DNA Triplexes: A Study of Their Hydrogen Bonds

Martin Peters,[†] Isabel Rozas,*,[†] Ibon Alkorta,[‡] and Jose Elguero[‡]

Centre for Synthesis and Chemical Biology, Department of Chemistry, Trinity College Dublin, Dublin 2, Ireland, and Instituto de Química Médica, CSIC, Juan de la Cierva 3, 28006-Madrid, Spain

Since their discovery, DNA triplexes have received a large amount of interest because of their potential as viable strategy for control of gene expression and their formation with some specific drugs. For that reason, we have studied, by means of *ab initio* calculations, the structure of the bases of these DNA triplexes in terms of their interaction energies and electron densities for the different hydrogen bonds, which is one of the interactions that determine the structure and dynamics of nucleic acid molecules. A good agreement has been found with experimental results and previous calculations. In general, when a third base is added to a dimer, the strength of the hydrogen bonds of the dimer becomes modified provoking changes in their structures by increasing the distance in the major groove and decreasing the space in the minor groove.

Introduction

Nucleic acid triplex structures were discovered in 1956 by Davies and Rich,[1] who observed an apparent increase in solution viscosity upon interacting equimolar mixtures of the synthetic ribo-homopolymers poly(A) and poly(U). Although originally interpreted as poly(A.U) duplex formation, stoichiometric analysis later indicated that in the presence of moderate concentrations of Mg^{2+} (or other cations) the predominant structure formed from this mixture was actually a three stranded complex, a triplex.[2] Since then, DNA triplexes have attracted considerable attention mainly because of their potentiality as therapeutic agents.

Triple helices can be classified into two structural motifs. The parallel YRY triplex (Y= pyrimidine and R= purine) forms when the third strand binds parallel to the Watson−Crick (WC) purine base via Hoogsteen base pairing scheme and the sugar groups are cis to each other. The antiparallel YRR motif forms when the third strand binds antiparallel to the WC purine base through a reverse Hoogsteen base pairing scheme; that is, the sugar groups are trans to each other (see Figure 1).

Oligonucleotides generally bind to the major groove of the double stranded DNA where they specifically recognize oligopyrimidine−oligopurine regions by forming triple helices. These triplex-forming oligonucleotides can penetrate the cell wall, can interact or modify the double-stranded DNA, and can inhibit transcription.[3] DNA triplexes have been found to be associated with the onset and development of several neurodegenerative diseases such as Huntington's disease and Friedrich's ataxia (FRDA).[4]

Another example of triplex formation is that of some peptide nucleic acids (PNAs) which are polyamide oligomers used as anticancer agents. These compounds can form very stable hybrids with both RNA and DNA.[5] PNAs that are composed of homopyrimidines form a triplex with double stranded DNA where two PNAs bind the complementary homopurine strand of DNA and the homopyrimidine DNA strand is displaced to
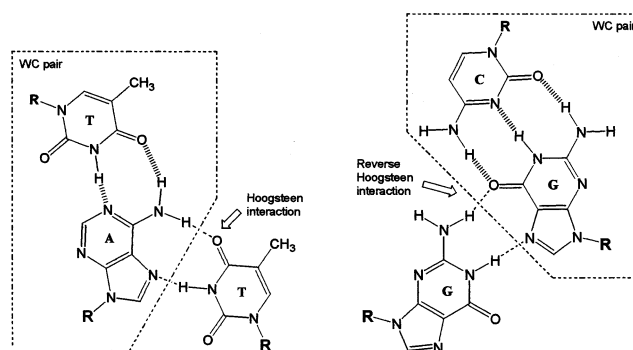

Figure 1.

form a single stranded loop (PNA$_2$DNA). Other less stable hybrid structures also form, depending on base sequence, triplexes where a cytosine rich PNA forms Hoogsteen base pairs with the DNA molecule.[6] PNAs have been shown to inhibit DNA polymerization, RNA polymerization, reverse transcription, and transcription factor binding.[6,7]

Therefore, it is clear that a better understanding of the forces that stabilize triplex structures will be important in curing gene-related diseases and possibly even controlling other biological activities such as cancer. For that reason, we have decided to study theoretically the structure of the bases of these DNA triplexes in terms of their electron densities and the different hydrogen bonding, which is one of the interactions that determine the structure and dynamics of nucleic acid molecules.

To study the intrinsic interactions of DNA bases experimentally, one needs to carry out accurate gas-phase experiments which provide large amounts of data giving insight into the physicochemical origin of hydrogen bonding. In that sense, the stability and structure of RNA duplexes containing isoguanosine and isocytidine and the stability and heat of hydration of DNA, RNA, and DNA/RNA oligomer duplexes have been measured by UV optical melting.[8] As well, the stability of DNA triplexes inside cells has been experimentally studied by iodine-125 radioprinting.[9] Moreover, duplex and triplex formation were monitored by thermal denaturation analisys, CD spectroscopy, and gel shift measurements.[10]

* To whom correspondence should be addressed. E-mail: rozasi@tcd.ie. Fax: +353 1 671 2826.
† Trinity College Dublin.
‡ Instituto de Química Médica, CSIC.

TABLE 1: Matrix of All of the Possible DNA Triplexes

	A−T	T−A	G−C	C−G
A	A.A−T	A.T−A	A.G−C	A.C−G
T	T.A−T	T.T−A	T.G−C	T.C−G
G	G.A−T	G.T−A	G.G−C	G.C−G
C	C.A−T	C.T−A	C.G−C	C.C−G

However, experiments on DNA base pairs are very difficult to perform. This means that the gas phase approaches remain of a primary relevance in studies of molecular interactions in nucleic acids. The DNA bases have been thoroughly computationally studied by Sponer, Leszczynski, and Hobza using DFT and MP2 methods. They have studied not only the potential energy and free energy surfaces of all of the canonical and methylated base pairs but also the hydrogen bonding, staking interactions, and nonrigidity of these pairs.[11]

In the particular case of DNA triplexes, and among several studies, it is worth mentioning the experimental work of Dervan et al.[12] in which the association constants for the formation of 16 pyridine triple helical complexes were determined, the theoretical study of the electrostatic potential of various trimers performed by Gadre et al.,[13] and the computational work on triplexes of Shields et al.[14] Very recently, an extensive review studying the stability effects of bases modifications in nucleic acid triplexes has been published by Luque et al.[15]

We are aware that triplexes are affected by numerous other contributions such as stacking, hydration, counterions, phosphate environment, backbone restrictions, and other effects, which are expected to have a large impact in the structure and stability of nucleic acids. However, in the present paper, we have concentrated in the analysis of the geometry of the possible DNA triplexes and mainly, by means of the atoms in molecules theory,[16] on the nature of all of the hydrogen bonds (HBs) formed within the complexes and how the addition of the third base to a WC base pair influences the nature of the HB pattern of such a pair.

