Py*Pu·Py Type Triplexes with Modified Bases: Ab Initio SCF-MO Studies toward Improved DNA Recognition

Divi Venkateswarlu*

Department of Medicinal Chemistry, University of Minnesota, Minneapolis, Minnesota 55455

Jerzy Leszczynski*

The Computational Center for Molecular Structure and Interactions, Department of Chemistry, Jackson State University, Jackson, Mississippi 39217

Received: November 9, 1998; In Final Form: February 24, 1999

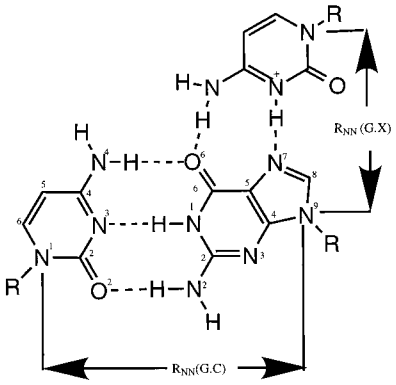
Ab initio quantum mechanical calculations were employed to examine the binding energies and geometries of several nonnatural base analogues of nucleic acid bases bound to the major groove of the G·C base-pair in an attempt to explore promising candidate bases for triplex-helix-forming oligonucleotides. Seven neutral base analogues together with guanine and protonated forms of cytosine and adenine are considered. The full geometry optimizations were carried out at the HF/6-31G(d) level, and single-point energies were obtained at the MP2/6-31G(d,p)//HF/6-31G(d) level. Our analysis reveals that protonated cytosine (C^+) has the highest binding energy over the other bases. C^5-methylation of protonated cytosine does not improve the base-pairing efficiency over protonated cytosine. Among the neutral bases, the binding efficiency of 6-thioguanine is the closest to that of guanine in the Py*Pu·Py type motif (1.4 kcal/mol less than guanine at MP2/6-31G(d,p)//HF/6-31G(d)), and 8-oxoadenine ranked the least. Semiempirical AM1 and PM3 methods compare qualitatively well with the energetic trends of that predicted by the HF/6-31G(d) method. At the MP2/6-31G(d,p)/HF/6-31G(d) (BSSE and ZPE corrected) level, the following order of major-groove binding affinity is established. C^+*G·C > $^{5M}C^+$*G·C > A^+*G·C > G*G·C > CX1*G·C > $^{6-thio}$G*G·C > $^{8-thio}$A*G·C > *$^{6-oxo}$C*G·C > $^{5M,6-oxo}$C*G·C > $^{8-oxo}$A*G·C. Close consideration of the structural isomorphism of these triplexes with native C^+*G·C and T*A·T and the major-groove binding efficiency of the bases suggests that the neutral form of 8-thioadenine ($^{8-thio}$A) may be considered as an alternative to protonated cytosine for efficient triple-helix formation within the Py*Pu·Py type parallel motif.

Introduction

The current interest in exploring the potential of triplex-helix-forming oligonucleotides (TFO) as a viable strategy for control of gene expression stems from the pioneering studies of the Helene and Dervan groups, who have shown that the short stretches of oligonucleotides can bind to a DNA duplex in a sequence-specific manner.[1−3] While a number of studies have been directed toward understanding the structural aspects of DNA triplexes,[4−6] the successful application of TFO's in antigene therapy is still elusive. The major hurdles for efficient use of TFO's for antigene applications at the cellular level are the prerequirement of protonation of nucleic acid bases at a biologically inaccessible low pH,[7] vulnerability of the backbone for nuclease degradation,[8] and poor affinity to the target.[9] To overcome several of these problems in a systematic and rational manner, a large number of studies are being directed toward improving the triple-helix-forming oligonucleotides.[10−12] A number of modifications to three major constituents of oligonucleotides such as nucleobase,[10] sugar, and the phosphate backbone[11,12] have been addressed in order to impart better "recognition" and stability. While it is difficult to ascertain the relative importance of the three modifications, clearly the specific base-pairing potential of the TFO base is, arguably, one of the predominant factors in govering the overall structure and stability of the hybrid triplex through specific hydrogen-bonding interactions.[13,15]

According to the third-strand orientation, in principle, two types of DNA triple-helices were characterized. In the purine*purine·pyrimidine (Py*Pu·Py) family, a purine-rich third strand is bound *antiparallel* to the purine strand of the target duplex. In the pyrimidine*purine·pyrimidine family, a homopyrimidine third strand is bound *parallel* to the purine strand of the target duplex (Pu·Py) in the major groove of DNA through isomorphous C^+*G·C or T*A·T base triplets. The major groove region of the G·C pair involves three hydrogen-bonding sites of which two are acceptors (N^7 and O^6 atoms of guanine), and one is donor (N^4-hydrogen of cytosine). The application of the Py*Pu·Py type motifs in homopurine triplexes is limited because of the prerequirement of the protonation of cytosine,[7] which is required in order to establish two hydrogen bonds with guanine. The pK_a of isolated cytidine is 4.3, and even though it could be slightly higher in oligonucleotides, it still becomes a major hurdle for effective binding at intracellular pH, which is about 7.3. Owing to such a limitation, attempts have been made to generate "*pyrimidine-like*" structures. A number of cytosine analogues and mimics have been synthesized and tested for their thermodynamic properties. For example, 5-methylcytosine, which has a slightly higher pK_a value than cytosine, generates triplexes that are more stable at a higher pH but are still not formed under physiological conditions.[8,9,16−18] Some of the neutral analogues of protonated cytosine that do not require protonation include 6-oxocytosine,[19] pseudoisocytosine,[20] N^7-

