J|A|C|S
ARTICLES
Published on Web 11/18/2003

True Stabilization Energies for the Optimal Planar Hydrogen-Bonded and Stacked Structures of Guanine···Cytosine, Adenine···Thymine, and Their 9- and 1-Methyl Derivatives: Complete Basis Set Calculations at the MP2 and CCSD(T) Levels and Comparison with Experiment

Petr Jurečka and Pavel Hobza*

Contribution from the J. Heyrovský Institute of Physical Chemistry, Academy of Sciences of the Czech Republic and Center for Complex Molecular Systems and Biomolecules, 182 23 Prague 8, Czech Republic

Received June 11, 2003; E-mail: hobza@indy.jh-inst.cas.cz

Abstract: Planar H-bonded and stacked structures of guanine···cytosine (G···C), adenine···thymine (A···T), 9-methylguanine···1-methylcytosine (mG···mC), and 9-methyladenine···1-methylthymine (mA···mT) were optimized at the RI-MP2 level using the TZVPP ([5s3p2d1f/3s2p1d]) basis set. Planar H-bonded structures of G···C, mG···mC, and A···T correspond to the Watson−Crick (WC) arrangement, in contrast to mA···mT for which the Hoogsteen (H) structure is found. Stabilization energies for all structures were determined as the sum of the complete basis set limit of MP2 energies and a ($\Delta E^{CCSD(T)} - \Delta E^{MP2}$) correction term evaluated with the cc-pVDZ(0.25,0.15) basis set. The complete basis set limit of MP2 energies was determined by two-point extrapolation using the aug-cc-pVXZ basis sets for X = D and T and X = T and Q. This procedure is required since the convergency of the MP2 interaction energy for the present complexes is rather slow, and it is thus important to include the extrapolation to the complete basis set limit. For the MP2/aug-cc-pVQZ level of theory, stabilization energies for all complexes studied are already very close to the complete basis set limit. The much cheaper D→T extrapolation provided a complete basis set limit close (by less than 0.7 kcal/mol) to the more accurate T→Q term, and the D→T extrapolation can be recommended for evaluation of complete basis set limits of more extended complexes (e.g. larger motifs of DNA). The convergency of the ($\Delta E^{CCSD(T)} - \Delta E^{MP2}$) term is known to be faster than that of the MP2 or CCSD(T) correlation energy itself, and the cc-pVDZ(0.25,0.15) basis set provides reasonable values for planar H-bonded as well as stacked structures. Inclusion of the CCSD(T) correction is essential for obtaining reliable relative values for planar H-bonding and stacking interactions; neglecting the CCSD(T) correction results in very considerable errors between 2.5 and 3.4 kcal/mol. Final stabilization energies (kcal/mol) for the base pairs studied are very substantial (A···T WC, 15.4; mA···mT H, 16.3; A···T stacked, 11.6; mA···mT stacked, 13.1; G···C WC, 28.8; mG···mC WC, 28.5; G···C stacked, 16.9; mG···mC stacked, 18.0), much larger than published previously. On the basis of comparison with experimental data, we conclude that our values represent the lower boundary of the true stabilization energies. On the basis of error analysis, we expect the present H-bonding energies to be fairly close to the true values, while stacked energies are still expected to be about 10% too low. The stacking energy for the mG···mC pair is considerably lower than the respective H-bonding energy, but it is larger than the mA···mT H-bonding energy. This conclusion could significantly change the present view on the importance of specific H-bonding interactions and nonspecific stacking interactions in nature, for instance, in DNA. Present stabilization energies for H-bonding and stacking energies represent the most accurate and reliable values and can be considered as new reference data.

Introduction

Noncovalent interactions among nucleic acid (NA) bases are used to assemble various architectures of DNA and RNA, and most important among them are H-bonding (hydrogen-bonding), stacking, electrostatic, and charge-transfer interactions.[1] The relative importance of the first two contributions plays a key role in determining the structure and dynamics, not only of DNA and RNA but also of various complexes of these biomacromolecules with intercalators, minor groove binder, etc.

Originally it was believed that specific H-bonding interactions are much stronger and contribute dominantly to the stability of DNA or RNA, but later, on the basis of advanced quantum chemical calculations,[2] it was realized that nonspecific stacking interactions are much more important than first expected. The

(1) Müller-Dethlefs, K.; Hobza, P. *Chem. Rev.* **2000**, *100*, 143.

