crick base pair binding energy of −25.0 kcal/mol. The DNA−GC−PNA base pair has a H-bond energy of −21.6 kcal/mol while PNA−GC−DNA has a H-bond energy of −27.6 kcal/mol. While unexpected, this

TABLE 3: Optimized H-Bond Lengths and Angles for GC Base Pairs[a]

	bond length (Å)			bond angle (deg)		
	O−HN	NH−N	NH−O	O−HN	NH−N	NH−O
isolated GC base pair	2.79	2.93	2.92	179.5	178.1	180.0
PNA_GC_PNA	2.79	2.94	2.91	178.8	177.4	178.7
DNA_GC_DNA	2.82	2.96	2.90	178.3	176.9	176.2
Na_DNA_GC_DNA_Na	2.81	2.93	2.91	178.7	177.1	178.7
PNA_GC_DNA	2.88	2.91	2.80	178.7	178.6	177.9
DNA_GC_PNA	2.74	2.96	3.00	177.7	176.3	173.9
PNA_GC_DNA_Na	2.81	2.93	2.90	179.9	177.6	177.9
Na_DNA_GC_PNA	2.78	2.94	2.92	179.0	177.7	178.5

[a] Bonds and bond angles are read from guanine to cytosine.

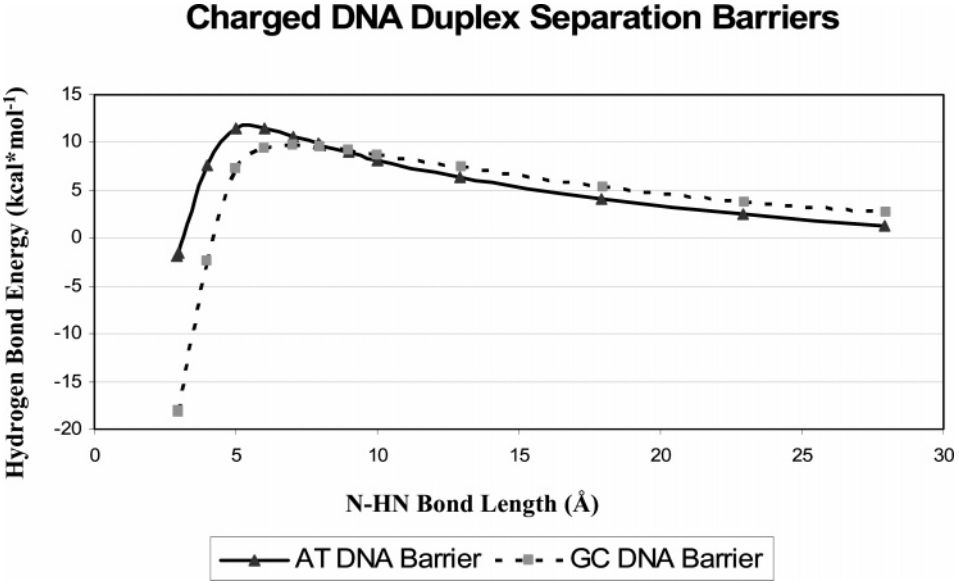


Figure 1. Potential energy surface of charged base pair separation for AT and GC. Hydrogen bonding energy (kcal/mol) vs central N−H···N hydrogen bond length (Å).

TABLE 4: H-Bonding Energies (kcal/mol) of the AT and GC Base Pairs[a]

	adenine (X)-thymine(Y) configuration	guanine(X)-cytosine(Y) configuration
isolated base pair	−11.7	−25.0
PNA_XY_PNA	−11.8	−24.7
DNA_XY_DNA	5.7	−11.0
Na_DNA_XY_DNA_Na	−10.9	−24.5
PNA_XY_DNA	−11.1	−27.6
DNA_XY_PNA	−10.8	−21.6
PNA_XY_DNA_Na	−11.8	−24.9
Na_DNA_XY_PNA	−10.7	−24.1

[a] Purines are represented by X, and pyrimidines are represented by Y. Negative numbers indicate stable conformations. Energies include zero-point corrections calculated using B3LYP/6-31G(d,p).

may be related to the change in the NH···O bond lengths noted earlier. In the case of PNA−GC−DNA, the bond is shortened by 0.1 Å, and the interaction energy for the base pair is ∼3 kcal/mol stronger than that of the neutral base pairs. Conversely, the NH···O bond length in the DNA−GC−PNA base pair increases by 0.1 Å, and the interaction is weakened by ∼3 kcal/mol with respect to the neutral base pairs.