Methods

The geometries of all the monomers, dimers and trimers have been fully optimized at a C_s geometry with the program Gaussian 98[17] using the hybrid method Becke3LYP[18] with the 6-31+G**[19] basis set. Currently, the use of DFT data is mostly considered of better accuracy for base pairing.

Even though C_s symmetry has been imposed, it is known that some triples are intrinsically nonplanar, such is the case of G.A pairing.[11] However, the C_s symmetry represents a meaningful approximation and the energy difference with the nonplanar pair can be considered negligible.

Considering the four DNA bases, adenine (A), thymine (T), guanine (G), and cytosine (C), and their paring according to the WC approach, it will be possible to build triplexes by adding the third base to such WC pairs (see Table 1). The convention used along the paper to define the triplexes will be as follows: the base preceding the dot will be the third strand base; the base following the dot would be the WC strand base that is involved in the Hoogsteen (or reverse Hoogsteen) pairing with the third strand; the bases connected by a dash will be those in a WC configuration. From all of the possible triplexes, we have only considered those that are sterically feasible. That means that all of those triplexes implying the C−G WC interaction ([A,T,G,C]·C−G) have been discarded because no third base can be possibly added to the C. As well, in the case of the T−A WC pair, only the G.T−A trimer has been considered because in the other three trimers ([A,T,C].T−A) only one HB could

TABLE 2: Dimer Interaction Energies (kcal mol⁻¹), Calculated BSSE (kcal mol⁻¹), and BSSE Corrected Dimer Interaction Energies (kcal mol⁻¹)

DIMER	E_I	BSSE	E_{I+BSSE}
A−T(WC)	−12.75	0.72	−12.03
A−T(RWC)	−11.09	0.73	−10.36
G−C(WC)	−26.68	1.02	−25.66
A−A(N1N7−RH)	−10.19	0.65	−9.54
A−A(N7N7−RH)	−8.69	0.61	−8.08
T−A(H)	−12.25	0.71	−11.53
T−A(RH)	−6.59	0.76	−5.83
G−A(N7N1)	−12.58	0.75	−11.83
A−G(N6−H)	−7.51	0.48	−7.02
C−A(RH)	−13.25	0.74	−12.51
G−T(W)	−9.00	0.71	−8.29
G−T(RW)	−15.71	0.75	−14.96
T−G(N3−H)	−5.73	0.56	−5.17
G−G(H)	−13.03	0.60	−12.40
G−G(RH)	−17.97	0.76	−17.21
C−G(H)	−11.91	0.58	−11.33
C+−G(H)	−42.23	0.98	−41.25

be established between the third base and the WC pair. Moreover, the protonated cytosine (C+) has been considered because the corresponding C+.G−C base triplet is one of the most stable experimentally found, probably because of the added proton and the charge−dipole interaction established.

To assess the relative stabilities of the triplexes, it is natural to consider the interaction energy (E_{TB}) of a triplex from its precursors, a single base and the corresponding duplex. Thus, the energy needed to add the third base has been evaluated as the difference between the energy of the third base (as the isolated monomer) and that of the WC dimer:

$$E_{TB} = E_{XYZ} - E_{XY} - E_Z$$

In addition, the total interaction energy, calculated as the difference of the total energy of the complex and the sum of the three isolated monomers, can give an insight into the total stability of the triplex rather than the stability of the third base binding:

$$E_I = E_{XYZ} - E_X - E_Y - E_Z$$

All of the interaction energies have been corrected of the inherent basis set superposition error (BSSE) using the Boys−Bernardi counterpoise technique.[20]

The topological properties of the electron density at the bond critical points (BCPs) have been characterized using the atoms in molecules methodology (AIM)[16] with the AIMPAC program package[21] at the B3LYP/6-31+G** level. On the basis of the AIMs theory, we have chosen the electron density at the bond critical point (ρ_{BCP}), and its Laplacian ($\nabla^2\rho_{BCP}$) as criteria to establish hydrogen bonding.

Results and Discussion

Interaction Energies of Dimers and Triplexes. All of the dimers involved in the possible triplexes studied in this article were optimized, and the results obtained for the interaction energies, BSSEs, and corrected interaction energies of all these dimers are gathered in Table 2.

In DNA, A binds to T and G to C, otherwise it is called mismatch base pairing. Of the complementary base pairing, the WC bonding scheme has the lowest corrected interaction energy (E_{I+BSSE}) which means it is the most stable followed by the Hoogsteen (H) interactions and then the reverse Hoogsteen (RH) which have a surprisingly low E_{I+BSSE}. According to Hobza et al.,[22] the electronic distributions of both A and T are character-
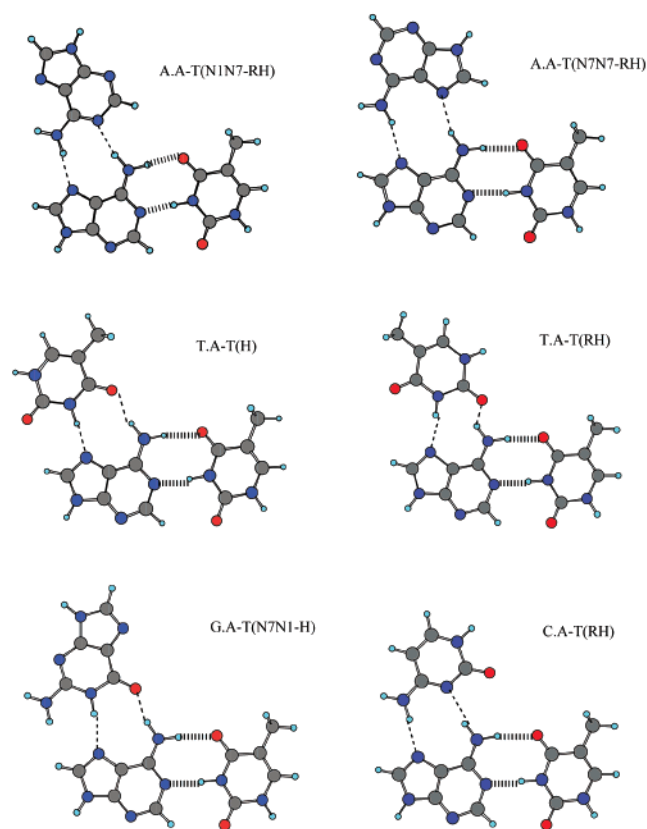
DNA Triplexes: A Study of Their Hydrogen Bonds

