

Gas-Phase DNA Oligonucleotide Structures. A QM/MM and Atoms in Molecules Study

Arturo Robertazzi and James A. Platts*

School of Chemistry, Cardiff University, Park Place, Cardiff CF10 3AT, U.K.

Received: November 16, 2005; In Final Form: January 17, 2006

QM/MM calculations have been employed to investigate the role of hydrogen bonding and π -stacking in single- and double-stranded DNA oligonucleotides. DFT calculations and Atoms in Molecules analysis on QM/MM-optimized structures allow characterization and estimation of the energies of π -stacking and hydrogen-bond interactions. This shows that π -stacking interactions depend on the number and the nature of the DNA bases for single-stranded nucleotides; for instance, guanines are found to be involved in strong hydrogen bonds, whereas adenines interact mainly via stacking interactions. The role of interbase hydrogen bonding was explored: the $-\text{NH}_2$ groups of guanine, adenine, and cytosine participate in $\text{N}-\text{H}\cdots\text{O}$ and $\text{N}-\text{H}\cdots\text{N}$ interactions. These are much stronger in single-strand oligonucleotides, where the $-\text{NH}_2$ groups are highly nonplanar. In double-stranded DNA, the strong base-pairing hydrogen bonds of complementary bases lead to more planar $-\text{NH}_2$ groups, which tend to be involved in π -stacking interactions rather than H-bonds. The use of AIM also allows us to evaluate the interplay of π -stacking and H-bonding, suggesting that cooperativity does occur, but is generally limited to about 1–2 kcal/mol.

Introduction

Intermolecular $\pi\cdots\pi$ stacking interactions play a significant role in many fields of chemistry, biology, and physics.^{1–8} Although the benzene dimer is considered the prototypical example of this motif,^{9–11} π -stacking interactions are fundamental in many other aspects of science. Perhaps the preeminent example is nucleic acids,¹² where H-bonds and stacking interactions¹³ lead to the final macromolecular structure, the interplay between forces playing an important role. Experimental evidence of this interplay includes the work of Gray et al.,¹⁴ who studied model systems to investigate the synergy between aromatic stacking and hydrogen bonding in the binding of a flavin derivative. Electrochemical analysis showed the interplay between H-bonding and stacking, which, in turn, influences the overall receptor affinity. Harris and co-workers¹⁵ studied the properties of 1:1 cocrystals formed between benzene and hexafluorobenzene, as well as related materials such as $\text{C}_6\text{H}_5\text{-OH}$ and $\text{C}_6\text{F}_5\text{OH}$, employing X-ray and neutron diffraction to conclude that crystal structures are stabilized by both H-bonding and stacking interactions.

Theoretical studies have led to similar conclusions: Geerlings and co-workers showed that, in stacked complexes of pyridine and benzene, the H-bonding capacity of the pyridine nitrogen is closely related to the interaction between the aromatic rings. In particular, they suggest that electron-donating substituents on benzene lead to charge transfer to pyridine and, hence, to a more basic nitrogen.¹⁶ More recent work¹⁷ on the influence of stacking on the H-bonding ability of cytosine showed similar results: the substituted benzene was able to modulate the donor/acceptor characters of N and O atoms on the pyrimidine base. Guo et al.¹⁸ studied the effects of π -stacking on multiply H-bonded dimers of ureidopyrimidinone, finding that both the strength of H-bonds and the stability of tautomers are influenced by π -stacking. This was explained in terms of charge-transfer enhancement between the H-bonded partners.

As noted above, H-bonding and π -stacking are the main interactions determining nucleic acid structure and are even more important in the gas phase, where intermolecular forces are dominant. Several experimental studies have focused on structural motifs of nucleic acid complexes in the gas phase: although the repulsion between negatively charged phosphate groups should significantly destabilize the helix structure,¹⁹ it has been shown that DNA duplexes can be transferred from solution to gas phase.^{20–30} It is clear from these studies that H-bonding between complementary nucleobases is preserved,³¹ and indeed, it has been suggested that the stability of DNA chains in the gas phase is proportional to the number of H-bonds present.³²

Bowers and co-workers have reported several studies on this problem: using ion mobility mass spectrometry as well as computational tools, they found three families of minima for dinucleotides in the gas phase, namely, an “open” structure, a hydrogen-bonded structure, and a “stacked” form closer to the DNA conformation in solution. The latter, although distorted, shows π -stacking interactions and $\text{N}-\text{H}\cdots\text{X}$ ($\text{X} = \text{O}, \text{N}$) bonds between bases. They also suggest that the preferred conformation of small nucleotides is related to their sequence.^{33,34} In another study, they suggested that, as the DNA structure is highly destabilized by repulsion between phosphates, the number of nucleotides is crucial in order to provide a regular structure.³⁵ For instance, $[\text{d}(\text{CG})\cdot\text{d}(\text{CG})]_n$ adopts a globular conformation when $n < 4$ and a mixture of globular and helical structures when $n > 4$, eventually adopting the regular DNA helix for $n > 10$. Orozco et al. reported gas-phase molecular dynamics calculations on several DNA structures,³⁶ finding that the helical structure, although severely distorted, retains both π -stacking and H-bonding between bases. In particular, DNA strands rich in G and C bases are more stable as most Watson–Crick GC pairs are preserved, whereas AT pairs are quickly disrupted. They also suggest that, in some cases, T-shaped stacking involving clusters of three bases occurs.

In this work, our use of a relatively inexpensive DFT method and Atoms in Molecules (AIM) analysis allows full optimization

* Corresponding author. Phone: +44-29-2087-4950. Fax: +44-29-2087-4030. E-mail platts@cardiff.ac.uk.

di- and trinucleotides and DNA duplexes. This allows us to address the possibility of interplay between π -stacking and H-bonding, first in simple models such as benzene/guanine/cytosine complexes and then in more realistic double-stranded dinucleotides.

Computational Methods

All DFT calculations were performed with the Gaussian 03 suite of programs.³⁷ Throughout this work, we have made extensive use of Becke's "half-and-half" functional, BH&H.³⁸ Recently, we showed that this functional, with polarized and diffuse basis sets, is able to reproduce results of correlated calculations for several archetypal π -stacked complexes.³⁹ In this work, the success of the BH&H functional is assigned to a cancellation of errors between Hartree–Fock and LDA exchange energies; nonetheless, it performed remarkably well in all cases tested. Comparison against literature CCSD(T) binding energies for 10 complexes and MP2 values for 22 complexes, including stacked dimers of substituted benzenes and pyridines and DNA bases, yielded average errors of less than 0.5 kcal/mol and a maximum error of less than 1 kcal/mol. However, despite this excellent performance for π -stacking, it was found that BH&H significantly overestimates the strength of the hydrogen bonds in Watson–Crick GC and AT pairs, in common with many hybrid DFT methods. MP2 is known to overestimate binding energies of π -stacked complexes in general, but Hobza and co-workers have proposed that, in combination with a modification of Pople's 6-31G* basis set in which the diffuse function on heavy atoms is more diffuse than normal, termed 6-31G(0.25)*, MP2 calculations give binding energies close to more accurate values in reasonable time scales. Larger complexes such as nucleotides were divided into high and low layers using the ONIOM method,^{40–44} with nucleobases entirely within the high layer, i.e., BH&H/6-311++G(d,p), and the sugar–phosphate backbone treated using AMBER potentials.⁴⁵ Subsequent single-point calculations on the entire structure were performed using BH&H/6-311++G(d,p) in order to carry out AIM analysis.

