Bifurcated Hydrogen Bonds in DNA Crystal Structures. An ab Initio Quantum Chemical Study

Jiří Šponer*,†,‡ and Pavel Hobza†,§

Contribution from the J. Heyrovský Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, Dolejškova 3, 182 23 Prague 8, Czech Republic, Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic, and Technical University of Munich, Institute of Physical and Theoretical Chemistry, 8046 Garching, Germany

Received May 27, 1993. Revised Manuscript Received October 5, 1993®

Abstract: An ab initio quantum chemical analysis is performed on the intrinsic deformability of the DNA base amino groups and their role in the base stacking interactions and conformational variability observed in the DNA crystal structures. The present calculations, made at the HF/6-31G(NH_2*) and MP2/6-31G* levels of theory, lead to results qualitatively different from the previous empirical potential studies and demonstrate limited applicability of the commonly used force fields. The amino groups of isolated DNA bases are nonplanar, and the deviation of the amino group hydrogens from the DNA base plane amounts to 0.1–0.5 Å. The largest amino group nonplanarity is found for guanine. In the case of cytosine containing complexes, modeling the isolated base pair, the amino group geometry is determined primarily by the intermolecular geometry of the hydrogen bonds. The flexibility of the amino groups facilitates optimization of the interaction energy under condition of nonplanar geometry of the complex. On the other hand, the DNA base amino groups are significantly nonplanar, if they participate in the interstrand bifurcated hydrogen bonds or in the interstrand contacts of amino groups. Both phenomena are observed in many DNA crystal structures. The nonplanar amino group geometry improves the interaction energy. It is demonstrated that the widespread idea of the interstrand repulsive amino group clashes in the DNA is not correct, because close contact between two amino groups results in an attractive interaction similar to that in the bifurcated hydrogen bonds. The only exception represents the steps having crystallographically identical base pairs. It is because attractive amino group interaction requires a highly asymmetric arrangement of the two amino groups, while any geometry with 2-fold symmetry is repulsive. The ab initio calculations are supplemented by an analysis of the contacts of amino groups in the available B-DNA crystals to show that the close amino group contacts are very frequent in the asymmetric steps. These close contacts are, however, absent in the central steps of the crystal structures with crystallographically identical strands. This finding agrees with the nonempirical calculations and shows that conformational variability of the symmetric steps is significantly restricted by the crystal packing forces.

1. Introduction

Studies of DNA conformation by single-crystal X-ray crystallography revealed that the DNA molecule deviates from canonical structures as known from fiber diffraction analysis. The conformational variability significantly affects the biological function of DNA, its deformability, and its interactions with proteins, drugs, and other molecules.

This study presents an ab initio quantum chemical analysis of the base stacking interactions observed in the DNA crystal structures. The basic advantage of the ab initio method is the fact that it is free of any empirical parameters. Further, as analyzed in detail elsewhere,[1a–d] various commonly used empirical potentials give very different results in the studies of base stacking in DNA. Two problems that cannot be analyzed unambiguously using empirical potential calculations[1b,d] are selected for the ab initio calculations. The first one is the bifurcated hydrogen bond between the adenine N6 amino group and O4 atoms of the two successive thymines occurring at the major groove of the ApA step of B-DNA. This three-center hydrogen bond is expected to stabilize the unique conformation of the long $(AA)_n$ stretches.[2] The second one is the most frequent close contact of neighboring base pairs existing in the DNA crystal structures, the N6···N6 mutual amino group interaction at the major groove of the ApT B-DNA steps.[3]

Our theoretical study consists of four different steps. Firstly, properties of the DNA base amino groups will be studied using a sufficiently large basis set with and without including correlation energy. Secondly, the interaction of cytosine amino group with carbonyl oxygens of two formamides will be analyzed. This van der Waals complex is believed to involve all the important intramolecular and intermolecular energy contributions participating in the interstrand bifurcated hydrogen bond in a DNA double helix. The amino group is bonded to the planar aromatic ring. Hydrogen bonds between the cytosine and the first formamide mimics the DNA base pair hydrogen bonds, and interaction between the cytosine and the carbonyl group of the other formamide represents the interstrand bifurcated hydrogen bond. Thirdly, the cytosine···formamide hydrogen bonded complex interacting with the methylamine amino group will be analyzed. This structure mimics the N6···N6 adenine or N4···N4 cytosine major groove contacts, frequently observed in the oligonucleotide crystals.[3] Finally, an empirical analysis of close amino group contacts will be made, using the available oligonucleotide crystal structures solved at high resolution. This

* To whom correspondence should be addressed at the J. Heyrovský Institute, Prague.
† J. Heyrovský Institute of Physical Chemistry.
‡ Institute of Biophysics.
§ Technical University of Munich.

® Abstract published in *Advance ACS Abstracts*, December 15, 1993.

(1) (a) Šponer, J.; Kypr, J. In *Theoretical Biochemistry and Molecular Biophysics*; Beveridge, D. L., and Lavery, R., Eds.; Adenine Press: New York, 1991; Vol. 1, pp 271–284. (b) Šponer, J.; Kypr, J. *J. Biomol. Struct. Dyn.* **1993**, *11*, 27–41. (c) Šponer, J.; Kypr, J. *J. Biomol. Struct. Dyn.* **1993**, *11*, 277–292. (d) Šponer, J.; Burcl, R.; Hobza, P. Submitted to *J. Biomol. Struct. Dyn.*

(2) Nelson, H. C. M.; Finch, J. T.; Luisi, B. F.; Klug A. *Nature (London)* **1987**, *330*, 221–225.