J. Phys. Chem. B, Vol. 107, No. 1, 2003 **325**
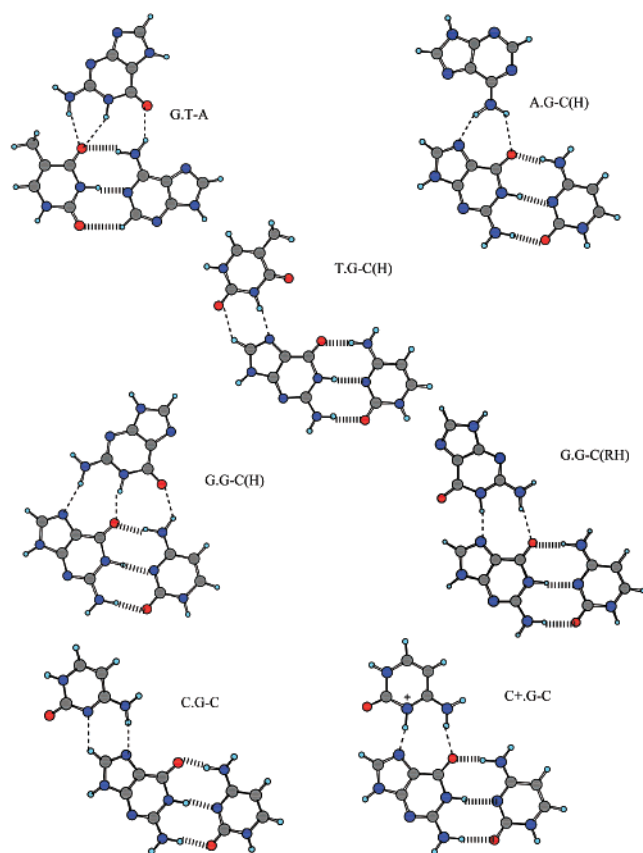

Figure 2.

Figure 3.

ized by a rather low polarity. This explains why the stability of all four A−T structures (WC, RWC, H, and RH) calculated for nucleic acids with the Amber force field is essentially identical.[23]

TABLE 3: Trimer Interaction Energies (kcal mol⁻¹), Calculated BSSE (kcal mol⁻¹), and BSSE Corrected Trimer Interaction Energies (kcal mol⁻¹)

TRIMER	E_I	BSSE	E_{I+BSSE}
A.A−T(N1N7−RH)	−22.28	1.38	−20.90
A.A−T(N7N7−RH)	−21.63	1.39	−20.24
T.A−T(H)	−25.01	1.42	−23.59
T.A−T(RH)	−24.69	1.45	−23.23
G.A−T(N7N1−H)	−23.46	1.40	−22.06
C.A−T(RH)	−21.51	1.46	−21.05
G.T−A	−21.97	1.47	−20.50
A.G−C(H)	−35.07	1.55	−33.51
T.G−C(H)	−33.11	1.62	−31.49
G.G−C(H)	−49.36	1.91	−47.45
G.G−C(RH)	−44.56	1.89	−42.67
C.G−C	−37.94	1.61	−36.33
C+.G−C	−70.38	2.06	−68.32

TABLE 4: Third Base Binding Energies (kcal mol⁻¹), Calculated BSSE (kcal mol⁻¹), and BSSE Corrected Third Base Energies (kcal mol⁻¹)

TRIMER	E_{TB}	BSSE	$E_{TB+BSSE}$
A.A−T(N1N7−RH)	−8.88	1.38	−7.51
A.A−T(N7N7−RH)	−9.53	1.39	−8.14
T.A−T(H)	−12.26	1.42	−10.84
T.A−T(RH)	−11.94	1.45	−10.49
G.A−T(H)	−10.71	1.40	−9.32
C.A−T(RH)	−9.76	1.46	−8.30
G.T−A	−9.22	1.47	−7.76
A.G−C(H)	−8.39	1.55	−6.84
T.G−C(H)	−6.43	1.62	−4.81
G.G−C(H)	−22.68	1.91	−20.77
G.G−C(RH)	−17.87	1.89	−15.99
C.G−C	−11.26	1.61	−9.65
C+.G−C	−43.69	2.06	−40.17

TABLE 5: Cooperativity (E_C, kcal mol⁻¹) Calculated as the Difference between the Interaction Energy of the Trimer XYZ (E_I, kcal mol⁻¹) Minus the Sum of the Interaction Energies of the Corresponding Dimers (E_I(XY) and E_I(YZ), kcal mol⁻¹)

TRIMER	E_C	TRIMER	E_C
A.A−T(N1N7−RH)	0.67	A.G−C(H)	−0.83
A.A−T(N7N7−RH)	−0.13	T.G−C(H)	−0.66
T.A−T(H)	−0.03	G.G−C(RH)	0.2
T.A−T(RH)	−5.37	C.G−C	0.66
G.A−T(H)	1.8	C+.G−C	−1.41
C.A−T(RH)	3.49		