To quantify intermolecular interactions, topological analysis of computed electron densities (ρ) was performed using the AIM2000 package.^{46,47,48} This is based on those critical points (CPs) where the gradient of the density, $\nabla\rho$, vanishes. Such points are classified by the curvature of the electron density; for example, bond or (3, -1) CPs have one positive curvature (in the internuclear direction) and two negative curvatures (perpendicular to the bond). Properties evaluated at such BCPs characterize the bonding interactions present⁴⁹ and have been widely used to study intermolecular interactions. Many studies have demonstrated approximately linear relations between H-bond stabilization energy and both the increase in density at H \cdots B BCP and the decrease at A–H for a wide range of A–H \cdots B systems.^{50,51} For instance, a linear combination of electron densities at A–H and H \cdots B BCPs can be used to accurately model basis set superposition error (BSSE)-corrected H-bond energies⁵² of a diverse set of organic and inorganic complexes,⁵³ an approach extensively employed in this work. We also recently suggested a similar AIM-based method for quantifying π -stacking interactions:³⁹ for a total of 55 complexes, BSSE-corrected binding energies are linearly related to the sum of the electron density collected between interacting molecules, $\sum\rho_\pi$. In both models, r^2 values are at least 0.95, and standard deviations are no more than 1.5 kcal/mol.

Results and Discussion

The computational efficiency of the BH&H/6-311++G(d,p) level opens up the possibility of studying large systems: our

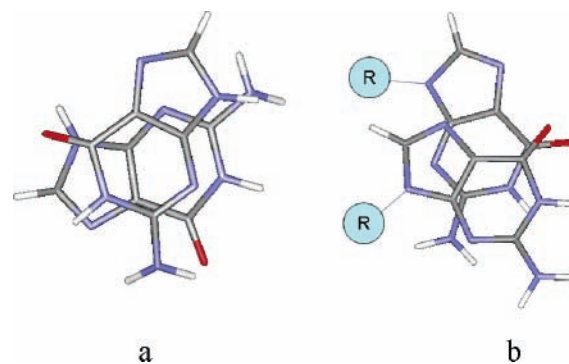


Figure 1. Conformations of (a) guanine dimer and (b) GpG.

TABLE 1: Hydrogen-Bond and π -Stacking Energies of DNA Oligonucleotides

	E_{HB} (kcal/mol)	E_π (kcal/mol)	E_π/E_{HB}
GpG	20.00	2.42	0.10
GpA	9.75	5.90	0.60
ApA	0.00	6.87	—
ApT	2.93	4.84	1.65
GpC	14.29	5.61	0.40
CpC	3.65	1.61	0.44
CpT	6.62	4.05	0.61
GpGpG	20.70	9.70	0.47
GpApG	17.93	10.40	0.58
ApApA	5.11	11.80	2.30
ApTpA	19.38	5.72	0.30
GpCpG	36.48	7.30	0.20

ultimate goal in these studies is to examine the stacking of DNA bases such as di- and trinucleotides, both with and without the deoxyribose–phosphate DNA linker. In accord with the literature,^{39,54–56} the geometries and binding energies of sugar/phosphate-linked complexes differ from those of free dimers and trimers because of the phosphate backbone, which tends to keep bases in a coplanar, coparallel orientation, whereas free complexes prefer coplanar bases but an almost perpendicular orientation between the axes of bases, which maximizes interactions (see Figure 1). Thus, π -stacking and hydrogen bonding in nucleotides are generally weaker than in free structures.

Table 1 summarizes our analysis of these nucleotides, and Figures 2 and 3 illustrate some examples. Generally, all DNA chains studied keep a coplanar, coparallel orientation of bases, with estimated π -stacking energies ranging from 2.5 to 7.0 kcal/mol. However, depending on the bases, the nucleotides show different features. For example, the GpG structure is clearly distorted (see Figure 2a) from the "ideal" geometry, with N–H atoms out of the plane, forming hydrogen bonds both between guanines and to the sugar backbone.

GpG contains four intermolecular H-bonds, including three between bases and one to oxygen in a sugar of the backbone, whose combined strength is the largest of all complexes studied (20.00 kcal/mol). It also has the second lowest estimated π -stacking energy (2.40 kcal/mol), giving an E_π/E_{HB} ratio of just 0.10. In contrast, GpA is much more coplanar (Figure 2b), indicating that H-bonding and π -stacking are more equally shared: AIM analysis shows two H-bonds (9.80 kcal/mol) and four π -stacking interactions (5.90 kcal/mol), giving an E_π/E_{HB} ratio of 0.60, a value that reflects the more regular structure of GpA.

π -stacking is the sole intermolecular interaction in ApA, with no H-bonding but five π -stacking BCPs found, corresponding to 6.87 kcal/mol. Interestingly, the $-\text{NH}_2$ nitrogens interact via

