Potential Energy Surface of the Cytosine Dimer: MP2 Complete Basis Set Limit Interaction Energies, CCSD(T) Correction Term, and Comparison with the AMBER Force Field

Petr Jurečka,[†] Jiří Šponer,[†,‡] and Pavel Hobza*[,†]

Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center for Complex Molecular Systems and Biomolecules, 166 10 Prague 6, Czech Republic, and Intitute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic

Received: January 5, 2004; In Final Form: February 12, 2004

The complete basis set (CBS) limit of the MP2 interaction energy and the CCSD(T) correction term determined as the difference between the CCSD(T) and MP2 interaction energies were evaluated for 17 stacked and 4 H-bonded structures of the cytosine dimer. Extrapolation to the MP2 CBS limit was done using the aug-cc-pVDZ and aug-cc-pVTZ results, and the CCSD(T) correction term was determined with the 6-31G*(0.25) basis set. Extrapolation to the CBS limit is essential in all parts of the potential energy surface and affects mainly the absolute MP2 stabilization energy. The effect on the relative stability is smaller but not negligible. The CCSD(T) correction term is for stacked structures with large overlaps of the monomers substantially repulsive but not uniform. Thus, when comparing the final estimate (abbreviated as CBS(T)) with previously used medium-level MP2/6-31G*(0.25) data, the CCSD(T) correction partially compensates for the enhanced absolute MP2 CBS stabilization but further increases the relative differences between different structures. The difference between the MP2/6-31G*(0.25) and the new CBS(T) reference values is in the range of +0.3 to −2.1 kcal/mol. The difference between the parallel and antiparallel structures is enhanced by 2 kcal/mol. The CCSD(T) correction is negligible in the H-bonded structure and in stacked structures with a minimal overlap of bases. The new reference CBS(T) data are compared with the AMBER force field testing both HF and MP2 electrostatic potential fitted atom-centered charges. The overall agreement of the force field with the CBS(T) data is very satisfactory. We nevertheless identify differences that are attributed to polarization and short-term effects not included in the force field. Interestingly, whereas MP2/6-31G*(0.25) calculations are better reproduced with the MP2 variant of the force field the new CBS(T) reference data are in better agreement with the use of HF charges in the force field, indicating the enhanced role of polarization.

1. Introduction

The determination of accurate stabilization energies of DNA and RNA base complexes represents a very important task because, for example, the stability of nucleic acids depends on the relative values of stabilization energies of planar H-bonded and stacked pairs of DNA and RNA bases.[1] Base-stacking energies provide a key contribution to the sequence-dependent conformational variability of DNA duplexes.[2] A proper description of the interactions helps us to understand the basic principles governing the formation of the 3D nucleic acid architectures and is of key importance in all molecular-modeling approaches. The energy information cannot be directly obtained, for example, from the oligonucleotide crystals, and experimental studies of gas-phase stabilization energies of base pairs are tedious if not impractical.[3] Solution thermodynamics experiments do not show the structure−energy relationship at the atomic level of resolution, and the energy data are largely influenced by other contributions such as the hydration energies.[4] In other words, the intrinsic base−base interaction terms cannot be unambiguously extracted. On the other side, ab initio computational chemistry offers us the possibility of consistently determining the intrinsic energies of different types of DNA base pairs and allows us to unambiguously assign reliable values of energy to each base−base configuration.[5] However, calculations should be performed at an adequate theoretical level where various contributions to the stabilization energy are properly considered.[5] It is, for example, well established that density functional methods, although suitable for planar base pairing, fail for base stacking.[6]

In the past decade, the base-stacking energies were extensively characterized using the second-order Møller−Plesset (MP2) method and the 6-31G*(0.25) diffusely polarized basis set of atomic orbitals.[5−10] This basis set has modified exponents of the d-polarization functions in order to optimize the calculated molecular polarizability.[11] The 6-31G*(0.25) basis set qualitatively improves the description of stacking energies compared to its 6-31G* standard variant.[5] MP2/6-31G*(0.25) (or closely related) calculations were carried out for several hundred configurations of nucleobase dimers including protonated and chemically modified bases,[9] base−intercalator complexes,[12] and other aromatic stacking interactions of biological or chemical interest.[4,13−17] The studies demonstrated, among other things, that aromatic stacking can be rather well described as a combination of the three most common contributions to the interaction energy: short-range exchange repulsion, dispersion attraction, and a Coulombic electrostatic term based on atom-centered point charges.[6,7,9,12]

Recent developments in computer hardware and software have allowed for further major improvements in aromatic

* Corresponding author. E-mail: pavel.hobza@uochb.cas.cz.
[†] Institute of Organic Chemistry and Biochemistry.
[‡] Intitute of Biophysics.