However, and having calculated the A−T(RWC) structure for the sake of comparison, according to our results, A−T WC, RWC, and H have similar stabilization (around −10 to −12 kcal mol⁻¹, see Table 2), and the corresponding RH pair is less stable (−6 kcal mol⁻¹). The protonated dimer C+−G(H) shows a very high interaction energy because of the molecular charge− molecular dipole interaction.[11h] In this dimmer, C presents an extra proton which establishes a very strong HB with the N7 of G. The mismatch pairs are slightly more complex. The traditional Wobble structure [G−T(W)] has a interaction energy (E_{I+BSSE}) of −8.29 kcal mol⁻¹ but the reverse Wobble structure [G−T(RW)] has a lower stability of −14.96 kcal mol⁻¹. However, the homopurines are rather different; they show a variable strength of base pairing, in agreement with the work of Sponer.[23]

All of the optimized triplexes studied are represented in Figures 2 and 3. The HBs present in the WC pairs are depicted by an intermittent line (‖‖‖), whereas those formed with the third base are represented by a dashed line (- - -). The results obtained for the interaction energies (E_I), BSSEs, and corrected interaction energies (E_{I+BSSE}) of all triplexes are gathered in

TABLE 6: Electron Density (ρ_{BCP}, e a.u.$^{-3}$), Laplacian ($\nabla^2\rho_{BCP}$, e a.u.$^{-5}$), and Energy Density (H_{BCP}, Hartree a.u.$^{-3}$) Calculated at the Bond Critical Points of the HBs Formed in the Complementary Base Pairs with the B3LYP/6-31+G Method[a]**

dimer	atom numbers	ρ_{BCP}	$\nabla^2\rho_{BCP}$	H_{BCP}	d(H...X)	a(D−H...X)
A−T(WC)	A(H6)−T(O4)	0.036	0.084	−0.002	1.92	173.9
	A(N1)-−(H3)	0.036	0.100	−0.001	1.83	179.1
	A(H2)−T(O2)	0.005	0.012	0.001	2.86	132.5
G−C(W)	C(N3)−G(H1)	0.033	0.079	−0.001	1.91	177.1
	C(O2)−G(H2)	0.027	0.077	0.000	1.91	178.3
	C(H4)−G(O2)	0.038	0.121	0.000	1.76	179.1
	G(H2)−C(N3)	0.045	0.141	−0.001	2.01	174.6
T−A(H)	A(H6)−T(O4)	0.024	0.071	0.000	1.95	170.8
	A(N7)−T(H3)	0.041	0.097	−0.002	1.81	175.5
T−A(RH)	A(H6)−T(O2)	0.025	0.072	0.000	1.95	170.3
	A(N7)−T(H3)	0.041	0.096	−0.002	1.81	176.3
C+−G(H)[b]	C+(H4)−G(O2)	0.040	0.120	0.000	1.73	168.4
	C+(H3)−G(N7)	0.053	0.107	0.005	1.70	174.3

[a] The distances (Å) and angles (°) of such interactions are also included. [b] Protonated cytosine.

TABLE 7: Electron Density (ρ_{BCP}, e a.u.$^{-3}$), Laplacian ($\nabla^2\rho_{BCP}$, e a.u.$^{-5}$), and Energy Density (H_{BCP}, Hartree a.u.$^{-3}$) Calculated at the Bond Critical Points of the HBs Formed in the Mismatch Base Pairs with the B3LYP/6-31+G Method[a]**

dimer	atom numbers	ρ_{BCP}	$\nabla^2\rho_{BCP}$	H_{BCP}	d(H...X)	a(D−H...X)
G−A(N7N1)	A(N7)−G(H1)	0.023	0.056	0.000	2.07	175.8
	A(H6)−G(O2)	0.032	0.099	0.001	1.83	165.6
C−A(RH)	A(N7)−C(H4)	0.028	0.069	0.000	1.97	176.0
	A(H6)−C(N3)	0.028	0.071	0.000	1.97	168.1
G−T(W)	G(O2)−T(H3)	0.035	0.098	0.000	1.81	178.5
	G(H1)−T(O2)	0.039	0.113	0.000	1.77	174.6
G−T(RW)	G(H1)−T(O4)	0.038	0.111	0.000	1.78	178.97
	G(O2)−T(H3)	0.035	0.099	0.000	1.80	174.91
A−G (H)	A(H6)−G(O2)	0.010	0.035	−0.001	2.42	128.0
	A(H6)−G(N7)	0.020	0.057	0.000	2.11	151.9

[a] The distances (Å) and angles (°) of such interactions are also included.

TABLE 8: Electron Density (ρ_{BCP}, e a.u.$^{-3}$), Laplacian ($\nabla^2\rho_{BCP}$, e a.u.$^{-5}$), and Energy Density (H_{BCP}, Hartree a.u.$^{-3}$) Calculated at the Bond Critical Points of the HBs Formed in the Homopurine Base Pairs with the B3LYP/6-31+G Method[a]**

dimer	atom numbers	ρ_{BCP}	$\nabla^2\rho_{BCP}$	H_{BCP}	d(H...X)	a(D−H...X)
A−A(N1N7−RH)	A(H1)−A(N6)	0.026	0.064	0.000	2.02	165.9
	A(N6)−A(H7)	0.027	0.069	0.000	1.98	175.3
A−A(N7N7−RH)	A(H6)−A(N7)	0.020	0.062	0.000	2.06	160.4
	A(N7)−A(H6)	0.020	0.062	0.000	2.06	160.4
G−G-H	G(O2)−G(H1)	0.029	0.085	0.000	1.95	176.2
	G(N7)−G(H2)	0.015	0.049	0.001	2.23	159.4
G−G-RH	G(O2)−G(H2)	0.017	0.049	0.000	2.11	167.6
	G(N7)−G(H1)	0.036	0.091	−0.001	1.86	172.1
	G(H8)−G(O2)	0.075	0.028	0.001	2.58	122.3

[a] The distances (Å) and angles (°) of such interactions are also included.

Table 3, and the results for the third base interaction energies (E_{TB}, BSSEs, and $E_{TB+BSSE}$) can be found in Table 4.