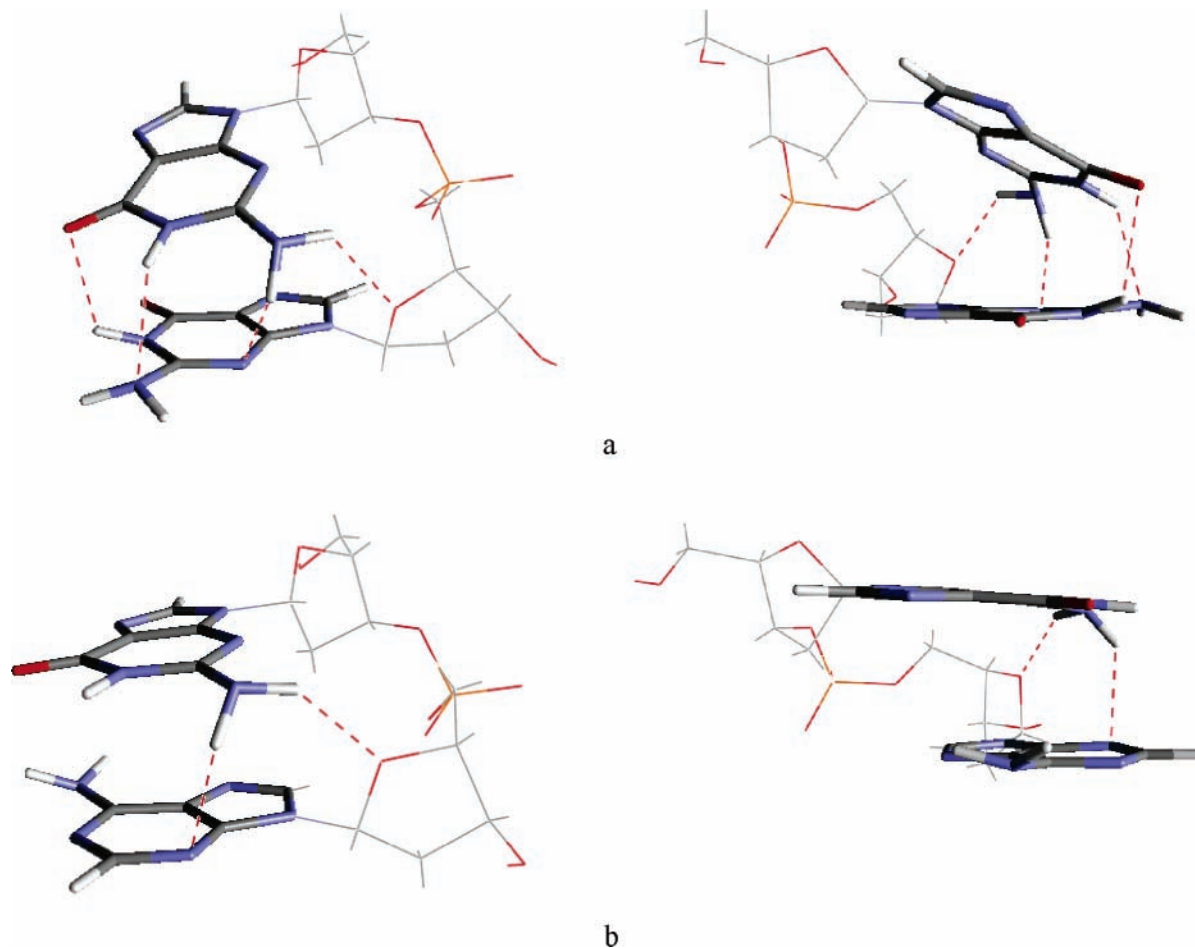


Figure 2. Two views of (a) GpG and (b) GpA.

π -stacking rather than H-bonding and contribute ca. 40% to $\Sigma\rho_{\pi}$, yielding the largest π -stacking energy among the dinucleotides considered. ApT shows similar properties, with just a single weak C–H $\cdots\pi$ interaction (2.93 kcal/mol) and three π -stacking BCP interactions, with the $-\text{NH}_2$ nitrogen in adenine again involved in π -stacking rather than H-bonding. Thus, for ApA and ApT, π -stacking interactions are more important than H-bonds. GpC contains three π -stacking and three H-bonding BCPs, contributing 5.61 and 14.20 kcal/mol, respectively ($E_{\pi}/E_{\text{HB}} = 0.40$), with $-\text{NH}_2$ groups of both guanine and cytosine involved in both H-bonds and π -stacking. Similarly, CpC is mainly stabilized by H-bonding, with just one H-bond and one stacking BCP, contributing 3.65 and 1.61 kcal/mol, respectively. CpT is more strongly bound than CpC, with H-bond and stacking energies of 6.62 and 4.05 kcal/mol, each from two BCPs.

Turning to the trinucleotides, optimization of GpGpG yields a much more regular structure than found in GpG, in which each pair contains two N–H $\cdots\text{N}$ and N–H $\cdots\text{O}$ H-bonds, with bases almost parallel (see Figure 3a), suggesting that H-bonding is less dominant than in GpG. Stacking interactions between each pair of bases are also similar (5.35 and 4.30 kcal/mol), giving an E_{π}/E_{HB} ratio of 0.47, indicating that both forms of interaction stabilize the final structure. Thus, it seems that GpG is unusually distorted by interbase H-bonding: the H-bond energy per pair in GpGpG is around one-half that found in the dinucleotide. GpApG (Figure 3b) has a balance of π -stacking and H-bonding ($E_{\pi}/E_{\text{HB}} = 0.58$), with some redistribution of energy compared to GpA: H-bonding is diminished in one G \cdots A pair and enhanced in the other. In contrast, ApApA is dominated by π -stacking just as in ApA ($E_{\pi}/E_{\text{HB}} = 2.31$), with only one relatively weak H-bond per pair.

Interestingly, ApTpA and GpCpG present unique features among the complexes studied here. As shown in Figure 3c,d, the structure of these trinucleotides is so distorted that H-bonding between the first and third bases occurs, while the combined π -stacking energies are the smallest found among trinucleotides. On the other hand, the hydrogen-bond energies are among the strongest, 19.38 kcal/mol for ApTpA and 36.48 kcal/mol for GpCpG, with the first base and third base (A \cdots A for ApTpA and G \cdots G for GpCpG) interacting via strong N–H $\cdots\text{N}$ and N–H $\cdots\text{O}$ bonds (ca. 30% of the overall H-bond energy). Thus, the E_{π}/E_{HB} ratios for ApTpA and GpCpG are 0.30 and 0.20, respectively. Therefore, as the hydrogen-bond interactions prevail by far over π -stacking, the overall structure assumes a globular conformation.

These studies show that $-\text{NH}_2$ groups interact via H-bonding to heavy atoms of other bases participate in π -stacking or, in some cases, both. $-\text{NH}_2$ groups of guanine are largely involved in H-bonding, mainly via N–H $\cdots\text{N}$ and N–H $\cdots\text{O}$ interactions, leading to a high degree of pyramidalization with the sum of bond angles at N (Σ°) equal to ca. 330° on average. $-\text{NH}_2$ groups of adenine are involved in both N–H $\cdots\text{N}$ H-bonds and π -stacking interactions. Adenine's H-bonds are generally weaker than those in guanine, and the $-\text{NH}_2$ groups are closer to planarity ($\Sigma^{\circ} = 346.7^{\circ}$ on average). Cytosine shows similar properties: $-\text{NH}_2$ groups interact via both H-bonding and stacking, and Σ° ranges between 347.0° and 352.0° . It is interesting to note that some correlation exists between the electron density of H-bonds of $-\text{NH}_2$ groups and Σ° , with $r^2 = 0.82$, supporting the idea that the nonplanar character of $-\text{NH}_2$ is related to the strength of base \cdots base H-bonds (see Supporting Information).