Charge−Charge Repulsions in DNA Base Pairs. As previously discussed, the addition of an anionic DNA backbone significantly decreases the stability of H-bonds in both AT and GC base pairs. The doubly anionic DNA−GC−DNA base pair has a binding energy of −11.0 kcal/mol. While significantly less stable than the corresponding isolated Watson−Crick GC base pair (binding energy of −25.0 kcal/mol), it remains bound. The DNA−AT−DNA base pair, on the other hand, was calculated to have a destabilized binding energy of +5.7 kcal/mol. Again, the optimized geometry indicates a typical bound environment. Clearly, there is an energy barrier preventing the DNA−AT−DNA base pair from separating into the respective bases (DNA−A and DNA−T).

We have performed a series of calculations at the B3LYP/6-31+G(d,p) level to investigate the potential energy surface for separation of the base pairs for both DNA−AT−DNA and DNA−GC−DNA. The base pair distances (as given by NH···N) were varied from 3 to 30 Å, while the remaining geometric parameters were held fixed, and the resulting relative energies are plotted in Figure 1. It is clear that both base pairs have significant energy barriers for separation into the component bases. Figure 1 shows that the AT base pair has a separation energy barrier of 13 kcal/mol, while the GC base pair has a substantially higher barrier of 28 kcal/mol. In fact, the binding energies relative to the transition state for separation are close to the values seen for the isolated base pairs.

The interpretation of Figure 1 is clear if we consider the reverse process (i.e., the approach of the isolated bases (e.g., DNA−A and DNA−T)). Since DNA−A and DNA−T are both negatively charged, there is Coulombic repulsion between them that increases the energy as they approach. As expected from Coulomb's Law, the potential energy curves at long distances, as seen in Figure 1, display inverse distance dependence. The repulsion continues as the isolated bases approach one another until they reach a distance at which H-bonding may occur. Now, strong H-bonding interactions take over, lowering the energy as expected. The overall smaller value of the binding reflects the two competing factors, the repulsive Coulombic interactions and the attractive H-bonding interactions. From the curve, it is clear that beyond 5 Å, the behavior is completely dominated by the Coulomb repulsion, while short distance scales are dominated by H-bond energetics. The separation barrier is close to the point at which the H-bond effects become very small. We note that the resulting competition of forces is analogous to the competing effects in dication chemistries.[70,71]

The optimized distance between the phosphate groups in the charged AT and GC DNA base pairs is ∼18 Å. Applying Coulombs Law, we expect a decrease in the stability in H-bond energies of ∼18 kcal/mol for the base pairs. This decrease in stability would result in binding energies of 6 and −7 kcal/mol for AT and GC DNA base pairs, respectively. Our computations predict binding energies of 5.7 and −11.0 kcal/mol for AT and GC, respectively, indicating that Coulombic effects are responsible for the decrease in H-bond stability in the charged DNA base pairs. In addition, there is some structural relaxation of the backbone geometry (vide infra) that also contributes to reduce the effects of charge repulsion.

Countercations and Point Charges. The doubly anionic DNA base pairs considered in the last section, while illustrative, do not represent a realistic environment since counterions are known in crystallographic structures of DNA helices. Compu-

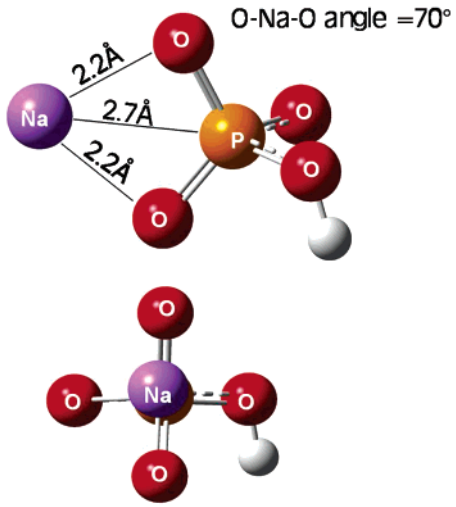


Figure 2. Neutralized phosphate group from DNA backbone: the Na^+ countercation is roughly at an equal distance from the two oxygens not involved in backbone stacking.