(3) Šponer, J.; Kypr, J. *Int. J. Biol. Macromol.* In press.

0002-7863/94/1516-0709$04.50/0

analysis will be compared with predictions based on the ab initio computations.

2. Method

(a) Optimization of Geometry of Isolated DNA Bases. Geometries of cytosine, guanine, and adenine were optimized at the Hartree–Fock (HF) level of theory using the standard 6-31G basis set augmented by the d polarization functions ($\alpha = 0.8$) on the amino group nitrogens (abbreviation 6-31G(NH_2*)). These polarization functions are necessary to properly describe the amino group nonplanarity.[4a-f] The cytosine geometry was also optimized using the full 6-31G* basis set (d polarization functions on all of the heavy atoms) with inclusion of the electron correlation using the second-order Møller–Plesset (MP2) theory. We are aware of the rather unbalanced character of the 6-31G(NH_2*) basis set. Our previous results,[4e,f] however, indicated that the use of the HF/6-31G* level led to a significant underestimation of nonplanarity of the amino groups of DNA bases, formamide and formamidine as compared to the more accurate MP2/6-31G*, MP2/6-31G**,[4f] and MP2/TZP[4f] levels. The HF/6-31G(NH_2*) amino group geometries were much closer to the above mentioned accurate calculations, although the nonplanarity was still somewhat underestimated.

Geometries of the DNA bases were optimized in three different ways: (1) The entire molecule was held planar (PLAN geometry); (2) The aromatic ring was held planar and only the amino group hydrogens were allowed to be nonplanar (NPA geometry); (3) The geometry of the entire molecule was completely optimized (FULL geometry). The gradient convergency criterion was equal to 0.00045.

(b) Optimization of the Amino Groups in the Complexes. Geometries of all the complexes considered (models of the base pair steps) were based on the base pair step geometries observed in the oligonucleotide crystals. The amino group interactions in these complexes were analyzed in the following way. The positions of the cytosine amino group hydrogens were relaxed, while the remaining intramolecular and all the intermolecular degrees of freedom were frozen. The calculations were made at the HF/6-31G(NH_2*) level of theory. The NPA HF/6-31G(NH_2*) optimized geometry was used as the starting geometry of cytosine.

(c) Interaction Energy. Interaction energy E_{INT} was calculated as a sum of the ab initio E_{HF} (HF, Hartree–Fock) energy and dispersion energy E_{DISP}

$$E_{INT} = E_{HF} + E_{DISP} \quad (1)$$

E_{HF} was evaluated as the difference of total energy of the complex ($E_{HF}^{R,T}$) and the sum of subsystem energies (E_{HF}^{R}, E_{HF}^{T})

$$E_{HF} = E_{HF}^{R,T} - (E_{HF}^{R} + E_{HF}^{T}) \quad (2)$$

The counterpoise method[5] was applied to eliminate the basis set superposition error (BSSE). All the orbitals of the "ghost" system were considered. E_{DISP} was estimated using the London-type expression.[6] Let us recall that this part of interaction energy is not included at the HF level of calculations and originates in electron correlation.[7a-d] The size of the present complexes prevents, however, the use of more accurate beyond HF methods such as MP2 theory. It is known that the MP2 interaction energies and interaction energies evaluated along eq 1 are, for small complexes, rather similar.[7a-d] The same was found in our preliminary study[1d] on the mutual amino group interactions (methylamine dimer).

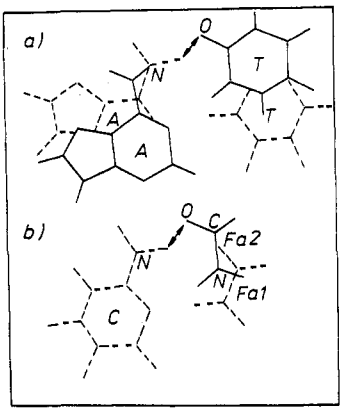

Figure 1. A sketch showing how the ApA base pair step (a) was substituted by the (C···Fa1)···Fa2 complex (b). The bifurcated hydrogen bond is indicated by an arrow.

Table 1. Intermolecular Geometry of the (C···Fa1)···Fa2, C···Fa2, and C···Fa1 Complexes[a]

cytosine···formamide1 HB pair		cytosine···formamide2 bifurcated HB	
N4–O(Fa)	3.01 Å	N4–O(Fa)	2.98 Å
C4–N4–O(Fa)	115.4°	C4–N4–O(Fa)	133.7°
N4–O(Fa)–C(Fa)	119.3°	N4–O(Fa)–C(Fa)	116.2°
N3–C4–N4–O(Fa)	13.6°	N3–C4–N4–O(Fa)	–74.9°
C4–N4–O(Fa)–C(Fa)	–31.3°	C4–N4–O(Fa)–C(Fa)	76.0°
N4–O(Fa)–C(Fa)–N(Fa)	10.3°	N4–O(Fa)–C(Fa)–N(Fa)	–65.2°

[a] Cf. Figure 1.