The triplex T.A−T has been found experimentally (using different techniques including UV Melting) to be one of the most stable triplexes in Py−Pu−Py oligomers.[24] Also, modeling studies on d(T.A−T)[27] by Cheng and Pettitt[25] showed T.A−T(H) to be slightly more stable than T.A−T(RH), which is similar to the findings of the present study (see Table 3). The results here obtained predict G.G−C to be the most stable neutral triplex. This is consistent with earlier work,[26] in which it was shown that the major contribution to this stability comes from the interaction of the G−C(WC) base pair (−25 kcal mol^{-1}) involving three hydrogen bonds. The stabilizing effect of this particular duplex is also observed in the energy values obtained for the C.G−C triplex that is the third with the lowest E_I. The corresponding charged triplex C+.G−C is the most

stable overall because of the additional stabilizing charge-dipole interaction. This has also been found by other authors.[12]

In 1992, Wang et al.[26] carried out UV melting experiments on some triplexes and showed that T.A−T, C.G−C, and G.T−A, in a 32 oligomer, melt in a single transition (i.e., yielding directly the corresponding three monomers), whereas A.T−A, T.T−A, and C.T−A melt in two stages (i.e., yielding first a monomer and a dimer and then, in a second step, the three monomers). This information can be compared to our *ab initio* calculations because, if the third base interaction energy ($E_{TB+BSSE}$, dissociation into a monomer and a dimer) is similar to that of the original isolated dimer to which it is attached (dissociation of the dimmer into two monomers), then the triplexes will possibly melt in one step. This implies that the experimental finding of having a single melting temperature for a triplex has a molecular origin.[13] This is true for both T.A−T,

DNA Triplexes: A Study of Their Hydrogen Bonds

J. Phys. Chem. B, Vol. 107, No. 1, 2003 **327**

TABLE 9: Electron Density (ρ_{BCP}, e a.u.$^{-3}$), Laplacian ($\nabla^2\rho_{BCP}$, e a.u.$^{-5}$) and Energy Density (H_{BCP}, Hartree a.u.$^{-3}$) Calculated at Bond Critical Points with B3LYP/6-31+G Method for All of the Trimers Studied[a]**

trimer	atom numbers	ρ_{BCP}	$\nabla^2\rho_{BCP}$	H_{BCP}	d(H...X)	a(D−H...X)
A.A−T(N1N7−RH)	A(H6)−T(O4)	0.025	0.071	0.000	1.95	177.6
	A(N1)−T(H3)	0.041	0.089	−0.002	1.82	179.7
	A(H2)−T(O2)	0.005	0.018	0.001	2.80	133.1
	A(H6)−A(N7)	0.021	0.055	0.000	2.10	161.0
	A(N1)−A(H6)	0.024	0.064	0.000	2.04	157.7
A.A−T(N7N7−RH)	A(H6)−T(O4)	0.025	0.069	−0.001	1.96	176.9
	A(N1)−T(H3)	0.040	0.089	−0.002	1.83	179.9
	A(H2)−T(O2)	0.005	0.018	0.001	2.80	134.0
	A(H6)−A(N7)	0.026	0.065	0.000	2.01	176.3
	A(N7)−A(H6)	0.025	0.063	0.000	2.03	163.9
T.A−T(H)	A(H6)−T(O4)	0.024	0.067	0.000	1.97	174.6
	A(N1)−T(H3)	0.040	0.090	−0.002	1.82	179.7
	A(H2)−T(O2)	0.005	0.019	0.001	2.79	134.4
	T(H4)−A(N6)	0.041	0.096	−0.001	1.81	175.9
	T(O3)−A(H7)	0.022	0.064	0.000	2.00	169.8
	T(O2)−A(H8)	0.005	0.019	0.001	2.81	122.6
T.A−T(RH)	A(H6)−T(O4)	0.023	0.064	0.000	1.99	174.4
	A(N1)−T(H3)	0.041	0.091	−0.002	1.82	178.9
	A(H2)−T(O2)	0.004	0.015	0.001	2.77	134.6
	T(O2)−A(H6)	0.022	0.064	0.000	2.00	169.6
	T(H3)−A(N7)	0.041	0.097	−0.002	1.81	175.7
	T(O4)−A(H8)	0.004	0.015	0.001	2.80	122.6
G.A−T(H)	A(H2)−T(O2)	0.005	0.019	−0.001	2.78	135.3
	A(N1)−T(H3)	0.039	0.089	0.002	1.83	179.6
	A(H6)−T(O4)	0.024	0.067	0.000	1.97	176.2
	G(H1)−A(N7)	0.023	0.057	0.000	2.07	176.8
	G(O2)−A(H6)	0.030	0.092	0.000	1.86	165.2
C.A−T(RH)	A(H6)−T(O4)	0.021	0.058	0.000	2.04	168.5
	A(N1)−T(H3)	0.040	0.091	−0.002	2.04	177.9
	C(H4)−A(N7)	0.029	0.073	0.000	1.96	180.0
	C(N3)−A(H6)	0.022	0.057	0.000	2.08	159.0
G.T−A	T(O2)−A(H2)	0.008	0.029	−0.001	2.53	134.5
	T(H3)−A(N1)	0.043	0.091	−0.003	1.79	173.9
	T(O4)−A(H6)	0.016	0.043	0.000	2.19	178.6
	G(H2)−T(O4)	0.006	0.030	−0.002	2.15	157.0
	G(H1)−T(O4)	0.009	0.035	−0.001	2.36	150.1
	G(O2)−A(H6)	0.021	0.061	0.000	2.05	146.9
A.G−C	G(H2)−C(O2)	0.027	0.078	0.000	1.91	177.5
	G(H1)−C(N3)	0.033	0.078	0.001	1.91	176.8
	G(O2)−C(H4)	0.036	0.105	0.000	1.79	178.5
	A(H6)−G(O2)	0.010	0.037	−0.001	2.39	129.0
	A(H6)−G(N7)	0.021	0.059	0.000	2.09	151.3
T.G−C(H)	G(H2)−C(O2)	0.027	0.077	0.000	1.92	177.7
	G(H1)−C(N3)	0.033	0.079	0.000	1.91	176.6
	G(O2)−C(H4)	0.039	0.114	0.000	1.75	178.6
	T(H3)−G(N7)	0.027	0.067	0.000	1.96	163.0
G.G−C(H)	G(H1)−C(N3)	0.035	0.080	−0.001	1.89	179.0
	G(H2)−C(O2)	0.032	0.093	0.000	1.84	179.7
	G(O2)−C(H4)	0.032	0.093	0.000	1.85	179.9
	G(H1)−G(O2)	0.026	0.082	0.000	1.89	176.7
	G(H2)−G(N7)	0.021	0.055	0.000	2.10	164.9
	G(O2)−C(H4)	0.017	0.064	0.001	2.22	125.6
G.G−C(RH)	G(H2)−C(O2)	0.029	0.083	0.000	1.88	178.4
	G(H1)−C(N3)	0.033	0.079	−0.001	1.90	177.3
	G(O2)−C(H4)	0.034	0.098	0.000	1.82	178.9
	G(H1)−G(N7)	0.037	0.094	−0.001	1.84	174.5
	G(H2)−G(O2)	0.018	0.050	−0.001	2.11	164.5
	G(O2)−G(H8)	0.008	0.027	0.001	2.60	124.3
C.G−C	G(H2)−C(O2)	0.027	0.077	0.000	1.91	178.2
	G(H1)−C(N3)	0.033	0.079	0.001	1.91	177.1
	G(O2)−C(H4)	0.038	0.110	0.000	1.77	179.4
	C(N3)−G(H8)	0.013	0.039	−0.001	2.41	131.6
	C(H4)−G(N7)	0.030	0.076	0.000	1.93	169.5
C+.G−C	G(H2)−C(O2)	0.036	0.106	0.000	1.79	179.4
	G(H1)−C(N3)	0.034	0.078	0.001	1.90	179.4
	G(O2)−C(H4)	0.025	0.070	0.001	1.95	177.6
	C+(H2)−G(O2)	0.041	0.118	0.000	1.74	166.4
	C+(H3)−G(N7)	0.065	0.107	0.012	1.62	175.4