10.1021/jp9843508 CCC: $18.00
Published on Web 04/15/1999



C.G*C$^+$

Figure 1. Numbering system for C$^+$*G·C triplet and some of the interbase configurational parameters used in the study.

guanine,[21] and 8-oxoadenine.[22] While many of these neutral bases have not been widely tested, the experimental studies on 8-oxoadenine indicate that the triplexes containing 8-oxoadenine (in syn conformation) are stable even at the elevated pH of about 7.4 and are structurally very similar to protonated cytosine.[22,23]

To better understand the influence of modified bases on the overall structure and stability of the triplex, we have studied a number of modified nucleic acid base analogues as possible G·C major-groove recognition candidates by employing accurate nonempirical ab initio methods. In this study, we examine the binding energies of several TFO bases at the HF/6-31G(d) and MP2/6-31G(d,p)//HF/6-31G(d) levels with correction of the basis-set superposition errors (BSSE) and the overall structural isomophism observed by the modified bases in comparison with native C$^+$*G·C triads.

Theoretical Methodology

Full geometry optimizations of the studied triplexes were carried out at HF/6-31G(d) without imposing any geometric restraints, and single-point energies were evaluated at the MP2/6-31G(d,p) level using HF/6-31G(d) reference geometries. All calculations are performed using the Gaussian94 suite of programs.[24] The interaction energies (E_p) were estimated as the energy required for binding of the third base to the major groove face of the G·C pair:

$$E_p = E_{\text{triplex}} - E_{\text{G·C}} - E_{\text{TFO base}} \qquad (1)$$

The final relative energies calculated at the MP2/6-31G(d,p)//HF/6-31G(d) level were corrected for both BSSE and zero-point energies (ZPE). Basis-set superposition errors (BSSE) were corrected using the counterpoise correction method.[25] The ZPE corrections were made on the basis of the HF/3-21G optimized geometries of the triplexes and individual TFO bases. All optimized structures were characterized as HF/3-21G equilibrium structures and true minima, as indicated by the only positive eigenvalues of the Hessian matrix. The zero-point energies were scaled by the recommended factor of 0.9.[26] We have also estimated the interaction energies of base trimers using the semiempirical AM1[27a] and PM3[27b] methods essentially to see how well they fare when compared with nonempirical ab initio methods. The intermolecular configurational parameters such as the interglycosidic bond distances among the three bases and the hydrogen-bonding distances were calculated to evaluate the structural perturbations caused by the modified bases. The configurational parameters evaluated in the present study are schematically represented for the C·G*C$^+$ trimer in Figure 1.

Results and Discussion

The base-pairing properties of protonated cytosine and adenine together with seven neutral forms of nucleic acid bases and analogues are studied. The TFO bases considered in the present study are protonated cytosine (C$^+$), C^5-methylated cytosine (^{5M}C$^+$), protonated adenine (A$^+$), guanine (G), 8-oxoadenine ($^{8-oxo}$A), 8-thioadenine ($^{8-thio}$A), 6-oxocytosine ($^{6-oxo}$C), 5-methyl-6-oxocytosine ($^{5M,6-oxo}$C), and a C-pyridine derivative (CX1). The structural forms adopted by these bases with the G·C major groove within the parallel Py*Pu·Py type motif is schematically represented in Figures 2 and 3.

