Nonempirical Calculations on All the 29 Possible DNA Base Pairs

Pavel Hobza*[1] and Camille Sandorfy*[2]

Contribution from the Dēpartement de Chimie, Universitē de Montrēal, Montrēal, Quēbec, Canada H3C 3J7. *Received June 25, 1986*

Abstract: A nonempirical ab initio SCF method employing Huzinaga's minimal MINI-1 basis set in combination with a London-type expression for the dispersion energy has been applied to the 29 different DNA base pairs. Fair agreement has been obtained between experimental gas-phase enthalpies and theoretical enthalpies. For complementary base pairs the heteropairs are more stable than the homopairs; the opposite is true for noncomplementary pairs. The different factors affecting the stability of the pairs are discussed. Dispersion energy represents for all the pairs an important part of the stabilization energy. The validity of empirical potentials and the electrostatic approximation is examined.

I. Introduction

Three main factors are responsible for keeping DNA in its unique structure:[3] (a) H-bonding between bases of the two strands (in the plane determined by a base pair), (b) stacking interactions between bases within each strand, and (c) interactions among phosphate ions (one in each of the two strands), counterions, and water molecules. The relative importance of these contributions is still not definitively known, but it seems that H-bonding between the bases plays the dominant role.

In DNA one finds only two base pairs: guanine (G)···cytosine (C) and adenine (A)···thymine (T), both in the Watson–Crick (WC) structure. Taking into account that H-bonding between bases cannot occur through the N9–H of purines and the N1–H of pyrimidines, there can be only 29 different hetero and homo base pairs; 28 of them were described by Donohue,[4] and the last one was described by Poltev and Shulyupina.[5] Are the two pairs occurring in DNA the most stable? This is an old question. There exists an extended body of papers[5-12] in which the structures and stabilization energies of different base pairs were studied by using semiempirical or empirical potentials. We do not intend to underestimate the usefulness of such potentials, but it must be stated very clearly that the results obtained in this way cannot be considered definite. Accumulated evidence[13,14] shows that such methods may give reasonable results (geometry, stabilization energy) in certain cases, but they fail in others.

The aim of the present paper is to investigate theoretically (ab initio SCF + dispersion energy) the structure and stabilization energies of the 29 different H-bonded pairs, formed by G, C, T, and A. In later papers the stacking complexes formed by these systems will be studied with the same theoretical approach.

II. Computational Strategy

Reliable data on the stabilization energy and geometry of a complex are obtained only by the ab initio SCF method in combination with an appropriate post-SCF method. The interaction energy (ΔE) is then constructed as the sum of the SCF interaction energy (ΔE^{SCF}) and the correlation interaction energy (ΔE^{COR}). Both contributions should be corrected for the basis set extention effect including the basis set superposition errors (BSSE) at the SCF and post-SCF levels (eq 1). The magnitude of correlation

$$\Delta E = \Delta E^{SCF} + \mathrm{BSSE(SCF)} + \Delta E^{COR} + \mathrm{BSSE(COR)} \quad (1)$$

interaction energy with various types of complexes is different, but in general, only with cation···neutral molecule complexes is it possible to neglect this energy completely. With H-bonded complexes this energy may amount to 20–50% of the total stabilization, and this portion becomes even larger for true van der Waals (vdW) complexes (e.g., Ar_2). Among different techniques suitable for evaluation of correlation interaction energy the many body perturbation treatment seems to be the most promising, and the technique has been used frequently in the last years. In order to obtain reasonable values of correlation interaction energy an extended basis set (at least of DZ+P quality) with flat polarization functions should be used, however. Using smaller basis sets, e.g., DZ or a minimal basis set, leads to strong underestimation of ΔE^{COR}, by a factor of 10 or even more.

With larger complexes, having 20 or more atoms, we are not able to work with extended basis sets, and in fact, it is only the minimal basis set which may be applied. Therefore another technique must be used for the estimation of ΔE^{COR}. It is known[13] that at larger distances ΔE^{COR} may be identified with the dispersion energy (E^D). ΔE^{COR} and E^D differ by the change of intrasystem correlation energy with varying distance between subsystems; this energy is usually of repulsive nature, and at the region of the vdW minimum it amounts to a fraction of E^D (10–20%). E^D may be evaluated with nonexpanded or expanded methods. In the former method an extended basis set with flat polarization functions is again required; the latter method, on the other hand, enables us to use experimentally determinable properties like ionization potentials, polarizabilities, etc. We can use the well-known exppressions developed by London,[15] Slater and Kirkwood,[16] or Müller.[17] It was found, however, that those equations underestimate dispersion energy. For example, the London expression, employing molecular polarizabilities and ionization potentials, is known to give about 50–75% of the accurate dispersion energy[14] (proportional to the sixth power of reciprocal distance). Recently a new method was proposed[18] for