10.1021/jp049956c CCC: $27.50
Published on Web 04/02/2004

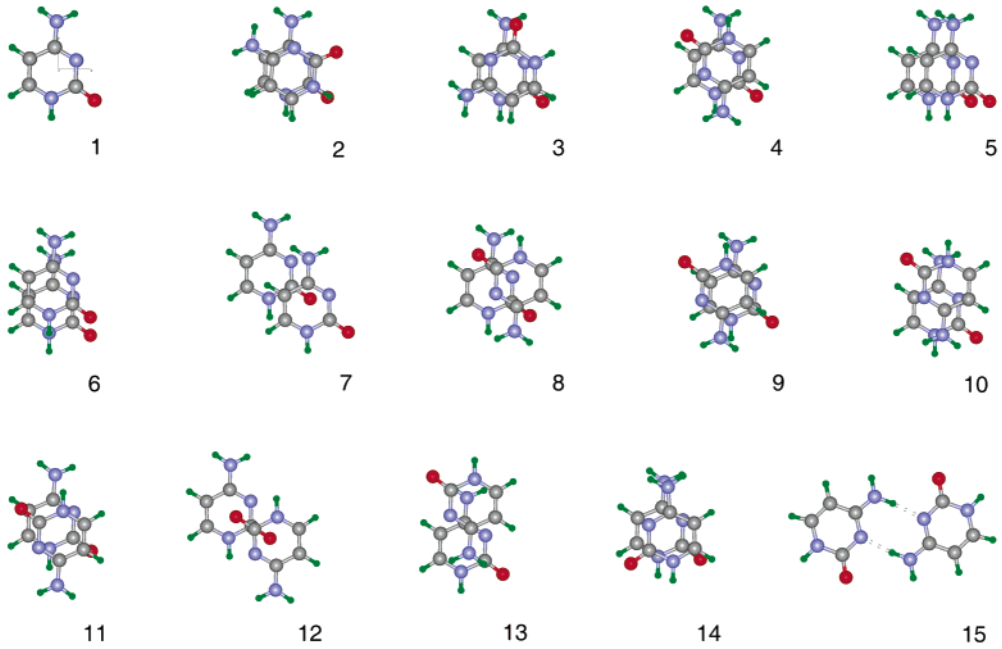

Figure 1. Cytosine dimer structures chosen.

stacking calculations. Contrary to widespread expectations, extended basis sets (e.g., triple-ζ quality containing f-polarization functions on heavy atoms and d-polarization functions on hydrogens) are still not capable of producing quantitatively accurate descriptions of different structural motifs of DNA base pairs, namely, planar H-bonded and stacked structures. And it is especially the aromatic stacking interaction that requires as accurate a description as possible because this interaction is known to be sensitive to the inclusion of higher-order electron correlation effects that lead to a repulsive correction compared to that at the MP2 level.[18]

Recently, several papers reported the stabilization energy of model complexes and stacked NA base pairs evaluated in the following way. First, the complete basis set (CBS) limit of the MP2 stabilization energy was calculated, and second, a correction term considering the higher-order correlation energy contributions was added.[19−21] The CBS limit was extrapolated from the SCF and MP2 energies evaluated at double-ζ and triple-ζ levels or, preferentially (but also more costly), at triple-ζ and quadruple-ζ levels. The second term was constructed as a difference between CCSD(T) and MP2 stabilization energies evaluated with basis sets smaller than those used at the MP2 level. The reason for the use of smaller basis set is, besides the enormous computer requirements of the CCSD(T) method, the fact that the respective difference converges to the CBS limit much faster than the MP2 and CCSD(T) stabilization energies themselves.[22] This term covers mostly triple excitations, which are included in the CCSD(T) treatment but are missing in the MP2 method. For most closed-shell molecules, triple excitations bring a nonnegligible contribution to the total correlation energy, and its omission is usually demonstrated by an overestimation of the dispersion contribution to the interaction energy.

Accurate experimental stabilization energies of extended molecular complexes that could be used for testing the above-mentioned technique are rather rare. Nevertheless, in the case of benzene...indole we were able to reproduce experimental stabilization enthalpy with high precision.[23] The situation with DNA base pairs was slightly more complicated, but final agreement with experimental stabilization enthalpies for guanine...cytosine and adenine...thymine was again very good.[24]

The recent studies confirmed the lowering of the aromatic stacking interaction energy by the inclusion of higher-order correlation contributions (i.e., the CCSD(T) correction can be considered to be repulsive or destabilizing) as anticipated earlier.[18] Whereas for H-bonded structures the CCSD(T) and MP2 stabilization energies were practically identical, these energies were very different for stacked structures. The repulsive CCSD(T) corrections were in some cases as large as several kcal/mol and quite variable.[19,20,24]

The recent studies evaluated only a limited number of geometries corresponding to gas-phase stacking energy minima.[19,20,24] In the present contribution, we for the first time analyzed the behavior of the MP2 CBS extrapolation and CCSD(T) correction over a large part of the potential energy surface (PES). Note that stacking geometries actually occurring in nucleic acids do not correspond to the minima of the isolated base dimers. This extension of reference calculations is of key importance for the overall assessment of the nature of stacking forces and the verification of other simplified models of stacking. The new reference values are carefully compared with the AMBER empirical force field to highlight the nature of base stacking and to estimate the accuracy of force fields used for molecular modeling.

2. Methods

Selection of Geometries. Figure 1 shows the structures of the cytosine dimer considered (graphics produced by POSMOL[25]).