Binding Energies. Because of their fundamental importance, the structure and base-pairing properties of nucleic acids have resulted in a number of theoretical studies.[28−31] In the absence of accurate experimental information, nonempirical ab initio methods have been shown to be reliable sources for estimating the base-pair interactions of nucleic acid duplexes[28,29] and triplexes.[30,31] The first quantum mechanical study of H-bonded triads of nucleobases appeared as early as 1967.[32] However, the first ab initio study using the minimal STO-3G basis set was carried out by Jiang et al.[33] They have studied both the neutral and protonated trimers, but the monomers within the triplex were not allowed to relax during complexation. Sponer et al have also studied similar systems at the HF and MP2 levels using the 6-31G(d) basis set.[30] Recently, the electrostatic-potential-based charge models have also been used to study the interaction energies of the canonical DNA triplexes.[34] While these studies provide good insight into the energy decomposition of DNA triads, they do not give direct estimates of the binding potential of the TFO base in the major groove of the G·C base-pair. However, most of these studies come to the general conclusion that the protonated bases have better interaction energies than neutral bases within the DNA triad. Our present studies on both natural and modified base trimers are performed without imposing any symmetry constraints, and all systems are fully optimized. The estimated binding energies are tabulated in Table 1 together with the BSSE and ZPE corrected energies. The data clearly point out that the protonated bases have higher binding energies than the neutral bases and base analogues. The influence of basis-set superposition errors and electron correlation may also be observed to play an important role in the calculation of binding energies. In all cases, inclusion of electron correlation leads to an overall increase in the binding energies of the TFO bases by up to 4 kcal/mol. It is evident from the data that at all levels of theory protonated cytosine has the highest binding energy than the rest of the bases considered in the present study. We estimate that at the HF/6-31G(d) level, protonated cytosine has a binding energy of about 41.0 kcal/mol while inclusion of the electron correlation and BSSE effects improve the binding efficiency by about 3.9 kcal/mol. Despite its high binding affinity as mentioned above, protonated cytosine is considered to be a poor choice for efficient triplex formation essentially because of two factors. Triplexes containing protonated cytosine are stable only when the pH of the medium is well below 6 and has been shown to strongly destabilize the triplex at a biological pH (about 7.1). Another destabilizing factor stems from the electrostatic repulsion in the contiguous appearance of protonated cytosines in triplexes of homopurine stretches.[35]

C^5-methylation of cytosine has been considered as a possible choice to increase the pK_a of cytosine.[8,9] It has been shown that C^5-methylation leads to a slight increase in the pK_a of cytosine. However, it is interesting to observe that C^5-methylation leads to a slight lowering in the binding than the corresponding

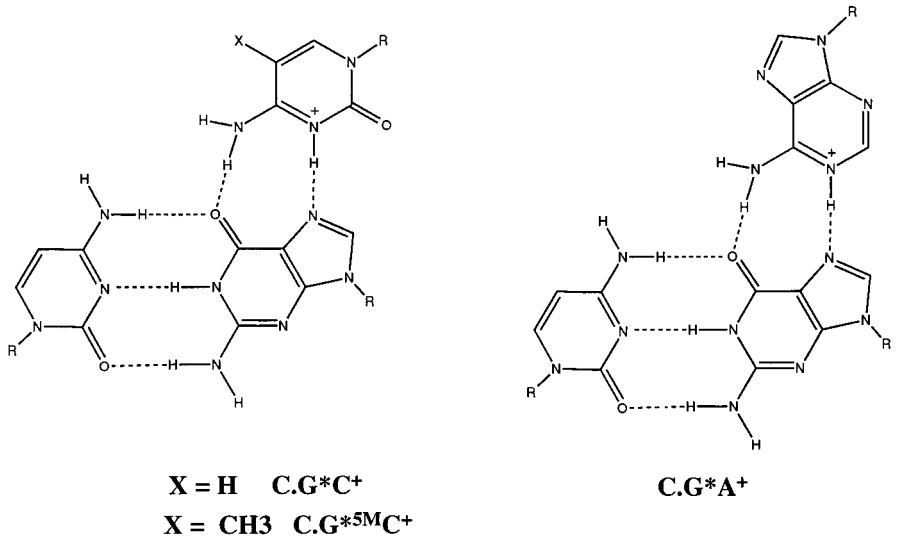


Figure 2. Schematic representation of the studied protonated base triplets.

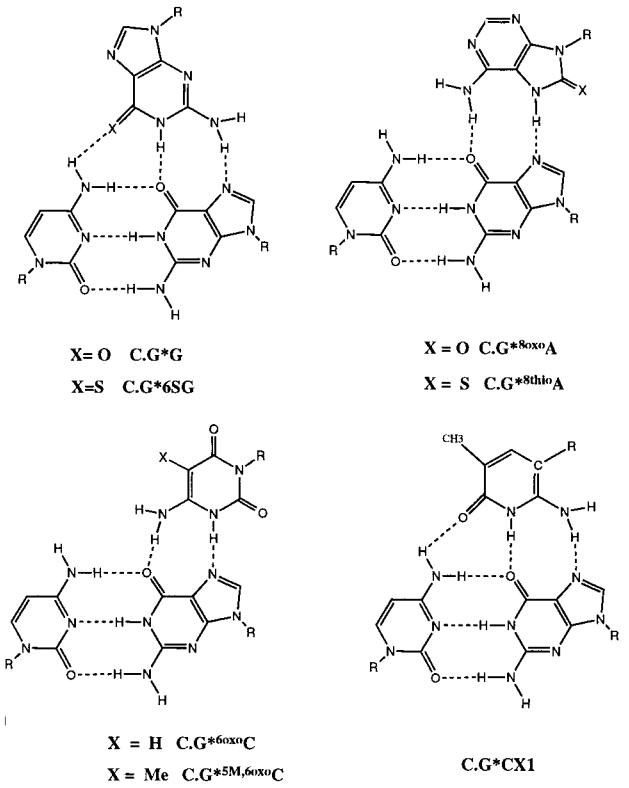


Figure 3. Schematic representation of the studied triplets containing neutral base candidates.