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

10.1021/ja036611j CCC: $25.00

relative importance of H-bonding and stacking interactions is important not only for the problem mentioned (which without doubt represents one of the key problems of today's science) but also for a wide variety of problems ranging from gas-phase physicochemical experiments to the condensed-phase experiments and simulations.

The question arises, is there any chance to evaluate directly the role of these interactions on the basis of experimental data? The (probably slightly surprising) answer is no. Despite enormous progress in various experimental techniques, relevant experiments for determining the relative importance of H-bonding and stacking without the presence of other effects are still missing. The only usable data on stabilization energies of NA base pairs in vacuo are the more than 20 years old field-ionization mass spectroscopy data of Sukhodub et al.[3] who determined the stabilization enthalpies of 9-methylguanine···1-methylcytosine, 1-methylcytosine···1-methylcytosine, 9-methyladenine···1-methylthymine, and 1-methylthymine···1-methylthymine complexes in the gas phase. It must be mentioned that these, up to now unique, experiments do not reveal the molecular structure of the complexes studied. Also, due to the fact that these experiments were performed at rather high temperature, several isomers of base pairs should in fact have been present.[4] We further stress that the state-of-the-art gas-phase experiments from the laboratories of de Vries and Kleinermanns provide evidence only about the spectrum of a selected NA base pair but do not allow one to extract information on its structure or stabilization energy.[5] Furthermore, to evaluate the role of H-bonding and stacking, complexes of both types should be simultaneously present and detectable. Thus, presently, the only chance to evaluate the relative importance of the H-bonding and stacking interaction between NA bases comes from quantum chemical ab initio calculations. Evidently, the calculations should be performed at a very high ab initio level excluding problems with e.g. the size of the AO basis set or the portion of correlation energy covered.

In this paper we evaluate for the first time the state-of-the-art stabilization energies and stabilization enthalpies of the most favorable H-bonded and stacked structures of the guanine···cytosine (G···C), adenine···thymine (A···T), 9-methylguanine···1-methylcytosine (mG···mC), and 9-methyladenine···1-methylthymine (mA···mT) NA base pairs. Data obtained for the last two complexes will then be compared with the relevant experiments of Sukhodub et al.[3]

Strategy of Calculations and Its Justification. The potential energy surface (PES) of NA base pairs is very rich and contains a large number of energy minima. To find the global minimum, it is necessary to use some efficient searching procedures. We have shown that molecular dynamics simulations together with the quenching technique (MD/Q) provide for a full description of the PES of various nonmethylated and methylated NA base pairs.[6] Because the simulations rely on the quality of the potential, the predicted energy minima were recalculated[6] at a nonempirical level using the combination of the HF/6-31** and MP2/6-31G*(0.25) techniques. This level is, however, not sufficient for the accurate evaluation of the structure of stacked pairs and a higher correlated level including second polarization functions on all the atoms is required. In the present study structures of all these complexes are determined using the approximate resolution of the identity MP2 (RI-MP2) method[7,8] with the extended TZVPP basis set ([5s3p2d1f/3s2p1d]). The stabilization energies for the minima found for the H-bonded and stacked motifs have then been constructed as the sum of the complete basis set limit of the MP2 energy and a ($\Delta E^{CCSD(T)} - \Delta E^{MP2}$) correction term covering the difference between the MP2 and CCSD(T) stabilization energies. Then, to be able to compare the theoretical results with experimental enthalpies, the zero-point vibration energy (ZPVE) and temperature-dependent enthalpy terms were estimated.

The single step of the procedure described was tested for systems for which accurate experimental data were available (it should be stressed that such studies for extended complexes having more than 24 atoms—the size of the benzene dimer—are very rare). First, the structure of the phenol dimer was determined for which experimental rotational constants were measured. Passing from the MP2/6-31G** to the RI-MP2/TZVPP level results in a significant improvement of the theoretical data, and the absolute average difference between experimental and theoretical rotational constants drops from 4.6 to 1.5%.[9] Second, the stabilization enthalpy of the indole···benzene complex was determined experimentally and evaluated theoretically along the procedure described above. The final agreement between experimental and theoretical energies was very close (~0.1 kcal/mol or less than 5%).[10]

Methods

Structure. All the calculations were carried out using the TURBOMOLE 5.6 program suite[11] with the extended TZVPP basis set ([5s3p2d1f/3s2p1d]) and standard (default) auxiliary basis sets. Recently we explored[12] the applicability of the RI-MP2 method for NA base pairs and larger DNA fragments and have shown that the method is capable of an accurate description of H-bonded and stacked DNA base interactions. The results obtained with the RI-MP2 method differ only marginally from those evaluated with the exact MP2 method, while the time saving is as large as 1 order of magnitude.