(1) Institute of Hygiene and Epidemiology, Šrobārova 48, 100 42 Prague 10, Czechoslovakia.
(2) Dēpartement de Chimie, Universitē de Montrēal, Montrēal, Quēbec, Canada H3C 3J7.
(3) Watson, J. D.; Crick, F. H. C. *Nature (London)* **1953**, *171*, 737.
(4) Donohue, J. *Proc. Natl. Acad. Sci. U.S.A.* **1956**, *42*, 60. Donohue, J.; Trueblood, K. N. *J. Mol. Biol.* **1960**, *2*, 363.
(5) Poltev, V. I.; Shulyupina, N. V. *J. Biomol. Struct. Dyn.* **1986**, *3*, 739.
(6) De Voe, H.; Tinoco, I., Jr. *J. Mol. Biol.* **1962**, *4*, 500.
(7) Pullman, B.; Claverie, P.; Caillet, J. *Proc. Natl. Acad. Sci. U.S.A.* **1966**, *55*, 904.
(8) Stamatiadou, M. N.; Swissler, T. J.; Rabinowitz, J. R.; Rein, R. *Biopolymers* **1972**, *11*, 1217.
(9) Kudritskaya, Z. G.; Danilov, V. *J. Theor. Biol.* **1976**, *59*, 303.
(10) Maclagan, R. G. A. R. *Aust. J. Chem.* **1979**, *32*, 1635.
(11) Rein, R. In *Intermolecular Interactions: From Diatomics to Biopolymers*; Pullman, B., Ed.; Wiley-Interscience: New York, 1978; pp 308–363.
(12) Langlet, J.; Claverie, P.; Caron, F.; Boeuve, J. C. *Int. J. Quantum. Chem.* **1981**, *20*, 299.
(13) Hobza, P.; Zahradnik, R. *Weak Intermolecular Interactions in Chemistry and Biology*; Elsevier: Amsterdam, 1980.
(14) Hobza, P.; Zahradnik, R. *Intermolecular* Complexes. *The role of van der* Waals *Systems in Physical Chemistry and Biodisciplines*; Elsevier: Amsterdam, in press.
(15) London, F. *Z. Phys. Chem., Abt. B* **1930**, *B11*, 222.
(16) Slater, J. C.; Kirkwood, J. G. *Phys. Rev.* **1931**, *37*, 682.
(17) Müller, A. *Proc. R. Soc. London, A* **1936**, *A154*, 624.

0002-7863/87/1509-1302$01.50/0

obtaining the atomic static polarizabilities for an atom in a particular valence state. Using a London-type formula with experimental atomic polarizabilities and atomic ionization potentials (also in a particular valence state), Kang and Jhon[18] have obtained very good agreement with the accurate dispersion energy (proportional to the sixth power of reciprocal distance) for series of different molecules.

The interaction energy for the complexes studied was constructed as the sum of ΔE^{SCF}, BSSE at SCF level, and E^D (eq 2).

$$\Delta E = \Delta E^{SCF} + \text{BSSE} + E^D \quad (2)$$

II. 1. SCF Interaction Energy. The choice of the basis set represents a very difficult problem. With the complexes under study it is impossible to work with larger than minimal basis sets (the largest complex we have studied, G···G, has 120 AOs in the minimal basis set). The basis set chosen should be applicable not only to the presently studied H-bonded complexes but also to stacking complexes of G, C, T, and A, which will be studied later. Among minimal basis sets, Pople's STO-3G[19] is the most popular, partially because of its very fast performance. Unfortunately, this basis set failed in the field of molecular interactions; the respective BSSE is extremely large, comparable or even larger than the ΔE^{SCF} values themselves. Inclusion of the BSSE leads to very small values of SCF interaction energy. This is true not only for small H-bonded complexes [$(H_2O)_2$:[13] ΔE^{SCF} = -27.6 kJ/mol; BSSE = 23.4 kJ/mol] but also for larger H-bonded complexes [adenine···thymine:[20] ΔE^{SCF} = -59.0 kJ/mol; BSSE = 36.5 kJ/mol]. Further, the STO-3G underestimates repulsion; consequently the optimized distances are too short. (With H-bonded complexes they are too short by 0.2–0.3 Å.) This failure is even more visible with stacking complexes,[21] where STO-3G may lead to no minimum on the potential energy curve (the absolute value of E^D is larger than that of ΔE^{SCF} at any distance). There is no reason to neglect the basis set superposition error. Very convincing evidences were accumulated recently,[22-24] demonstrating the necessity to include the BSSE at the SCF level as well as at the post-SCF level. On the other hand, several papers criticized the concept of inclusion of the BSSE. Schwenke and Truhlar[25] have studied $(HF)_2$ and have found that the inclusion of BSSE does not improve the accuracy obtained with different basis sets. It is hardly possible to expect this from the mere inclusion of the BSSE, however, the main reason for the discrepancy being the incorrect description of subsystem multipole moments. Only after correcting for both multipole moments and BSSE may one expect a systematic improvement in the accuracy of ΔE^{SCF}. Collins and Gallup[26] concluded that inclusion of BSSE overcorrects the basis set extension error. These authors[26] have supposed that E^I (the sum of the Coulombic and exchange-repulsion energies) is not affected by basis set extension; this is not generally so, however (see, e.g., discussion in ref 24). Frisch et al.[27] compared the theoretical ΔH for $(H_2O)_2$ with its experimental value. Because of the close agreement between the uncorrected (BSSE not included) value of ΔH (-15.1 kJ/mol) and its experimental value (-15.5 kJ/mol), the authors[27] have concluded that it is not necessary to take the BSSE into account. Two important factors should be kept in mind, however: (i) the respective experimental value[28] is not -15.5 kJ/mol but is -15.5 ± 2.1 kJ/mol, and (ii) despite the inclusion of higher polarization functions (6-311G++(3df,3pd)), the basis set is still not saturated with respect to the calculation of ΔE^{COR} (see, e.g., ref 29 and 30). Hence, the actual (theoretical) ΔH should be larger than the value calculated by Frisch et al. In the light of these evidences their corrected value (for the BSSE) of ΔH (-12.1 kJ/mol) is not in disagreement with experiment either.

The Huzinaga MINI-1 basis set[31] gives, contrary to STO-3G, rather small values of BSSE. This basis set, contracted as the STO-3G, was constructed in a way to minimize the BSSE. We have tested this basis set for H-bonded complexes,[32] cation···neutral molecule complexes, and anion···neutral molecule complexes[33] as well as for stacking complexes.[21] In all the cases the corrected ΔE^{SCF} values (BSSE included) as well as the geometries were close to their values obtained with extended basis sets; BSSE was for all the complexes rather small. Computational time has increased (with respect to (STO-3G)) by no more than 20%.

