H-Bonded and Stacked DNA Base Pairs: Cytosine Dimer. An Ab Initio Second-Order Møller-Plesset Study

Pavel Hobza,*,† Jiří Šponer,†,‡ and Martin Polášek†

Contribution from the J. Heyrovský Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, Dolejškova 3, 182 23 Prague 8, Czech Republic, and Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic

Received June 6, 1994[⊗]

Abstract: H-Bonded and stacked structures of the cytosine dimer were studied at the MP2/6-31G* ab initio level. In addition, the electrostatic energy was estimated separately using correlated distributed multipole analysis (CDMA) with atomic multipoles calculated at the MP2/6-31G* level. The H-bonded structure is more stable; the stabilization of the optimal stacked structure is, however, quite large and constitutes almost 50% of the H-bonded pair stabilization. The dominant part of the H-bonded stabilization originates in the electrostatic interaction. The dispersion energy is responsible for the stabilization in the stacked pair, while the mutual orientation of the stacked cytosines is governed by the electrostatic term. Due to the dipole-dipole interactions, antiparallel arrangement of the two stacked cytosines in the isolated cytosine dimer is strongly favored. No significant stabilization originating in the interactions between the polar exocyclic groups and the delocalized electrons of the aromatic rings was found. The CDMA calculations gave good insight into the nature of the stabilization of various complexes; however, some regions of the potential energy surface were not satisfactorily reproduced by this method.

Introduction

An important problem in molecular biophysics arises in attempts to provide a reliable description of interactions of DNA bases: the planar, H-bonded interactions and the vertical, stacking interactions. The relative importance of these two types of interactions for the stability and conformational variability of DNA is still not known, and it is possible to find both extremes, considering one type as dominant and the other as negligible. In addition, their physical nature and origin are also rather unclear. The H-bonded interactions are believed to be governed mainly by the electrostatic term with non-negligible contributions from the delocalization (induction) and dispersion terms. The stacking interactions are thought to be controlled mainly by the dispersion term and, to a smaller extent, by the electrostatic term. Due to the rather different characters of the dominant energy contributions for both types of interactions, no reliable theoretical study comparing the H-bonded and stacked DNA pairs has yet been performed. The H-bonded and stacked interactions require different theoretical treatment. While the interaction energies for the H-bonded systems are satisfactorily described at the Hartree-Fock (HF) level, the stacking interactions require a description at the beyond-HF level.

In order to obtain reasonable interaction energies at the beyond-HF level, the basis set used should contain diffuse polarization functions.[1] This is strictly necessary for stacking interactions because here the dispersion energy is crucial. (For H-bonded pairs, basis sets with standard polarization functions could be used.)

The H-bonded DNA base pairs have been studied at the ab initio level in the past.[2a-l] Recently, MP2/DZP//HF/6-31G* calculations were reported for three DNA base pairs.[2h] On the other hand (to our best knowledge) only a few papers have investigated the base stacking at the ab initio level.[2i-k]

Here we present the first comparison of the H-bonded and stacked DNA base pair based on reliable beyond-HF calculations with a basis set large enough to cover a significant part of the electron correlation.

Calculation

The interaction energy of the pair was determined as the sum of the SCF interaction energy and the correlation interaction energy; the latter was evaluated employing the second-order Møller-Plesset (MP2) theory,

$$\Delta E = \Delta E^{SCF} + \Delta E^{MP2} \quad (1)$$

All the calculations were performed with finite basis sets. The basis set superposition effects for ΔE^{SCF} and ΔE^{MP2} were therefore eliminated using the counterpoise procedure of Boys and Bernardi.[3] All the occupied and virtual orbitals of the "ghost" system were used. It was recently demonstrated that there is no overcorrection[4] in the original function counterpoise procedure.

The MP2 calculations were performed with the "frozen core" approximation; i.e., the 1s electrons of the non-hydrogen atoms were not considered in the calculation of the correlation energy.

The 6-31G* basis set with modified exponents of the polarization functions (0.25 instead of the standardly used value of about 0.8) was employed (abbreviation 6-31G*(0.25)). The use of more diffuse functions (taken from ref 5) results in more realistic values of the correlation interaction energy. This is especially true for stacked

† J. Heyrovský Institute of Physical Chemistry.

‡ Institute of Biophysics.

⊗ Abstract published in *Advance ACS Abstracts,* December 15, 1994.

(1) Hobza, P.; Mehlhorn, A.; Čársky, P.; Zahradník, R. *J. Mol. Struct.: THEOCHEM* **1986**, *138*, 387.

(2) (a) Clementi, E.; Mehl, J.; von Niessen, W. *J. Chem. Phys.* **1971**, *54*, 508. (b) Del Bene, J. *J. Mol. Struct.: THEOCHEM* **1985**, *124*, 201. (c) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc.* **1987**, *109*, 1302. (d) Aida, M. *J. Comput. Chem.* **1988**, *9*, 362. (e) Dive, G.; Dehareng, D.; Ghuysen, J. M. *Theor. Chim. Acta.* **1993**, *85*, 409. (f) Hrouda, V.; Florián, J.; Hobza, P. *J. Phys. Chem.* **1993**, *97*, 1542. (g) Colson, A.-O.; Besler, B.; Sevilla, M. D. *J. Phys. Chem.* **1993**, *97*, 13852. (h) Gould, I. R.; Kollman, P. A. *J. Am. Chem. Soc.* **1994**, *116*, 2493. (i) Nagata, C.; Aida, M. *J. Mol. Struct.: THEOCHEM* **1988**, *179*, 451. (j) Aida, M. *J. Theor. Biol.* **1988**, *130*, 327. (k) Aida, M.; Nagata, C. *Chem. Phys. Lett.* **1982**, *86*, 44. (l) Florián, J.; Leszczynski, J. Submitted to *Chem. Phys. Lett.*

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

(4) (a) Cybulski, M. S.; Chalasinski, G. *Chem. Phys. Lett.* **1992**, *197*, 591. (b) Gutowski, M.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H.; van Duijneveldt, F. B. *J. Chem. Phys.* **1993**, *98*, 4728. (c) Gutowski, M.; Chalasinski, G. *J. Chem. Phys.* **1993**, *98*, 5540.