Complete Basis Set Limit of the MP2 Stabilization Energies. Whereas the HF interaction energy can be considered as converged with respect to the one-electron basis set already for relatively small basis sets, the correlation part of the interaction energy converges to its complete basis set limit (CBS) unsatisfactorily slow. To correct the computed results for bases set incompleteness error, several extrapolation schemes have been successfully employed in the literature. We have chosen the schemes of Helgaker and co-workers[13]

(3) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149.
(4) Kratochvil, M.; Engkvist, O.; Šponer, J.; Hobza, P. *J. Phys. Chem. A* **1998**, *102*, 6921.
(5) Nir, E.; Kleinermanns, K.; de Vries, M. S. *Nature* **2000**, *408*, 949. Nir, E.; Janzen, C.; Imhof, P.; Kleinermanns, K.; de Vries, M. S. *J. Chem. Phys.* **2001**, *115*, 4604. Nir, E.; Plützer, Ch.; Kleinermanns, K.; de Vries, M. S. *Eur. Phys. J. D* **2002**, *20*, 317. Nir, E.; Janzen, C.; Imhof, P.; Kleinermanns, K.; de Vries, M. S. *Phys. Chem. Chem. Phys.* **2002**, *4*, 732. Nir, E.; Janzen, C.; Imhof, P.; Kleinermanns, K.; de Vries, M. S. *Phys. Chem. Chem. Phys.* **2002**, *4*, 740. Plützer, Ch.; Hünig, P.; Kleinermanns, K. *Phys. Chem. Chem. Phys.* **2003**, *5*, 1158.
(6) Kabeláč, M.; Hobza, P. *J. Phys. Chem. B* **2001**, *105*, 5804.
(7) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359.
(8) Weigend, F.; Häser, M. *Theor. Chem. Acc.* **1997**, *97*, 331.
(9) Hobza, P.; Riehn, Ch.; Weichert, A.; Brutschy, B. *Chem. Phys.* **2002**, *283*, 331.
(10) Braun, J.; Neusser, H.-J.; Hobza, P. *J. Phys. Chem. A* **2003**, *107*, 3918.
(11) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165.
(12) Jurečka, P.; Nachtigall, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4578.
(13) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243.

$$E_X^{HF} = E_{CBS}^{HF} + Ae^{-\alpha X} \quad \text{and} \quad E_X^{corr} = E_{CBS}^{corr} + BX^{-3} \tag{1}$$

and Truhlar:[14]

$$E_X^{HF} = E_{CBS}^{HF} + AX^{-\alpha} \quad \text{and} \quad E_X^{corr} = E_{CBS}^{corr} + BX^{-\beta} \tag{2}$$

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 and β are parameters fitted by the authors.[13,14] These schemes were chosen because (i) both of the approaches extrapolate HF and correlation energy separately and (ii) both use the two-point form (they extrapolate two successive basis sets results). The two-point extrapolation form is preferable because it was shown[13,15] that inclusion of additional (lower quality basis set) points results in extrapolations of lower quality fit, especially when the smallest basis set (cc-pVDZ) is used. We utilized systematically augmented Dunning's basis sets rather than nonaugmented in order to reduce the extrapolation error (note that the aug-cc-pVDZ basis set gives absolute energies as well as interaction energies comparable to those calculated with the TZVPP basis). BSSE counterpoise correction[16] and frozen-core approximation were applied throughout this study. The CBS extrapolation was applied to all calculated energies (dimer, monomers in both monomer- and dimer-centered basis sets, and monomers in vacuo); i.e., both deformation and BSSE corrections were extrapolated.