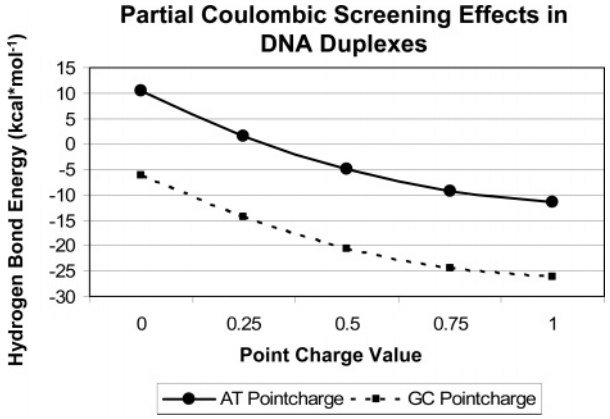


Figure 3. Coulombic effects in AT and GC DNA base pairs: hydrogen bonding enthalpy (kcal/mol) vs point charge value.

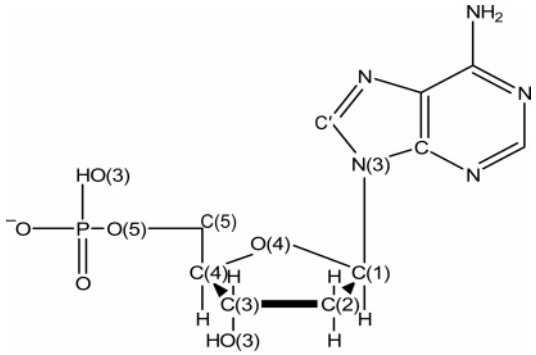
tationally, the effects of such counterions can be considered by explicitly associating individual monovalent countercations with each anionic phosphate group. As mentioned earlier, stabilizing the anionic charge from the phosphate groups with a sodium cation causes the DNA−DNA base pair binding energy to become roughly equivalent with that of the isolated base pairs and the PNA−PNA base pairs. Each sodium countercation interacts with two phosphate oxygens in a roughly symmetric manner as illustrated in Figure 2.[14−16]

Crystallographic evidence suggests that monovalent and divalent countercations are present and contribute to an overall charge neutralization, but do not involve a one-to-one association as stated previously. If the countercations are distributed around the DNA helices, the effective charge felt by the phosphate groups may be better modeled as partial charges. We have investigated the effects of partial charge stabilization by a series of calculations where a point charge of varying magnitudes (0.0, 0.25, 0.50, 0.75, and 1.0 au) is placed at the location of the sodium countercation. Base pair binding energies were calculated for each point charge value and plotted in Figure 3 for both AT and GC base pairs. Each curve shows smooth decay from the neutral DNA binding energy to the charged DNA binding energy as the point charge goes from +1 to 0 au. Table 5 compares the optimized binding energies of the charged and neutral DNA base pairs with the energies calculated using the point charge model. In the point charge calculations, the geometry has been held rigid in the optimized neutral DNA backbone conformation. Recall that this conformation is not the conformation of the charged DNA backbone. We observe that the neutral backbone conformations with a doubly anionic charge cause the interaction to be about 6.5 kcal/mol less stable for the AT and GC base pairs, suggesting that relaxation of the backbone partially offsets the Coulombic repulsions of the charged DNA base pairs.[72,73] It is important to note that while the addition of partial charges to the phosphate group increases the stability of the DNA−DNA base pairs, an interaction energy equivalent to that of PNA−PNA base pairs is only achieved at a 1:1 stoichiometry between the counterion and the phosphate group. This is in agreement with experimental observations reported by Gräslund and co-workers who observed stabilities for DNA comparable to that for PNA only at high salt concentrations.[32]

TABLE 5: Comparison of H-Bond Energies for Charged and Neutral Base Pairs Using Na^+ Counter Ions and +1 Point Charges[a]

	optimized	point charge	Δ
AT charged DNA	4.1	10.5	−6.4
AT neutral DNA	−11.7	−11.5	−0.2
GC charged DNA	−12.8	−6.1	−6.7
GC neutral DNA	−25.8	−26.2	0.4

[a] Δ = difference in stabilities between the neutral DNA structures and the charged DNA optimized structures. See text for details. Energies are presented in kcal/mol. Energies do not contain zero-point corrections.