unmethylated cytosine. The weak electron-donating nature of the methyl group might be attributed to the overall lowering of the positive charge on nitrogen (N^3-site), which would consequently reduce the hydrogen-bonding potential of the N^3-H^3 bond in the major groove (Figure 1). We estimate that methylation leads to a lowering in the binding efficiency by about 0.8 kcal/mol at the MP2/6-31G(d,p)//HF/6-31G(d) level. It is worth mentioning that experimental studies point out that C^5-methylation increases the stability of the triplex. However, considering the destabilizing effect of C^5-methylation on the major groove, the binding affinity of the TFO base indicates that the stability of these triplexes perhaps originates from other sources such as the favorable entropic contributions due to the presence of a hydrophobic methyl group at the C^5 position.[18] Adenine has a basic pK_a value of about 5.5 and is considered as an alternative to cytosine. Interestingly, despite being a purine base, protonated adenine (A^+) is about 2.0 kcal/mol less efficient in binding than protonated cytosine.

Out of the seven neutral analogues considered in the present study, guanine possesses the highest binding energy in the major groove than the rest of the bases. However, it is immediately evident from the structural parameters shown in Table 2 that guanine cannot be a good base candidate within the Py*Pu·Py type parallel motif because of its marked deviations from the structural patterns shown by the C^+*G·C triad. Moreover, guanine-rich TFO's are known to be involved in self-association by a stacking of K^+/guanine quartets that are stabilized by a combination of mutual bidendate H-bonding and coordination of the four O^6 atoms of guanosine to K^+ located at the center of the quadruplex.[36]

Introduction of the oxo group at the C^6-position of cytosine would render the N^3-site of cytosine to be protonated in the neutral form (Figure 3). Despite the encouraging structural similarity observed between 6-oxocytosine[19] and protonated cytosine, both 6-oxocytosine and its C^5-methylated analog[37] are shown to have very weak binding in the major groove of the C·G base-pair. Our calculations estimate that $^{6-oxo}$C and $^{5M,6-oxo}$C are 4.3 and 4.8 kcal/mol less efficient than guanine. As is observed in the case of cytosine, here too C^5 methylation of $^{6-oxo}$C leads to only a destabilization in binding (about 0.4 kcal/mol when compared with 6-oxocytosine). Apparently, the methylation of nucleic acid bases turns out to weaken the overall binding efficiency of any TFO base in the major groove. Another interesting base analogue is the substituted C-nucleoside, 2-aminopyridine (CX1), which is an isocytosine analogue. The chief interest in using this base is that it is more basic than either cytosine or its C^5-methylated analogue. Also, the replacement of the C−N glycosidic bond with the less labile C−C bond makes it less vulnerable to deglycosylation. As shown in Figure 1, this base can be expected to form three hydrogen bonds within the G·C major groove, similar to guanine. It has good binding energy which is only about 1.5 kcal/mol less than that of guanine. While this base has slightly lower binding than 6-thioguanine, inclusion of ZPE corrections render it to have more binding than 6-oxoguanine by about 0.3 kcal/mol. However, as is evident from the configurational parameters shown in Table 2, this analogue may be expected to slightly destabilize the triplex because of its short interglycosidic distances with the guanine base of the Watson−Crick duplex.

TABLE 1: Major-Groove Binding Energies of Nucleic Acid Bases (in kcal/mol) Estimated at Various Levels of Theory[a]

triplex motif	HF/6-31G(d)	HF/6-31G(d)[b] (BSSE)	MP2/6-31G(d,p)[b] (BSSE)	AM1	PM3
protonated base					
C^+*G·C	−41.01 (0.0)	−41.51 (0.00)	−44.87 (0.00)	−35.38 (0.00)	−26.88 (0.00)
$^{5M}C^+$*G·C	−40.41 (0.6)	−40.83 (0.68)	−44.03 (0.84)	−34.67 (0.81)	−26.02 (0.86)
A^+*G·C	−39.01 (2.0)	−39.49 (2.02)	−42.85 (2.02)	−33.61 (1.77)	−25.27 (1.61)
neutral base					
G*G·C	−22.88 (0.00)	−21.08 (0.00)	−23.29 (0.00)	−13.45 (0.00)	−16.44 (0.00)
$^{6-thio}$G*G·C	−20.49 (2.39)	−20.48 (0.60)	−21.89 (1.40)	−13.11 (0.34)	−15.96 (0.48)
CX1*G·C	−20.34 (2.54)	−20.04 (1.04)	−21.82 (1.47)	−10.92 (2.53)	−15.95 (0.49)
$^{8-thio}$A*G·C	−18.80 (4.08)	−18.19 (2.89)	−21.19 (2.10)	−9.30 (4.15)	−9.37 (7.07)
$^{6-oxo}$C*G·C	−18.99 (3.89)	−17.50 (3.58)	−18.96 (4.33)	−10.68 (2.77)	−10.48 (5.96)
$^{5M,6-oxo}$C*G·C	−17.87 (5.01)	−16.95 (4.13)	−18.54 (4.75)	−9.67 (3.78)	−9.93 (6.51)
$^{8-oxo}$A*G·C	−16.22 (6.66)	−15.33 (5.75)	−18.08 (5.21)	−8.96 (4.49)	−8.03 (8.41)

[a] The energies relative to either C^+*G·C or G*G·C are in parentheses. [b] Computed at the HF/6-31G(d) geometries.