Correction for Higher Order Correlation Effects. Assuming the difference between CCSD(T) and MP2 interaction energies ($\Delta E^{CCSD(T)} - \Delta E^{MP2}$) exhibits only a small basis set dependence,[17] the CBS CCSD(T) interaction energy can be approximated as

$$\Delta E^{CCSD(T)} = \Delta E^{MP2}\text{CBS} + (\Delta E^{CCSD(T)} - \Delta E^{MP2})|_{\text{small basis set}} \tag{3}$$

Hobza and Šponer[18] investigated the CCSD(T) − MP2 difference for DNA bases and found that the 6-31G*(0.25) basis set already yields a satisfactory value of this difference. In the present paper we use the larger cc-pVDZ(0.25,0.15) basis set to evaluate the ($\Delta E^{CCSD(T)} - \Delta E^{MP2}$) correction term for nonmethylated base pairs and for reasons of computational cost the cc-pVDZ(0.25) basis set for the methylated ones.

Zero-Point Vibration Energy. Zero-point vibrational energies were computed numerically at the MP2/cc-pVDZ level.

CCSD(T) calculations were carried out by the Molpro 2002 suite of programs,[19] while all remaining calculations were performed with the TURBOMOLE 5.6[11] program package.

Results and Discussion

Guanine···Cytosine and 9-Methylguanine···1-Methylcytosine Complexes. MD/Q calculations[6] clearly show that planar H-bonded structures of both complexes correspond to the Watson−Crick (WC) structural motif known from DNA. The other planar H-bonding structure for the nonmethylated base pair is by 3.6 kcal/mol less stable, while other structures are considerably less stable. In the case of the methylated base pair energy difference between the Watson−Crick H-bonded structure and the other planar H-bonded structure is even more pronounced (~9 kcal/mol). On the other hand methylation increases the stability of stacked base pairs. RI-MP2 optimized structures of the Watson−Crick and stacked complexes considered are visualized in Figure 1. The upper part of Table 1 shows the RI-MP2 stabilization energies determined using the TZVPP and the aug-cc-pVXZ (X = D, T, Q) basis sets. Apparently, the TZVPP energies are reliable and are comparable to those evaluated with augmented basis sets. In the case of Watson−Crick complexes the TZVPP values lie between the ones for the aug-cc-pVDZ and aug-cc-pVTZ basis sets (for mG···mC the TZVPP values agree with the ones for aug-cc-pVTZ), while for stacking complexes the TZVPP values are (in absolute values) slightly smaller than the aug-cc-pVDZ results. This finding is important since the geometry of all complexes is evaluated at the TZVPP level. From the table it is further seen that passing from the aug-cc-pVDZ to the aug-cc-pVTZ basis set is connected with a rather large stabilization energy increase (on average, by 1.3 kcal/mol), while passing from aug-cc-pVTZ to aug-cc-pVQZ reduces the respective difference to about half (0.6 kcal/mol). From the aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ stabilization energies, it is, however, evident that they are still not converged and extrapolation to the complete basis set limit is essential for obtaining accurate values.

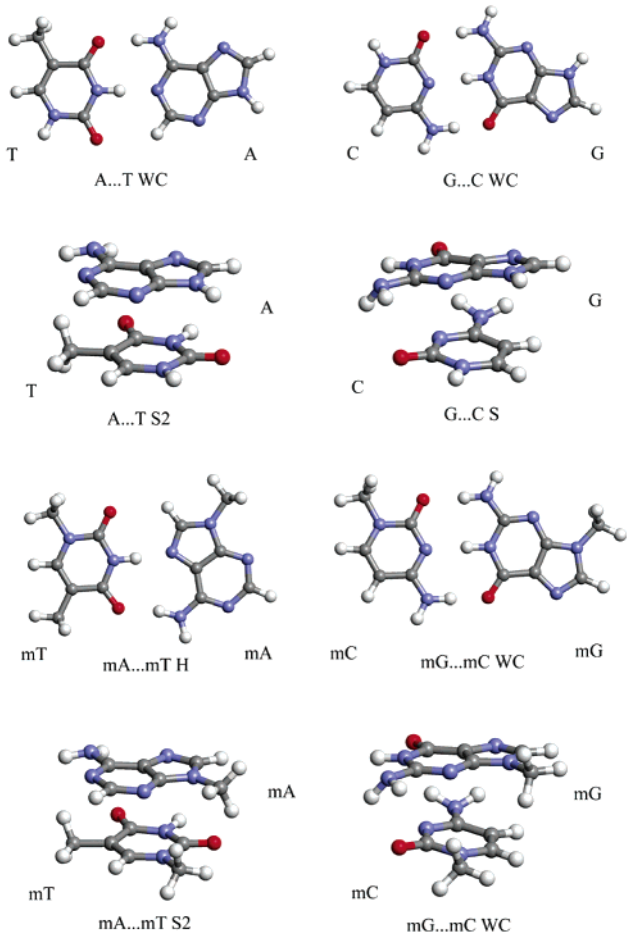

Figure 1. Structures of the nucleic acid base pairs (G, guanine; mG, 9-methylguanine; C, cytosine; mC, 1-methylcytosine; A, adenine; mA, 9-methyladenine; T, thymine; and mT, 1-methylthymine).