0002-7863/95/1517-0792$09.00/0

structures.[1] We are aware of the limitations of the theoretical procedure used, but the size of the clusters studied prevents going beyond the MP2 level and the use of larger basis sets. Our calculations on smaller molecular complexes indicated[6a-c] that the MP2/6-31G*(0.25) level gives reliable results.

The gradient optimization techniques do not adopt the function counterpoise method, and the respective interaction energies and equilibrium distances are spoiled by the basis set superposition effects. This is even more serious because we are going to compare the H-bonded and stacked structures. The basis set superposition error (BSSE) will be much larger for the stacked structures than for the planar ones.[1] Therefore, stacked as well as H-bonded structures were optimized by a point-by-point procedure.

The cytosine geometry was kept rigid during the optimization; the planar MP2/DZ(2d) optimized geometry[7] was used. Only planar H-bonded pairs and stacked complexes with two coplanar bases were considered.

The GAUSSIAN 92 set of programs[8] was used, utilizing the direct mode for both the SCF and MP2 parts.

In addition to the total interaction energy, the electrostatic energy was evaluated by using point charges, dipoles, quadrupoles, octopoles, and hexadecopoles localized at the atom sites; the calculations were performed with the ORIENT code.[9] Its present version includes all the interactions up to r^{-5}. The point multipoles for the isolated cytosine were calculated with the distributed multipole analysis (DMA) implemented in the CADPAC 5.0 set of programs[10] using the 6-31G*(0.25) basis set; both the SCF and MP2 multipoles were considered.

Further, the electrostatic energy was also calculated with the point charges evaluated using the molecular electrostatic potential (MEP), generated with the 6-31G* basis set.[11] The resulting charges were multiplied by 0.85.[11] We also tested the frequently used MEP charges, obtained with the STO-3G basis set.[12]

In order to mimic the dispersion and short range repulsion contributions, the electrostatic terms were combined with the 6-9 Lifson-Hagler (6-9LH) Lennard-Jones empirical potential,[13] which provides a satisfactory description of base stacking in DNA.[14a-d]

Results and Discussion

A. H-Bonded Structures. The H-bonded pair was considered symmetrical (i.e., both the H-bonds were equal) and perfectly planar. Although somewhat nonsymmetrical geometry was reported in our previous study[2c] for the C···C pair, gradient optimization with the minimal basis set (MINI-1, not shown) fully justified the symmetrical structure as more stable than the nonsymmetrical one. It was also demonstrated (e.g., see refs 7, 15a-e) that the DNA base amino groups are intrinsically nonplanar, which could result in somewhat nonplanar geometries of the pairs.[2e,15a] It is to be noticed, however, that a formation of the H-bonded structure decreases or even eliminates the amino group nonplanarity due to the fact that H-bond energy is considerably larger than the energy associated with pyramidalization of the amino groups.[15b] We therefore performed gradient optimization of the H-bonded pair at the HF/6-31G-(NH_2*) level,[16] but the optimized geometry was almost perfectly planar (not shown). This is not surprising, because among the DNA bases the amino group hydrogen atoms of cytosine that participate in the C···C base pairing exhibit the smallest nonplanarity.[7,15a,b]

Table 1. Interaction Energies (ΔE), SCF Interaction Energies (ΔE^{SCF}), MP2 Correlation Interaction Energies, (ΔE^{MP2}) and Electrostatic Energies (E^{EL}) for the H-Bonded Structure[a] of the Cytosine Dimer (Energies in kcal/mol)

R_{HB} (Å)[b]	β (deg)[c]	ΔE^{SCF}	E^{EL} [d]	ΔE^{MP2}	ΔE
2.80	123.5	−9.31		−3.91	−13.21
2.90	123.5	−11.59		−3.46	−15.04
3.00	123.5	−12.83		−3.03	−15.85
3.03	123.5	−13.06		−2.89	−15.95
3.07	123.5	−13.22	−14.39	−2.76	−15.98
3.10	123.5	−13.32		−2.63	−15.95
3.07	126.0	−12.26		−2.80	−15.06
3.07	121.0	−13.91		−2.75	−16.66
3.07	119.0	−14.27		−2.76	−17.03
3.07	117.0	−14.46		−2.80	−17.26
3.07	115.0	−14.50		−2.85	−17.35
3.07	113.0	−14.39		−2.93	−17.32
3.10	115.0	−14.56		−2.71	−17.29
3.03	115.0	−14.35	−14.83	−3.00	−17.35

[a] Cf. Figure 1a. [b] N4 (cyt1)···N3 (cyt2) distance. [c] N3 (cyt1)···N4 (cyt2)−C4 (cyt2) angle. [d] MP2 distributed multipoles.

The stabilization energy and its components are summarized in Table 1. Because we considered a planar and symmetrical structure, there were only 2 degrees of freedom: the length of the hydrogen bond (N3(Cyt1)···N4(Cyt2)) and the N3(Cyt1)···N4-(Cyt2)−C4(Cyt2) angle, designated β throughout this article (see also Figure 1a). We first optimized the N3···N4 bond length for a fixed β angle of 123.5°. Then the β angle was optimized with the fixed length of the hydrogen bonds. The latter procedure did not affect the hydrogen bond length. The optimal stabilization energy of 17.35 kcal/mol corresponds to the H-bond distance of 3.05 Å and β angle of 115°. The electrostatic energy evaluated with the MP2 multipoles is similar to the HF interaction energy and thus favors the electrostatic nature of the H-bonding in the cytosine pair. In our previous paper,[2c] the electrostatic energy was also slightly larger than the MINI-1 SCF stabilization energy.