The BSSE was evaluated by using the function counterpoise procedure introduced by Boys and Bernardi.[34] Contrary to the case of evaluation of BSSE for smaller H-bonded complexes, with the present complexes we have met serious problems with the convergency of the SCF procedure. Only after applying level shifting[35] were we able to overcome the divergency. In the first SCF iterations we have used a level shift parameter equal to 2.0 or 3.0 au; this value was decreased by half in each of the eight following iterations.

II. 2. Dispersion Energy. Dispersion energy was evaluated with a London-type formula employing the experimental atomic polarizabilities (α) and experimental ionization potentials (I)[18] in the particular valence state.

$$E^D = -\frac{3}{2}\sum_i^R\sum_j^T \frac{I_i I_j}{I_i + I_j}\alpha_i\alpha_i r_{ij}^{-6} \quad (3)$$

r means the distance between subsystems

Summations are running over all the atoms of subsystems R and T. Using atomic polarizabilities, Kang and Jhon[18] calculated the total polarizabilities for guanine, adenine, cytosine, and thymine; fair agreement with experimental values was obtained. Because of working with atomic characteristics, this approach takes into account the anisotropy of dispersion energy.

II. 3. Geometry Optimization. The geometries of G, C, T, and A were taken from Del Bene,[36] who obtained them through gradient optimization. We prefer theoretical geometries over experimental ones because (i) the positions of all the atoms are known (compare with X-ray experiments where the positions of the hydrogens are unknown), and (ii) in this way we shall be able to work consistently with other subsystems for which experimental geometries may not be known.

The intramolecular geometry was frozen for all the complexes, and only intermolecular coordinates were optimized. Existence of two or three H-bonds with the complexes studied ensures the coplanarity of the complexes; the number of intermolecular coordinates which should be optimized is therefore reduced from six to three. All the complexes possess X–H···Y–Z type H-bonds; we have optimized (point by point) the H···Y bond length and the XHY and HYZ angles. These three parameters describe the mutual positions of the two subsystems providing for the coplanarity of the complexes.

(18) Kang, Y. K.; Jhon, M. S. *Theor. Chim. Acta* **1982**, *61*, 41.
(19) Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657.
(20) Förner, W.; Otto, P.; Ladik, J. *Chem. Phys.* **1984**, *86*, 49.
(21) Hobza, P.; Mehlhorn, A.; Čársky, P.; Zahradník, R. *THEOCHEM* **1986**, *138*, 387.
(22) Gutowski, M.; van Lenthe, J. H.; Verbeek, J.; van Duijneveldt, F. B.; Chalasinki, G. *Chem. Phys. Lett.* **1986**, *124*, 370.
(23) Gutowski, M.; van Duijneveldt, F. B.; Chalasinski, G.; Piela, L. *Chem. Phys. Lett.* **1986**, *129*, 325.
(24) Szczesniak, M. M.; Scheiner, S. *Chem. Phys. Lett.* **1986**, *131*, 230.
(25) Schwenke, D. W.; Truhlar, D. G. *J. Chem. Phys.* **1985**, *82*, 2418.
(26) Collins, J. R.; Gallup, G. A. *Chem. Phys. Lett.* **1986**, *123*, 56.
(27) Frisch, M. J.; Del Bene, J. E.; Binkley, J. S.; Schaefer, H. F., III. *J. Chem. Phys.* **1986**, *84*, 2279.
(28) Curtiss, L. A.; Frurip, D. J.; Blander, M. *J. Chem. Phys.* **1973**, *71*, 2703.
(29) Chalasinski, G.; van Lenthe, J. H.; Groen, T. P. *Chem. Phys. Lett.* **1984**, *110*, 369.
(30) Knowles, P. J.; Meath, W. J. *Chem. Phys. Lett.* **1986**, *124*, 164.
(31) Tatewaki, H.; Huzinaga, S. *J. Chem. Phys.* **1979**, *71*, 4339.
(32) Hobza, P.; Sauer, J. *Theor. Chim. Acta* **1984**, *65*, 279.
(33) Sauer, J.; Hobza, P. *Theor. Chim. Acta* **1984**, *65*, 291.
(34) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
(35) Saunders, V. R.; Hillier, I. H. *Int. J. Quantum Chem.* **1973**, *7*, 699.
(36) Del Bene, J. E. *J. Phys. Chem.* **1983**, *87*, 367.

Table I. Geometrical Characteristics of Different DNA Base Pairs (Lengths in Å, Angles in deg)