Angle (°)	
α	O3-P-O5-C5
β	P-O5-C5-C4
γ	O5-C5-C4-C3
δ	C5-C4-C3-O3
χ	O4-C1-N3-C

Figure 4. DNA dihedral angles: angles are listed to define the backbone from top to bottom of the figure. The base shown here is adenine.

Backbone Geometrical Parameters. The structure of the DNA backbone is classified according to a series of five dihedral angles. Schneider et al.[13] have compiled a statistical sampling of the DNA backbone geometry from crystal structures and reported their results for each of the relevant backbone dihedral angles and one defining the association of the backbone to the base (Figure 4).

Neutral DNA−DNA and DNA−PNA base pair dihedral values are in good agreement with experiment, except δ (C5−C4−C3−O3) and χ (O4−C1−$N3_{base}$−C_{base}). In both cases, the deviation from the crystal structure is more than 40° and is attributed to H-bonding between the phosphate group and the

TABLE 6: Charged DNA Base Pair Geometry Results[a]

angle (deg)	exp. value	av. opt. value	DNA-A			DNA-T			DNA-G			DNA-C		
			(deg)	|Δ|(exp)	|Δ|(opt)	(deg)	|Δ|(exp)	|Δ|(opt)	(deg)	|Δ|(exp)	|Δ|(opt)	(deg)	|Δ|(exp)	|Δ|(opt)
a	−62	−66. 3	**−107.5**	**45.5**	**41.2**	−65.9	**3.9**	**0.4**	**−107.2**	**45.2**	**40.9**	−45.4	16.6	**20.9**
b	176	−172.0	**−99.0**	**85.0**	**73.0**	−178.7	**5.3**	**6.7**	**−101.0**	**83.0**	**71.0**	172.5	3.5	15.5
g	48	54.9	59.6	11.6	4.7	56.0	8.0	1.1	58.5	10.5	3.6	53.9	5.9	1.0
d	**128**	**82.4**	**147.2**	**19.2**	**64.8**	**84.9**	**43.1**	**2.5**	**147.6**	19.6	**65.2**	145.8	17.8	**63.4**
c	−102	−152.0	**−96.1**	**5.9**	**55.9**	**−149.5**	**47.5**	**2.5**	**−95.1**	**6.9**	**56.9**	**−116.7**	**14.7**	**35.3**

[a] |Δ|(exp) gives the absolute difference from the experimental value, and |Δ|(opt) gives the difference from the average value from all optimized geometries, excluding charged DNA base pairs.

TABLE 7: H-Bonding Energies in kcal/mol for AT and GC Base Pairs Using CPCM and IEF-PCM Solvation Models[a]

	A (X)-T (Y) base pairs	G (X)-C (Y) base pairs	A (X)-T (Y) base pairs	G (X)-C (Y) base pairs
	CPCM	CPCM	IEF-PCM	IEF-PCM
isolated	−1.88	−6.89	−2.20	−7.01
PNA	−2.01	−6.07	−2.92	−7.85
charged DNA	0.64	−3.09	0.55	−3.82
neutral DNA	−1.88	−6.67	−1.94	−6.93
PNA_XY_DNA	−2.04	−5.23	−2.33	−5.86
DNA_XY_PNA	−0.65	−6.06	−0.61	−6.99
PNA_XY_DNA_Na	−2.43	−6.19	−3.01	−6.45
Na_DNA_XY_PNA	−1.38	−6.73	−1.99	−7.51

[a] X represents the purine and Y the pyrimidine pair constituents. Energies do not contain zero-point corrections.

potential H-bond donors (O3 hydroxyl for δ and the CH group directly adjacent to N3 on the base for χ). As seen in Table 6, the charged DNA base pairs exhibit considerable deviation from experimental dihedrals and from our neutral DNA structures (for four optimized dihedrals: α, β, δ, and χ), suggesting that charge−charge interactions play an important role in backbone conformation as well as H-bond stability. Relaxation of the charged DNA backbone maximizes interactions between the charged phosphate groups and the potential H-bond donors. Clearly, such a structural rearrangement would not be possible in extended helical conformations.