The calculated H-bonds are longer (by about 0.2 Å) than these obtained previously[2c] at the SCF/MINI-1 level. This confirms the known fact that the SCF/MINI-1 intermolecular distances are underestimated. Comparing the energy components evaluated with the MINI-1[2c] and 6-31G*(0.25) basis sets in the respective energy minima we find that the SCF interaction energies are almost identical while the 6-31G*(0.25) MP2

(5) Kroon-Batenburg, L. M. J.; van Duijneveldt, F. B. *J. Mol. Struct.* **1985**, *121*, 185.
(6) (a) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Phys. Chem.* **1993**, *97*, 3937. (b) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Am. Chem. Soc.* **1994**, *116*, 3500. (c) Hobza, P.; Selzle, H. L.; Schlag, E. W. *Chem. Rev.* **1994**, *94*, 1767.
(7) Šponer, J.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 3161.
(8) Frisch, M. J.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomberts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzales, J. S.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A. *GAUSSIAN 92*; Gaussian, Inc.: Pittsburgh, PA, 1992.
(9) Stone, A. J. *ORIENT 2.2*; Cambridge, 1992.
(10) Amos, R. D.; Alberts, I. L.; Andrews, J. S.; Colwell, S. M.; Handy, N. C.; Jayatilaka, D.; Knowles, P. J.; Kobayashi, R.; Koga, N.; Laidig, K. E.; Maslen, P. E.; Murray, C. W.; Rice, J. E.; Sanz, J.; Simandiras, E. D.; Stone, A. J.; Su, M.-D. *CADPAC 5.0*; Cambridge, 1992.
(11) Cieplak, P.; Bash, P.; Singh, U. C.; Kollman, P. *J. Am. Chem. Soc.* **1987**, *109*, 6283.
(12) Weiner, S. J.; Kollman, P. A.; Nguyen, T. Z.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.
(13) (a) Lifson, S.; Hagler, A. T.; Dauber, P. *J. Am. Chem. Soc.* **1979**, *101*, 5111. (b) Hagler, A. T.; Lifson, S.; Dauber, P. *J. Am. Chem. Soc.* **1979**, *101*, 5122.
(14) (a) Šponer, J.; Kypr, J. *J. Mol. Biol.* **1991**, *217*, 201. (b) Šponer, J.; Kypr, J. In *Theoretical Biochemistry & Molecular Biophysics*; Beveridge, D. L., Lavery, R., Eds.; Adenine Press: Guilderland, NY, 1990; p 271. (c) Šponer, J.; Kypr, J. *J. Biomol. Struct. Dyn.* **1993**, *11*, 27. (d) Šponer, J.; Kypr, J. *J. Biomol. Struct. Dyn.* **1993**, *11*, 277.
(15) (a) Šponer, J.; Hobza, P. *J. Mol. Struct.: THEOCHEM* **1994**, *304*, 35. (b) Šponer, J.; Hobza, P. *J. Am. Chem. Soc.* **1994**, *116*, 709. (c) Gould, I. R.; Vincent, M. A.; Hillier, I. H. *Spectrochim. Acta* **1993**, *49C*, 1727. (d) Leszczynski, J. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1992**, *19*, 43. (e) Ha, T.-K.; Gunthard, H. H. *J. Mol. Struct.: THEOCHEM* **1992**, *276*, 209.
(16) The HF/6-31G(NH_2*) level (standard 6-31G basis set augmented by d-polarization functions on the amino group nitrogen atoms) provides a satisfactory description of the amino group nonplanarity.[15a] Here a smaller exponent of 0.65 was used for the d-polarization functions, which leads to DNA base amino group nonplanarity (both energy and geometry) very close to that obtained at the MP2/6-31G* level.[7,15a]

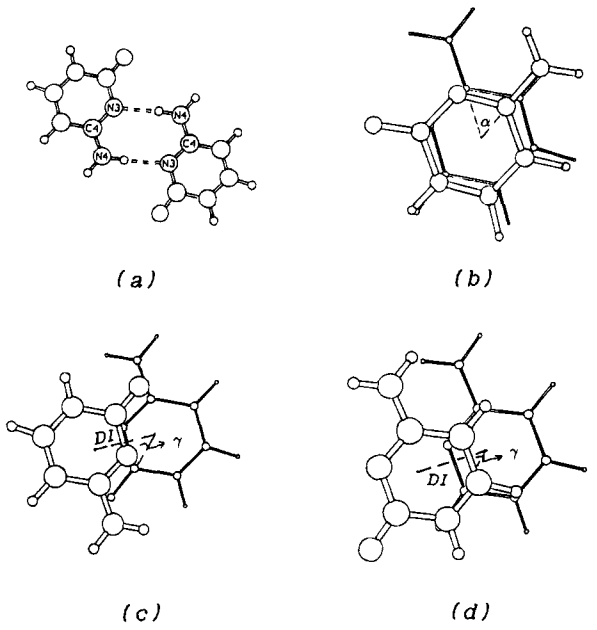

Figure 1. H-bonded (a) and stacked (b–d) structures of the cytosine dimer. The twist in the stacked complex is introduced by rotation of the upper cytosine around its center of mass by the twist angle α, as indicated in Figure 1b (the present geometry is for the α angle of 60°). The displaced stacked structures (c, α = 180°; d, α = 0°) are determined by the displacement (DI) of the center of mass of the upper base in the direction determined by the displacement angle γ. (The DI, γ reference frame is fixed with the lower base.)

correlation interaction energy is considerably smaller (by about 5 kcal/mol) than the London dispersion energy. This is because the MP2 correlation interaction energy consists of intersystem and intrasystem parts. The former contribution corresponds roughly to the attractive London dispersion energy, while the change of intrasystem correlation energy is frequently repulsive. Among all the 29 possible DNA pairs it was the C···C pair which exhibited[17a] the largest (repulsive) change of the intrasystem correlation energy. In other words, the frequently used combination of the ab initio SCF interaction energy and dispersion energy may result in inaccurate values of the total stabilization energy because the dispersion energy might be (in absolute value) considerably larger than the correlation interaction energy. Comparing clusters of different types, and also different structures of one cluster, the total interaction energy should be constructed as the sum of the SCF and correlation interaction energies.