pair[a]	H-bond[b]	R(X···Y)[c]	α(X–H···Y)	α(H···H–Y)
GC(WC)[d]	O6···H–N4	2.96		166
	N1–H···N3	2.94	164	
	N2–H···O2	2.91	171	
GG(I)	O6···H–N1	2.69		174
	N1–H···O6	2.84	180	
CC	N4–H···N3	2.91	175	
	N3···H–N4	2.87		174
GG(III)	N1–H···N7	2.84	177	
	N2–H···O6	3.06	153	
GC(II)	N3···H–N4	2.99		177
	N2–H···N3	2.83	171	
AC(I)	N6–H···N3	2.86	176	
	N1···H–N4	2.96		177
GA(I)	O6···H–N6	2.81		176
	N1–H···N1	2.96	180	
GG(IV)	N3···H–N2	2.89		180
	N2–H···N3	2.90	180	
GT(I)	O6···H–N3	2.77		177
	N1–H···O4	2.80	180	
GC(I)	N1–H···O2	2.85	180	
	N2–H···N3	3.03	169	
AT(RWC)[d]	N1···H–N3	2.88		177
	N6–H···O2	2.90	180	
GT(II)	O6···H–N3	2.76		170
	N1–H···O2	2.82	175	
AT(RH)[d]	N7···H–N3	2.87		176
	N6–H···O2	2.89	165	
AT(WC)[d]	N6–H···O4	2.97	177	
	N1···H–N3	2.91		180
AT(H)[d]	N6–H···O4	2.97	163	
	N7···H–N3	2.87		180
AA(I)	N6–H···N1	2.89	166	
	N1···H–N6	2.84		168
GG(II)	N1–H···O6	2.88	180	
	N2–H···N7	3.16	152	
AA(II)	N6–H···N7	2.95	180	
	N1···H–N6	2.98		163
GA(II)	N3···H–N6	2.98		165
	N2–H···N7	2.91	180	
GA(III)	O6···H–N6	2.88		159
	N1–H···N7	3.00	180	
GA(IV)	N3···H–N6	2.90		171
	N2–H···N1	2.94	171	
TC(I)	O4···H–N4	2.85		177
	N3–H···N3	2.91	180	
TC(II)	N3–H···N3	2.97	180	
	O2···H–N4	2.78		165
TT(III)	N3–H···O2	2.85	175	
	O2···H–N3	2.77		177
TT(II)	O4···H–N3	2.78		177
	N3–H···O4	2.85	175	
AC(II)	N7···H–N4	3.13		160
	N6–H···N3	2.95	175	
TT(I)	O4···H–N3	2.75		180
	N–H3···O2	2.88	176	
AA(III)	N7···H–N6	3.19		177
	N6–H···N7	2.95	176	

[a] Cf. Figure 2. [b] Cf. Figure 1. [c] Distance between heavy atoms. [d] WC, Watson–Crick; RWC, reversed Watson–Crick; H, Hoogsteen; RH, reversed Hoogsteen.

Upon H-bond formation the X–H bond is prolongated while the Y–Z bond is practically unchanged. The increase in X–H bond length is proportional to the stabilization energy of the complex; the changes evaluated with STO-3G are much smaller than those evaluated with MINI-1.[32,33] Optimizing the X–H bond length for the present complexes within MINI-1 (instead of keeping it rigid) would result in a small systematic increase in stabilization energy. Because of the proportionality of the increase in X–H bond length to the stabilization energy of the complex, the optimization of the X–H bond length with the MINI-1 basis set is not expected to change the order of stability of the complexes investigated. (Cf. ref 32, 33, and 34 and unpublished exploratory work in this laboratory.)

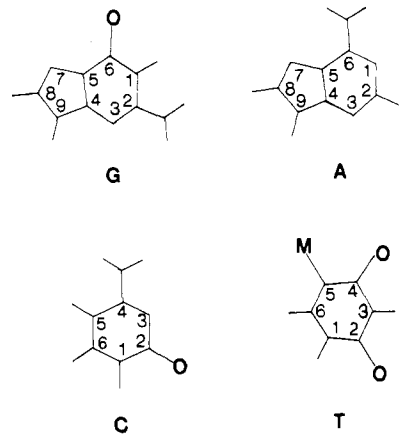

Figure 1. Optimal structures of guanine (G), adenine (A), cytosine (C), and thymine (T); the standard numbering is presented. M means methyl group.

Table II. SCF Interaction Energy (ΔE^{SCF}), Basis Set Superposition Error (BSSE), Dispersion Energy (E^D), Electrostatic Energy (E^{ES}), and Interaction Energy (ΔE) for the DNA Base Pairs (Energies in kJ/mol)

pair[a]	ΔE^{SCF}	BSSE	E^D	BSSE + E^D	E^{ES}	ΔE
GC(WC)[b]	−97.9	16.0	−28.6	−12.6	−85.5	−110.6
GG(1)	−97.1	19.4	−28.7	−9.3	−80.5	−106.4
CC	−66.9	9.5	−32.8	−23.3	−70.9	−90.3
GG(III)	−70.7	10.3	−26.5	−16.2	−66.5	−86.9
GC(II)	−59.6	9.6	−32.1	−22.5	−58.9	−82.1
AC(I)	−56.9	9.1	−31.7	−22.6	−61.2	−79.5
GA(I)	−59.6	13.0	−30.0	−17.0	−54.6	−76.6
GG(IV)	−48.1	9.1	−33.2	−24.1	−45.1	−72.2
GT(I)	−62.6	16.0	−24.5	−8.5	−50.8	−71.1
GC(I)	−59.6	10.7	−21.4	−10.7	−60.4	−70.3
AT(RWC)[b]	−54.4	12.2	−26.8	−14.6	−48.1	−69.0
GT(II)	−61.0	16.9	−23.9	−7.0	−52.8	−68.0
AT(RH)[b]	−52.2	10.7	−26.1	−15.4	−50.3	−67.6
AT(WC)[b]	−53.9	11.0	−24.0	−13.0	−45.1	−66.9
AT(H)[b]	−52.2	10.7	−26.1	−15.4	−47.6	−66.7
AA(I)	−42.0	10.3	−34.5	−24.2	−47.8	−66.2
GG(II)	−56.7	7.0	−16.5	−9.5	−51.8	−66.2
AA(II)	−42.7	7.8	−26.8	−19.0	−44.1	−61.7
GA(II)	−39.5	8.2	−30.2	−22.0	−39.4	−61.5
GA(III)	−45.6	10.0	−25.8	−14.8	−41.5	−61.4
GA(IV)	−38.1	9.1	−31.5	−22.4	−44.3	−60.5
TC(I)	−45.6	12.8	−26.8	−14.0	−41.4	−59.6
TC(II)	−45.2	13.6	−26.0	−12.4	−41.1	−57.6
TT(III)	−46.8	15.8	−22.1	−6.3	−37.2	−53.2
TT(II)	−45.7	15.0	−21.9	−6.9	−35.6	−52.6
AC(II)	−38.2	13.7	−28.0	−14.3	−36.4	−52.4
TT(I)	−44.6	15.1	−22.5	−7.4	−35.2	−52.0
AA(III)	−21.3	10.0	−27.5	−17.5	−25.7	−38.8

[a] Cf. Figure 2 and Table I. [b] WC, Watson–Crick; RWC, reversed Watson–Crick; H, Hoogsteen; RH, reversed Hoogsteen.