(d) Bifurcated Hydrogen Bond. The ApA(TpT) step was generated to form the bifurcated hydrogen bond between the adenine N6 amino group and the O4 oxygen of the thymine of the neighboring A···T pair. Both base pairs were nonplanar with a large propeller twist[8,9] –20°, while their buckles[9] were 0°. Helical twist[10] was 36°, roll[11] 5°, and slide[12] –0.3 Å. The size of the complete base pair step prevents the relevant ab initio analysis being performed. We have therefore prepared a smaller model complex with equivalent amino group interactions (Figure 1). The adenine forming the bifurcated hydrogen bond was replaced by smaller cytosine (deformability of the adenine and cytosine amino group is similar—see below), while the other adenine was removed. The two thymines were replaced by formamides Fa1 and Fa2. Fa1 and Fa2 mimics the "base pair" hydrogen bonds (HB) with the cytosine[13a,b] and the bifurcated HB, respectively. Table 1 contains all the 12 independent distances and angles defining the intermolecular geometry of the (C···Fa1)···Fa2 complex.

(4) (a) Boggs, J. E.; Niu, Z. *J. Comput. Chem.* **1985**, *6*, 46–55. (b) Niu, Z.; Boggs, J. E. *J. Mol. Struct. (Theochem)* **1984**, *109*, 381–389. (c) Leszczynski, J. *Int. J. Quantum Chem. Quantum Biol. Symp.* **1992**, *19*, 43–55. (d) Riggs, N. V. *Chem. Phys. Lett.* **1991**, *177*, 447–450. (e) Šponer, J.; Hobza, P. *J. Mol. Struct. (Theochem)* In press. (f) Šponer, J.; Hobza, P. Submitted to *J. Phys. Chem.*

(5) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.

(6) Kang, Y. K.; Jhon, M. S. *Theor. Chim. Acta* **1982**, *61*, 41–48.

(7) (a) Hobza, P.; Zahradník, R. *Chem. Rev.* **1988**, *88*, 871–897. (b) Hobza, P.; Zahradník, R. *Intermolecular Complexes. The Role of van der Waals Systems in Physical Chemistry and in the Biodisciplines*; Elsevier: Amsterdam, 1988. (c) Sokalski, W. A.; Hariharan, P. C.; Kaufman, J. J. *Int. J. Quantum Chem. Quantum Chem. Symp.* **1987**, *21*, 645–660. (d) Szczesniak, M. M.; Scheiner, S. *J. Chem. Phys.* **1984**, *80*, 645–655.

(8) For a definition of the DNA base pair and base pair step parameters, see: *J. Biomol. Struct. Dyn.* **1989**, *4*, 627–634.

(9) Base pair propeller twist is the angle between the two base normals caused by counterrotation of the bases around the long base pair axis (it connects the C6-pyrimidine and C8-purine atoms). Base pair buckle originates in counterrotation of the bases around the short base pair axis. Short axis is perpendicular to C6–C8 axis and contains its midpoint.

(10) Helical twist is a counterrotation of the two successive base pairs around the helical axis, or, in the case of irregular double helix structure, around the base pair step axis.

(11) Roll is the wedge angle between two successive base pairs. Positive (negative) roll opens the angle toward the minor (major) groove, while the major (minor) groove is compressed.

(12) Slide is a relative displacement of the two successive base pairs in the direction perpendicular to the grooves. Slide is near zero in the B-DNA and adopts negative values about –1.5 A in A-DNA.

(13) (a) Hobza, P.; Sandorfy, C. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 2859–2860. (b) Hobza, P.; Sandorfy, C. *Biophys. Chem.* **1984**, *19*, 201–209.

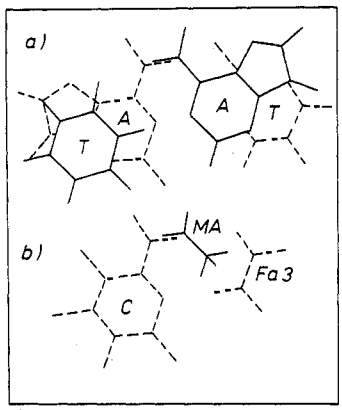

Figure 2. A sketch showing how the ApT base pair step (a) was replaced by the (C···Fa3)···MA complex (b).

The influence of the intermolecular interactions on the amino group geometry was investigated in the following way.

The geometry of the two cytosine amino group hydrogens (i.e., their bond lengths and valence and dihedral angles) was optimized, while intermolecular geometry of the complex and the cytosine ring were frozen (see above). The optimization was made for the following complexes: (C···Fa1)···Fa2 complex with the three centered bifurcated HB, C···Fa1 HB pair (Fa2 removed), and C···Fa2 interstrand contact (Fa1 removed).

Finally, the effect of amino group nonplanarity on the base stacking energy (the interaction between two successive base pairs) was estimated in the following way. Interaction energies of the C···Fa2 (Fa1 removed) dimer was calculated, using optimized cytosine amino group geometries obtained for both the (C···Fa1)···Fa2 and C···Fa2 complexes. These energies were compared with the interaction energy obtained by replacing the nonplanar cytosine by the planar one.