Structures 1−14 were carefully selected to map the conformational space of the stacked dimer and are composed of rigid monomers (isolated planar cytosines optimized at the MP2/6-31G** level). The two cytosines initially coincide with each other and are placed in the *xy* plane of the coordinate system. Their centers of mass coincide with the origin of the coordinate system, and the C1−H1 bonds are parallel to the *y* axis. Then stuctures 1−13 are constructed using four parameters: vertical separation along the *z* axis, twisting of the upper base along its normal passing through its center of mass, and horizontal shifts of the upper base along the *x* and *y* axes. The vertical separation is 3.4 Å in case of the most repulsive structure (1) and 3.3 Å

for the remaining structures. This corresponds to the vertical distance of stacked bases in DNA and RNA. Note that extended stacked aromatic systems have always essentially optimized the vertical intermonomer separation because any significant vertical compression or extension would be associated with very large energy gradients.[26]

Structures 1−4 map the twist dependence of the stacking for undisplaced dimers, and structures 5−7 and 8−13 show the mutual displacement of parallel and antiparallel cytosine dimers. The intermolecular separation of the centers of mass of both subsystems varies from 3.3 Å in structures 2−4 to 4.35 Å in structures 7 and 12. The orientation of subsystem dipoles is parallel in structures 1, 5, and 6 and antiparallel in structures 4 and 8−13. Structures 7, 12, and 13 were selected to test geometries with negligible overlap of the aromatic rings and interactions between the polar exocyclic groups and the aromatic system. Note that the interactions between polar exocyclic groups and the aromatic rings were suggested to provide extra stabilization; however, no such effect was detected with the MP2/6-31G*(0.25) reference calculations.[7]

Whereas structures 1−13 represent "face-to-back" dimers, structure 14 represents the "face-to-face" arrangement and was obtained by flipping the upper cytosine of structure 1 along its C1−H1 bond (with a vertical separation of 3.3 Å).

Structures 15 and 16 correspond to the optimal H-bonded and face-to-back stacked C···C structures derived via full geometry optimization (thus including the monomer relaxation) performed at the resolution of the identity[27] MP2 (RI-MP2) level using the cc-pVTZ basis set. Complex 15 represents the only neutral C···C H-bonded base pair that can be embedded into nucleic acids. Structure 16 differs from rigid monomer dimer 4 only by a small deformation and by a small change in the intermolecular distance. (This structure is therefore not included in Figure 1.) It is not guaranteed that structure 16 is the global stacking minimum but it certainly belongs to the most stable ones, which is entirely sufficient for the purpose of this study.

In Figure 1, two additional geometries were derived from the stacked optimum geometry by contracting and elongating the intermolecular distance by 0.25 and 0.5 Å, respectively (structures 17 and 18), and geometries 19, 20, and 21 correspond to optimal H-bonded structure with the intermolecular distance compressed by 0.15 Å and extended by 0.35 and 0.7 Å, respectively. Geometries 17−21 are obviously not included in Figure 1. All geometries are listed in the Supporting Information section.

Complete Basis Set Limit of the MP2 Stabilization Energies. The RI-MP2 calculations used for extrapolation to the CBS limit were performed with the standard aug-cc-pVDZ and aug-cc-pVTZ basis sets of atomic orbitals. In the RI approximation, the calculation speedup (only about 10% of the time of standard MP2 is necessary) is achieved by an approximate evaluation of four-index, two-electron integrals. According to our previous study on adenine...thymine and guanine...cytosine, H-bonded and stacked complexes' interaction energies differ from standard MP2 data by at most 0.03 kcal/mol.[28] Because no extra integral neglects are introduced (only the integrals are evaluated in an approximate way), the RI-MP2 method is highly preferable for intermolecular interaction energy calculations over the standard MP2 variant. For all RI-MP2 calculations, the TurboMole 5.6[29] program was used.

The HF interaction energy converges with respect to the one-electron basis setfor relatively small basis sets, but the correlation part of the interaction energy converges to its complete basis set limit much more slowly. Several extrapolation schemes have been suggested in the literature to correct the computed results for basis set incompleteness error.[30,31] In the present paper, we used the schemes of Helgaker and co-workers:[30]

$$E_X^{\mathrm{HF}} = E^{\mathrm{HF}}{}_{\mathrm{CBS}} + A\mathrm{e}^{-\alpha X} \text{ and } E_{\mathrm{X}}^{\mathrm{corr}} = E^{\mathrm{corr}}{}_{\mathrm{CBS}} + BX^{-3} \quad (1)$$

where E_X and E_{CBS} are energies for the basis set with the largest angular momentum X and for the complete basis set, respectively, and α is the parameter fitted by the authors.[30] More details about the scheme and results for some other DNA base pairs can be found in our previous papers.[20,22,24] We utilized the augmented Dunning correlation-consistent basis sets rather than nonaugmented ones to reduce the extrapolation error. Note that the aug-cc-pVNZ basis set gives stacking interaction energies better than those calculated with the cc-pV(N+1)Z basis set because stacking calculations require one to use diffuse functions.[20] The BSSE counterpoise correction[32] and frozen-core approximation were applied throughout this study. The CBS extrapolation was applied to all calculated energies (i.e., the BSSE correction was also extrapolated).