B. Stacked Complex. 1. Effect of Twist and Vertical Separation of Bases. For the stacked pair we do not have any previous (reliable) evidence on the most stable structure. First, we investigated the dependence of the stacking energy on twist angle α (the upper base was rotated around its center of mass, see Figure 1b) for the expected optimum vertical separation of bases, 3.4 Å,[14d] and maximum geometrical overlap of the bases.[18] The respective stabilization energies and their components are given in Table 2. The antiparallel structure (α = 180°) is the most stable while the parallel structure (α = 0°) is the least stable. This confirms the important role of the electrostatic energy; the dipole–dipole contribution is optimum for the antiparallel structure. Investigation of the relative values

(17) (a) Szczesniak, M. M.; Scheiner, S.; Hobza, P. *J. Mol. Struct.: THEOCHEM* **1988**, *179*, 177. (b) Hobza, P.; Havlas, Z. *Int. J. Quantum Chem.* **1989**, *36*, 287. (c) Friedman, R. A.; Honig, B. *Biopolymers* **1992**, *32*, 145.

(18) Centers of mass of the two bases were stacked one directly above the other, i.e., with minimum distance between them.

Table 2. Interaction Energies (ΔE), SCF Interaction Energies (ΔE^{SCF}), MP2 Correlation Interaction Energies (ΔE^{MP2}), and Electrostatic Energies (E^{EL}) for the Stacked Structures[a] of the Cytosine Dimer (Energies in kcal/mol, Displacement = 0 Å)

R (Å)[b]	α (deg)[c]	ΔE^{SCF}	E^{EL} [d]	ΔE^{MP2}	ΔE
3.4	0	13.70	5.45 (8.45)	−11.44	2.26
3.4	30	9.77	3.02 (5.91)	−10.62	−0.85
3.4	60	5.68	−0.36 (1.89)	−9.04	−3.35
3.4	90	2.31	−2.49 (−1.22)	−7.94	−5.63
3.4	120	0.37	−3.06 (−2.86)	−7.54	−7.20
3.4	150	−0.11	−3.29 (−3.44)	−7.26	−7.36
3.4	180	−1.12	−3.83 (−4.12)	−6.96	−8.08
3.0	180	6.90	−5.01	−13.41	−6.50
3.1	180	3.77	−4.67	−11.42	−7.63
3.2	180	1.54	−4.36	−9.71	−8.17
3.3	180	−0.03	−4.08	−8.23	−8.26
3.4	180	−1.12	−3.83	−6.96	−8.08
3.8	180	−2.87	−3.04	−3.50	−6.37
4.0	180	−3.00	−2.73	−2.38	−5.38

[a] Cf. Figure 1b. [b] Vertical separation of bases. [c] Twist angle α. [d] MP2 distributed multipoles; values in parentheses were obtained with HF distributed multipoles.

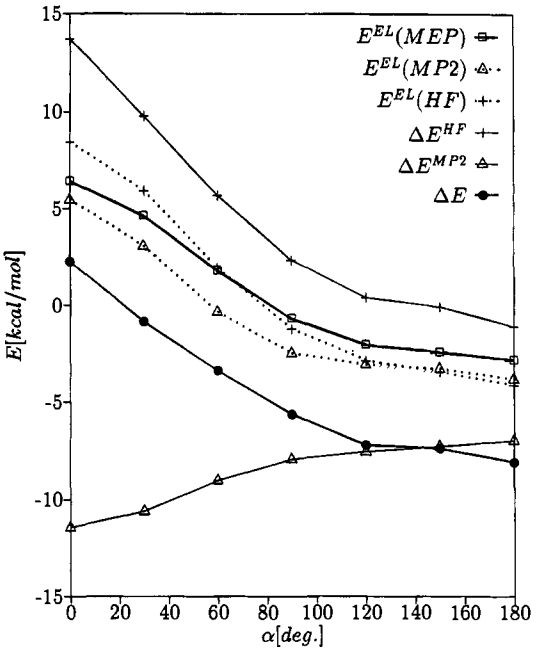

Figure 2. Twist angle dependence (displacement 0 Å, vertical separation 3.4 Å) of various energies for the stacked cytosine dimer: total ab initio interaction energy, its HF and MP2 components, DMA electrostatic energies (HF, MP2) and MEP electrostatic energy with scaled 6-31G* charges.

of the energy components reveals that it is the HF interaction energy which is angle dependent; the MP2 interaction energy is (as expected) considerably less angle dependent (cf. Figure 2). It should be noted that practically all the twist dependence of MP2 interaction energy originates in the change of intrasystem correlation energy (i.e., mainly the dipole–dipole electrostatic energy, see below). The dispersion energy and 6–9LH Lennard-Jones energy vary less than 0.3 kcal/mol with the twist angle. As a result the ΔE^{HF} and ΔE^{MP2} curves run in opposite phases; i.e., the largest repulsion of the former term is correlated with the largest attraction of the latter term and vice versa. This effect has nothing in common with the intermolecular dispersion energy, but it is mainly due to reduction of the cytosine dipole moment when passing from the HF (7.32 D) to the MP2 (6.27 D) level. In the system studied here, the correlation contribution reduces the dipole–dipole interaction obtained at the HF level; i.e., the change in the intrasystem correlation energy is attractive

for structures that have repulsive dipole−dipole interaction energy and repulsive for structures for which this energy is attractive.

As explained above, the twist dependence of the base stacking energy of cytosine dimer is controlled by the electrostatic energy. It should be emphasized, however, that in the case of the helical twist between base pairs in the DNA the situation may be quite different. There are three reasons for it: (i) The base pair is, in contrast to the isolated cytosine, anisotropic (i.e., it has an oblong, rod-like shape). It leads to a strong dependence of the dispersion (of Lennard-Jones) contribution on the twist angle. Geometrical overlap of two stacked cytosines with twist angles 0° and 90° is the same, whereas base pair overlap of a base pair dimer with twist angles 0° and 90° is dramatically different. (ii) Increasing the number of atoms in interacting systems reduces the role of electrostatic interactions and increases the role of dispersion contribution.[17b] (iii) It is well-known that solvent effects reduce (via screening) the role of electrostatic contribution[17c] while their effect on the dispersion energy is considerably smaller. Therefore, the present ab initio calculations for gas-phase cytosine dimer do not contradict a recent investigation by Friedman and Honig,[17c] who found that due to the environmental effects the Lennard-Jones interactions, and not the electrostatic forces, determine the helical twist dependence of base stacking in a DNA double helix.