(e) Close Amino Group Contact. The ApT(ApT) step with close amino group contact was generated in the following way: The base pair propeller twist was −8°, buckle 0°, helical twist 39°, and roll 0°. The base pair vertical separation was adjusted to make the N6···N6 amino groups distance 3.15 Å, in agreement with the mean value observed in the crystal structures.[3,14] Then one of the A···T pairs was replaced by a cytosine···formamide (C···Fa3) HB complex (Figure 2). The other thymine was removed, and the remaining adenine (its amino group) was replaced by a methylamine (MA). The methylamine amino group was constrained to have fixed dihedral angles of the amino group hydrogens consistent with geometries of the DNA bases (optimized MA exhibits too large pyramidalization of the amino group[1d]). The MA amino group hydrogens pointed outward of the cytosine amino group. This geometry was chosen on the basis of our previous ab initio analysis made on methylamine dimer and cytosine dimer.[1d] It was shown that the mutual amino group interaction was optimized, if one of the amino groups was oriented away from the stacked complex (as MA here), and the hydrogens of the other amino group pointed toward the lone electron pair of the first amino group. Table 2 presents the intermolecular geometry of the (C···Fa3)···MA complex and the two dihedral angles of the MA amino group.

The geometry of the amino group hydrogens was then optimized (see above), while the intermolecular geometry of the complex and geometry of the rest of cytosine was kept rigid. The optimization was performed for the (C···Fa3)···MA complex, C···Fa3 HB pair, and C···MA amino group contact. Finally, the interaction energies of the C···MA dimer were calculated for various cytosine amino group geometries. These energies were compared with the interaction energy obtained by replacing the nonplanar cytosine by the planar one.

(14) As demonstrated elsewhere on methylamine dimer, the interaction between two amino groups does not depend much on the mutual orientation of the two C–NH_2 bonds.[1d] Therefore the base pair step geometry used is representative of all the major groove amino group contacts in the available DNA crystals.

Table 2. Intermolecular Geometry of the (C···Fa3)···MA, C···MA, and C···Fa3 Complexes

cytosine···formamide3 HB pair		cytosine···methylamine amino group contact	
N4–O(Fa)	2.96 Å	N4–N(MA)	3.15 Å
C4–N4–O(Fa)	118.2°	C4–N4–N(MA)	104.3°
N4–O(Fa)–C(Fa)	121.2°	N4–N(MA)–C(MA)	104.3°
N3–C4–N4–O(Fa)	4.1°	N3–C4–N4–N(MA)	−80.8°
C4–N4–O(Fa)–C(Fa)	−7.6°	C4–N4–N(MA)–C(MA)	73.2°
N4–O(Fa)–C(Fa)–N(Fa)	3.2°	N4–N(MA)–C(MA)–X	−80.8°
		H1(MA)–N(MA)–C(MA)–X	158.0°
		H2(MA)–N(MA)–C(MA)–X	10.0°

[a] X is a dummy atom replacing the adenine N1 nitrogen; it defines the DNA base plane for the methylamine amino group. Cf. Figure 2.

(f) Phenomenological Analysis of the Amino Group Contacts. Atomic coordinates of the base pair steps were extracted from the Brookhaven database.[15] Subsequently, the N···N amino group distances between the neighboring base pairs were calculated. The interstrand amino group contacts are possible in the alternating CpG, ApT, and ApG steps.[3] The analysis included the B-DNA hexamers d(CGATCG) and d(CGCGCG) complexed with intercalators (PDB codes 1D15,[16] 1D37,[17] 1D38,[17] 1D10,[18] 1D12,[18] and 1D33[19]) with exclusion of the covalently modified N2(G) of 1D33, and the B-DNA decamers d(CCAACGTTGG)[20] (5DNB), d(CCGGCGCCGG)[21] (1GCG), d(CGATCGATCG)[22] (1D56 and 1D57), d(CGATTAATCG)[23] (1D4 9), d(CGATCGATCG)[24] (1D23), d(CCAAGATTGG)[25] (3DNB), d(CCAGGCCTGG)[26] (1BD1), and d(CCAGGC^{5M}CTGG)[27a,b] (2D25). The A-DNA and Z-DNA crystal structures and the B-DNA dodecamers were not analyzed here. As shown elsewhere, the B-DNA dodecamers exhibit the same frequent amino group contacts as the B-DNA decamers.[3] The Z-DNA structures do not contain close interatomic contacts, while the number of available A-DNA crystal structures is still insufficient to make reasonable analysis of close contacts between neighboring base pairs.[3]

3. Results and Discussion

(a) Optimized Geometries of Isolated DNA Bases. The results obtained for the isolated DNA bases are summarized in Table 3.

The table shows the dihedral angles between the amino group hydrogens and the cytosine ring and the out of plane deviations of the amino group atoms.