[a] The distance (Å) and angle (°) of such interactions are also included.

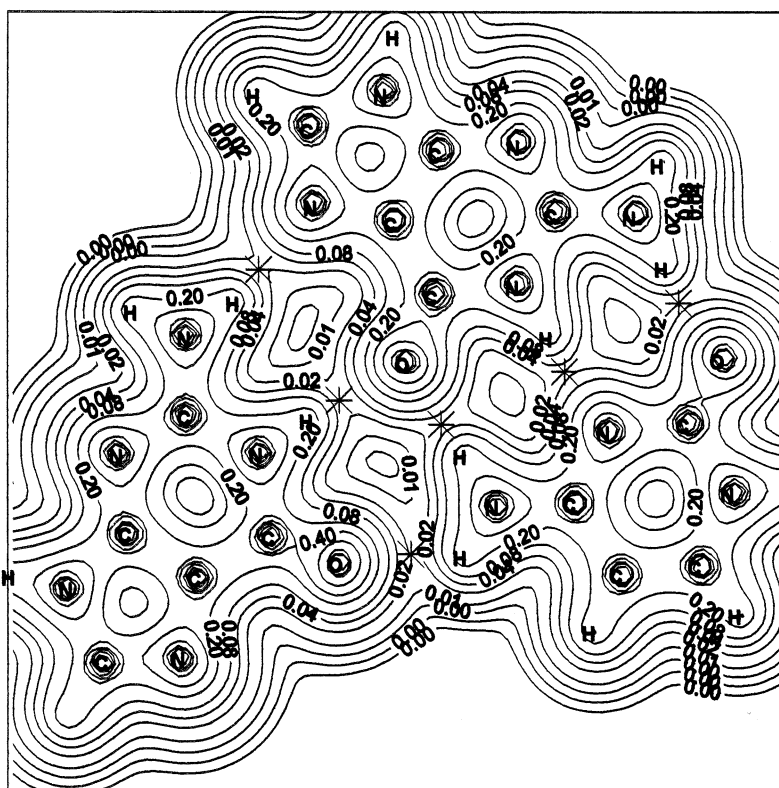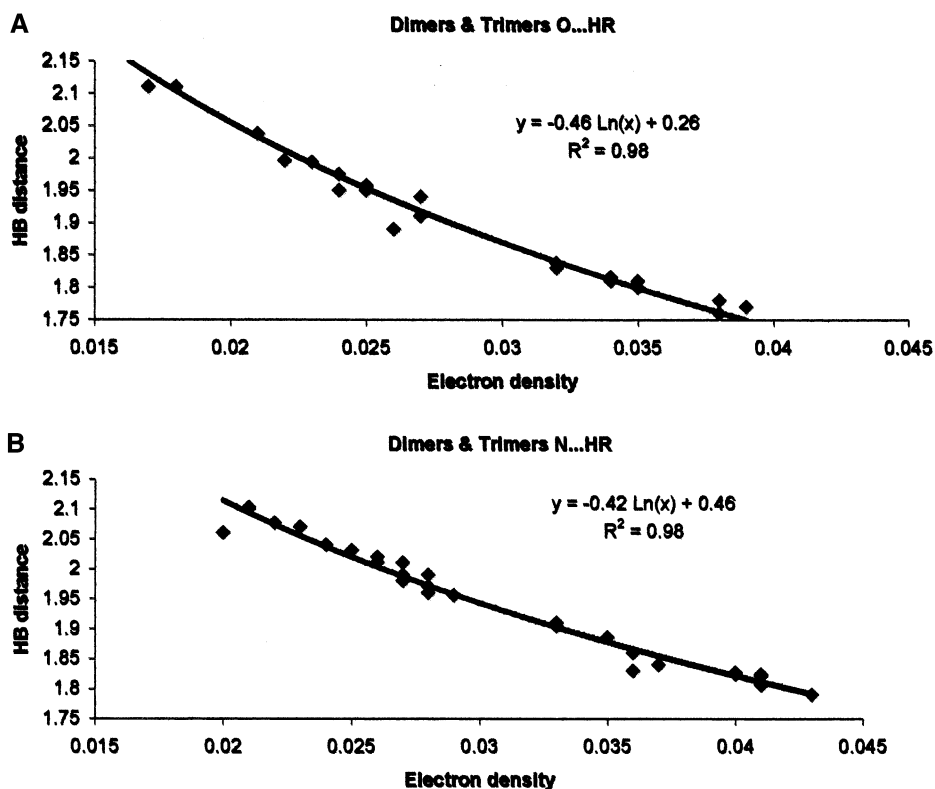**Figure 4.**

Figure 5.

which has a third base interaction energy of about −11 kcal mol⁻¹ and a A−T(WC) interaction energy of −12 kcal mol⁻¹, and for G.G−C, with a third base energy around −21 kcal mol⁻¹ and a G−C(WC) energy around −26 kcal mol⁻¹.