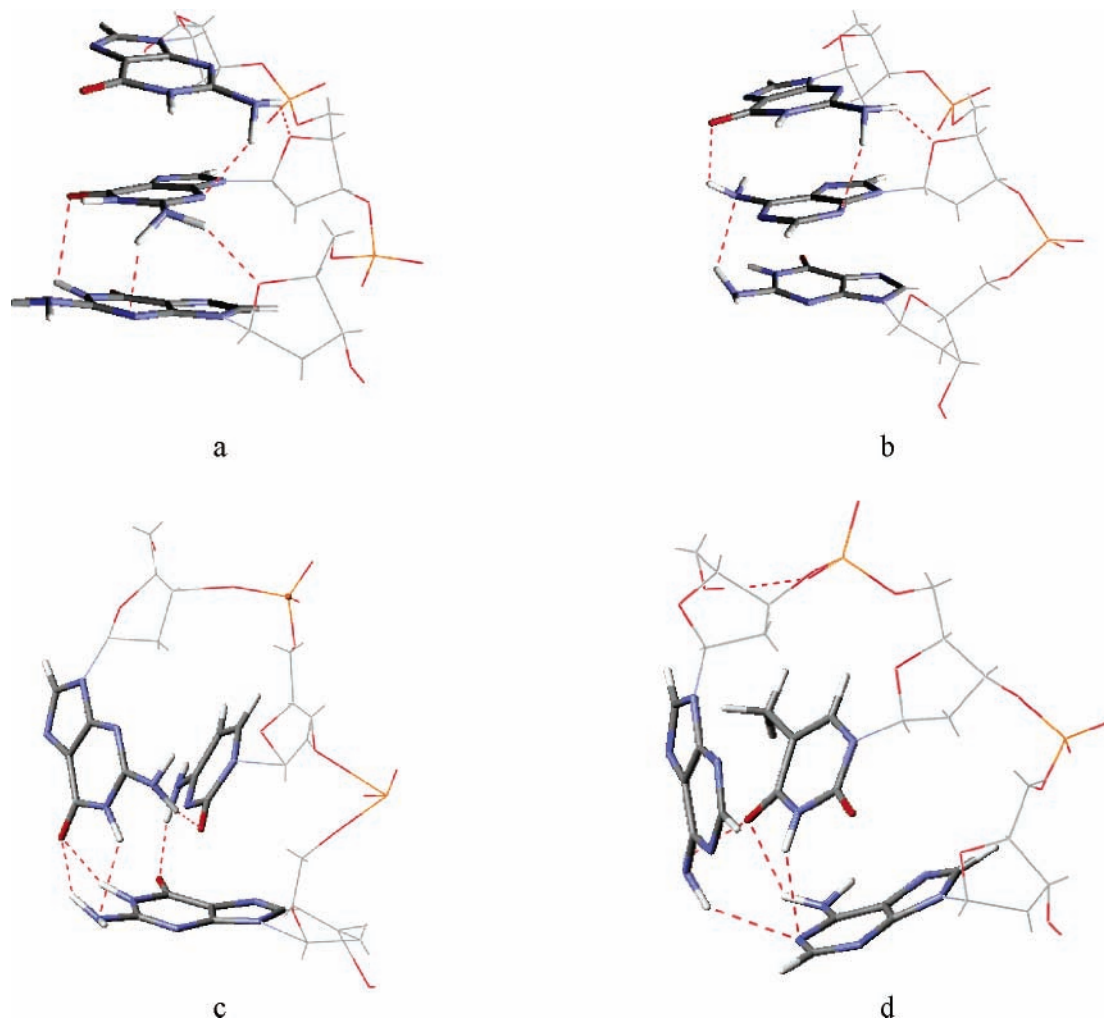


Figure 3. (a) GpGpG, (b) GpApG, (c) GpCpG, and (d) ApTpA nucleotides.

Considering the intricacy of interactions in these DNA oligonucleotides, it is not trivial to quantify the interplay between π -stacking and H-bonding: when these forces act together, the whole structure changes, and separation of the two effects becomes impossible. However, it is clear that the ratio of these energies is important in determining the final structure: when $E_{\pi}/E_{\text{HB}} \ll 0.5$, as in GpG, GpCpG, and ApTpA, for instance, the geometry is highly nonplanar, with the nucleotides “pointing toward” each other, leading to a loss of π -stacking energy. When this ratio approaches or exceeds 0.5, for instance, on adding a guanine to GpG, where $E_{\pi}/E_{\text{HB}} = 0.47$, the bases tend to adopt a more parallel conformation. This is evidence of cooperativity in π -stacking, with the second base providing additional stabilization of the regular, parallel structure.

To further investigate any possible cooperativity between π -stacking and hydrogen bonding, we studied a prototypical system of benzene/guanine/cytosine, comparing density properties to those in the corresponding bimolecular complexes. We first considered benzene \cdots GC and benzene \cdots CG (see Figure 4): topological analysis shows no qualitative difference from bimolecular complexes. Moreover, in neither case do ρ_{π} nor ρ_{HB} differ significantly from values found in the analogous dimers (guanine–cytosine, benzene–guanine, and benzene–cytosine), with maximum variations of 0.001 au, suggesting that little interplay between stacking and H-bonding occurs here. Following Geerlings et al.’s recent work,¹⁷ we then considered the effect of substitution on benzene, including groups such as

–NO₂, –F, –CH₃, –CHO, –OH, and –NH₂. Table 2 reports data for these complexes. (See Figure 5 for atom numbering scheme.)

Geerlings et al. suggested that the mutual influence of π -stacking and H-bonding depends on the hardness of the substituted benzene, i.e., benzenes with electron-withdrawing groups stacked over guanine lead to lower charge transfer to cytosine. Thus, from –NO₂ to –NH₂, cytosine acts as a progressively better H-bond acceptor (through N₃ and O₂) and a worse H-bond donor (through H₄), confirming that π -stacking does influence the H-bonding of GC. However, individual variations in H-bonds are small (generally no more than ca. 1.5 kcal/mol), and because the trend for H₄ \cdots O₆ is opposite that for H₁ \cdots N₃ and H₂ \cdots O₂, the total pairing energy hardly changes, even though distortion of the GC pair occurs.

A similar treatment of more realistic models of DNA chains, namely, the double-stranded dinucleotides GpC·CpG, CpT·GpA, and GpG·CpC, is shown in Figures 6 and 7. To our knowledge, this is the first attempt to fully optimize such systems using ab initio or DFT methods, as opposed to classical force fields. Compared to solution and crystal structures, little is known about the structure of DNA in the gas phase, but experimental and computational studies agree on some important aspects, especially that base pairing and π -stacking are preserved, but strong distortion of DNA occurs.^{33,34}

Our results are consistent with these findings: π -stacking and H-bonding are evident, as are large distortions of the “ideal”

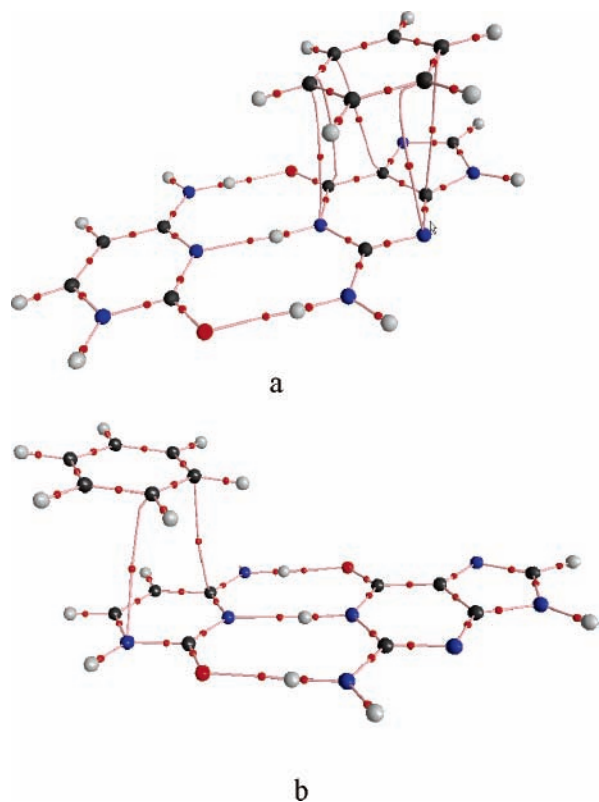


Figure 4. Topologies of (a) benzene...GC and (b) benzene...CG.