(15) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Mayer, G. F.; Brice, M. D.; Rotgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535–542.
(16) Williams, L. D.; Frederick, C. A.; Ughetto, G.; Rich, A. *Nucleic Acids Res.* **1990**, *18*, 5533–5540.
(17) Gao, Y.-G.; Wang, A. H.-J. *Anti Cancer Drug Des.* **1991**, *6*, 137–144.
(18) Frederick, C. A.; Williams, L. D.; Ughetto, G.; van der Marel, G. A.; van Boom, J. H.; Rich, A.; Wang, A. H.-J. *Biochemistry* **1990**, *209*, 2538–2549.
(19) Wang, A. H.-J.; Gao, Y.-G.; Liaw, Y.-C.; Li, Y.-K. *Biochemistry* **1991**, *30*, 3812–3815.
(20) Privé, G. G.; Yanagi, K.; Dickerson, R. E. *J. Mol. Biol.* **1991**, *217*, 177–199.
(21) Heinemann, U.; Alings, C.; Bansal, M. *EMBO J.* **1992**, *11*, 1931–1939.
(22) Yuan, H.; Quintana, J.; Dickerson, R. E. *Biochemistry* **1992**, *31*, 8009–8021.
(23) Quintana, J. R.; Grzeskowiak, K.; Yanagi, K.; Dickerson, R. E. *J. Mol. Biol.* **1992**, *225*, 379–395.
(24) Grzeskowiak, K.; Yanagi, K.; Privé, G. G.; Dickerson, R. E. (1991) *J. Biol. Chem.* **1991**, *266*, 8861–8883.
(25) Privé, G. G.; Heinemann, U.; Chandrasegaran, S.; Kan, L.-S.; Kopka, M. L.; Dickerson, R. E. *Science* **1987**, *238*, 498–504.
(26) Heinemann, U.; Alings, C. *J. Mol. Biol.* **1989**, *210*, 369–381.
(27) (a) Heinemann. U.; Alings, C. *EMBO J.* **1991**, *10*, 35–43. (b) Heinemann, U.; Hahn, M. *J. Biol. Chem.* **1992**, *267*, 7332–7341.

Table 3. Optimized Geometries of the Isolated DNA Bases[b]

DNA base	level	geometry	δE	DN (Å)	dihedral[a]	TH (deg)	DH (Å)
cytosine	MP2/6-31G*	FULL	0.38	-0.04	C5C4N4H41	26.2	0.31
					N3C4N4H42	-14.7	0.17
cytosine	MP2/6-31G*	NPA	0.28	0	C5C4N4H41	22.7	0.34
					N3C4N4H42	-12.7	0.20
cytosine	HF/6-31G(NH_2*)	FULL	0.15	-0.03	C5C4N4H41	20.8	0.24
					N3C4N4H42	-11.7	0.12
cytosine	HF/6-31G(NH_2*)	NPA	0.10	0	C5C4N4H41	18.4	0.27
					N3C4N4H42	-10.5	0.16
adenine	HF/6-31G(NH_2*)	NPA	0.09	0	C5C6N6H61	13.0	0.20
					N1C6N6H62	-14.2	0.21
guanine	HF/6-31G(NH_2*)	NPA	0.63	0	N3C2N2H21	-11.8	0.18
					N1C2N2H22	31.1	0.46

[a] Cf. Figure 3. [b] δE (kcal/mol), energy stabilizing the nonplanar geometry compared to the planar one; DN, deviation of the amino group nitrogen from the DNA base plane; DH, deviation of the amino group hydrogen from the DNA base plane; TH, dihedral angle of the amino group hydrogen; FULL geometry, optimization made without any constraints; NPA geometry, only the amino group hydrogens are nonplanar.

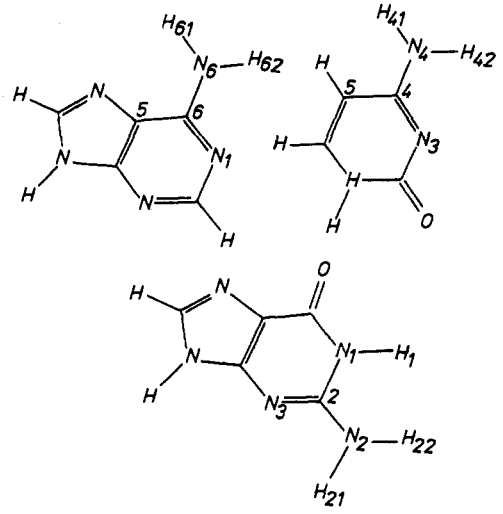

Figure 3. The molecular structure and atom numbering of cytosine, guanine, and adenine.

The first row of Table 3 presents the geometry of isolated cytosine evaluated at the MP2/6-31G* level. The cytosine adopts nonplanar geometry with significant pyramidalization[4a-f] of the amino group hydrogens, while the cytosine ring is almost planar (not shown). The amino group hydrogens deviate out from the DNA base plane (in the same direction) by 0.17 and 0.31 Å. The amino group nitrogen is slightly shifted into the opposite direction. The dihedral angle N3C4N4H42 of the amino group hydrogen H42 (Figure 3) is significantly smaller than the dihedral angle C5C4N4H41 of the other amino group hydrogen. The nonplanar geometry is stabilized by 0.38 kcal/mol compared to the planar geometry. The second row of Table 3 shows that the cytosine geometry optimization made with nonplanar amino group hydrogens only (NPA) gives similar results as the full geometry optimization.

The next four rows of Table 3 present a comparison of the amino group nonplanarity of cytosine, adenine, and guanine at the HF/6-31G(NH_2*) level. The deformation of the amino groups of adenine and cytosine is similar, except that the cytosine amino group is deformed in an asymmetric way, as discussed above. Guanine exhibits significantly larger nonplanarity of the amino group with a high degree of asymmetry. The dihedral angle N1C2N2H22 of its H22 hydrogen is almost three times larger than that of the other amino group hydrogen. This is probably due to a repulsive interaction of the H22 hydrogen with the neighboring H(N1) hydrogen.

The main goal of this article is to study the bifurcated hydrogen bonds observed in the DNA crystal structures and not the isolated DNA bases. Therefore, a thorough study of the properties of the isolated DNA bases will be published in future.