The geometries of all the complexes were optimized with respect to ΔE^{SCF}; the BSSE and E^D were evaluated at the potential energy minimum.

III. Results

Optimized[36] structures of G, A, C, and T are shown in Figure 1; in this figure the standard numbering of atoms is also given.

Optimized geometries of all the complexes studied are given in Figure 2. The geometrical characteristics describing unambiguously the different pairs are presented in Table I; the pairs are listed in order of decreasing stability (see Table II). With 28 complexes the optimization leads to an energy minimum having two or three (G···C(WC)) H-bonds. Only with the C···C pair with two C=O2···H—N4 H-bonds does the optimization lead to an open structure with just one H-bond. This is not surprising in light of the rather large negative charge located on N3; the electrostatic repulsion between the charges on the two N3's prevents the formation of the complex with two H-bonds. The

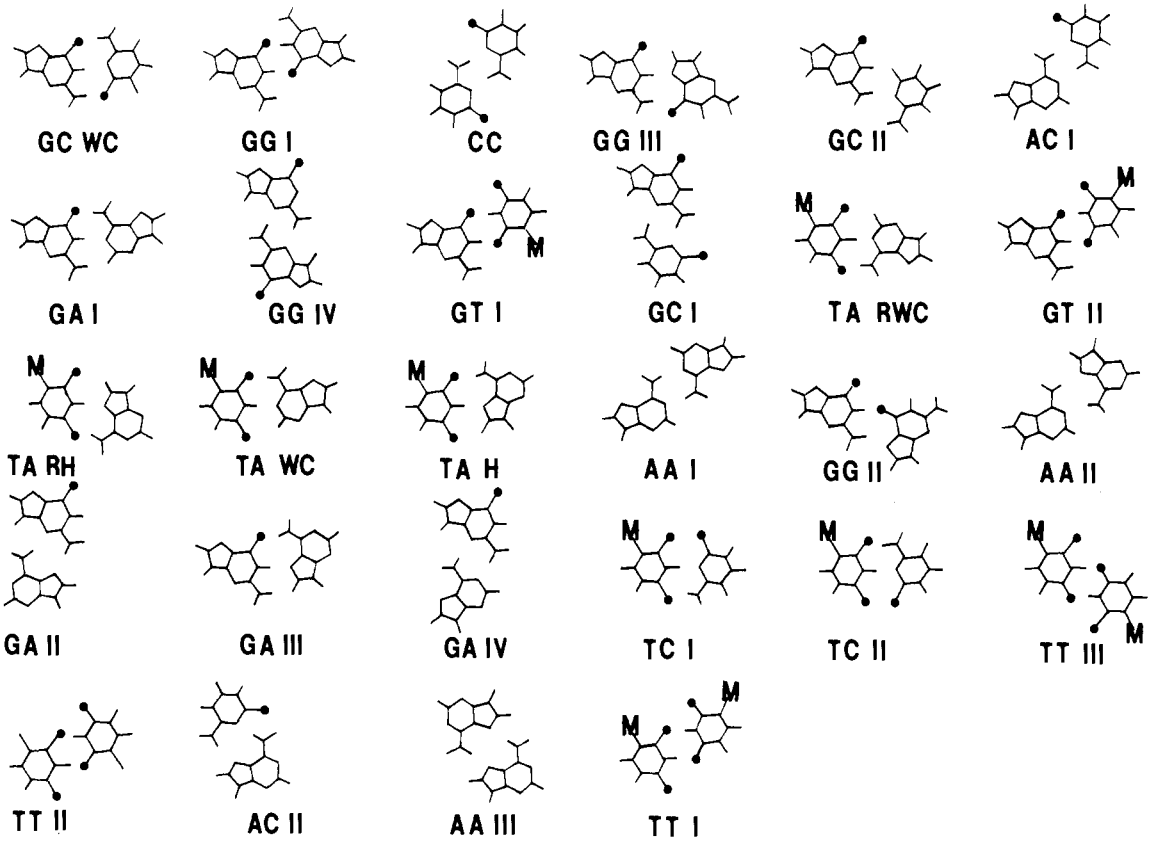

Figure 2. Optimal structures of different DNA pairs; the pairs are ordered according to their decreasing stability. Points represent oxygen atoms; M means methyl group.

energetic characteristics of the 28 pairs are collected in Table II.

IV. Discussion

Let us first discuss the energetic characteristics (Table II). The values of ΔE^{SCF} with different pairs differ greatly—the ΔE^{SCF} with the strongest complex (GC(WC)) is more than 4 times larger than that of the weakest complex (AA(III)). The values of BSSE differ less [from 19.4 (GG(I)) to 7.0 kJ/mol (GG(II))]. On the average, the ratio BSSE/ΔE^{SCF} for our large H-bonded complexes is not too different from the ratio found previously[32] with smaller H-bonded complexes.