The electrostatic energy was evaluated with the HF as well as MP2 multipoles; these energies differ especially for the range of the twist angle α 0−60° (cf. Figure 2). Here the electrostatic energy evaluated with the MP2 multipoles reaches only about 65% of the value of the electrostatic energy calculated with the HF multipoles.

The HF DMA electrostatic energy closely follows the HF interaction energy while the MP2 DMA electrostatic energy nicely follows the total interaction energy. This is an important conclusion for the construction of the intermolecular potentials: the use of electrostatic energy utilizing correlated multipoles gives much better angular dependence of the total (empirical) energy.

The DMA and correlated DMA (CDMA) electrostatic energies include all the interactions up to r^{-5}, i.e., up to quadrupole−quadrupole, dipole−octopole, and charge−hexadecopole terms. Investigation of the convergency of the expansion indicated that including all the interactions up to r^{-4} resulted in satisfactory values. However, using only those up to r^{-3} led to larger differences compared to the electrostatic energy evaluated up to r^{-5}. For small systems it was shown[19] to be adequate to evaluate the electrostatic energy up to r^{-5}. For the present (larger) systems this should be even more adequate since higher terms in the electrostatic energy expansion decay much faster.

The electrostatic energy calculated with the modified 6-31G* MEP charges has angle dependence similar to that evaluated with the MP2 point multipoles (Figure 2). The STO-3G MEP charges also provide satisfactory dependence (not shown).

In the second step, the dependence of the interaction energy on the vertical separation of the bases was studied for the optimal antiparallel structure. The optimal distance (cf. Table 2) differs only slightly from the assumed one.[14b,d] At the energy minimum the HF interaction energy is negligible and all the stabilization comes from the electron correlation. At larger distances the HF interaction energy becomes equally important or even more important than the MP2 correlation interaction energy. This is due to the long-range dipole−dipole electrostatic contribution, included in the HF interaction energy.

(19) Price, S. L.; Stone, A. J. *J. Chem. Phys.* **1987**, *86*, 2859.

Table 3. Interaction Energies (ΔE), SCF Interaction Energies (ΔE^{SCF}), MP2 Correlation Interaction Energies (ΔE^{MP2}), and Electrostatic Energies (E^{EL}) for the Parallel-Displaced and Antiparallel-Displaced Structures. γ and DI Are the Displacement Angle and Distance (Energies in kcal/mol)

structure[a]	γ (deg)	DI (Å)[b]	ΔE^{SCF}	E^{EL} [c]	ΔE^{MP2}	ΔE
parallel-displaced	0	0.0	16.79	5.92	−13.35	3.44
		1.0	12.26	4.48	−11.75	0.51
		2.0	8.97	3.80	−9.29	−0.32
	45	1.0	11.77	4.09	−11.52	0.25
		2.0	7.30	2.71	−8.52	−1.23
	90	1.0	11.94	4.40	−11.60	0.34
		2.0	7.69	3.88	−8.73	−1.04
	135	1.0	12.31	4.62	−11.79	0.52
		2.0	8.64	3.26	−9.40	−0.77
antiparallel-displaced	0	0.0	−0.03	−4.08	−8.23	−8.26
		1.0	−0.01	−1.96	−7.00	−7.01
		2.0*	−1.25	−0.96	−5.02	−6.27
	45	1.0	−0.61	−1.71	−6.64	−7.26
		2.0	−0.68	−0.78	−4.50	−5.20
	90	1.0	−0.22	−2.37	−7.15	−7.37
		2.0	−0.96	−1.77	−5.09	−6.05
	135	1.0	1.62	−2.84	−8.25	−6.63
		2.0$	−0.62	−3.66	−6.76	−7.38
	180	1.0	2.94	−3.09	−8.88	−5.94
		2.0	2.22	−1.48	−7.69	−5.47
	225	1.0	0.53	−5.56	−8.08	−7.55
		2.0	0.13	−3.17	−7.29	−7.16
	270	1.0	−1.04	−6.30	−7.34	−8.39
		2.0&	−1.08	−4.60	−6.35	−7.43
	315	1.0	−0.28	−4.24	−7.36	−7.63
		2.0	−1.97	−3.79	−5.39	−7.36

[a] Cf. Figure 1c,d; distance between molecular planes equal to 3.3 Å. [b] Structures designated by *, $, and & are shown in Figure 5, parts a, b, and c, respectively. [c] MP2 point multipoles.

2. Effect of Displacement. Investigation of the benzene dimer[6a,b] revealed that a stabilization results upon displacing the upper molecule in the stacked C_{6v} structure; this was due to the quadrupole−quadrupole electrostatic interaction. We therefore also investigated the displaced stacked structures for the present complex. 16 different antiparallel structures (α = 180°) with displacement angles γ (see Figure 1c,d for a definition) of 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315° and displacement distances (DI) of 1.0 and 2.0 Å were considered, together with eight different parallel structures (α = 0°, γ angles[20] 0°, 45°, 90°, and 135°, and displacement distances 1.0 and 2.0 Å). The results are summarized in Table 3.

Displacing the upper molecule in the parallel structure reduces the overall repulsion. The decrease is larger for the larger displacement. This effect is reproduced by the electrostatic energy. Displacing the upper molecule in the antiparallel structure is connected with modest energy destabilization. The antiparallel structures were always much more stable than the parallel ones. The destabilization increases with increasing displacement from 1.0 to 2.0 Å; the only exception consists of the structure with a displacement angle of 135°. This structure exhibits direct interaction of the amino groups with the aromatic ring (see below).

Figure 3 compares the MP2/6-31G*(0.25) interaction energies for the stacked structure of the cytosine dimer with the energies obtained using various electrostatic models, combined with the 6−9LH van der Waals terms. These terms mimic the short-range repulsion and dispersion energy contributions. Neither of the electrostatic models provides satisfactory results. The CDMA terms predict a deep minimum near γ = 270°, which is much less pronounced at the ab initio calculation level. (The

(20) Due to symmetry, for parallel arrangements the structures with γ within 0−180° are equivalent to those with γ within 180−360°.