Table 4. Optimized Dihedral Angles of the Cytosine Amino Group Hydrogens within Various Complexes

complex[a]	level	C5C4N4H41 (deg)	N3C4N4H42 (deg)
C	HF/6-31G(NH_2*)	+18.4(18.4)	-10.5(+10.5)
C···Fa1	HF/6-31G(NH_2*)	-11.4	+13.2
C···Fa2	HF/6-31G(NH_2*)	+28.2	-12.6
(C···Fa1)···Fa2	HF/6-31G(NH_2*)	+16.8	+4.7

[a] Cf. Figure 1.

The MP2/6-31G* optimization could not be applied to larger van der Waals complexes of biological relevance, analyzed in the next two sections. The following analysis is therefore based on the HF/6-31G(NH_2*) calculations. It was shown[4a,b] that the use of the 4-21G(NH_2*) basis set caused the amino groups of formamide and aniline to be excessively nonplanar. We have therefore tested the HF/6-31G(NH_2*) level of calculations for both formamide and aniline (not shown). Fortunately, the HF/6-31G(NH_2*) calculations led to significantly smaller amino group nonplanarity than the HF/4-21G(NH_2*) calculations. In addition, a comparison of the HF/6-31G(NH_2*) data with the more reliable MP2/6-31G* data for cytosine (Table 3) shows that the amino group nonplanarity could still be underestimated at the HF/6-31G(NH_2*) level. It is therefore believed that the present level is sufficient for a qualitative nonempirical analysis of the interactions in the DNA, which is necessarily accompanied by many other approximations (small model system, rigid intermolecular geometry instead of the flexible base pairs, missing solvent effects, etc.).

(b) Bifurcated Hydrogen Bond. Table 4 summarizes the results obtained for the C···Fa1, C···Fa2, and (C···Fa1)···Fa2 van der Waals complexes used as a model of the bifurcated hydrogen bond within the ApA B-DNA step.

The deformation of the cytosine amino group within the isolated C···Fa1 HB pair is nonnegligible; dihedral angles of the two amino group hydrogens are larger than 10°. At the same time the O(Fa1)–N4(C)–H42(C) angle is only 3.1° (not shown), so that the hydrogen bond is almost linear. It indicates that the intermolecular geometry and intermolecular interaction energy (and not the intrinsic nonplanarity of the cytosine amino group) determine the geometry of the cytosine amino group within the C···Fa1 HB complex. (The C···Fa1 HB pair is nonplanar due to the propeller twist—see Method.)

The amino group is deformed even more in the C···Fa2 complex. The opposite signs of the amino group hydrogen dihedral angles of C···Fa1 and C···Fa2 complexes (Table 4) show that the formamide Fa2 bends the cytosine amino group hydrogens in the opposite direction compared to the formamide Fa1.

The amino group geometry within the (C···Fa1)···Fa2 bifurcated HB complex is influenced by both formamides. The hydrogen bonded amino group hydrogen H42 interacts more strongly with the Fa1 than with the Fa2 formamide. The sum

Table 5. Optimized Dihedral Angles of the Cytosine Amino Group Hydrogens within Various Complexes[a]

complex	level	C5C4N4H41 (deg)	N3C4N4H42 (deg)
C	HF/6-31G(NH_2*)	+18.4(−18.4)	−10.5(+10.5)
C···Fa3	HF/6-31G(NH_2*)	−4.2	+4.1
C···MA	HF/6-31G(NH_2*)	+27.6	−18.5
C···MA	HF/6-31G	11.5	−11.8
(C···Fa3)···Ma	HF/6-31G(NH_2*)	+19.9	−5.9

[a] Cf. Figure 2.

Table 6. Interaction Energies of the C···Fa2 and C···MA Complexes[f]

complex	E_{HF}	E_{DISP}	E_{INT}	δE
C···Fa2[a]	−2.19	−1.19	−3.38	
C···Fa2[b]	−4.18	−1.38	−5.56	−2.18
C···Fa2[c]	−3.13	−1.20	−4.33	−0.95
C···MA[a]	+1.98	−1.81	+0.17	
C···MA[d]	−0.71	−2.15	−2.86	−3.03
C···MA[e]	+0.65	−2.00	−1.35	−1.52

[a] Planar cytosine. [b] Cytosine amino group geometry taken from the C···Fa2 complex (Table 4, row 3). [c] Cytosine amino group geometry taken from the (C···Fa1)···Fa2 complex (Table 4, row 4). [d] Cytosine amino group geometry taken from the C···MA complex (Table 5, row 3). [e] Cytosine amino group geometry taken from the (C···Fa3)···MA complex (Table 5, row 5) [f] The cytosine amino group adopts various geometries, as indicated. E_{HF} (kcal/mol) is the ab initio HF/6-31G(NH_2*) interaction energy, E_{DISP} (kcal/mol) is the dispersion energy, and $E_{INT} = E_{HF} + E_{DISP}$ is the total interaction energy of the complex. δ_E (kcal/mol) is the difference between the interaction energies for a given amino group geometry and for the planar amino group geometry.

of the N4(C)–H42–O(Fa1), N4(C)–H42–O(Fa2), and O(Fa1)–H42–O(Fa2) angles is 359.8°, while the individual angles are larger than 90°. It means that the optimized amino group geometry satisfies the criteria of the bifurcated hydrogen bonds.[28] Position of the H41 hydrogen is determined mainly by its interaction with Fa2. The surprisingly large value of its dihedral angle indicates that this hydrogen participates nonnegligibly in the interaction energy optimization.