Correction for Higher-Order Correlation Effects. The difference between CCSD(T) and MP2 interaction energies ($\Delta E^{\mathrm{CCSD(T)}} - \Delta E^{\mathrm{MP2}}$) exhibits only a small basis set dependence.[22] Therefore, we utilized the 6-31G*(0.25) basis set for this purpose. We studied the CCSD(T) − MP2 difference for DNA base pairs using various basis sets and found that the 6-31G*(0.25) basis set provides reasonable results.[20]

Note that no significantly larger basis set could be utilized with the present computer resources for those structures lacking symmetry. Thus, we approximate the CBS CCSD(T) interaction energy as follows:

$$\Delta E_{\mathrm{CBS}}{}^{\mathrm{CCSD(T)}} = \Delta E_{\mathrm{CBS}}{}^{\mathrm{MP2}} + (\Delta E^{\mathrm{CCSD(T)}} - \Delta E^{\mathrm{MP2}})|_{6-31\mathrm{G}^*(0.25)} \quad (2)$$

The last term in this equation was computed using the Molpro 2002.6[33] program package.

Empirical Potential Calculations. To characterize the nature of base stacking, we compared the reference QM data with the pair-additive empirical force field. It consisted of a van der Waals term taken from the Cornell et al. force field[34] commonly used for nucleic acid simulations and the Coulombic term with atom-centered point charges. The charges were derived using the electrostatic potential (ESP) fitting for the isolated monomers at the HF/6-31G*, MP2/6-31G*(0.25), and MP2/aug-cc-pVDZ levels.

Abbreviations. The following abbreviations are used below to compact the text and tables. SV, VDZ, and VTZ stand for the MP2 or RI-MP2 interaction energies evaluated with the 6-31G*(0.25), aug-cc-pVDZ, and aug-cc-pVTZ basis sets, respectively. CBS is the MP2 basis set limit, Δ(T) is the difference between CCSD(T) and MP2 data evaluated with the 6-31G*(0.25) basis set; SV(T) is the CCSD(T)/6-31G*(0.25) interaction energy; and CBS(T) is the final CBS estimate with the CCSD(T) correction (eq 2 above). AMB-HF and AMB-MP2 stand for the force field data obtained with HF/6-31G* and MP2/aug-cc-pVDZ ESP charges. Note that the AMB-HF force field is essentially identical to the standard Cornell et al. force[34] field except for being derived for the isolated base rather than for the base-backbone segment.

3. Results and Discussion

Table 1 summarizes the interaction energies of all complexes. The first four columns show the MP2-level data (i.e., the SV,

TABLE 1: Energy Characteristics (kcal/mol) for the Investigated Structures

	MP2[b]				CCSD(T)[c]		
structure[a]	SV[d]	VDZ[e]	VTZ[f]	CBS	Δ(T)[g]	SV(T)[d]	CBS(T)[h]
1	2.20	1.56	0.87	0.58	1.87	4.07	2.45
2	−3.10	−4.48	−5.14	−5.40	1.55	−1.55	−3.84
3	−7.19	−9.05	−9.72	−9.99	1.11	−6.08	−8.88
4	−8.26	−10.15	−10.78	−11.03	1.11	−7.15	−9.93
5	0.24	−0.60	−1.26	−1.52	1.84	2.08	0.32
6	0.52	−0.27	−0.95	−1.24	1.88	2.40	0.64
7	−0.52	−1.46	−1.83	−1.98	1.00	0.48	−0.98
8	−7.26	−9.07	−9.58	−9.81	0.71	−6.55	−9.10
9	−7.58	−9.38	−10.03	−10.31	1.20	−6.38	−9.11
10	−6.63	−8.42	−9.11	−9.39	1.12	−5.51	−8.27
11	−7.59	−9.44	−10.05	−10.29	0.86	−6.73	−9.43
12	−5.47	−7.20	−7.45	−7.55	0.12	−5.35	−7.43
13	−7.38	−8.45	−9.37	−9.71	0.91	−6.47	−8.80
14	−7.04	−9.02	−9.71	−10.0	0.89	−6.15	−9.11
15		−20.8			0.19		
16		−11.9			1.36		
17		−8.4			1.99		
18		−11.6			0.89		
19		−18.7			0.53		
20		−19.2			−0.18		
21		−15.1			−0.26		

[a] Compare with Figure 1 and supplemetary material; structures 17 and 18 were derived from structure 16 by contracting and elongating the vertical distance 0.25 and 0.50 Å, respectively. Structures 19, 20, and 21 were derived from the optimal H-bonded structure (15) by contracting the intermolecular bonds 0.15 Å and elongating these bonds 0.35 and 0.70 Å. [b] MP2 interaction energy. [c] Higher-order electron correlation effects are included. [d] 6-31G*(0.25) basis set. [e] aug-cc-pVDZ. [f] aug-cc-pVTZ. [g] Difference between CCSD(T) and MP2 interaction energies evaluated with the 6-31G*(0.25) basis set. [h] MP2 CBS stacking energy corrected for the CCSD(T) term with the 6-31G*(0.25) basis set.