The absolute values of dispersion energy vary less than the values of ΔE^{SCF}, the ratio between the largest (−34.5 kJ/mol, AA(I)) and the smallest (−16.5 kJ/mol, GG(II)) value of E^D being 2.1. This is not surprising, because the values of the total polarizabilities of G, C, T, and A differ by no more than 30%.[37] The dispersion energy is, for all the pairs, very important. With the strongest pairs it forms about 25% of the stabilization energy, with weaker pairs this portion increases. With some pairs the dispersion energy amounts to even more than 50% of the total stabilization energy! The importance of E^D is somewhat surprising; with smaller H-bonded complexes the dispersion energy forms no more than 20% of the stabilization energy. Analyzing the quality of the components of ΔE, we find that the accurate dispersion energy will cover an even larger portion of the real stabilization energy. The present values of the corrected ΔE^{SCF} (BSSE included) are not too far from the values of corrected ΔE^{SCF} evaluated with extended basis sets. Our dispersion energy matches satisfactorily the first term ($\sim R^{-6}$) of the accurate dispersion energy. It is known, however, that higher terms ($\sim R^{-8}$, R^{-10}...) are important, especially with larger subsystems. For example,[38] with $(H_2O)_2$ the higher terms of E^D account for 30% of E^D; this proportion increases to 40% with $CCl_3H\cdots H_2O$ (damping included).

For these reasons we must change, therefore, the traditional viewpoint of characterizing the DNA base pairs as H-bonded complexes where the dispersion energy is only a small fraction of the stabilization energy.

It was suggested several times in the literature that with H-bonded complexes it is possible to identify the stabilization energy with ΔE^{SCF}, simply because the BSSE and E^D compensate for each other. Examining the fifth column in Table II we find that this condition is not fulfilled. The values of BSSE + E^D differ greatly, from −6.3 (TT(III)) to −24.2 kJ/mol (AA(I)), i.e., by almost 400%. Furthermore, the addition of BSSE + E^D to ΔE^{SCF} changes the order of stability of the DNA base pairs. For example, GT(I) and GT(II) should be the 5th and the 6th in the order of the ΔE^{SCF}, but they are the 9th and the 12th in the order of the ΔE.

Taking with each pair only the energetically most favorable structure into account, the following order is obtained:

$$GC > GG \gg CC > AC > GA > GT > AT > AA > TC > TT$$

The heteropairs are more stable with G and C and with A and T, i.e., with complementary pairs. With the noncomplementary pairs the homopairs (at least one of them) are more stable than the heteropair. This is known from experiment.[39] The order proposed was further partially confirmed by direct experimental measurements of interaction enthalpy of base pairs in vacuo.[40] Due to investigation (by field mass spectrometry) of the pairs formed by G, C, T, and A, the following relative stabilities were found:[40]

$$GC(-88) \gg CC(-67) > AT(-54) > TT(-38)$$

(37) LeFèvre, C. G.; LeFèvre, R. J. W. *Rev. Pure Appl. Chem.* **1955**, *5*, 261.
(38) Hobza, P.; Mulder, F.; Sandorfy, C. *J. Am. Chem. Soc.* **1981**, *103*, 1360.
(39) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New-York, 1984; p 131.
(40) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149.

The numbers in parentheses indicate the values (in kJ/mol) of interaction enthalpies at 300 K (ΔH^{300}). Our values of ΔE are systematically larger; this is not surprising because of the relation between ΔH^{300} and ΔE given in eq 4. ΔZPE and $\Delta H^{0\rightarrow 300}$ stand

$$\Delta H^{300} = \Delta E + \Delta \mathrm{ZPE} + \Delta H^{0\rightarrow 300} \quad (4)$$

for the changes in zero-point energy and the temperature dependence of ΔH from 0 to 300 K. The former term is positive, i.e., decreases the value of ΔE; the last term is negative, but its absolute value is much smaller than that of the former term. To evaluate the ΔZPE it is necessary to know the complete set of intramolecular frequencies as well as all the intermolecular frequencies. Experimental or theoretical evaluation of the complete set of frequencies would be very difficult. For a broad series of H-bonded complexes we have found[41] that there exists a linear relationship between ΔZPE and ΔE; using this relationship yields the following values of ΔZPE (in kJ/mol) for GC, CC, AT, and TT: 16.8, 14.6, 12.3, and 10.6. Fair agreement between experimental and theoretical interaction enthalpies (−94, −76, −57, and −43 kJ/mol for GC, CC, AT, and TT, respectively) was obtained when these values were added to the ΔE (Table II). On the basis of this comparison we may state that our theoretical values of ΔE for the DNA base pairs are reasonable both as to their relative order and numerically.

There exist several papers in which single DNA base pairs were studied with ab initio SCF methods. The A···T(WC) pair was investigated mostly with the STO-3G basis set in ref 42–46, and the following stabilization energies (in kJ/mol) were found: 66.5,[42] 61.1,[43] 59.0,[20] 59.8,[44] and 58.6[45] (the BSSE was not included). The G···C(WC) pair was studied with the STO-3G basis set in ref 20, 45, and 46, and the following stabilization energies (in kJ/mol) resulted: 124.3,[20] 113,[45] and 130[46] (the BSSE was not included). These values, especially with C···C, are overestimated; on the other hand, taking the BSSE into account resulted[20] in underestimated values.