(c) Close Amino Group Contact. Table 5 summarizes the results obtained for the C···Fa3, C···MA, and (C···Fa3)···MA van der Waals complexes modeling the close amino group contact.

The cytosine amino group is almost planar in the C···Fa3 HB complex and the O(Fa3)–N4(C)–H42(C) angle is 1.9°. It indicates that the amino group geometry is again determined by geometry of the HB complex, which is almost planar (Table 2). Some additional calculations were made using the C···Fa complex to show that fixed planar geometry of the HB complex planarized the cytosine amino group (not shown).

The amino group is significantly deformed in the C···MA complex (Table 5, row 3). For such large deformations the intrinsic deformability of the amino group is very important. It is demonstrated on the next row of Table 5, showing the same result, but obtained using the 6-31G basis set (i.e., without the polarization functions). This basis set penalizes any nonplanarity of the cytosine amino group, similar to the classical empirical potentials, and it results in a strong underestimation of the amino group pyramidalization.

The cytosine amino group geometry within the (C···Fa3)···MA complex is affected by both formamide and methylamine. The sum of the N4(C)–H42–O(Fa3), N4(C)–H42–N(MA), and O(Fa1)–H42–N(MA) angles is 360.0°, while the individual angles are larger than 90°. Such a perfect coplanarity of the four atoms demonstrates that the close amino group contact results in an interaction closely related to the bifurcated hydrogen bonds. Again, the non-HB hydrogen H41 is significantly deformed. Table 6 compares the interaction energies E_{INT} of the C···Fa2 and C···MA complexes, calculated using the cytosine amino group geometry optimized within the C···Fa2, C···MA, (C···Fa1)···Fa2, and (C···Fa3)···MA complexes. To estimate the interaction energy improvement (δE) caused by the nonplanar amino group geometry, the calculations were made using the planar cytosine as well. The comparison shows two interesting things. Firstly, the interaction is attractive for both NH_2···NH_2 and NH_2···O contacts; however, the bifurcated NH_2···O hydrogen bond leads to larger stabilization energies. On the other hand, δE is larger in the case of the NH_2···NH_2 contact. It indicates that the deformability of the amino groups is more important in the case of close amino group contact than in the case of the O···NH_2···O bifurcated HB. It confirms our preliminary calculations made on smaller van der Waals complexes[1d] showing that the optimization of the mutual amino groups orientation leads to a significant improvement (several kcal/mol) of the interaction energy.

The importance of the amino group clashes for the DNA conformational variability was proposed more than one decade ago and conformational manuvers (roll, twist, slide, and reduced propeller[8]) were found, eliminating the amino group clashes.[30] Later, on the basis of the empirical potential calculations, more efficient manuvers were proposed: stagger[1b] and positive cup.[1b,31] However, the present nonempirical calculations give a different picture of the mutual amino group interactions in the DNA crystal structures. The most efficient way to eliminate the repulsive interstrand contact of amino groups is the optimization of mutual interaction of the amino groups, enabled by the intrinsic deformability of the amino groups. It results in an attractive interaction between the amino groups. In contrast to the previous theories,[30,31] the ab initio calculations are capable of explaining the mutual amino group contacts observed in the oligonucleotide crystals.[1a,3]

The empirical potentials qualitatively underestimate the role of the mutual interactions of amino groups in the DNA conformational variability. These potentials penalize the nonplanar geometries of the DNA base amino groups and underestimate the effect of the amino group nonplanarity on the intermolecular interaction energy.[1d]

(d) ApT and CpG Steps with 2-Fold Symmetry. Most of the available oligonucleotide crystal structures consist of two self-complementary strands. These strands could, in principle, be crystallographically identical. However, the DNA crystals usually have two nucleotide strands per asymmetric unit. The mechanisms, stabilizing the asymmetry, are expected to be relatively subtle.[27,33] We propose, on the basis of the present calculations, that the close amino group contacts can contribute to the crystallographic asymmetry of the self-complementary oligonucleotides. This can be explained in the following way.

Figure 4 illustrates the four possible geometries of amino groups within the ApT step. (The analysis also holds for any other interstrand amino group contact.) Figures 4a,b show the unfavorable arrangements of the amino groups. The amino groups (their hydrogens) both point either away from the step or toward each other. This arrangement is symmetric, i.e., the step could also be symmetric. The other two possible amino group arrangements (Figure 4c,d) enable a bifurcated bond to be made between the amino groups. Hydrogens of one of the amino groups interact with the lone electron pair region of the nitrogen of the other amino group. However, the two adenines are no longer equivalent, and therefore the step cannot be symmetric. At the same time, it follows from Figure 4 that the central ApT, CpG,

(28) The N–H···O1, N–H···O2, and O1···H···O2 angles should all be greater than or equal to 90°, and the sum of these angles should be close to 360°, i.e., the four atoms should be coplanar.[26,29a,b]

(29) (a) Jeffrey, G. A.; Mitra, J. *J. Am. Chem. Soc.* **1984**, *106*, 5546–5553. (b) Fritsch, V.; Westhof, E. *J. Am. Chem. Soc.* **1991**, *113*, 8271–8277.
(30) Calladine, C. R. *J. Mol. Biol.* **1982**, *161*, 343–352.
(31) Šponer, J.; Kypr. J. *J. Biomol. Struct. Dyn.* **1990**, *7*, 1211–1220.
(32) Cup is a difference of buckles of two successive base pairs. It is positive, if the two base pairs are bent toward each other like two cupped hands.[20]
(33) Schneider, B.; Ginell, S. L.; Jones, R.; Gaffney, B.; Berman, H. M. *Biochemistry* **1992**, *31*, 9622–9628.