TABLE 2: Electron Density (au) at H-Bond CPs in Benzene...G≡C Complexes

	H ₄ ...O ₆ ^a	H ₁ ...N ₃ ^a	H ₂ ...O ₂ ^a
free GC	0.0520	0.0436	0.0365
-NO ₂	0.0512	0.0438	0.0371
-CHO	0.0512	0.0439	0.0371
-F	0.0507	0.0442	0.0375
-H	0.0506	0.0443	0.0377
-CH ₃	0.0505	0.0443	0.0378
-OH	0.0506	0.0444	0.0378
-NH ₂	0.0501	0.0446	0.0383

^a Numbering scheme shown in Figure 5.

DNA chain structure. Table 3 reports data for intra- (S) and interstrand (IS) π -stacking, as well as Watson-Crick (WC) H-bonding, as schematized in Figure 7a. In general, calculated energies are similar to those for the single nucleotide strands (Table 1). For instance, the stacking energies between C...T and G...A in CpT·GpA are 2.92 and 5.22 kcal/mol, respectively, within 1 kcal/mol of the corresponding single-strand energies. Similarly, the intrastrand G...C π -stacking energies for GpC·GpC

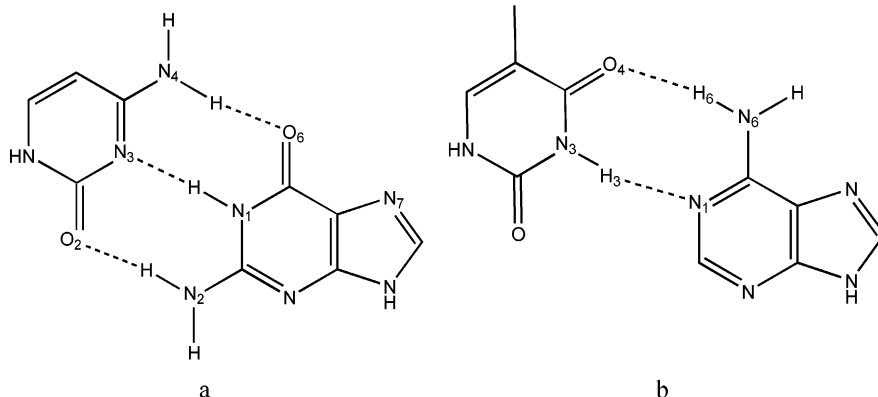


Figure 5. Numbering schemes for (a) GC and (b) AT.

TABLE 3: H-Bonding and π -Stacking Energies of the Duplexes (kcal/mol)

step ^a	E_{HB}^b	E_{π}
	CpT·GpA	
CT _S	0.0	2.92
CA _{IS}	0.0	2.15
GA _S	0.0	5.22
	GpC·CpG	
GC _S	4.44	6.38
CC _{IS}	0.0	0.88
CG _S	4.12	6.35
	GpG·CpC	
GG _S	0.0	7.06
GC _{IS}	0.0	2.28
CC _S	0.0	3.79

^a Subscripts refer to intrastrand (S) and interstrand (IS) interactions, as shown in Figure 7. ^b Only N-H...O interactions in GpC·CpG are found, the oxygen atom belonging to the sugar-phosphate backbone.

are ca. 6 kcal/mol, cf. 5.61 for single-stranded nucleotides. Interestingly, in GpG·CpC, both G...G and C...C are rather larger than the corresponding single strands, 7.00 and 3.79 kcal/mol, respectively, compared to ca. 5.00 and 1.61 kcal/mol.

Topological analysis also reveals evidence for interstrand (IS) stack interactions between bases belonging to two different oligonucleotides. As shown by Hobza et al.,⁵⁷ these interactions are weak, generally not greater than 2 kcal/mol. For GpC·CpG, we find two CC_{IS} BCPs with very small electron density (0.0049 au in total), corresponding to less than 1kcal/mol. In contrast, the GC_{IS} and CA_{IS} interactions, in GpG·CpC and CpT·GpA, respectively, are slightly stronger, equal to 2.28 and 2.15 kcal/mol. Thus, although weak, these interactions contribute between 10% and 25% of the overall π -stacking energy and therefore play a role in the structure of these chains.

As noted above, the flexibility of -NH₂ groups allows H-bonds to stabilize single-stranded nucleotides, especially in guanine and adenine strands. However, our analysis of the duplexes CpT·GpA, GpC·CpG, and GpG·CpC suggests a different scenario: comparatively few intrastrand H-bonds are found here (although, of course, such points are found between strands) involving solely guanine N-H...O, with estimated energies of ca. 4 kcal/mol each. Moreover, -NH₂ groups are much closer to planarity than in the single strands and are involved in π -stacking interactions rather than intrastrand H-bonds. This appears to be an effect of *interstrand* pairing: the bases are paired via strong H-bonds in the plane of the base, acting to constrain -NH₂ groups to this plane, which are thus less able to deform. This is evident in the average values of