Before comparing our ΔE values with those obtained by empirical potentials let us discuss the use of the electrostatic approximation, i.e., the possibility to identify the interaction energy of a complex with just an electrostatic energy (E^{ES}). The method gives surprisingly good results providing (i) E^{ES} is the dominant attractive term (this requirement is fulfilled with H-bonded molecules but not with true vdW molecules) and (ii) E^{ES} is calculated accurately, i.e., the higher terms of the expansion are taken into account.[47] The problem may be overcome by working with atomic charges, effectively including higher point multipole moments. The charges calculated by using Mulliken populations are not suitable; the charges derived from calculated electrostatic potentials, are, on the other hand, useful for the evaluation of E^{ES}. We have used the atom-centered charges for G, C, A, and T derived[48] from electrostatic potentials calculated with a minimal basis set (all atom model). The resulting values of E^{ES} are presented in Table II; these values were evaluated at the energy minimum. Values of E^{ES} should be compared with those of ΔE^{SCF} + BSSE. We cannot, of course, expect an absolute agreement between these values for all the structures of all the DNA base pairs; this is clearly beyond the possibilities of this simple approach. The method gives, however, the correct order of relative stabilities for the different pairs (with each pair only the energetically most favorable structure was taken into account). The order, predicted by ΔE^{SCF} + BSSE and by E^{ES}, is the same as that given by ΔE (see above). This is, without doubt, a success for the electrostatic approximation. Let us stress once more, however, that the electrostatic approximation (as well as ΔE^{SCF} or the corrected ΔE^{SCF} approximation) is not able to give full agreement with ΔE when relative stabilities of different structures of different DNA base pairs are concerned.

Analyzing the results obtained with different empirical potentials[5-12] we have found that all of them are in disagreement with our results. This is not surprising; empirical potentials are simply not able to take properly into account all the different energy contributions as does the approach used here (ab initio SCF + E^{D}). In the second step we have analyzed the ability of those potentials to give the correct order of relative stabilities for different base pairs (only the energetically most stable structure of each pair was taken into account). From all the potentials[5-12] only that of Poltev and Shulyupina[5] agreed with our results. The success of this potential may probably be attributed to the parametrization of the electrostatic term of the potential; the atomic charges used[5] make it possible to reproduce the experimentally determined dipole moments of molecules.

Among the different base pairs those occurring in DNA (GC and AT) are the most important. There exist three structures of GC and four structures of AT; in DNA only Watson–Crick structures of both pairs were observed. For the GC pair the Watson–Crick structure with three H-bonds is by far the most stable. With the AT pair all four structures are energetically very similar (the difference being less than 3 kJ/mol). Because it is not ΔE but ΔG which is responsible for the complex formation, we have tried to explain the preference for some of these structures in terms of entropy. Translational and vibrational contributions to the entropy term are supposed to be identical with all four structures, so the only difference may come from the rotational contribution. We have found, however, that differences in the rotational contribution are very small, less than 0.2 kJ/mol. Hence, not only ΔE, but also ΔG, describing the formation of the different structures of the AT pair, is very similar.

From Table II it is evident that the GC(WC) pair, having three H-bonds, is the most stable. The GG(1) pair, however, possessing just two H-bonds, is almost as stable as GC(WC). Seemingly the presence of a higher number of H-bonds does not ensure the stability of a pair. The H-bonds in GG(WC) are not perfectly linear (see Figure 2 and Table I); on the other hand, those of GG(1) are much more linear. Is this linearity of the H-bonds the decisive factor for the stability of the pair? This does not appear to be the case. It is possible to find different pairs with the same type of "linear" H-bonds, differing considerably in stability. GG(1), GT(1), and TT(1) have two C=O···H—N H-bonds. In all these cases (see Table I) the linearity of the H-bonds is high; the stability of the pairs is, however, very different (−106.4, −71.1, and −52.0 kJ/mol). Also, the length of these bonds is very similar (see Table I). Conditions are the same for complexes having the N–H···N type of H-bond. From the analysis of energetic characteristics (see above) the crucial role of the electrostatic contribution becomes evident. Investigating the charges (derived from electrostatic potentials[48]) on atoms forming the C=O···H—N bonds in GG(1), GT(1), and TT(1) we find that they do not differ significantly. Evidently *all* the atoms of both subsystems (and not only those forming the H-bonds) should be taken into consideration to explain the large differences in the stability of the above mentioned pairs. The dominant stabilizing forces are electrostatic and forces of dispersion, but *all* the atoms of the subsystems must be considered.

Let us now investigate the structures of different pairs by using the geometrical characteristics of H-bonds presented in Table I. Within the 28 DNA base pairs (and any higher oligomers of the DNA bases) there exist only four types of H-bonds: C=O···H—N(A), C=O···H—N(B), N(A)—H···N(A), and N(A)—H···N(B), where N(A) and N(B) mean the $tr^2trtr\pi$ and te^2tete valence states of nitrogen. The fourth type of H-bond appeared most frequently (18×); the third (14×), the first (13×), and the second (12×) types follow. The H-bond length is, contrary to the stabilization energy, almost constant for different pairs. For example, the O···N distances (in Å) for the above mentioned pairs GG(1), GT(1), and TT(1) are 2.69, 2.84; 2.77, 2.80; and 2.75, 2.88; respectively. Let us recall that stabilization energies of these

(41) Szczesniak, M. M.; Hobza, P. *J. Phys. Chem.* **1983**, *87*, 2608.
(42) Sagarik, K. P.; Rode, B. M. *Inorg. Chim. Acta* **1983**, *78*, 177.
(43) Sarai, A.; Saito, M. *Int. J. Quantum Chem.* **1984**, *25*, 527.
(44) Hobza, P.; Sandorfy, C. *J. Biomol. Struct. Dyn.* **1985**, *2*, 1245.
(45) Del Bene, J. E. *THEOCHEM* **1985**, *124*, 201.
(46) Sarai, A.; Saito, M. *Int. J. Quantum Chem.* **1985**, *28*, 399.
(47) Buckingham, A. D.; Fowler, P. W. *Can. J. Chem.* **1985**, *63*, 2018.
(48) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.

pairs differ considerably (see above). The mean values and the standard deviations of the H-bond length (distances between the heavy atoms, in Å) for the four different classes of H-bonds are as follows: 2.81, 0.06 (n = 13); 2.90, 0.08 (n = 12); 2.91, 0.05 (n = 14); and 2.96, 0.11 (n = 18). The C=O···H—N($tr^2trtr\pi$) H-bond is the shortest, and the N($tr^2trtr\pi$)—H···N($te^2tetete$) one is the longest. The mean value and the standard deviation of H-bond length for all the H-bonds are (in Å) 2.90 and 0.10 (n = 57). The mean values and standard deviations of the X–H···Y (or X···H–Y) angles (in degrees) for the four different types of H-bonds are as follows: 176.8, 2.9 (n = 13); 169.5, 8.8 (n = 12); 177.6, 4.3 (n = 14); and 170.3, 7.3 (n = 18). The mean values and standard deviations of the X–H···Y (or X···H–Y) angles for all the H-bonds are (in degrees) 173.4 and 7.1 (n = 57).