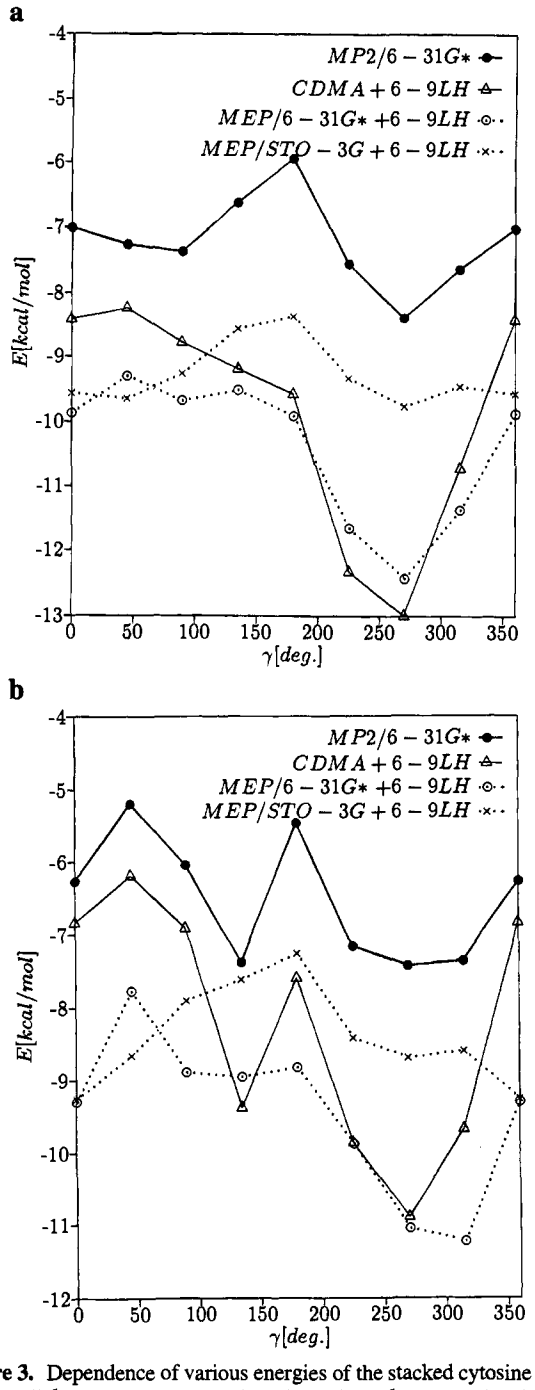

Figure 3. Dependence of various energies of the stacked cytosine dimer (antiparallel structure) on the direction of displacement (γ) for two values of the displacement DI (a, 1.0 Å; b, 2.0 Å): the ab initio MP2/6-31G*(0.25) interaction energies, CDMA electrostatic energy + the 6−9LH van der Waals terms, MEP electrostatic energy with scaled 6-31G* charges + 6−9LH potential, and MEP electrostatic energy with STO-3G charges + the 6−9LH potential.

shapes of the empirical potential curves in Figure 3 are almost exclusively due to the electrostatic term, while the van der Waals terms determine only the position of a given curve along the vertical (energy) axis.) The electrostatic energy evaluated with the scaled 6-31G* charges fails in the same region, and in a similar fashion, as the CDMA calculations. The electrostatic energy evaluated with the STO-3G (AMBER) charges is surprisingly successful for a displacement of 1.0 Å (Figure 3a), but it fails for a displacement of 2.0 Å and displacement angles of 0−180° (Figure 3b).

In order to analyze the differences between the ab initio and empirical potential data, the calculations (for displacement = 1.0 Å) were repeated with increased vertical separation of the bases of 3.8 Å. The increased distance between the bases results in elimination of all the short-range effects. The comparison of the results for 3.3 and 3.8 Å reveals two interesting points. First, the maximum on the ab initio curve near the displacement angle $\gamma = 180°$ is due to short-range repulsion because it vanishes completely when passing to the higher vertical separation. The effect described should therefore be covered by the van der Waals part of the empirical potential. However, the 6−9LH Lennard-Jones potential (as well as any related van der Waals potential[14d]) varies only insignificantly with the γ angle at any vertical separation of the bases and does not reproduce the mentioned effect at all. Second, the deep electrostatic energy minimum near displacement angle $\gamma = 270°$ is reduced when passing from 3.3 Å to 3.8 Å but it is still considerably more pronounced than this on the respective ab initio curve. The most likely explanation could be sought in the penetration term (not included in the electrostatic energy) and/or in the fact that, for this particular dimer geometry, the convergency of the multipole expansion may still not be good enough. This paragraph could be concluded by warning that the accurate reconstruction of the ab initio potential energy surface of DNA bases using empirical potentials, even containing a sophisticated electrostatic term, represents a difficult task.

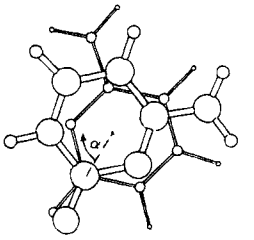

Figure 4. Face-to-face structure of the cytosine dimer with twist angle $\alpha = 0°$. The twist is introduced by rotation of the upper cytosine around its center of mass by the angle α, as indicated.

Despite the above-mentioned discrepancies, the electrostatic energy qualitatively determines the interaction energy changes for both twist and displacement. We have therefore calculated it for several hundreds of other conformations including all the twists of the bases and displacements up to 3.0 Å. Again, the antiparallel structures were always much more stable than the parallel ones. For the parallel structures, the electrostatic energy never decreased below 2.0 kcal/mol. For displacements larger than 1.0−1.5 Å, the absolute value of the electrostatic energy gradually decreased. The optimum electrostatic energy (−6.4 kcal/mol) was found for the twist angle 180°, displacement 1.05 Å, and displacement angle 280°. This structure agrees with the optimum structure obtained by the ab initio procedure (Table 3). The fact that optimum electrostatic energy results for a somewhat displaced structure favors the role of quadrupole−quadrupole interaction (cf. the benzene dimer[6a,b]).

The dispersion energy depends almost exclusively on the overlap of the bases (the displacement should be minimal). Therefore, optimum stacking of two cytosines requires antiparallel or nearly antiparallel geometry of the two bases with their significant mutual overlap.