VDZ, VTZ, and CBS values; see above for abbreviations). The last three columns show the CCSD(T) correction, the CCSD-(T)/6-31G*(0.25) interaction energy, and the final estimate composed of the MP2 CBS value and CCSD(T) correction.

Table 2 summarizes the differences between selected QM methods. Of special interest is to compare the MP2/6-31G*-(0.25) (SV) values with the final CBS(T) estimates because the MP2/6-31G*(0.25) level was the leading reference method in studies of aromatic stacking in the past decade.[4−17] It is unlikely that all of the available MP2/6-31G*(0.25) data will soon be reevaluated with better methods, especially for extended systems.

Let us first analyze the MP2 data. For reference structures 1−14, the 6-31G*(0.25) data underestimate the CBS MP2 limit on average by 2.3 kcal/mol. The individual CBS − SV differences are in the range of −0.72 to − 2.96 kcal/mol. Although the extrapolation to the CBS limit affects the absolute values of the stabilization energies rather significantly, the relative MP2/6-31G*(0.25) order of stability of the individual structures remains unchanged. However, the difference is far from being uniform, and it affects different parts of the PES surface in distinct ways. The CBS calculations obviously have a larger effect on structures with a significant geometrical overlap of the cytosines. However (and this is surprising), the effect of the CBS extrapolation is considerably larger for the antiparallel dimers than for the parallel ones. In other words, the CBS calculation increases the energy gap between parallel and antiparallel structures.

The average difference in the aug-cc-pVDZ results compared to the CBS data is 0.9 kcal/mol, and the aug-cc-pVTZ values are on average 0.3 kcal/mol below the CBS limit. This indicates that the results are swiftly converging with the aug-cc-pVTZ basis set.

The $\Delta(T) = \Delta E^{CCSD(T)} - \Delta E^{MP2}$ correction term is repulsive for stacked structures 1−14, being in the range of 0.12 to 1.88 kcal/mol. Its value for the gradient-optimized stacked structure is 1.3 kcal/mol. The Δ(T) corrections for gradient-optimized stacked adenine...thymine, 9-methyladenine...1-methylhymine, guanine...cytosine, and 9-methylguanine...1-methylthymine dimers are 2.8, 3.5, 1.9, and 2.5 kcal/mol, respectively,[24] which are above the value of any of the 14 cytosine dimers studied here. The corrections for the uracil dimer and the guanine dimer (rigid monomers, face-to-back orientation, antiparallel undisplaced geometry) are 0.9 and 1.9 kcal/mol, respectively.[20] The correction term for the face-to-face gradient-optimized uracil dimer is 1.4 kcal/mol.[19] This data survey clearly shows that the Δ(T) term for optimal stacked dimers is highly composition-dependent. Nevertheless, in general the Δ(T) correction tends to increase with the size of the system studied and is modestly enhanced when the gradient optimization is applied. The later effect may well be caused by reduced intermonomer separation (caused by uncorrected BSSE during the optimization) and an increase in the effective mutual overlap of bases because the bases are markedly nonplanar in gradient-optimized gas-phase structures.[35] For a given dimer, however, the correction also evidently depends on the geometry.

The data for cytosine dimers 1−14 also show that the Δ(T) term is greatly reduced for structures with a minimized overlap of the aromatic rings (represented here by structure 12, +0.12 kcal/mol). The Δ(T) correction is also evidently larger for parallel structures (1, 5, 6) compared to that for the antiparallel ones (4, 8−11). This is interesting because this term, similar to the CBS extrapolation, increases the energy difference between the parallel and antiparallel structures compared to the MP2/6-31G*(0.25) level. As a result, the energy difference between structures 4 and 1 deepens from −10.46 at the MP2/6-31G*-(0.25) level to −12.38 kcal/mol at the CBS(T) level, which means a −1.92 kcal/mol or 18% shift. Although the average difference between the new (CBS(T)) and old (SV) reference data is only −1.2 kcal/mol, the range of the difference is from +0.25 to −2.07 kcal/mol. For parallel dimers, the CBS(T) method actually shifts the stacking energy to marginally more positive values compared to the values obtained from the MP2/6-31G*(0.25) method, which was entirely unanticipated.

Taking all of the results together, we suggest that the CBS extrapolation and CCSD(T) correction do not qualitatively change the picture (nature) of base stacking as revealed by the MP2/6-31G*(0.25) data. For most applications, the later method appears to be sufficiently accurate and almost always modestly underestimates the base-stacking stabilization. However, there are rather significant quantitative changes depending on the composition of the dimer and its geometry, and no unifying correction term can be proposed for the MP2/6-31G*(0.25) data. Substantially closer to the CBS(T) data is the RI-MP2/aug-cc-pVDZ level (VDZ). Taking into consideration the fact that the cost of the MP2/6-31G*(0.25) calculation is similar to that of the RI-MP2/aug-cc-pVDZ calculation, the RI-MP2/aug-cc-pVDZ method is well suited to be used for a medium-cost analysis of aromatic stacking instead of the MP2/6-31G*(0.25) method.