Using the mean values of H-bond lengths and H-bond angles one may construct the geometries of higher oligomers of the DNA base pairs. The stabilization energies of these complexes may be deduced from the values of the electrostatic energy. This approach of estimation of geometries and stabilization energies of DNA base oligomers will be, in our opinion, more accurate than that based on empirical potentials.

V. Conclusions

The nonempirical ab initio SCF method employing the minimal MINI-1 basis set in combination with a London-type expression for the dispersion energy has been applied to all possible DNA base pairs. The following conclusions may be drawn.

(i) Fair agreement between experimental gas-phase enthalpies and theoretical enthalpies at 300 K has been obtained for GC, CC, AT, and TT pairs (experimental values are available only for the pairs mentioned).

(ii) The GC pair, having three H-bonds, is the most stable; the GG pair with just two H-bonds is, however, comparably stable. Furthermore, different pairs with the same types of H-bonds (which are almost linear) may differ considerably in stability. It follows, therefore, that neither the number of H-bonds nor their linearity is primarily responsible for the stability of the pairs. The stability of the pairs is impossible to explain by using only the atoms forming the H-bond; all the atoms of both subsystems must be taken into account.

(iii) The heteropairs are more stable than the homopairs with complementary base pairs; homopairs, on the other hand, are more stable than heteropairs with the noncomplementary base pairs.

(iv) Stabilization energies of different pairs differ greatly; the opposite is true about H-bond lengths.

(v) Formation of all four structures of the AT pair is equally probable; this is true even if the entropy is taken into account.

(vi) Empirical potentials in general fail to predict absolute as well as relative values of stabilization energies. On the other hand, the electrostatic approximation proved to be useful.

(vii) Dispersion energy is responsible for a very important part of total stabilization for all the pairs.

(viii) The basis set superposition error and dispersion energy do not compensate each other; the sum of these terms changes the order of relative stability as predicted by the SCF interaction energy.

(ix) Using the mean values of H-bond lengths and H-bond angles should make it possible to construct the structures of larger DNA base oligomers; their stabilization energies may be estimated on the basis of electrostatic energies.

Registry No. G, 73-40-5; C, 71-30-7; T, 65-71-4; A, 73-24-5.

Theoretical Investigations of the Anaerobic Reduction of Halogenated Alkanes by Cytochrome P450. 1. Structures, Inversion Barriers, and Heats of Formation of Halomethyl Radicals

Brian T. Luke,[†,‡] Gilda H. Loew,*[‡] and A. D. McLean[§]

Contribution from the Molecular Research Institute, Palo Alto, California 94304, SRI International, Menlo Park, California 94025, and IBM Research Laboratory, San Jose, California 95193. *Received June 17, 1985*

Abstract: For understanding the properties of the radicals produced in the anaerobic reduction of halogenated methanes by cytochrome P450, the geometries of all chlorofluoromethyl radicals are optimized with both MNDO and ab initio methods. The ab initio calculations employ unrestricted Hartree–Fock theory with 3-21G and 6-31G* basis sets. In addition, the structures of CH_3, CH_2F, and CH_2Cl are optimized with second-order Møller–Plesset perturbation theory (MP2) with 6-31G*, 6-31G**, and 6-31+G* basis sets. MP2 structures of CH_3 and CH_2F are also obtained with 6-311G** and 6-311+G** basis sets. The degree of nonplanarity, and the inversion barrier, increases in the order H < Cl < F. An examination of the results shows a correlation between the ab initio equilibrium geometry and the inversion barrier for these radicals. A plot of the inversion barrier vs. the degree of nonplanarity produces a single curve for all levels of calculation. This suggests that the equilibrium geometry of the radical determines the magnitude of the inversion barrier. MP2/6-31G*//HF/3-21G energies and HF/3-21G vibrational frequencies of all chlorofluoromethanes and methyl radicals are used with the available experimental data to calculate theoretical heats of formation for these compounds. These theoretical values are used to determine the strengths of C–H bonds in halogenated methanes, which correlate to the activity of the radicals produced in anaerobic reduction toward abstraction of lipid hydrogens. The theoretical heats of formation serve as a guide in deciding between conflicting experimental values. After a reliable set of experimental heats of formation is determined, these values are used to extend the list of HF/3-21G and HF/6-31G* atom equivalents presented by Ibrahim and Schleyer. In addition, MP2/6-31G*//HF/3-21G atom equivalents are determined.

Under anaerobic reducing conditions, various halogenated hydrocarbons are known to be reductively dehalogenated[1] by cytochrome P450. This activity is in contrast to that observed in the presence of O_2 and NADPH, where these heme proteins behave as mixed function oxidases, transfering an oxygen atom from O_2 to a variety of substrates. Initial evidence for this al-

[†] Molecular Research Institute.
[‡] SRI International.
[§] IBM Research Laboratory.

(1) Anders, M. W.; Pohl, L. R. In *Bioactivation of Foreign Compounds*; Anders, M. W., Ed.; Academic: New York, 1985.

0002-7863/87/1509-1307$01.50/0