TABLE 2: Structural Parameters for Triple-Helical Motifs (All Distances Are in Angstroms)

motif	RNN(G·C)[a] RNN(G·X) RNN(C·X)	C·G, N2G·O2C N1G·N3C O6G·N4C	C·G·X, N7G·D−X[b] O6G·D−X N4C·A−X
G·C	9.05	2.02/3.02	
		2.04/3.04	
		1.92/2.93	
C^+*G·C	9.03	1.90/2.91	1.82/2.84
	6.78	2.04/3.05	1.85/2.84
	12.31	2.10/3.10	-
$^{5M}C^+$*G·C	9.03	1.91/2.91	1.82/2.84
	6.76	2.04/3.05	1.86/2.85
	12.31	2.09/3.09	
G*G·C	9.03	1.95/2.95	2.28/3.23
	10.04	2.01/3.03	1.99/2.99
	11.31	1.99/2.99	2.17/2.87
A^+*G·C	9.02	1.89/2.90	1.91/2.93
	8.92	2.04/3.05	1.84/2.84
	12.85	2.10/3.10	
$^{6-thio}$G*G·C	9.05	1.94/2.95	2.62/3.24
	10.08	2.02/3.03	2.47/3.35
	11.72	2.02/3.02	2.66/3.56
CX1*G·C	9.05	1.96/2.97	2.55/3.38
	7.95	2.02/3.03	2.12/3.07
	10.09	1.99/2.99	2.05/2.89
$^{8-thio}$A*G·C	9.04	1.98/2.98	2.00/3.01
	6.84	2.04/3.04	2.14/3.02
	12.58	1.98/2.99	
$^{6-oxo}$C*G·C	9.04	1.98/2.98	1.95/2.95
	6.77	2.03/3.04	2.22/3.19
	12.52	1.98/2.98	
$^{5M,6-oxo}$C*G·C	9.04	1.98/2.99	1.96/2.96
	6.75	2.03/3.04	2.24/3.20
	12.50	1.98/2.98	
$^{8-oxo}$A*G·C	9.04	1.99/2.99	1.96/2.95
	6.70	2.03/3.04	2.29/3.21
	12.54	1.97/2.98	

[a] R_{NN} is the interglycosidic nitrogen distance between any two bases in the triad, and X corresponds to the glycosidic atom of the TFO base. [b] D and A represent donor and acceptor sites in the TFO base in the C·G major groove, if present.

As an alternative to guanine, 6-thioguanine may be considered as a possible TFO base for triplex formation with the C·G strand in a parallel motif. One of the main advantages of using 6-thioguanine would be to repel the K^+ ion aggregation around the 6-oxo group of guanine,[36] which is known to increase the possibility of the unsought guanine quadruplex formation[39] as discussed above. Though the incorporation of the thio group at the O^6-position would result in better binding than the rest of the neutral analogues (excluding guanine), as discussed in the following section, it may be expected to disturb the DNA triplex because of the wider interglycosidic distances similar to that of guanine (Table 2). As an alternative, the use of 8-oxoadenine (^{8oxo}A) would significantly reduce the problem of structural dissimilarity when compared with guanine. It has been demonstrated that 8-oxoadenine prefers to adopt a syn conformation due to the presence of a bulkier oxo group at the C^8 position[38] and would easily bind to the G·C major groove in syn fashion, which is much closer to the C^+*G·C type motif. Recent experimental studies indicate that the TFOs containing 8-oxoadenine bind to the DNA duplex with good affinity even at an elevated pH.[39,40] Encouragingly, as shown in Table 2, ^{8oxo}A has remarkable structural similarity to the C^+*G·C triad. However, the binding of ^{8oxo}adenine is about 5.2 and 3.8 kcal/mol less efficient than guanine and 6-thioguanine, respectively, making it a poorer choice. Interestingly a thio substitution at the C^8 position of adenine turns out to be a good choice due to its remarkable structural similarity to C^+*G·C and better binding than 8-oxoadenine. Incorporation of the thio group at the C^8 position instead of the oxo group improves the binding by about 3.1 kcal/mol (at the MP2/6-31G(d,p)//HF/6-31G(d) level). Apart from this, as discussed in the following section, its structural similarity to the C^+*G·C triplet makes it a promising candidate base for triplex formation, perhaps without imparting any global structural distortion to the triplex.