Solvation Effects. Applications of PNA as an analytic tool or as a therapeutic agent do not generally take place in the gas phase. Furthermore, while the stability of a fully charged DNA−DNA base pair is questionable in the gas phase, it exists quite naturally in vivo, and therefore, the effects of solvation are clearly important.[74] Polarizable continuum models (PCM) have been employed in this work to investigate each base pair in an aqueous environment and to include the effects of dielectric screening in charged DNA base pairs. The two models employed here are CPCM[56,57] and IEF-PCM.[58−61] A recent review by Tomasi et al.[74] has suggested that CPCM is an excellent model for highly polar solvents (such as water) and that IEF-PCM performs reasonably well for modeling most solvent environments. Molecular geometries have been taken from gas-phase minimizations in all cases. Results from these calculations are presented in Table 7.

CPCM and IEF-PCM predict similar H-bond strengths for isolated base pairs as well as PNA−PNA and Na−DNA−DNA−Na base pairs, although the calculated values in solution are significantly reduced from the gas-phase value. This is consistent with solvent interactions with the solute dipoles that decrease the strength of H-bonding. Interestingly, the DNA−DNA base pairs are calculated to have weaker H-bond interactions than Na−DNA−DNA−Na and PNA−PNA base pairs by 2−3 kcal/mol. These results indicate that charge screening from water is not as effective as direct association with a countercation in decreasing Coulombic repulsion forces between DNA backbones. Nevertheless, the dielectric of water (78.39) is sufficient to screen a large portion of the repulsive Coulombic interactions. Overall, the CPCM and IEF-PCM methods predict equivalent bonding energies for all base pairs in solution, yielding H-bond strengths of approximately −2 and −6 kcal/mol for all AT and GC base pairs, respectively.

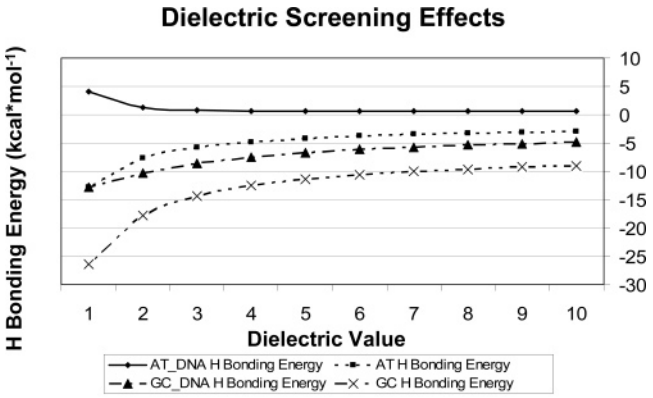


Figure 5. Solvent screening effects in AT and GC base pairs: plot of hydrogen bonding energy vs dielectric constant for both isolated Watson−Crick base pairs and charged DNA base pairs. Calculations were carried out using CPCM.

The previous analysis was done using the dielectric constant of pure water (78.39), which can easily screen most charges. However, in a more realistic environment where the base pairs and DNA are present, it is likely that the effective dielectric constant experienced by the H-bond region of the base pairs will be significantly diminished from that of pure water.[75,76] This may partly be due to solvent inaccessibility to some of these regions because of the presence of stacked base pairs. Thus, it is necessary to investigate lower dielectric constants to determine if our results are still meaningful when the effective dielectric constant is much less than that of water.

To determine the dependence on the dielectric constant, CPCM calculations were carried out on DNA−AT−DNA and DNA−GC−DNA base pairs and the isolated Watson−Crick AT and GC pairs. A range of dielectric constant values was considered, and Figure 5 shows the resulting plots from each of the four sets of calculations. Our results indicate that a dielectric constant of 10 yields a converged value in screening the anion charges. Therefore, a dielectric constant range of 1−10

is shown in Figure 5. Note that when the dielectric is very small, the H-bond energy approaches the gas phase results in all cases, as expected. Clearly, a completely nonpolar solvent would play little to no role in stabilizing DNA. On the other hand, water, or any other solvent with a high dielectric constant, will effectively screen the charges in the DNA base pairs.