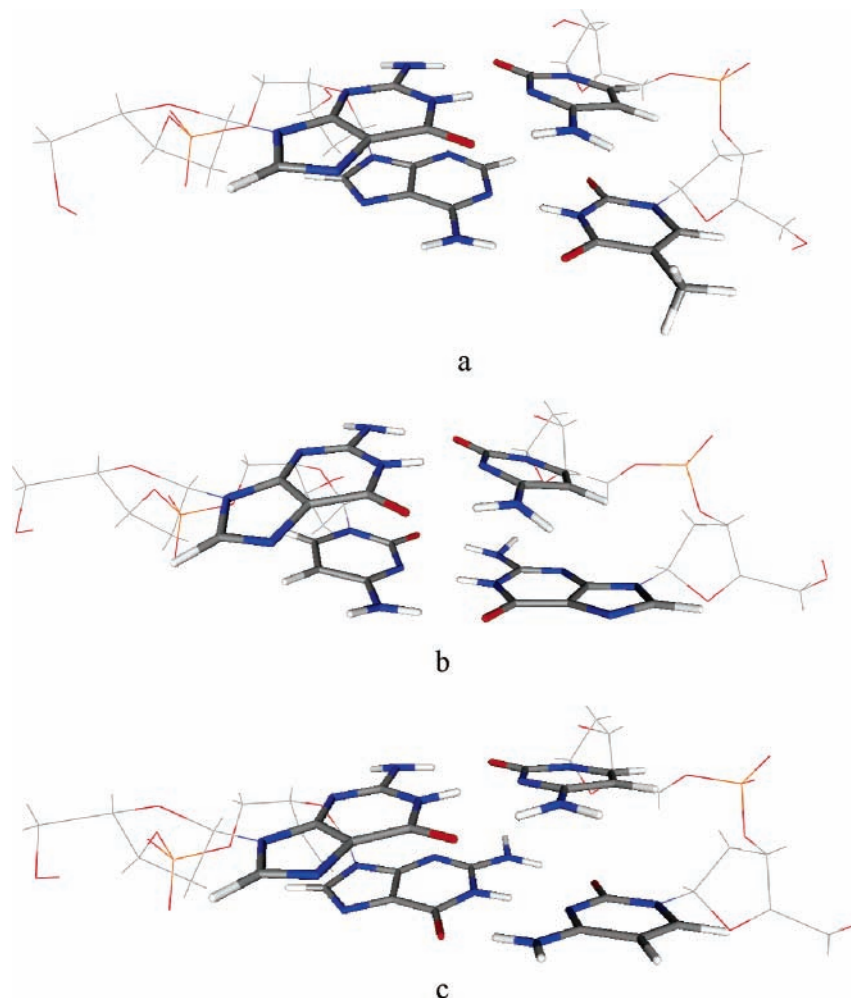


Figure 6. (a) CpT·GpA, (b) GpC·CpG, and (c) GpG·CpC duplexes.

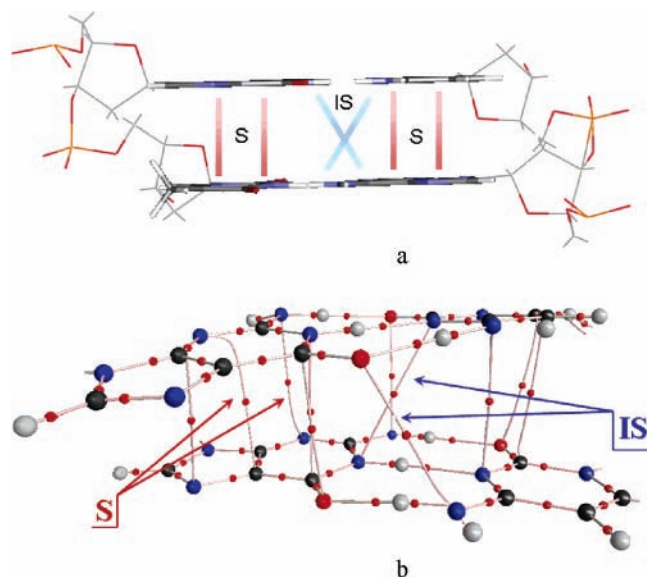


Figure 7. (a) Schematic drawing of the intrastrand stack (S) and interstrand stack (IS) interactions; (b) detail of the GpG·CpC topology. Σ° , where C = 358.3° > A = 357.1° > G = 351.2°, and the increased π -stacking interactions of these nitrogens seen in Table 4.

Although the strength of H-bonding is known to be overestimated using BH&H,³⁹ it is still possible to compare GC and AT pairing in various environments (Table 4). In all cases considered, the overall pairing energy is close to that found in

TABLE 4: Electron Density (au) in GC and AT Pairs in Duplexes

DNA	base pair	H ₄ ···O ₆ ^a (H ₃ ···N ₁)	H ₁ ···N ₃ ^a (H ₆ ···O ₄)	H ₂ ···O ₂ ^a	ρ_{TOT}
free	GC	0.0520	0.0436	0.0365	0.132
	AT	0.0586	0.0285		0.087
CpT·GpA	GC	0.0506	0.0433	0.0370	0.131
	AT	0.0644	0.0302		0.095
GpC·CpG	GC ₁	0.0409	0.0443	0.0431	0.129
	GC ₂	0.0429	0.0453	0.0427	0.131
GpG·CpC	GC ₁	0.0534	0.0410	0.0329	0.128
	GC ₂	0.0550	0.0440	0.0346	0.134

^a See Figure 5 for numbering; the numbering in parentheses refers to AT.

isolated GC and AT. The largest change is for AT in CpT·GpA at +0.008 au (ca. +2 kcal/mol), whereas GC in GpC·CpG is reduced by 0.003 au (ca. -1 kcal/mol). However, individual H-bonds differ substantially from their values in the free base pairs: Table 4 and Figure 8 display the electron densities at the H-bonds of GC pairs in the studied duplexes. Whereas the electron densities of the single H-bonds are very close to those of free GC for the GC pairs in CpT·GpA and GpG·CpC, in the case of GpC·CpG, H₄···O₆ is stronger and H₂···O₂ weaker than in free GC.

Table 5 summarizes our results in terms of G···G, G···A, and G···C interactions, excluding those from the highly distorted GpCpG and ApTpA. Thus, we compared the estimation of the π -stacking and H-bond energies to BSSE binding energies calculated as shown in Hobza et al.'s works.^{57,58} Nucleobase

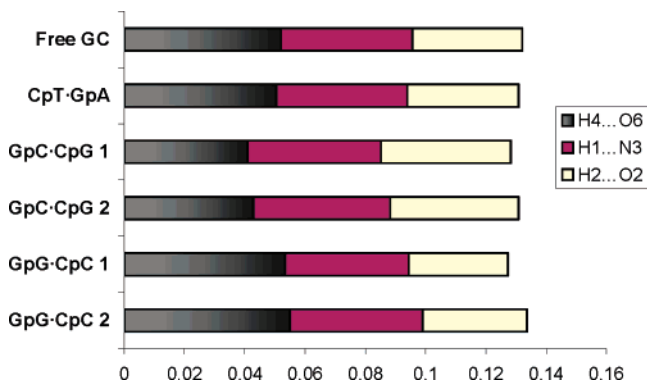


Figure 8. Electron density in H-bonds of GC (au).

TABLE 5: Summary of G···G, G···A, and G···C Interactions

	E (kcal/mol)			ΔE^b	
	E_π^a	E_{HB}^a	$E_\pi + E_{HB}$	BH&H	MP2
G···G					
GpG	2.42	9.97	12.37	8.85	10.89
GpGpG	4.32	5.92	10.24	6.52	9.51
GpGpG	4.33	5.92	10.25	4.85	8.61
GpG·CpC	7.06	0.0	7.06	3.21	4.25
G···A					
GpA	5.90	4.42	10.32	11.10	12.64
GpApG	5.57	4.10	9.67	6.75	10.07
GpApG	4.83	8.53	13.36	9.27	11.30
CpT·GpA	5.22	0.0	5.22	5.14	8.85
G···C					
GpC	5.61	9.94	15.55	15.94	14.68
GpC·CpG	6.38	0.0	6.38	6.20	9.57
GpC·CpG	6.35	0.0	6.35	6.05	9.40

^a E_π and E_{HB} are calculated from topological analysis of the electron density. ^b BSSE-corrected binding energies of stacked base pairs using the 6-311++G(d,p) basis set for BH&H and 6-31G(0.25)* for MP2.

geometries were extracted from the optimized nucleotide structures, and the backbone was replaced by hydrogen atoms on N₉; then the BSSE-corrected energy was evaluated at the BH&H/6-311++G(d,p) and MP2/6-31G(0.25)* levels.^{59,60} This approach was used to (i) test the capability of BH&H in reproducing π -stacking energies away from the equilibrium of the gas-phase dimers and (ii) clarify whether cooperativity arising from interplay of H-bonding and intra- and interstrand stack interactions might play a role in these complexes.