Finally, we add the discussion about the performance of semiempirical methods in estimating the binding affinity of the TFO bases in the GC major groove in comparison with the ab initio data. Semiempirical methods have long been used as a good choice for compromise between the computational efficiency and accuracy of the electronic properties.[41,42] The binding energies are computed by the AM1 and PM3 SCF-MO methods according to eq 1 by a similar procedure used above and are tabulated in the last two columns of Table 1. It has been shown by several groups that the hydrogen-bond strength of van der Waals complexes, particularly the nucleic acid bases, is systematically underestimated by about 6 or 7 kcal/mol with a corresponding overestimation of the hydrogen-bond distances.[43] However, as shown in Table 1, the relative order of the binding energies predicted by both AM1 and PM3 methods deserves close attention. In the case of protonated bases, both AM1 and PM3 methods predict the same order of binding energy as that predicted by the HF and MP2/6-31G(d) methods. Encouragingly, even among the triplexes containing the neutral bases, the relative order is largely preserved except in case of $^{8-thio}$A*G·C, which is predicted by both AM1 and PM3 methods to possess less binding efficiency over $^{6-oxo}$C. Indeed the calculations by ab initio methods using the HF/6-31G(d) basis set too leads to the prediction of an erroneous trend, which may be seen to have reversed upon considering the basis-set superposition errors. While it is obvious from the data that

though caution should be exercised when using the AM1 or PM3 methods for calculating the hydrogen-bonding energies of nucleic acid triplexes, it is encouraging to see a qualitatively good trend in the estimation of the binding energies of nucleic acid base triplets.

Structural Isomorphism among Base Triads. The structural isomorphism in DNA base triads is one of the predominant factors that are responsible for the overall stability of the triplex structure.[44,45] The T*A·T and C^+*G·C triplets are isohelical,[5,44] and in both cases the third strand base (T or C^+) is held by two hydrogen bonds. As shown in Table 2, the interglycosidic bond distances of T*A·T and C^+*G·C triads are quite similar and are thus conformationally isomorphic. It should be borne in mind that the design of any new nonnatural base analogue should be structurally isomorphic with C^+*G·C (for parallel orientation to the G strand). Triplets are isomorphous if their glycosidic bonds are superimposable. Any nonisomorphism induced by nonnatural/natural bases would conceivably lead to profound triplex destabilization for random purine/pyrimidine sequences due to significant ribose−phosphate backbone distortion. To understand the overall structural perturbations caused by the modified bases, the configurational parameters such as the intermolecular hydrogen bond distances and interglycosydic bond distances of the base triads are evaluated and presented in Table 2. Close observation of the data points out that most of the pyrimidine bases are conformationally close to both T*A·T or C^+*G·C while the purine base analogues show larger deviations. It is interesting to observe that both 8-oxoadenine and 8-thioadenine show remarkably structural similarity to C^+*G·C. Among the purine bases, both guanine and 6-thioguanine show large deviations in interglycosidic bond distances. Although the guanine base in the third strand has the highest binding energy among the neutral bases, its conformational dissimilarity with the C^+*G·C triad makes it a poor choice for TFO formation within the parallel motif. Replacement of the exocyclic oxygen atom of guanine with sulfur weakens the binding strength of guanine. It is reflected through the longer hydrogen-bond distances in the major groove of $^{6-thio}$G*G·C. Despite the overall lowering of binding efficiency upon C^5-methylation of protonated cytosine or 6-oxocytosine, neither of the bases leads to any noticeable changes in the intermolecular parameters. The binding of any of the bases in the major groove of the C·G base-pair does not appear to show any destabilizing effect on the Watson−Crick base-pairing of the target C·G duplex pair, as revealed by the intermolecular hydrogen-bond distances of the C·G base-pair that seem very similar in all base-triads considered in the present study.

Conclusions

The structure and binding properties of several nucleic acid bases and their neutral analogues in the major groove of the guanine·cytosine base-pair, within the pyrimidine*purine·pyrimidine parallel motif, have been studied by using ab initio SCF-MO studies at the HF/6-31G(d) level. The energies are estimated with inclusion of the basis-set superposition errors and the electron correlation effects at the MP2 level of theory. These studies reveal some important base-pairing properties of modified nucleic acid base derivatives. Although purine bases such as guanine do not fit within the pyrimidine*purine·pyrimidine type parallel triplexes, the striking structural similarity between the triplexes containing 8-thioadenine or 8-oxoadenine and the canonical C^+*G·C or T *A·T triplets is particularly promising. While the stability of the triplexes containing 8-thioadenine has to be yet tested, our studies clearly suggest that 8-thioadenine binds to the GC major groove by about 3.1 kcal/mol more than 8-oxoadenine, indicating that it might be a potential candidate base for triple-helix-forming oligonucleotides. Among all the bases considered in the present study, at the MP2/6-31G(d,p)//HF/6-31G(d) (BSSE and ZPE corrected) level the following order of binding efficiency is obtained: C^+*G·C > $^{5M}C^+$*G·C > A^+*G·C > G*G·C > CX1*G·C > $^{6-thio}$G*G·C > $^{8-thio}$A *G·C > $^{6-oxo}$C *G·C > $^{5M,6-oxo}$C*G·C > $^{8-oxo}$A*G·C.