Overall, two effects are seen in our PCM calculations. While the charged DNA−DNA base pairs are stabilized to make their interaction energies closer to those of the isolated and PNA−PNA base pairs, the strength of the H-bond in the neutral base pairs is significantly reduced. This decrease in stability is approximately 10 kcal/mol for the AT base pair and approximately 20 kcal/mol for the GC base pair and is caused by the polarizable continuum interacting with the H-bond region and decreasing the effective dipole of each H-bond. The GC pair shows a more significant decrease in stability because it contains three H-bonds as compared to the two H-bonds in the AT pair and is innately more polar than the AT pair.

In the case of charged DNA−DNA base pairs, two competing effects are evident from the different behaviors seen in Figure 5. As the dielectric constant increases, the increase in stability due to charge screening is in competition with the decrease in stability due to interaction with the H-bond region. Figure 5 shows that the H-bond strengths of DNA−AT−DNA and DNA−GC−DNA base pairs have opposite curvature with increasing dielectric constant. These plots indicate that, in DNA−GC−DNA, the decreased stability due to polar interactions in the H-bond region is greater than the increase in stability due to charge screening, while the opposite is true for DNA−AT−DNA. For the isolated base pairs, however, there is only a cumulative decrease in the interaction energy due to solvent−dipole interactions for both AT and GC pairs. Overall, while the GC base pair remains more stable than AT, the degree of stabilization is decreased in both upon solvation.

Conclusions

DNA and PNA have an analogous central region, composed of H-bonded Watson−Crick base pairs. The molecules differ from one another only in the composition of their backbones. We observe that the addition of a PNA backbone maintains the H-bond strength of the isolated Watson−Crick base pairs. However, the addition of an anionic DNA backbone to the same base pair causes a significant decrease in the stability of the H-bond region, suggesting that H-bonding effects may not be the only contribution to the total bonding energy of the charged DNA base pairs. In particular, we have identified and described competing short-range H-bond effects and long-range Coulombic repulsions. The competition of these two effects results in an energy barrier for the separation of charged DNA base pairs. The proximal addition of countercations to the anionic phosphate groups negates Coulombic repulsion between DNA backbone subunits and causes the H-bond strengths to be roughly equivalent to isolated and PNA base pair H-bond strengths. Point-charge calculations indicate that association of a partial positive charge with the phosphate group increases the H-bond stability. However, only when the countercation is present in 1:1 stoichiometry with respect to the phosphate group does DNA achieve the H-bond stability of the PNA base pairs.

Solvating the base pairs destabilizes the isolated Watson−Crick base pairs, as well as neutral DNA and PNA base pairs. Screening effects are responsible for this observation, as the solvent diminishes polar H-bond interactions (with respect to gas-phase results). The H-bond strength of the charged DNA−DNA base pair comes much closer to the neutral Na−DNA−DNA−Na and PNA−PNA base pair H-bond strengths but is nevertheless weaker by 2−3 kcal/mol. These results suggest that solvent effects, on their own, are unable to fully screen the anionic phosphates. By exploring dielectric effects, we showed that the stability of each base pair, even in the decreased dielectric environment of helical DNA, is substantial. In addition to van der Waals interactions between stacked based pairs, we expect that the inclusion of additional base pairs will stabilize these systems by minimizing interactions of the solvent with internal H-bond regions. While our conclusions are based on studies of single base pairs, they offer a reasonable first approximation to the strength of each Watson−Crick pair and provide important insights into the underlying effects responsible for different binding strengths observed in neutral and charged DNA and PNA base pairs (homogeneous as well as heterogeneous). Studies of stacked base pair systems and the interactions contained therein are currently ongoing and will be the subject of future work.

Acknowledgment. The authors gratefully acknowledge computational resources provided by the University Information Technology Services at Indiana University and from NSF high-performance computing (Grant CHE030049).