Two schemes were used for extrapolation to the CBS limit, and the respective values differ slightly. Comparing the D→T extrapolations, we found that both Truhlar and Helgaker

(14) Truhlar, D. G. *Chem. Phys. Lett.* **1998**, *294*, 45.
(15) Lara-Castells, M. P.; Krems, R. V.; Buchachenko, A. A.; Delgado-Barrio, G.; Villarreal, P. *J. Chem. Phys.* **2001**, *115*, 10438.
(16) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
(17) Jurečka, P.; Hobza, P. *Chem. Phys. Lett.* **2002**, *365*, 89.
(18) Hobza, P.; Šponer, J. *J. Am. Chem. Soc.* **2002**, *124*, 11802.
(19) *MOLPRO* is a package of ab initio programs written by H.-J. Werner and P. J. Knowles, with contributions from J. Almlof, R. D. Amos, A. Berning, P. Celani, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, S. T. Elbert, C. Hampel, G. Hetzer, R. Lindh, A. W. Lloyd, W. Meyer, A. Nicklass, T. Korona, K. Peterson, R. Pitzer, G. Rauhut, A. J. Stone, P. R. Taylor, M. E. Mura, P. Pulay, M. Schutz, H. Stoll, and T. Thorsteinsson.

Table 1. Interaction Energies of Methylated and Nonmethylated A···T and G···C DNA Base Pairs

method/basis set	A···T WC	mA···mT H	A···T S2	mA···mT S2	G···C WC	mG···mC WC	G···C S	mG···mC S
RI-MP2/TZVPP	−14.3	−14.8	−12.1	−14.4	−25.8	−25.6	−16.3	−17.7
RI-MP2/aDZ[a]	−13.8	−15.2	−12.8	−14.9	−25.6	−25.4	−16.9	−18.3
RI-MP2/aTZ[a]	−14.7	−15.9	−13.8	−16.2	−27.0	−26.8	−18.1	−19.6
RI-MP2/aQZ[a]	−15.1	−16.2	−14.1	−16.4	−27.7	−27.5	−18.5	−20.2
D→T	−15.0	−16.2	−14.3	−16.8	−27.5	−27.4	−18.6	−20.1
T→Q	−15.4	−16.4	−14.4	−16.6	−28.2	−27.9	−18.8	−20.5
D→T Truhlar	−15.3	−16.4	−14.6	−17.3	−27.9	−27.7	−19.0	−20.6
MP2→CCSD(T)	0.0	0.1	2.8	3.5	−0.6	−0.6	1.9	2.5
$\Delta E_{tot.}$	−15.4	−16.3	−11.6	−13.1	−28.8	−28.5	−16.9	−18.0
ΔH_0°	−14.6	−15.8	−11.0	−12.2	−27.5	−27.2	−15.7	−16.8
ΔH_0^T	-14.0_{323K}	-15.1_{323K}	-10.3_{323K}	-11.9_{323K}	-27.0_{381K}	-27.0_{381K}	-15.0_{381K}	-16.5_{381K}

[a] aug-cc-pVXZ basis sets (X = D,T,Q) were abbreviated as aXZ.

schemes provide the same maximum error (0.7 kcal/mol in comparison with T→Q values), whereas the mean error is smaller for the Truhlar extrapolation (we would like to mention that the Truhlar exponents are optimized for cc-pVDZ and cc-pVTZ basis sets and not for their augmented modifications as used here). On the other hand the Helgaker D→T extrapolation gives an excellent agreement (difference smaller than 0.2 kcal/mol) with the results for the aug-cc-pVQZ basis set. In the present study we will further use the T→Q extrapolations on the basis of the Helgaker scheme. The difference between the CBS limit and the aug-cc-pVQZ values is definitively not negligible and amounts to about 0.3 kcal/mol. Much larger differences (1.1 and 2.6 kcal/mol) result when the CBS limit is compared with the aug-cc-pVTZ or aug-cc-pVDZ values. This supports the idea that the aug-cc-pVTZ level represents the first level for obtaining reliable data. It is, however, evident that even in this case the inclusion of the D→T extrapolation is essential. If any lower level stabilization energies would agree with an experimental value, then it would likely be a consequence of a compensation of theoretical or experimental errors (e.g. neglect of higher order correlation energy contributions or inaccurate consideration of some structural type only).