3. Face-to-Face Orientation. DNA bases can be stacked in two ways, "face-to-back" or "face-to-face". The face-to-back orientation means that the two cytosines can be superimposed by a simple translation and reduction of the twist angle to 0°. In order to get face-to-face orientation, one of the cytosines must be inverted as shown in Figure 4. Both orientations of bases can be found in the DNA double helix; the face-to-back stacking occurs within one strand while the

Table 4. Interaction Energies (ΔE), SCF Interaction Energies (ΔE^{SCF}), MP2 Correlation Interaction Energies (ΔE^{MP2}), and Electrostatic Energies (E^{EL}) for the Stacked Face-to-Face[a] Undisplaced Structures of the Cytosine Dimer (Energies in kcal/mol)

α (deg)[b]	ΔE^{SCF}	E^{EL} [c]	ΔE^{MP2}	ΔE
0	8.46	2.03	−11.03	−2.58
60	11.97	3.16	−12.02	−0.05
120	8.40	0.54	−11.10	−2.70
180	3.23	−3.40	−9.67	−6.44
240	0.70	−4.36	−7.98	−7.28
300	0.75	−4.36	−8.18	−7.43
285[d]	0.94	−4.42	−7.99	−7.05
45[e]	11.15	3.57	−11.98	−0.83

[a] Cf. Figure 4; vertical separation of bases equal to 3.3 Å. [b] Twist angle. [c] MP2 point multipoles. [d] Minimum of the electrostatic energy. [e] Maximum of the electrostatic energy.

face-to-face orientation is characteristic for the cross-strand interaction between neighboring base pairs.

All the above calculations were made for the face-to-back orientation. For the face-to-face orientation, the ab initio calculations were limited only to the dependence of stacking energy on the twist angle (Table 4). Minimum and maximum CDMA electrostatic energies were found for twist angles 285° and 45°, respectively. These points did not accurately correspond to the minimum (maximum) values of the ab initio energies. The electrostatic energy thus reproduced the dependence of the total interaction energy on the twist angle somewhat less satisfactorily than for the face-to-back arrangement (Figure 2).

The displaced structures were investigated using only the CDMA approach (not shown). The potential energy surface was similar to that for the face-to-back arrangement, and thus more extensive ab initio analysis was not performed.

C. Comparison of H-Bonded and Stacked Structures. The H-bonded structure of the cytosine pair is more stable than the optimal stacked structure. Whereas the stabilization of the H-bonded structure originates predominately in the HF interaction energy, the MP2 correlation interaction energy is almost exclusively responsible for the stabilization of the stacked pair. Displacing the subsystems from the optimum H-bond structure leads to rather dramatic energy destabilization whereas sliding the stacked pair leads to moderate energy changes in the whole conformational space. The H-bonded interactions are thus specific whereas the stacking interactions are more nonspecific and flexible.[14a-c]

D. Nature of Stacked Structure Stabilization. Two contradictory opinions on the optimum stacking exist. According to the first, optimal stacking is attributed to structures with large geometrical overlap of the bases and maximal dispersion energy. Another view is that the optimal stacking is characterized by only a small overlap of the bases, and the stabilizing contributions originate mainly in the interactions of the permanent dipoles of the exocyclic groups of one base with the delocalized electrons of the aromatic rings of the other base (induction interaction).[21] The latter opinion is supported by the crystal structures of DNA bases, nucleosides, and nucleotides, frequently exhibiting the polar exocyclic groups (amino group, carbonyl group) stacked above the rings of adjacent bases.[21] The mechanism described was used, i.e., to explain the unusual conformational properties of the CpA dinucleotide, known from

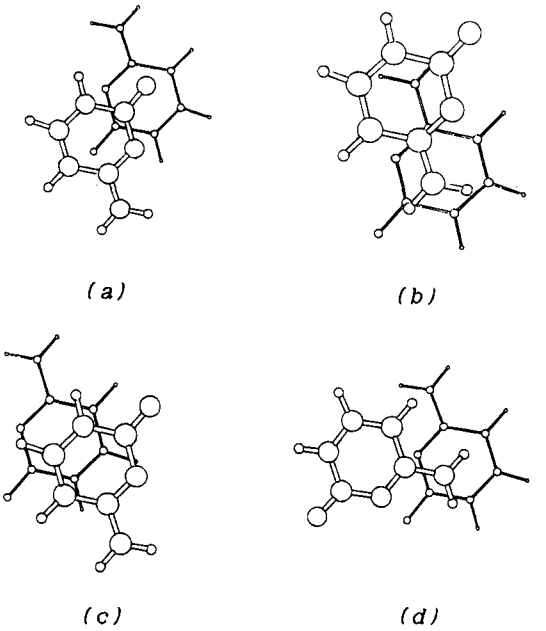

Figure 5. Four stacked geometries of cytosine dimer. The respective energies are shown in Table 3 or in the text.

the X-ray studies on B-DNA decamers.[22a-c] The empirical potential calculations cannot be used to discriminate between these two theories, because they mostly do not include the induction term at all.

As discussed above, our calculations undoubtedly favor the antiparallel structures with maximum or nearly maximum overlap of the bases. These ab initio calculations also included several cytosine dimer arrangements, exhibiting exocyclic group−aromatic ring stacking. Figure 5a shows the displaced (displacement 2 Å) antiparallel structure with both carbonyl oxygens directly interacting with the rings. This structure (denoted with an asterisk in Table 3) is, however, energetically rather unfavorable. The displaced antiparallel structure with the amino groups interacting with the rings (Figure 5b, denoted with a dollar sign in Table 3) exhibits a significantly larger stabilization energy. This enlargement is satisfactorily reproduced by the electrostatic term, and therefore it is rather due to the electrostatic interaction, and not the induction interaction. Figure 5c shows the most stable arrangement with displacement 2 Å (designated with an ampersand in Table 3). In this dimer neither the amino groups nor the carboxyl groups interact with the rings. This stable structure is also reproduced by the electrostatic term. Both the induction and electrostatic interactions are fully included in the HF interaction energy. Thus, the present ab initio calculations rule out the induction theory of stacking, at least for the cytosine dimer.