References and Notes

(1) Jakimov, D.; Cetojevic-Simin, D.; Kojic, V.; Mrdanovic, J. *Arch. Oncol.* **2000**, *8*, 11−14.
(2) Ray, A.; Norden, B. *FASEB J.* **2000**, *14*, 1041−1060.
(3) Demers, D. B.; Curry, E. T.; Egholm, M.; Sozer, A. C. *Nucleic Acids Res.* **1995**, *23*, 3050−3055.
(4) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Science* **1991**, *254*, 1497− 1500.
(5) Faruqi, A. F.; Egholm, M.; Glazer, P. M. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1398−1403.
(6) Hyrup, B.; Egholm, M.; Nielsen, P. E.; Wittung, P.; Norden, B.; Buchardt, O. *J. Am. Chem. Soc.* **1994**, *116*, 7964−7970.
(7) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Anti-Cancer Drug Des.* **1993**, *8*, 53−63.
(8) Seeman, N. C. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 225−248.
(9) Seeman, N. C. *Nano Lett.* **2001**, *1*, 22−26.
(10) Seeman, N. C. *Nature* **2003**, *421*, 427−431.
(11) Seeman, N. C. *Sci. Am.* **2004**, *290*, 64.
(12) Seeman, N. C.; Belcher, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6451−6455.
(13) Schneider, B.; Neidle, S.; Berman, H. M. *Biopolymers* **1997**, *42*, 113−124.
(14) Gorb, L.; Shishkin, O.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **2005**, *22*, 441−454.
(15) Grzeskowiak, K.; Goodsell, D. S.; Kaczor-Grzeskowiak, M.; Cascio, D.; Dickerson, R. E. *Biochemistry* **1993**, *32*, 8923.
(16) Schneider, B.; Kabelac, M.; Hobza, P. *J. Am. Chem. Soc.* **1996**, *118*, 12207− 12217.
(17) Subirana, J. A.; Soler-Lopez, M. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 27−45.
(18) Varnai, P.; Zakrzewska, K. *Nucleic Acids Res.* **2004**, *32*, 4269−4280.
(19) Wittung, P.; Nielsen, P. E.; Buchardt, O.; Egholm, M.; Norden, B. *Nature* **1994**, *368*, 561−563.
(20) Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Freier, S. M.; Driver, D. A.; Berg, R. H.; Kim, S. K.; Norden, B.; Nielsen, P. E. *Nature* **1993**, *365*, 566−568.
(21) Egholm, M.; Behrens, C.; Christensen, L.; Berg, R. H.; Nielsen, P. E.; Buchardt, O. *J. Chem. Soc.−Chem. Commun.* **1993**, 800−801.
(22) Egholm, M.; Buchardt, O.; Nielsen, P. E.; Berg, R. H. *J. Am. Chem. Soc.* **1992**, *114*, 1895−1897.
(23) Egholm, M.; Nielsen, P. E.; Berg, R. H.; Buchardt, O. *J. Cell. Biochem.* **1993**, 212−212.
(24) Egholm, M.; Nielsen, P. E.; Buchardt, O.; Berg, R. H. *J. Am. Chem. Soc.* **1992**, *114*, 9677−9678.
(25) Nielsen, P. E.; Egholm, M. *Bioorg. Med. Chem.* **2001**, *9*, 2429−2434.
(26) Nielsen, P. E.; Egholm, M.; Buchardt, O. *Bioconjugate Chem.* **1994**, *5*, 3−7.