In the case of stacked structures the MD/Q calculations predicted only a few stacked structures with very similar stabilization energies, and the most stable ones for the nonmethylated and methylated pair (cf. Figure 1) were considered. In the case of stacking, extrapolation to the CBS limit is equally important as in the case of H-bonded complexes and, for nonmethylated and methylated pairs, results in differences of 0.3, 0.7, and 1.9 kcal/mol and 0.3, 0.9, and 2.2 kcal/mol with respect to the aug-cc-pVQZ, aug-cc-pVTZ, and aug-cc-pVDZ data. The aug-cc-pVDZ stacking energies are thus again lower than the respective CBS limit by more than 10%. Finally, we stress one important difference compared to planar H-bonded pairs. While the stabilization energies of planar nonmethylated and methylated pairs were comparable (the latter pair was even surprisingly slighly less stable), in the case of stacked pairs the methylation increases the stability significantly, by about 1.5 kcal/mol (or by slightly less than 10%). The ratio of CBS limits of H-bonded and stacked structures is 1.5 and 1.4 for nonmethylated and methylated systems, which means that methylation slightly reduces the difference between H-bonding and stacking.

The ($\Delta E^{CCSD(T)} - \Delta E^{MP2}$) correction term for H-bonded structures is slightly negative which implies that the CCSD(T) stabilization energy is larger than the MP2 one. The difference is, however, small and supports the known fact[20] that in the case of H-bonded systems both stabilization energies are practically identical. The situation with stacking is, however, different, and the CCSD(T) stabilization energy is now found to be significantly smaller (by 1.9 and 2.5 kcal/mol for the nonmethylated and the methylated pairs, respectively). One important consequence immediately follows—not considering the ($\Delta E^{CCSD(T)} - \Delta E^{MP2}$) correction term will produce a large error (2.5 and 3.1 kcal/mol) for the relative stabilities of H-bonded and stacked structures of nonmethylated and methylated GC pairs.

Adenine···Thymine and 9-Methyladenine···1-Methylthymine Complexes. For the adenine···thymine base pair, MD/Q calculations predicted and ab initio calculations confirmed that the global minimum possesses the (3192) structure[6] (H-bonds between N3 and N9 of adenine and N1 and O2 of thymine). The Watson−Crick and Hoogsteen motifs known from DNA and RNA were found to be less stable by as much as 2.8 and 4.1 kcal/mol, respectively (with respect to the 3192 structure). The situation with the methylated pairs is different, and now the Hoogsteen, reversed Hoogsteen, and two stacked structures are almost equally stable, while the Watson−Crick structure is by about 1.5 kcal/mol less stable. In the case of the methylated systems we consider also the planar Hoogsteen structure, while for nonmethylated systems we take into consideration the planar Watson−Crick structure with the aim being to mimic the situation in DNA. The RI-MP2 structures of both H-bonded complexes are visualized in Figure 1, and Table 1 contains the respective energy data. In this case we cannot compare the effect of methylation since we investigate different structures. Similarly as in the case of guanine···cytosine complexes the RI-MP2/TZVPP stabilization energy agrees well with that evaluated with augmented basis sets. Very similar effects were found also when passing from the aug-cc-pVDZ to larger basis sets, and even the stabilization energy enlargements were comparable. The same is true about the CBS limits, and D→T values agree well with the T→Q ones. The Truhlar D→T values are again close to the T→Q values of Helgaker with the exception of the mA···mT stacked structure where the Truhlar value is larger. A conclusion similar to that in the case of G···C pairs can be drawn about the importance of the CBS limit extrapolation. When performing calculations "only" at the aug-cc-pVDZ level, the final MP2 stabilization energy will be underestimated by 1.6 and 1.2 kcal/mol for the nonmethylated and methylated species, respectively. This corresponds to about 10% of the stabilization energy. The CCSD(T) stabilization energy of both H-bonded complexes is practically identical to the MP2 one,

(20) Hobza, P.; Zahradnik, R. *Chem. Rev.* **1988**, *88*, 871.

and this again supports the conclusion about the negligible effect of higher order correlation energy contributions on stabilization energy in the case of H-bonded complexes.