Acknowledgment. We thank Dr. Jiri Sponer for kindly providing some of the starting coordinates of DNA triplexes and for helpful discussions. The authors thank the Mississippi Center for Supercomputing Research for the computational facilities. This work was facilitated in part by the NSF Grant OSR-94527857 and by the Office of Naval Research, Grant N00014-95-1-0049 and the Army High Performance Computing Research Center under the auspices of the Department of the Army, Army Research Laboratory Cooperative Agreement DAAH04-95-2-0003/Contract DAAH04-95-C-0008, the content of which does not necessarily reflect the position or policy of the government, and no official endorsement should be inferred. D.V. thanks Prof. David M. Ferguson of the University of Minnesota for his encouragement. The Mississippi Center for Supercomputing Research is acknowledged for a generous allocation of computer time.

References and Notes

(1) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645.

(2) Le Doan, T.; Perrouault, L.; Praseuth, D.; Habhub, N.; Decout, J. L.; Thuong, N. T.; Lhomme, J.; Helene, C. *Nucleic Acids Res.* **1987**, *15*, 7749.

(3) Sun, J.-S.; Garestier, T.; Helene, C. *Curr. Opin. Struct. Biol*. **1996**, *6*, 327.

(4) (a) Lee, J. S.; Woodsworth, M. L.; Latimer, L. J. P.; Morgan, A. R. *Nucleic Acids Res*. **1984**, *12*, 6603. (b) Broitman, S. L.; Im, D. D.; Freso, J. R. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 5120. (c) Htun, H.; Dahlberg, J. E. *Science* **1988**, *241*, 1791. (d) Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science* **1988**, *241*, 456. (e) Griffin, L. C.; Dervan, P. B. *Science* **1989**, *245*, 967. (f) Durland, R. H.; Kessler, D. J.; Gunnel, S.; Duvic, M.; Pettitt, B. M.; Hogan, M. E. *Biochemistry* **1991**, *30*, 9250.

(5) (a) Arnott, S.; Bond, P. J. *Nature (London), New Biol*. **1973**, *244*, 99. (b) Arnott, S.; Selsing, E. *J. Mol. Biol*. **1974**, *88*, 509.

(6) (a) Raghunathan, G.; Miles, H. T.; Sasisekharan, V. *Biochemistry* **1993**, *32*, 455. (b) Howard, F. B.; Miles, H. T.; Liu, K.; Frazier, J.; Raghunathan, G.; Sasisekharan, V. *Biochemistry* **1992**, *31*, 10671. (c) Liu, K.; Frazier, J.; Miles, H. T.; Sasisekharan, V. *Biochemistry* **1993**, *9*, 11802.

(7) Felsenfeld, G.; Davies, D. R.; Rich, A. *J. Am. Chem. Soc.* **1957**, *79*, 2023.

(8) Lee, J. S.; Wordsworth, M. L.; Latimer, L. J.; Morgan, A. R. *Nucleic Acids Res.* **1984**, *12*, 6603.

(9) Boussif, O.; Lezoualc'h, F.; Zanta, M. A.; Mergny, M. D.; Scherman, D.; Demeneix, B.; Behr, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 7297.

(10) (a) Froehler, B. C.; Ricca, D. J. *J. Am. Chem. Soc.* **1992**, *114*, 8320. (b) Huang, C.-Y.; Bi, G.; Miller, P. S. *Nucleic Acids Res*. **1996** *24*, 2606. (c) Ono, A.; Ts'o, P. O. P.; Kan, L. *J. Am. Chem. Soc*. **1991**, *113*, 4032. (d) Miller, P. S.; Bhan, P.; Cushman, C. D.; Trapane, T. L. *Biochemistry* **1992**, *31*, 2999. (e) Jetter, M. C.; Hobbs, F. W. *Biochemistry* **1993**, *32*, 3249.

(11) Sun, J.-S.; Garestier, T.; Helene, C. *Curr. Opin. Struct. Biol*. **1996**, *6*, 327.

(12) (a) Torres, R. A.; Almarsson, O.; Bruice, T. C. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 6875. (b) Almarsson, O.; Bruice, T. C.; Kerr, J.; Zuckermann, R. N. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 7518.

(13) Doronina, S. O.; Behr, J.-P. *Chem. Soc. Rev*. **1997**, 63.

(14) Cieplak, P.; Kollman, P. A. *J. Am. Chem. Soc*. **1988**, *110*, 3734.

(15) (a) Nakano, N. I.; Igarashi, S. J. *Biochemistry* **1970**, *9*, 577. (b) Crothers, D. M.; Zimm, B. H. *J. Mol. Biol*. **1964**, *9*, 1. (c) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *19*, 1149.

(16) Koh, J. S.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 1470.

(17) Povsic, T. J.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 3059.

(18) Xodo, L. E.; Manzini, G.; Quadifoglio, F.; Van der Marel, G.; Van Boom, J. *Nucleic Acids Res.* **1991**, *19*, 5625.