Besides the systematic search of the stacked cytosine dimer PES, we carried out some additional calculations mainly to see the behavior of the CCSD(T) correction term. Structures 15 and 16 correspond to gradient-optimized C···C H-bonded and stacked structures. The Δ(T) is negligible for H-bonded base

TABLE 2: Stacking-Energy Difference (kcal/mol) Calculated between Selected Quantum Chemical Methods[a]

structure	CBS-SV	CBS(T)-SV	CBS(T)-VDZ	structure	CBS-SV	CBS(T)-SV	CBS(T)-VDZ
1	−1.62	+0.25	+0.89	8	−2.55	−1.84	−0.03
2	−2.30	−0.74	+0.64	9	−2.73	−1.53	+0.27
3	−2.80	−1.69	+0.17	10	−2.76	−1.64	+0.15
4	−2.77	−1.67	+0.22	11	−2.70	−1.84	+0.01
5	−1.28	+0.08	+0.92	12	−2.08	−1.96	−0.23
6	−0.72	+0.12	+0.91	13	−2.33	−1.42	−0.35
7	−1.46	−0.46	+0.48	14	−2.96	−2.07	−0.09

[a] CBS, CBS(T), SV, and VDZ stand for MP2 CBS extrapolation, MP2 CBS extrapolation with the CCSD(T) correction, and MP2/6-31G*(0.25) and RI-MP2/aug-cc-pVDZ values, respectively.

pair 15 and is essential for stacked dimer 16; this appears to be common behavior.[24] Structures 17−21 test the variation of the intermonomer separation for stacked and H-bonded C···C complexes. A reduction of the intermolecular distance from the optimal structure leads generally to an increase in the correction term; this is true for stacked as well as planar H-bonded systems (cf. structures 16 vs 17 and 15 vs 19).

However, on passing from the optimal intermolecular separation to the larger distances, the value of the correction term is decreasing and is even becoming negative. This conclusion is valid not only for planar H-bonded structures (cf. structures 15, 20, and 21) but also for stacked complexes (cf. structures 12, 16, and 18). We can thus conclude this section by stating that in the short-range (repulsive) parts of the PES the Δ(T) correction makes the PES more repulsive (in comparison with the MP2 PES) than in the energy minimum and in the long-range part of the PES the effect is far smaller or even reversed. As a result, the optimal vertical separation may differ at the MP2 and CCSD(T) levels, and this could explain why, in preceding studies, the vertical interbase separation appeared to be a little underestimated.[6,7] However, more data will be needed to clarify this issue.

Empirical Force Field. Table 3 compares the new reference CBS(T) data with the force field. Such a comparison is important for two reasons. First, it gives us an idea of how accurately base-stacking forces are described in the course of molecular modeling and MD studies. Second, the force field consists of just an isotropic Lennard-Jones term and a simple Coulombic term. Thus, the analysis of the difference patterns helps to reveal the presence of additional contributions (such as polarization) not included in the force field.

The Table confirms that the AMBER force field on average well reproduces the magnitude of the stacking energy. There are, however, interesting differences. The stacking energy of parallel structure 1 is well reproduced with the HF charges, but the utilization of MP2 charges leads to overstabilization by −1.5 kcal/mol. This can be explained by two effects or their combination. First, the error coming from the noninclusion of the polarization of cytosine monomers in the force field cancels with the error coming from the exaggeration of the polarity of the cytosine charge distribution due to the use of HF charges. Second, the difference is caused by anisotropic short-range repulsion that is maximized for parallel undisplaced structure 1.[6] Then the agreement with HF charges would be due to a compensation of errors and the disagreement of the AMB-MP2 values due to the deficiency of the isotropic Lennard-Jones term.

Interestingly, the AMB-HF method performs better than its MP2 variant when investigating the twist dependence (structures 1−4) regarding both the absolute and relative energies. On the other side, the AMB-HF performance worsens for parallel displaced structures. We believe that this pattern of difference can be best explained as a combination of both polarization and short-range effects. This is fully supported when comparing

TABLE 3: Stacking Energies (kcal/mol) Calculated by ab Initio Data and the AMBER Force Field[a]

structure	VS	twist	*x* shift	*y* shift	CBS(T)	AMB-HF	AMB-MP2
1	3.4	0	0	0	2.45	2.86	**0.94**
2	3.3	60	0	0	−3.84	−3.18	−3.80
3	3.3	120	0	0	−8.88	−8.72	**−7.86**
4	3.3	180	0	0	−9.93	−9.88	**−8.59**
5	3.3	0	1[b]	0	0.32	**1.47**	−0.14
6	3.3	0	0	1[b]	0.64	**2.11**	0.36
7	3.3	0	2	−2	−0.98	**0.11**	−1.00
8	3.3	180	1	0	−9.10	−8.48	**−7.42**
9	3.3	180	−1	0	−9.11	**−10.30**	−8.91
10	3.3	180	0	1	−8.27	−8.82	−7.90
11	3.3	180	0	−1	−9.43	**−10.45**	−8.87
12	3.3	180	2	−2	−7.43	−7.39	**−6.23**
13	3.3	180	0	2	−8.80	−8.36	**−7.66**
14	3.3	f...f	0	0	−9.11	−9.70	**−7.81**
difference[c]							
4−1	0.1	180	0	0	−12.38	−12.74	**−9.53 (2.85)**
5−1	0.1	0	1	0	−2.13	−1.39	**−1.08**
6−1	0.1	0	0	1	−1.81	**−0.75**	**−0.58**
9−8	0	0	−2	0	−0.01	**−1.82**	**−1.49**
11−8	0	0	−1	−1	−0.33	**−1.97**	**−1.45**
4−5	0	180	−1	0	−10.25	**−11.35**	**−8.45**
4−6	0	180	0	−1	−10.57	**−11.99**	**−8.95**