BH&H binding energies are generally smaller than MP2 values by between 1 and 4 kcal/mol: only in GpC (the most strongly bound complex in Table 5) does BH&H binding exceed the MP2 result. In the other cases, two factors might be responsible for this discrepancy: (i) MP2 is known to overestimate π -stacking energies^{39,61} and (ii) the BH&H functional might be less effective in predicting the energy of nonequilibrium geometries than was found for fully optimized species. The difference between BH&H and MP2 is greatest for the most weakly bound complexes, and over all 11 complexes, the average absolute error is 2.6 kcal/mol. Interestingly, despite these differences, there is a reasonable correlation ($r^2 = 0.85$) between BH&H and MP2 binding energies (see Supporting Information). However, to definitively state whether differences are due to deficiencies in BH&H or MP2 (or both) would require extrapolation to infinite basis sets and CCSD(T) corrections, which are beyond our computational resources.

Comparison of directly calculated and AIM-estimated ($E_\pi + E_{HB}$) binding energies shows similar accuracy, with average absolute differences of 2.3 and 2.0 kcal/mol vs BH&H and MP2,

respectively. However, these differences are less systematic than above, such that there is no statistically significant correlation between AIM and supermolecule binding energies. Closer inspection shows that substantially smaller errors (1.30 kcal/mol on average) are found for the single-stranded oligonucleotides, whereas for double-stranded DNA structures, the difference between AIM and MP2 is generally much larger. This suggests that the approach of taking stacked base pairs from optimized oligonucleotides works well for single strands but apparently fails for duplexes because of the intricacy of the interactions. In other words, the interaction, for instance, of G···A in CpT·GpA is strongly affected by the environment and, particularly, by the complementary bases C and T, which interact with G and A via strong H-bonds, as well as inter- and intrastrand stack interactions. AIM analysis, which takes into account the effects of environment on the electron density, therefore complements the supermolecule approach, allowing study of the subtle interplay arising from the complexity of H-bonding and stacking interactions.

Hobza et al.⁵⁷ and Geerlings et al.¹⁷ have performed high-level ab initio calculations on stacked base structures extracted from experimental DNA geometries. Despite the use of different geometries and theoretical methods, agreement between their values and our BH&H data is qualitatively good: stacking in GpG is rather weak (2.42 kcal/mol), slightly less than Geerlings et al.'s MP2 value of 3.39 kcal/mol, but almost doubles to 5.35 and 4.33 kcal/mol in GpGpG. G···A interactions are also affected by the length of the chain, with stacking energies between 3.70 and 5.90 kcal/mol: the latter value is close to Hobza et al.'s value of 6.5 kcal/mol. Finally, the π -stacking energy estimated from the topology in GpC is rather large, between 5.61 and 6.38 kcal/mol, within ca. 1–2 kcal/mol of Hobza et al.'s values of 7.7 and 7.9 kcal/mol, with less pronounced differences from environmental factors.

Similarly, direct comparison with experimental studies is not possible because of difficulties in obtaining accurate experimental gas-phase DNA structures: nonetheless, our models match several known facts. Bowers and co-workers have provided many fascinating results, e.g., that single-stranded nucleotides exist in three different conformations, with important interplay between π -stacking and H-bonding, confirming that base···base N–H···N/O H-bonding occurs, cf. Table 1 and Figures 2 and 3.^{33,34} They also show that the conformations of di- and trinucleotides are largely determined by the sequence, even for such small DNA chains. Our work supports this conclusion; for instance, whereas the GpG structure is strongly distorted, with a stacking energy of 2.42 kcal/mol, GpA is almost parallel with a π -stacking energy of ca. 5 kcal/mol.

Several studies indicate that, in the gas-phase, Watson–Crick pairs are better preserved in G···C than A···T,^{22,36} because of the stronger H-bonds here. Orozco and co-workers³⁶ showed that G···C stacked pairs also are better preserved, suggesting that these interactions are largely responsible for the maintenance of the structural features of DNA in the gas phase. Our results indicate that G···C stacking interactions are strong, similar to G···G interactions. Moreover, Orozco and co-workers stressed that, although many known DNA features are retained in the gas phase, some interesting differences emerge. Molecular dynamics simulations found that T-shaped π -stacking occurred in some DNA chains. Similarly, we find structures of GpCpG and ApTpA that present such features, with two bases interacting via parallel stacking and the third in a T-shaped conformation (see Figure 3c,d). In particular, our analysis indicates that this conformation is principally due to N–H···O/N and C–H··· π

hydrogen-bonding interactions. Thus, even if direct comparison of DNA geometry obtained with this hybrid functional to experiment is not practicable, literature theoretical and experimental data support our estimations of stacking in these nucleotides, confirming the validity of this approach.

Conclusions

Combined BH&H/6-311++G(d,p) and AIM analysis has allowed study of the intermolecular forces and their mutual interplay in DNA chains and some model systems. Single-stranded di- and trinucleotides show that the conformation adopted is related to the number of and type of bases, with a balance of H-bonding and π -stacking needed to obtain regular structures. Trinucleotides in which the central base is cytosine or thymine have highly distorted structures, closer to "T-shaped" complexes rather than the more normal parallel stacking structure, in which two bases interact with the third via hydrogen bonds and C-H $\cdots\pi$ interactions. $-\text{NH}_2$ groups play an important role, involved in both H-bonds and π -stacking, and are found to be significantly nonplanar in many structures.

Furthermore, the interplay of π -stacking and H-bonding was explored: simple models such as benzene/guanine/cytosine confirm that benzene molecules can modulate H-bonding capacity, leading to distortion in the guanine-cytosine pair, but barely perturbing the overall binding energy. In DNA duplexes, Watson-Crick pairing of GC and AT is hardly affected by stacking partners, in accord with literature. Base pairing also leads to increased planarity of $-\text{NH}_2$ groups, which now interact mainly via π -stacking rather than H-bonding, leading to an overall gain of π -stacking energy of 1–2 kcal/mol per stacked pair, compared to single-stranded chains. In such studies, AIM analysis is particularly useful, as the intricacy of inter- and intrastrand interactions means that pairwise analysis of base-base interactions inevitably ignores the wider environment. Where comparison is possible, energies and structures at least qualitatively match experimental and theoretical data reported in the literature.