When studying the interaction of at least three molecules, cooperativity effects are often observed. Therefore, the addition of a third base to a WC pair could result in an energetic gain

by those cooperativity effects (E_C). Thus, the comparison of the interaction energies of all of the dimers involved in the triplexes of DNA (see Table 2) and those of the triplexes themselves (see Table 3, except those in which the third base is interacting with both of the WC bases, G.T−A and G.G−C(H)) show that in the case of A.A−T(N1N7−RH), G.A−T(H), C.A−T(RH), G.G−C(RH), and C.G−C the values are smaller

DNA Triplexes: A Study of Their Hydrogen Bonds

J. Phys. Chem. B, Vol. 107, No. 1, 2003 **329**

in the latter case as an indication of no cooperativity effects (see Table 5). However, in the rest of the trimers, small energetic gains are generally observed. Exceptions are C+.G−C with a cooperativity effect of −1.41 kcal mol⁻¹ and especially T.A−T(RH) with an energetic gain of −5.37 kcal mol⁻¹. In this last case, it is interesting that, even though T.A−T(H) and T.A−T(RH) are structurally very similar, their corresponding E_C are very different being almost negligible in the first case and the largest value in the second. This could be explained by the fact that the third thymine changes from a Hoogsteen interaction to a reverse Hoogsteen interaction inverting the dipole moment orientation of the molecule and this will affect the final cooperativity effects.

Electron Density Analysis of the Dimers and Triplexes. Another way to investigate the molecular binding of the third base is to look at the effect on the duplex upon binding. By using Bader's theory of "atoms in molecules" (AIM),[16] the properties of the established HBs were studied, first for the dimers and then for the triplexes to explore the true nature of the binding.

The presence of BCPs between a H atom and an acceptor one, with the characterization of electron density in such a point given in the Methods section, can be taken as an absolute indicator of a bonding interaction and more particularly of a HB. In addition, interatomic distances and angles have also been considered. The H···A distance should be smaller than the sum of the their atomic radii (according to Gavezzoti's atomic radii,[27] a O···H−R HB should have a maximum length of 2.57 Å and a N···H−R one 2.72 Å).

Using the above limits, A−T(WC) has two strong HBs, but the interaction in A−T(WC):A(H2)-T(O2) cannot be considered a HB, as can be seen in Table 6, because the bond distance is too large and the electron density is too small. G−C(WC) has three HBs, the center HB (C(N3)-G(H1)) is strong, whereas the other two are of medium strength. The other complementary base pairs in Table 6 have two HBs each, and they are medium to strong HBs. In particular, the duplex formed by the protonated cytosine shows two very strong HBs with extremely short distances.

According to the electron density characteristics at the BCPs and the interatomic distances and angles, the mismatch base pairs have medium strength HBs (see Table 7). The homopurine base pairs (see Table 8) show similar results to that of the mismatch base pairs, most pairs having medium strength HBs. The G−G(RH) base pair has the lowest interaction energy (−17.21 kcal mol⁻¹) because it has two medium/strong HBs and one weak HB.

According to the interaction energy data (see Table 3), the G.G−C(H) is the most stable of the neutral triplexes. The major contribution to this stability comes from the weak HB between the first and third base in the triplex, C(H4)−G(O2). This interaction is missing in the G.G−C(RH) (see Figure 3 and Table 9). G.G−C(H)'s electron density plot can be seen in Figure 4, and the BCPs at which the electron and energy densities and the Laplacian of the electron density were calculated are denoted by an asterisk. The G.G−C(H) triplex has six HBs; G.G−C(RH), C.G−CA.G−C(H), and T.G−C(H) have five; and T.A−T(H), T.A−T(RH), G.A−T(N7N1−H), and C.A−T(RH) have four HBs each (see Table 9), following the same trend seen in their interaction energies. A.A−T(N1N7−RH) has two strong HBs and two medium strength ones, and A.A−T(N7N7−RH) has one strong HB and three medium strength HBs (see Table 9), again correlating with the interaction energies, (A.A−T(N1N7−RH) −20.90 kcal mol⁻¹ and A.A−

T(N7N7−RH) −20.24 kcal mol⁻¹). In other words, the stronger the HBs the lower the interaction energy. C.A−T and G.T−A have medium to strong HBs also with similar interaction energies to that of A.A−T.

Interestingly, the major groove angle seems to be increasing, possibly allowing more space for the third base to bind, e.g., the G.T−A:T(O6)−A(H4) bond distance increases from 1.92 to 2.19 Å. T.G−C has a surprisingly low interaction energy (−33.49 kcal/mol) but it is justified by looking at the strength of the HBs, all four of them being medium to strong (high electron density and short bond lengths). As well, C.G−C has a very low interaction energy (−36.33 kcal mol⁻¹) justifiable by the four strong HB and another HB of medium strength (see Table 9).

However, the corresponding protonated triplex, C+.G−C, shows the same number of HBs, but according to the electron density, all of them are strong or very strong as in the case of C+(H3)-G(N7), which corresponds to the added proton, with very high electron and energy density and extremely short bond distance (1.62 Å).

By looking at the data obtained and gathered in Table 9, it was observed that the HBs which include a N have a slightly higher electron density at the BCP when in the trimers than in the dimers. Good logarithmic relationships were found between all of the ρ_{bcp} and the HB distances from Tables 6−9 for the oxygen HBs (represented in Figure 5 A) and for the nitrogen HBs (see Figure 5 B). This logarithmic relationship and the dependence on the acceptor atom have been observed in our previous studies.[28]

Conclusions

The structures of the bases of these DNA triplexes have been studied, by means of *ab initio* calculations, in terms of their interaction energies and electron densities for the different hydrogen bonds established, which is one of the interactions that determines the structure and dynamics of nucleic acid molecules. A good agreement has been found with experimental results and previous calculations. The trends observed for the interaction energies are consistent with those observed by experimental and/or by molecular dynamics studies in the case of T.A−T, C+.G−C, and G.G−C. As for G.T−A, the *ab initio* predictions are different than the experimental ones, and this may be due to the absence of stacking effects.