[a] Vertical separation (Å), twist (deg), *x* shift (Å), and *y* shift (Å) are the geometrical parameters of the dimers. CBS(T) is the final QM estimate (Table 1), and AMB-HF and AMB-MP2 are the force field values with HF and MP2 charges. Differences of 1.0−1.6 kcal/mol are shown in bold, those within 1.6−2.0 kcal/mol interval are underlined, and differences >2.0 kcal/mol are listed in parentheses explicitly. [b] Note that for parallel dimers structures with shits of −1 Å would be identical. [c] Energy difference calculated between the structures indicated.

parallel-displaced structures 5 and 6 (where one can assume that the anomalous short-range effects are reduced; see Figure 1) with antiparallel structure 4. Here the AMB-HF force field exaggerates the difference, and the AMB-MP2 force field underestimates the difference (last two rows of Table 3). This suggests that the mutual cytosine polarization effect is not completely canceled by the errors due to the HF charge distribution used.

The most interesting region, however, is that between quite similar antiparallel structures 8, 9, and 11. Quantum chemistry predicts that all three are isoenergetic whereas the force field destabilizes structure 8 by 1.5 (MP2 charges) and 1.9 (HF charges) kcal/mol. This difference cannot be explained by the electrostatic term and must be attributed to some other contribution neglected by the force field. Further analysis will be necessary to pinpoint its exact origin.

In summary, the force field well reproduces the ab initio data; however, there are modest contributions not captured by the force field that may affect the conformational analysis of nucleic acids. It is instructive to compare the CBS(T) data with the preceding reference MP2/6-31G*(0.25) ab initio-level data. The MP2/6-31G*(0.25) data are in a better agreement with the AMB-

MP2 force field (not shown in detail). Note that the MP2/6-31G*(0.25) and MP2/aug-cc-pVDZ charge distributions provide interaction energies identical within 0.3 kcal/mol and are thus close to equivalent. However, the improved CBS(T) potential energy surface is less smooth and is closer to the AMB-HF force field. This is an important observation that deserves further analysis for other stacked dimers.

4. Conclusions

(i) Extrapolation to the CBS limit is essential not only in the potential energy minimum but also in other areas of the PES. Although the absolute values of the MP2 stabilization energy are significantly affected by extrapolation, the relative order of stability for various structures is correct even with the smallest 6-31G*(0.25) basis sets. Nevertheless, the balance between various parts of the PES is modestly changed.

(ii) The CCSD(T) correction term is considerably repulsive for the optimized stacked structure, but that for the optimized H-bonded structure is negligible. This fully agrees with previous data on other DNA base pairs. Regarding the other parts of the PES, the CCSD(T) correction term is repulsive for stacked structures with a large overlap of the monomers but is not uniform. This term tends to vanish for structures with minimal aromatic ring overlap. Thus, the CCSD(T) term on one side partially cancels the energy gains by the MP2 CBS extrapolation; however, it also further enhances the relative differences mainly between parallel and antiparallel structures. As a result, the energy difference between the old MP2/6-31G*(02.5) and new reference CBS(T) data ranges from +0.3 to −2.1 kcal/mol, and no unique scaling factor exists. Although the new results do not change our view on the nature of base stacking, the differences are large enough to affect quantitative modeling.

(iii) The AMBER force field on average reproduces the ab initio data very well; however, there also are occasional nonnegligible differences due to short-range and polarization contributions not captured by the force field. Interestingly, the new reference data are somewhat better reproduced when considering HF charges instead of the MP2 ones in the force field calculations.

Acknowledgment. This paper was supported by project LN00A032 (Center for Complex Molecular Systems and Biomolecules) by the Ministry of Education of the Czech Republic and partially by Wellcome Trust International Senior Research Fellowship in Biomedical Science in Central Europe GR067507 (J.S.).

Supporting Information Available: Geometries of all of the studied systems. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) (a) Mathews, D. H.; Sabina, J.; Zuker, M.; Turner, D. H. *J. Mol. Biol.* **1999**, *288*, 911. (b) SantaLucia, J. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1460.

(2) Yanagi, K.; Prive, G. G.; Dickerson, R. E. *J. Mol. Biol.* **1991**, *217*, 201.

(3) (a) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149. (b) Nir, E.; Plutzer, C.; Kleinermanns, K.; de Vries, M. *Eur. Phys. J. D* **2002**, *20*, 317.

(4) Luo, R.; Gilson, H. S. R.; Potter, M. J.; Gilson, M. K. *Biophys. J.* **2001**, *80*, 140.

(5) Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247.

(6) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Comput. Chem.* **1996**, *17*, 841.