Supporting Information Available: Figures showing the nonplanarity of $-\text{NH}_2$ groups versus $\sum\rho_{\text{HB}}$ and BH&H versus MP2 binding energies. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- Hunter, C. A.; Sanders, J. K. M. *J. Am. Chem. Soc.* **1990**, *112*, 5525.
- Hunter, C. A.; Singh, J.; Thornton, J. M. *J. Mol. Biol.* **1991**, *218*, 837.
- Burley, S. K. P.; G. A. *Science* **1985**, *23*.
- Claessens, C. G.; Stoddart, J. F. *J. Phys. Org. Chem.* **1997**, *10*, 254.
- Askew, B.; Ballester, P.; Buhr, C.; Jeong, K. S.; Jones, S.; Parris, K.; Williams, K.; Rebek, J. *J. Am. Chem. Soc.* **1989**, *111*, 1082.
- Smithrud, D. B.; Diederich, F. *J. Am. Chem. Soc.* **1990**, *112*, 339.
- Hunter, C. A. *Chem. Soc. Rev.* **1994**, *23*, 101.
- Rebek, J. *Chem. Soc. Rev.* **1996**, *25*, 255.
- Hobza, P.; Sponer, J. *Chem. Phys. Lett.* **1998**, *288*, 7.
- Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Phys. Chem.* **1996**, *100*, 18790.
- Jaffe, R. L.; Smith, G. D. *J. Chem. Phys.* **1996**, *105*, 2780.
- Watson, J. D.; Crick, H. C. D. *Nature* **1953**, *171*, 737.
- Saenger, W. *Principles of Nucleic Acid Structure*; Springer: New York, 1988.
- Gray, M.; Goodman, A. J.; Carroll, J. B.; Bardon, K.; Markey, M.; Cooke, G.; Rotello, V. M. *Org. Lett.* **2004**, *6*, 385.
- Meejoo, S.; Kariuki, B. M.; Harris, K. D. M. *ChemPhysChem* **2003**, *4*, 766.
- Mignon, P.; Loverix, S.; Geerlings, P. *Chem. Phys. Lett.* **2005**, *401*, 40.
- Mignon, P.; Loverix, S.; Steyaert, J.; Geerlings, P. *Nucleic Acids Res.* **2005**, *33*, 1779.
- Guo, D. W.; Sijbesma, R. P.; Zuilhof, H. *Org. Lett.* **2004**, *6*, 3667.
- Sinden, R. R. *DNA Structure and Function*; Academic Press: San Diego, CA, 1994.
- Ganem, B.; Li, Y. T.; Henion, J. D. *J. Am. Chem. Soc.* **1991**, *113*, 6294.
- Lightwahl, K. J.; Springer, D. L.; Winger, B. E.; Edmonds, C. G.; Camp, D. G.; Thrall, B. D.; Smith, R. D. *J. Am. Chem. Soc.* **1993**, *115*, 803.
- Ganem, B.; Li, Y. T.; Henion, J. D. *Tetrahedron Lett.* **1993**, *34*, 1445.
- Doktycz, M. J.; Habibigoudarzi, S.; McLuckey, S. A. *Anal. Chem.* **1994**, *66*, 3416.
- Bayer, E.; Bauer, T.; Schmeer, K.; Bleicher, K.; Maler, M.; Gaus, H. *J. Anal. Chem.* **1994**, *66*, 3858.
- Banoub, J. H.; Newton, R. P.; Esmans, E.; Ewing, D. F.; Mackenzie, G. *Chem. Rev.* **2005**, *105*, 1869.
- Hofstadler, S. A.; Sannes-Lowery, K. A.; Hannis, J. C. *Mass Spectrom. Rev.* **2005**, *24*, 265.
- Pothukuchy, A.; Mazzitelli, C. L.; Rodriguez, M. L.; Tuesuwan, B.; Salazar, M.; Brodbelt, J. S.; Kerwin, S. M. *Biochemistry* **2005**, *44*, 2163.
- Mariappan, S. V. S.; Cheng, X.; van Breemen, R. B.; Silks, L. A.; Gupta, G. *Anal. Biochem.* **2004**, *334*, 216.
- Hofstadler, S. A.; Griffey, R. H. *Chem. Rev.* **2001**, *101*, 377.
- Cheng, X. H.; Gao, Q. Y.; Smith, R. D.; Jung, K. E.; Switzer, C. *Chem. Commun.* **1996**, 747.
- Gabelica, V.; De Pauw, E. *J. Mass Spectrom.* **2001**, *36*, 397.
- Yang, M.; Thompson, R. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1354.
- Gidden, J.; Bowers, M. T. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 161.
- Gidden, J.; Bushnell, J. E.; Bowers, M. T. *J. Am. Chem. Soc.* **2001**, *123*, 5610.
- Gidden, J.; Baker, E. S.; Ferzoco, A.; Bowers, M. T. *Int. J. Mass Spectrom.* **2005**, *240*, 183.
- Rueda, M.; Kalko, S. G.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2003**, *125*, 8007.
- Frisch, M. J. T.; G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; 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.; 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, Revision B.05*; Gaussian, Inc., Pittsburgh, PA, 2003.
- Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
- Waller, M. P.; Robertazzi, A.; Platts, J. A.; Hibbs, D. E.; Williams, P. A. *J. Comput. Chem.* **2006**, *27*, 491.
- Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357.
- Matsubara, T.; Sieber, S.; Morokuma, K. *Int. J. Quantum Chem.* **1996**, *60*, 1101.
- Svensson, M.; Humbel, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 3654.
- Humbel, S.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 1959.
- Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170.
- Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- Bader, R. F. W. *Atoms in Molecules—A Quantum Theory*; Oxford University Press: Oxford, U.K., 1990.
- Popelier, P. L. A. *Atoms in Molecules: An Introduction*; Prentice Hall: Harlow, U.K., 2000.
- Friedrich Biegler-König, J. S. *J. Comput. Chem.* **2002**, *23*, 1489.
- Bader, R. F. W.; Essen, H. *J. Chem. Phys.* **1984**, *80*, 1943.
- Grabowski, S. J. *Chem. Phys. Lett.* **2001**, *338*, 361.
- Espinosa, E.; Souhassou, M.; Lachekar, H.; Lecomte, C. *Acta Crystallogr. B: Struct. Sci.* **1999**, *55*, 563.
- Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
- Robertazzi, A.; Platts, J. A. *Inorg. Chem.* **2005**, *44*, 267.
- Reha, D.; Kabelac, M.; Ryjacek, F.; Sponer, J.; Sponer, J. E.; Elstner, M.; Suhai, S.; Hobza, P. *J. Am. Chem. Soc.* **2002**, *124*, 3366.
- Burda, J. V.; Sponer, J.; Leszczynski, J. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4404.
- Hobza, P.; Sponer, J. *Chem. Rev.* **1999**, *99*, 3247.
- Dabkowska, I.; Gonzalez, H. V.; Jurecka, P.; Hobza, P. *J. Phys. Chem. A* **2005**, *109*, 1131.

(58) Hobza, P.; Kabelac, M.; Sponer, J.; Mejzlik, P.; Vondrasek, J. *J. Comput. Chem.* **1997**, *18*, 1136.

(59) Reha, D.; Kabelac, M.; Ryjacek, F.; Sponer, J.; Sponer, J. E.; Elstner, M.; Suhai, S.; Hobza, P. *J. Am. Chem. Soc.* **2002**, *124*, 3366.

(60) Sponer, J.; Gabb, H. A.; Leszczynski, J.; Hobza, P. *Biophys. J.* **1997**, *73*, 76.

(61) Sinnokrot, M. O.; Sherrill, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 7690.