The electron analysis results agree rather well with the calculated interaction energies. When a third base is added to a dimer, the strength of the HBs of the dimer becomes modified. These changes in the dimers HBs provoke modifications in their structures by increasing the distance in the major groove and decreasing the space in the minor groove. This modification in the original double helix may contribute to explain why transcription is inhibited upon triplex formation. The electron density analysis here performed gives insights of the nature of the interactions established among the bases and, therefore, would allow for a better understanding of the nucleic acid triplexes formation and how their structure and HB patterns could have biological implications.

References and Notes

(1) Rich, A.; Davies, D. R. *J. Am. Chem. Soc.* **1956**, *78*, 3548.
(2) Felsenfeld, G. *J. Am. Chem. Soc.* **1957**, *79*, 2023.

(3) Faria, M.; Wood, C. D.; White, M. R. H.; Helene, C.; Giovannangeli, C. *J. Mol. Biol.* **2001**, *306*, 15.

(4) Kraulis, P. J. *J. Appl Cryst.* **1991**, *24*, 946.

(5) Hanvey, J. C.; Peffer, N. J.; Bisi, J. E.; Thomson, S. A.; Cadilla, R.; Josey, J. A.; Ricca, D. J.; Hassman, C. F.; Bonham, M. A.; Au, K. G.; Carter, S. G.; Bruckenstein, D. A.; Boyd, A. L.; Noble, S. A.; Babiss, L. E. *Science* **1992**, *258*, 1481.

(6) Nielsen, P. E. *Curr. Opin. Struct. Biol.* **1999**, *9*, 353.

(7) (a) Boulme, F.; Freund, F.; Moreau, S.; Nielsen, P. E.; Gryaznov, S.; Toulme, J. J.; Litvak, S. *Nucleic Acids Res.* **1998**, *26*, 5492. (b) Vickers, T. A.; Griffith, M. C.; Ramasamy, K.; Risen, L. M.; Freier, S. M. *Nucleic Acids Res*. **1995**, *23*, 3003.

(8) (a) Chen, X.; Kierzek, R.; Turner, D. H. *J. Am. Chem. Soc.* **2001**, *123*, 1267. (b) Kankia, B. I.; Marky, L. A. *J. Phys. Chem. B* **1999**, *103*, 8759.

(9) Sedelnikova, O. A.; Panyutin, I. G.; Luu, A. N.; Neumann, R. D. *Nucleic Acids Res.* **1999**, *27*, 3844.

(10) Hoyne, P. R.; Gacy, A. M.; McMurray, C. T.; Maher, L. J., III. *Nucleic Acids Res*. **2000**, *28*, 770.

(11) Between others: (a) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Mol. Struct. (THEOCHEM)* **2001**, *573*, 43. (b) Sponer, J.; Berger, I.; Spackova, N.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn. Conversation* **2000**, *11*, 383. (c) Kabelac, M.; Hobza, P. *J. Phys. Chem. B* **2001**, *105*, 5804. (d) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149. (e) Ryjacek, F.; Engkvist, O.; Vacek, J.; Kratochvil, M.; Hobza, P. *J. Phys. Chem. A* **2001**, *105*, 1197. (f) Sponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers* **2001−2002**, *61*, 3. (g) Sponer, J.; Leszczynski, J.; Hobza, P. *Theoretical and Computational Chemistry*, *Vol. 8, Computational Molecular Biology*; Elsevier: Amsterdam, 1999. (h) Sponer, J.; Burda, J. V.; Mejzlik, P.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1997**, *14*, 613. (i) Sponer, J.; Sabat, M.; Burda, J. V.; Doody, A. M.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1998**, *16*, 139.

(12) Best, G. C.; Dervan, P. B. *J. Am. Chem. Soc*. **1995**, *117*, 1184.

(13) Pundlik, S. S.; Gadre, S. R. *J. Phys. Chem. B* **1997**, *101*, 9657.

(14) Shields, G. C.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc*. **1997**, *119*, 7463.

(15) Robles, J.; Grandas, A.; Pedroso, E.; Luque, F. J.; Eritja, R.; Orozco, M. *Curr. Org. Chem*. **2002**, in press.

(16) Bader, R. F. W. *Atoms in Molecules. A Quantum Theory*; Oxford University: New York, 1990.

(17) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.5; Gaussian, Inc.: Pittsburgh, PA, 1998.

(18) (a) Becke, A. D. *J. Chem. Phys*. **1993**, *98*, 5648. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(19) Hariharan, P. A.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.

(20) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(21) Bieger-Konig, F. W.; Bader, R. F. W.; Tang, T. H. *J. Comput. Chem.* **1982**, *3*, 317.

(22) Kratochvil, M.; Sponer, J.; Hobza, P. *J. Am. Chem. Soc.* **2000**, *122*, 3495.

(23) Sponer, J. *J. Phys. Chem.* **1996**, *100*, 1965.

(24) Soyfer, V. N.; Potaman, V. N. *Triple-Helical Nucleic Acids*; Springer: Berlin, 1996.

(25) Cheng, Y.-K.; Pettitt, B. M. *J. Am. Chem. Soc.* **1992**, *114*, 4465.

(26) Wang, E.; Malek, S.; Feigon, J. *Biochemistry* **1992**, *31*, 4838.

(27) Gavezzotti, A. *J. Am. Chem. Soc.* **1983**, *105*, 5220.

(28) (a) Alkorta, I.; Rozas, I.; Elguero, J. *Struct. Chem*. **1998**, *9*, 243. (b) Alkorta, I.; Barrios, L.; Rozas, I.; Elguero, J. *J. Mol. Struct. (THEOCHEM)* **2000**, *496*, 131. (c) Espinosa, E.; Souhassou, M.; Lachekar, H.; Lecomte, C. *Acta Crystallogr.* **1999**, *B55*, 563. (d) Knop, O.; Rankin, K. N.; Boyd, R. J. *J. Phys. Chem. A* **2001**, *105*, 6552. (e) Espinosa, E.; Alkorta, I.; Elguero, J.; Molins, E. *J. Chem. Phys.* **2002**, *117*, 5529.