(7) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590.

(8) (a) Šponer, J.; Gabb, H. A.; Leszczynski, J.; Hobza, P. *Biophys. J.* **1997**, *73*, 76. (b) Šponer, J.; Florian, J.; Ng, H. L.; Šponer, J. E.; Špačková, N.; *Nucleic Acids Res.* **2000**, *28*, 4893.

(9) (a) Šponer, J.; Leszczynski, J.; Vetterl, V.; Hobza, P. *J. Biomol. Struct. Dyn.* **1996**, *13*, 695. (b) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. A* **1997**, *101*, 9489. (c) Cysewski, P.; Czyznikowska-Balcerak, Z.; Szefler, B. *Pol. J. Chem.* **2003**, *77*, 1287.

(10) Florian, J.; Šponer, J.; Warshel, A. *J. Phys. Chem. B* **1999**, *103*, 884.

(11) Kroonbatenburg, L. M. J.; VanDuijneveldt, F. B. *Theochem. J. Mol. Struct.* **1985**, *22*, 185.

(12) (a) Bondarev, D. A.; Skawinski, W. J.; Venanzi, C. A. *J. Phys. Chem. B* **2000**, *104*, 815. (b) Řeha, D.; Kabeláč, M.; Ryjáček, F.; Šponer, J.; Sponer J. E.; Elstner, M.; Suhai, S.; Hobza, P. *J. Am. Chem. Soc.* **2002**, *124*, 3366.

(13) (a) Wintjens, R.; Biot, C.; Rooman, M.; Lievin, J. *J. Phys. Chem. A* **2003**, *107*, 6249. (b) Biot, C.; Buisine, E.; Kwasigroch, J. M.; Wintjens, R.; Rooman, M. *J. Biol. Chem.* **2002**, *277*, 40816. (c) Wintjens, R.; Lievin, J.; Rooman, M.; Buisine, E. *J. Mol. Biol.* **2000**, *302*, 395.

(14) (a) Gervasio, F. L.; Chelli, R.; Procacci, P.; Schettino, V. *Proteins* **2002**, *48*, 117. (b) Gervasio, F. L.; Procacci, P.; Cardini, G.; Guarna, A.; Giolitti, A.; Schettino, V. *J. Phys. Chem. B* **2000**, *104*, 1108.

(15) Mao, L. S.; Wang, Y. L.; Hu, X. C. *J. Phys. Chem. B* **2003**, *107*, 3963.

(16) Lee, N. K.; Park, S.; Kim, S. K. *J. Chem. Phys.* **2002**, *116*, 7910.

(17) (a) Rauwolf, C.; Mehlhorn, A.; Fabian, J. *Collect. Czech. Chem. Commun.* **1998**, *63*, 1223. (b) Alagona, G.; Ghio, C.; Monti, S. *Int. J. Quantum Chem.* **2001**, *83*, 128.

(18) Šponer, J.; Hobza, P. *Chem. Phys. Lett.* **1997**, *267*, 263.

(19) Leininger, M. L.; Nielsen, I. M. B.; Colvin, M. E.; Janssen, C. L. *J. Phys. Chem. A* **2002**, *106*, 3850.

(20) Hobza, P.; Šponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802.

(21) (a) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Phys. Chem. A* **2002**, *106*, 4423. (b) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Phys. Chem. A* **1999**, *103*, 8265. (c) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2000**, *122*, 3746.

(22) Jurečka, P.; Hobza, P. *Chem. Phys. Lett.* **2002**, *365*, 89.

(23) Braun, J.; Neusser, H. J.; Hobza, P. *J. Phys. Chem. A* **2003**, *107*, 3918.

(24) Jurečka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608.

(25) Lee, S. J.; Kim, K. S. *POSMOL Molecular Viewer*; Pohang University of Science and Technology:Pohang, Korea, 1999.

(26) Šponer, J.; Kypr, J. *J. Biomol. Struct. Dyn.* **1993**, *11*, 277.

(27) (a) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359. (b) Vahtras, O.; Almlöf, J.; Feyereisen, M. *Chem. Phys. Lett.* **1993**, *213*, 514. (c) Bernholdt, D. E.; Harrison, R. J. *Chem. Phys. Lett.* **1996**, *250*, 470.

(28) Jurečka, P.; Nachtigall, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4578.

(29) (a) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165. (b) Weigend, F.; Häser, M. *Theor. Chem. Acc.* **1997**, *97*, 331.

(30) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243.

(31) Truhlar, D. G. *Chem. Phys. Lett.* **1998**, *294*, 45.

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

(33) Amos, R. D.; Bernhardsson, A.; Berning, A.; Celani, P.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Knowles, P. J.; Korona, T.; Lindh, R.; Lloyd, A. W.; McNicholas, S. J.; Manby, F. R.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, R.; Rauhut, G.; Schütz, M.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, A.; Thorsteinsson, T.; Werner, H.-J. *MOLPRO*, version 2002.6 (a package of ab initio programs designed by H.-J. Werner and P. J. Knowles).

(34) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(35) Hobza, P.; Šponer, J. *Chem. Phys. Lett.* **1998**, *288*, 7.