(27) Dueholm, K. L.; Egholm, M.; Behrens, C.; Christensen, L.; Hansen, H. F.; Vulpius, T.; Petersen, K. H.; Berg, R. H.; Nielsen, P. E.; Buchardt, O. *J. Org. Chem.* **1994**, *59*, 5767−5773.
(28) Leijon, M.; Graslund, A.; Nielsen, P. E.; Buchardt, O.; Norden, B.; Kristensen, S. M.; Eriksson, M. *Biochemistry* **1994**, *33*, 9820−9825.
(29) Lukeman, P. S.; Mittal, A. C.; Seeman, N. C. *Chem. Commun.* **2004**, 1694−1695.
(30) Rasmussen, H.; Kastrup, J. S.; Nielsen, J. N.; Nielsen, J. M.; Nielsen, P. E. *Nat. Struct. Biol.* **1997**, *4*, 98.
(31) Griffin, T. J.; Smith, L. M. *Anal. Biochem.* **1998**, *260*, 56−63.
(32) Tomac, S.; Sarkar, M.; Ratilainen, T.; Wittung, P.; Nielsen, P. E.; Norden, B.; Graslund, A. *J. Am. Chem. Soc.* **1996**, *118*, 5544−5552.
(33) Sponer, J.; Berger, I.; Spackova, N.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **2000**, 383−407.
(34) Sponer, J.; Burda, J. V.; Mejzlik, P.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1997**, *14*, 613−628.
(35) Sponer, J.; Florian, J.; Hobza, P.; Leszczynski, J. *J. Biomol. Struct. Dyn.* **1996**, *13*, 827−833.
(36) Sponer, J.; Gabb, H. A.; Leszczynski, J.; Hobza, P. *Biophys. J.* **1997**, *73*, 76−87.
(37) Sponer, J.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 3161−3164.
(38) Sponer, J.; Hobza, P. *Chem. Phys.* **1996**, *204*, 365−372.
(39) Sponer, J.; Hobza, P. *Chem. Phys. Lett.* **1997**, *267*, 263−270.
(40) Sponer, J.; Hobza, P. *Collect. Czech. Chem. Commun.* **2003**, *68*, 2231−2282.
(41) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1996**, *14*, 117−135.
(42) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965−1974.
(43) Sponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers* **2001**, *61*, 3−31.
(44) Sponer, J.; Leszczynski, J.; Hobza, P. *THEOCHEM* **2001**, *573*, 43−53.
(45) Wesolowski, S. S.; Leininger, M. L.; Pentchev, P. N.; Schaefer, H. F. *J. Am. Chem. Soc.* **2001**, *123*, 4023−4028.
(46) Richardson, N. A.; Wesolowski, S. S.; Schaefer, H. F. *J. Am. Chem. Soc.* **2002**, *124*, 10163−10170.
(47) Richardson, N. A.; Wesolowski, S. S.; Schaefer, H. F. *J. Phys. Chem. B* **2003**, *107*, 848−853.
(48) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
(49) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.
(50) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.
(51) Hariharan, P. C.; Pople, J. A. *Mol. Phys.* **1974**, *27*, 209.
(52) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149- 1170.
(53) Evaluation of counterpoise correction for the PNA_GC_PNA base pair using B3LYP/6-31+G(d,p) gives a basis set superposition error <1 kcal/mol.
(54) Sponer, J.; Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2004**, *126*, 10142−10151.
(55) Berman; Westbrook; Feng; Gilliland; Bhat; Weissig; Shindyalov; Bourne. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(56) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995−2001.
(57) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669−681.
(58) Cances, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032−3041.
(59) Mennucci, B.; Cances, E.; Tomasi, J. *J. Phys. Chem. B* **1997**, *101*, 10506−10517.
(60) Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *106*, 5151−5158.
(61) Rega, N.; Cossi, M.; Barone, V. *J. Comput. Chem.* **1999**, *20*, 1186−1198.
(62) Takano, Y.; Houk, K. N. *J. Chem. Theory Comput.* **2005**, *1*, 70−77.
(63) Li, H.; Jensen, J. H. *J. Comput. Chem.* **2004**, *25*, 1449−1462.
(64) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. GAUSSIAN 03; Gaussian, Inc.: Pittsburgh, PA, 2003.
(65) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *J. Am. Chem. Soc.* **2000**, *122*, 4117−4128.
(66) Asensio, A.; Kobko, N.; Dannenberg, J. J. *J. Phys. Chem. A* **2003**, *107*, 6441−6443.
(67) Kawahara, S.; Uchimaru, T. *PCCP Phys. Chem. Chem. Phys.* **2000**, *2*, 2869−2872.
(68) Kawahara, S.; Uchimaru, T. *Eur. J. Org. Chem.* **2003**, 2577−2584.
(69) Hobza, P.; Sponer, J. *Chem. Rev.* **1999**, *99*, 3247−3276.
(70) Olah, G. A. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 767−788.
(71) Lammertsma, K.; Schleyer, P. V.; Schwarz, H. *Angew. Chem., Int. Ed. Engl.* **1989**, *28*, 1321−1341.
(72) Gurlie, R.; Zakrzewska, K. *J. Biomol. Struct. Dyn.* **1998**, *16*, 605.
(73) Gurlie, R.; Zakrzewska, K. *Theor. Chem. Acc.* **2001**, *106*, 83−90.
(74) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999−3093.
(75) Arora, N.; Jayaram, B. *J. Phys. Chem. B* **1998**, *102*, 6139−6144.
(76) Hasegawa, K.; Noguchi, T. *Biochemistry* **2005**, *44*, 8865−8872.