The crystal structures of nucleic acid constituents provide a number of very different stacking patterns.[21] For example, the cytosine dimer[23] exhibits a parallel face-to-back arrangement with displacement of approximately 1.5 Å and displacement angle of about 45° (see Figure 7d in ref 21). The stacking stabilization of this structure is only negligible (less than 1 kcal/mol; cf. Table 3). The crystal structure of 1-methylcytosine[23] exhibits an antiparallel displaced face-to-back arrangement, similar to that shown here in Figure 5a (see Figure 6c in ref 21). Although this structure is much more stable, it probably

(21) Bugg, C. E.; Thomas, J. M.; Sundaralingam, M.; Rao, S. T. *Biopolymers* **1971**, *10*, 175.

(22) (a) Heinemann, U.; Alings, C. *J. Mol. Biol.* **1989**, *210*, 369. (b) Privé, G. G.; Yanagi, K.; Dickerson, R. E. *J. Mol. Biol.* **1991**, *217*, 177. (c) Privé, G. G.; Heinemann, U.; Chandrasegaran, S.; Kan, L.-S.; Kopka, M. L.; Dickerson, R. E. *Science* **1987**, *238*, 498.

(23) See ref 21 and the references therein.

still does not correspond to an optimized arrangement of stacked bases. Finally, the crystal structure of cytosine monohydrate exhibits a face-to-face arrangement (see Figure 5b in ref 21). We also considered this geometry (Figure 5d); the respective interaction energy was −0.3 kcal/mol. Evidently, stacking patterns in the crystal structures of DNA constituents are governed mainly by the crystal packing forces and hydrogen-bonding interactions and cannot be used to study the base-stacking interactions. Some crystal structures of pyrimidines exhibit the presumably most unfavorable face-to-back parallel undisplaced geometry. This is accompanied (see Figure 15 in ref 21) by dramatically increased vertical separation between the bases, 3.8 Å, which gives a strong repulsion between the stacked bases.

Our optimized structures for the cytosine stacked pair agree surprisingly well with those found by Poltev and Shylyupina[24] using an empirical potential. On the other hand, the optimized stacked structures, determined by using the perturbation energy terms,[25] are rather different.

E. Accuracy of the Evaluated Interaction Energies. We are certainly aware of the fact that the theoretical level reached is still far from accurate. The size of the cluster studied prevents, however, the use of a more sophisticated method and/or considerably larger basis sets. The accuracy of the stabilization energies calculated here could thus be evaluated only using smaller clusters for which the stabilization energy was determined at the actual as well as higher theoretical levels and where experimental data are available. The interaction energy of the stacked benzene···Ar cluster was determined[26] at a comparable theoretical level (MP2/6-31+G*, 7s4p2d1f). The resulting stabilization energy agreed well with the experimental value. The one-particle basis set limit[27] of the MP2 procedure was, however, considerably larger, by as much as about 60%. But it was further shown[6c,27] that the higher-order correlation energy contributions are repulsive. As frequently happens in the world of molecular clusters, the MP2 stabilization energy evaluated with the medium basis set is quite reasonable as a result of the compensation of errors. It is not straightforward to expect the same type of compensation for the presently studied cluster, but experience shows[28] that it is surprisingly generally valid.

For comparison of the H-bonded and stacked pairs we believe that the MP2 level is sufficient to provide reliable relative stabilization energy values. We are at least not aware of any evident failure of the MP2 procedure in this respect.

(24) Poltev, V. I.; Shulyupina, N. V. *J. Biomol. Struct. Dyn.* **1986**, *2*, 739.

(25) Langlet, J.; Claverie, P.; Caron, F.; Boeuve, J. C. *Int. J. Quantum Chem.* **1981**, *20*, 299.

(26) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Chem. Phys.* **1991**, *95*, 391.

(27) Bauder, A.; Brupbacher, T.; Klopper, W.; Luthi, H. P. *J. Chem. Phys.*, submitted.

(28) Hobza, P.; Zahradník, R. *Chem. Rev.* **1988**, *88*, 871.

Conclusion

(i) The H-bonded structure of the cytosine dimer is more stable than the optimal stacked structure. The stabilization energy of the latter structure, however, corresponds to almost 50% of that of the H-bonded pair. The antiparallel arrangement of the two stacked cytosines is always significantly favored over the parallel one (this concerns the isolated cytosine dimer).

(ii) H-bonded interactions are specific while stacking interactions are nonspecific and more flexible and, therefore, contribute significantly to the conformational flexibility of DNA.

(iii) The electrostatic energy contributes a dominant part of the H-bonded stabilization while the decisive part of the stabilization in the stacked pair comes from electron correlation.

(iv) The twist dependence of the stabilization energy of the stacked structure is reproduced by the electrostatic energy. However, for some orientations of the dimer, the ab initio interaction energy considerably differs from the electrostatic energy. These discrepancies are not removed by augmenting the electrostatic term by the Lennard-Jones−van der Waals (dispersion and repulsion) contributions.

(v) Electrostatic energy evaluated with correlated multipoles reproduces the dependence of the total interaction energy on the twist and displacement better than that evaluated with the HF multipoles. It is therefore recommended to use electrostatic energy evaluated with correlated multipoles in construction of the empirical potentials.

(vi) No significant attraction originating in the interactions between the polar exocyclic groups of one cytosine and delocalized electrons of the aromatic ring of the other cytosine was found in the stacked dimer.

Note Added in Proof: We have recently studied the rotation and displacement dependence of the stacking interaction energy of the cytosine dimer with various sets of MEP charges. Electrostatic energy evaluated with MP2/6-31G* (0.25) MEP charges nicely reproduces the present MP2/6/31G* (0.25) interaction energy with the exception of the positive peak of nonelectrostatic origin near the displacement angle of 180°. The MEP electrostatic energy agreed with ab initio data considerably better than the correlated distributed multiple analysis (CDMA) used in the present study, specifically concerning the negative peak at a displacement angle of about 270°, which is clearly overestimated by the ORIENT calculations. It indicates that CDMA calculations carried out with the present cluster still do not converge and supports the use of the MEP approach.

JA941750+