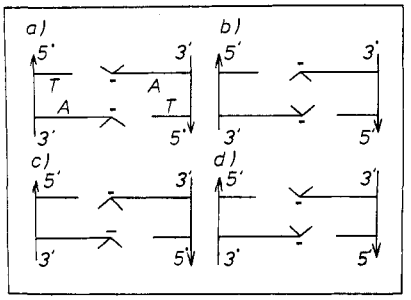

Figure 4. Four possible mutual arrangements of the amino group hydrogens of the interstrand amino group contact. The negative charges (–) represent the amino group nitrogen lone electron pairs. For details see the text.

Table 7. Interstrand Amino Group Distances at the CpG, ApT, and ApG Steps with and without Symmetry[a]

	step	structures	*N*	*R* (Å)	**R** (Å)
Asymmetric Steps					
N6···N6	ApT	B-decamers	10	2.90–3.27	3.13
N6···N4(N6)	ApG	B-decamers	4	3.19–3.35	3.24
N2···N2	CpG	B-decamers	12	2.92–3.55	3.25
N4···N4	CpG	B-decamers	12	3.18–3.53	3.39
total			38	2.90–3.55	3.26
Steps with 2-Fold Symmetry					
N6···N6	ApT	B-hexamers	5	3.26–3.53	3.41
N2···N2	CpG	B-decamer	1	3.99	
N4···N4	CpG	B-decamer	1	3.49	
N4···N4	CpG	B-hexamer	1	3.64	
total			8	3.26–3.99	3.52

[a] *N*, a number of the available independent steps; *R*, the range of the observed N···N distances in Å; and **R**, mean N···N distance.

and ApG[34] steps of oligonucleotide crystals having crystallographically identical strands should not contain close interstrand contacts of amino groups. It is because the attractive interaction between the amino groups (Figure 4c,d) is excluded by the 2-fold symmetry. This theoretical prediction inspired us to compare the amino groups' distances in the symmetric and asymmetric steps of the available DNA crystal structures, solved at high resolution (Table 7). Comparison of eight symmetric steps with possible close amino group contact[34] with 38 asymmetric steps shows that the mean N···N amino groups distance in the symmetric steps is by 0.25 Å larger than is the mean N···N distance for the remaining steps. None of the analyzed symmetric steps has the N···N distance shorter than is the mean N···N amino groups distance in the steps without symmetry. The crystal data analysis thus supports the theoretical calculations and emphasizes the importance of the mutual amino group interactions on the DNA conformational variability. It has been proposed, on the basis of empirical potential studies, that a positive cup[32] combined with positive base pair roll represents the most efficient way of how to eliminate the amino group contact (clash) at the minor groove of the CpG B-DNA step.[1b,31] The central CpG step of the d(CCAACGTTGG) monoclinic decamer,[20] having a large positive cup 22° combined with positive roll, agrees with the empirical potential theory.[1b,31] However, the other CpG B-DNA steps exhibit mostly negative cups[21-24] accomapanied with close amino group contacts.[1b,3] The present results offer the following explanation of this observation. Since the above mentioned CpG step consists of crystallographically identical base pairs,[20] the mutual amino group interaction is unfavorable (Figure 4). The step, therefore, eliminates both the N2···N2 and N4···N4 contacts using the cup–roll manuver derived from empirical potential studies.[1b,31] On the other hand, the remaining CpG steps are asymmetric[21-24] and can contain the amino group contacts.

(34) Close amino group contacts are possible in the ApT, CpG, and ApG (CpT) B-DNA steps but not in the remaining steps.[3]

4. Conclusions

The ab initio calculations predict, in contrast to the empirical potentials, that the DNA base amino groups are intrinsically nonplanar. The dihedral angles between the amino group hydrogens and the DNA base ring range from approximately 10°–15° up to 30°–40° with the largest value obtained for the N1C2N2H22 dihedral angle of guanine. The energy stabilizing the nonplanar geometry of the DNA bases is (at the HF level) less than 1 kcal/mol, hence the amino groups are very flexible. In case of hydrogen bonded complexes like isolated base pairs the amino group geometry is primarily determined by intermolecular geometry of the complex. Providing that the complex is nonplanar, the flexible amino groups facilitate optimization of the interaction energy. On the other hand, the nonplanar amino group geometry significantly improves the interaction energy in the case of the interstrand bifurcated hydrogen bonds and mainly in the case of the close contacts between two amino groups. The close amino group contact, existing in the ApT and CpG steps of many B-DNA crystal structures, forces the two interacting amino groups to adopt asymmetric geometry. On the other hand, 2-fold symmetry destabilizes the mutual amino group contact. The theoretical prediction is consistent with the analysis of the available oligonucleotide crystal structures because the close amino group contacts are absent in the steps having 2-fold symmetry.

Acknowledgment. J.Š. thanks Professor V. Kapička and Professor V. Vetterl for permanent support of this work.