(19) Huang, C.-Y.; Bi, G.; Miller, P. S. *Nucleic Acids Res*. **1996**, *24*, 2606.
(20) Ono, A.; Ts'O, P. O. P.; Kan, L. *J. Am. Chem. Soc*. **1991**, *113*, 4032.
(21) Koshlap, K. M.; Schultze, P.; Brunar, H.; Dervan, P. B.; Feigon, J. *Biochemistry* **1997**, *36*, 2659.
(22) Miller, P. S.; Bhan, P.; Cushman, C. D.; Trapane, T. L. *Biochemistry* **1992**, *31*, 2999.
(23) Jetter, M. C.; Hobbs, F. W. *Biochemistry* **1993**, *32*, 3249.
(24) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C; Pople, J. A. *Gaussian 94*, revision E.2; Gaussian, Inc.: Pittsburgh, PA, 1995.
(25) Boys, S. F.; Bernardi, F. *Mol. Phys*. **1970**, *19*, 553.
(26) (a) Pople, J. A.; Schlegel, H. B.; Krishnan, R.; DeFrees, D. J.; Binkley, J. S.; Frisch, M. J.; Whiteside, R. A.; Hout, R. F.; Hehre, W. J. *Int. J. Quantum Chem., Quantum Chem. Symp*. **1989**, *15*, 269. (b) DeFrees, D. J.; McLean, A. D. *J. Chem. Phys*. **1985**, *82*, 333.
(27) (a) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc*. **1985**, *107*, 3902. (b) Stewart, J. J. P. *J. Comput. Chem*. **1989**, *10*, 209. (c) Stewart, J. J. P. *J. Comput. Chem*. **1989**, *10*, 221.
(28) (a) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn*. **1996**, *14*, 117 and references therein. (b) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem*. **1996**, *100*, 196.
(29) Gould, I. R.; Kollman, P. A. *J. Am. Chem. Soc*. **1994**, *116*, 2493.
(30) Sponer, J.; Sabat, M.; Burda, J. V.; Doody, A. M.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn*. **1998**, *16*, 139.
(31) Sponer, J.; Burda, J. V.; Mejzlik, P.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn*. **1997**, *14*, 613.
(32) Pullman, B.; Claverie, P.; Caillet, J. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *57*, 1663.
(33) Jiang, S.-P.; Jernigan, R. L.; Ting, K.-L.; Syi, J.-L.; Raghunathan, G. *J. Biomol. Struct. Dyn*. **1994**, *12*, 383.
(34) Gadre, S. R.; Pundlik, S. S. *J. Phys. Chem*. **1996**, *101*, 3298. Gadre, S. R.; Pundlik, S. S. *J. Phys. Chem. A* **1997**, *101*, 9657.
(35) Kiessling, L. L.; Griffin, L. C.; Dervan, P. B. *Biochemistry* **1992**, *31*, 2829.
(36) Olivas, W. M.; Maher, L. J. *Biochemistry* **1995**, *34*, 278.
(37) Xiang, G. B.; Soussou, W.; McLaughlin, L. W. *J. Am. Chem. Soc*. **1994**, *116*, 11155.
(38) (a) Culp, S. J.; Cho, B. P.; Kadlubar, F. F.; Evans, F. E. *Chem. Res. Toxicol*. **1989**, *2*, 416. (b) Cho, B. P.; Evans, F. E. *Nucleic Acids Res*. **1991**, *19*, 1041.
(39) Olivas, W. M.; Maher, L. J. *Nucleic Acids Res*. **1995**, *23*, 1936.
(40) (a) Jetter, M. C.; Hobbs, F. W. *Biochemistry* **1993**, *32*, 3249. (b) Miller, P. S.; Bhan, P.; Cushman, C. D.; Trapane, T. L. *Biochemistry* **1992**, *31*, 6788.
(41) (a) Leach, A. R.; Kollman, P. A. *J. Am. Chem. Soc*. **1992**, *114*, 3675. (b) Juranic, I.; Rzepa, H. S.; Yi, M. *J. Chem. Soc., Perkin Trans. 2* **1990**, 877.
(42) (a) Venkateswarlu, D.; Lyngdoh, R. H. D. *J. Chem. Soc., Perkin Trans. 2* **1995**, 839. (b) Venkateswarlu, D.; Lyngdoh, R. H. D. *J. Chem. Soc., Perkin Trans. 2* **1997**, 621. (c) Jurema, M. W.; Shields, G. C. *J. Comput. Chem*. **1993**, *14*, 89.
(43) Ford, G. P. *J. Mol. Struct.: THEOCHEM* **1997**, *401*, 253.
(44) Sekharudu, C. Y.; Yathindra, N.; Sundaralingam, M. *J. Biomol. Struct. Dyn*. **1993**, *11*, 225.
(45) Giovannangeli, C.; Rougee, M.; Garestier, T.; Thuong, N. T.; Helene, C. *Proc. Natl. Acad Sci. U.S.A.* **1992**, *89*, 8631.