Stacked structures investigated were found by MD/Q calculations, and their structures optimized at the RI-MP2/TZVPP level are shown in Figure 1. Similarly as in the case of G···C pairs the RI-MP2/TZVPP stabilization energies are smaller than the aug-cc-pVDZ ones. Passing from aug-cc-pVDZ to larger basis sets produces very similar effects such as in the case of planar systems, and a similar conclusion can be made about the CBS limits. At any theoretical level the adenine···thymine planar structure is by about 1 kcal/mol more stable than the respective stacked structure, while in the case of methylated systems this difference is smaller. Interestingly, the stacked structure of methylated systems is sometimes even slightly more stable than the H-bonded structure. The only important difference between stacked structures of the guanine···cytosine and adenine···thymine complexes were found to arise from higher order correlation energy contributions. In the case of G···C complex and its methylated analogue the CCSD(T) stabilization energies were larger by 1.9 and 2.5 kcal/mol than the MP2 ones. In the case of A···T complexes this difference is enlarged to 2.8 and even 3.5 kcal/mol for nonmethylated and methylated pairs. In the case of nonmethylated and methylated guanine···cytosine complexes we have shown that not considering the ($\Delta E^{CCSD(T)} - \Delta E^{MP2}$) correction term will produce a large error of 2.5 or 3.1 kcal/mol for the relative stabilities of the H-bonded and stacked structures. In the case of A···T complexes this error will become even larger (2.8 and 3.4 kcal/mol for nonmethylated and methylated pairs). These values unambiguously support the importance of higher order correlation energy contributions.

Final Stabilization Energies. Final stabilization energies for planar H-bonded and stacked pairs of nonmethylated and methylated adenine···thymine and guanine···cytosine pairs are collected in Table 1. The planar structure of the mA···mT pair is by about 1 kcal/mol more stable than the nonmethylated one, and both H-bonded structures are more stable than the stacked structures. Here it must be recalled again that the larger stability of H-bonded structures comes from the CCSD(T) part and that the MP2 stabilization energies determined at various levels are comparable for nonmethylated and methylated base pairs. The G···C WC pair is almost twice more stable than the A···T WC pair. The stacked structure of the guanine···cytosine pair is considerably less stable than the respective H-bonded structure, and the difference (∼12 kcal/mol) is much larger than this of the adenine···thymine base pair. The most important conclusion on these final stabilization energies is the fact that **they are large, much larger than the first reliable correlated ab initio calculations**[21,22] **and even larger than our previous estimates** based on medium-sized correlated calculations combined with the estimate of the larger basis sets effects.[23] Estimated stabilization energies for the A···T and G···C WC structures were[23] 14.3 and 26.3 kcal/mol, while the present values from Table 1 are larger by 1.1 and even by 2.5 kcal/mol. It should be mentioned that the stability of the G···C WC structure is especially high. The stacked structure of the guanine···cytosine pair is, as mentioned previously, significantly less stable than the respective WC structure. Nevertheless, it is still more stable than the A···T WC structure! The stability of the adenine... thymine stacked pair is comparable to that of the cytosine and uracil homodimers whose stacking energies were determined recently[23] in a similar way like in the present study.

Methylated base pairs were considered because for these systems experimental data exist (cf. the next paragraph). Passing from Watson−Crick structures of base pairs to the Watson−Crick pairs of nucleosides is surprisingly connected with only negligible energy changes.[24]

Comparison with Experimental Values. Experimental results on methylated adenine···thymine and methylated guanine···cytosine pairs were obtained from the temperature dependence of equilibrium constants measured at rather high temperatures (average temperatures were 323 and 381 K for mA···mT and mG···mC, respectively).[3] Under these conditions not only the most stable structures of base pairs but also many other structures are indispensably populated. According to our previous calculations[6] stacked structures of mG···mC and mA···mT are populated by about 21 and 81% ($T = 300$ and 400 K, respectively, which is close to experimental conditions). It is thus evident that the experimental stabilization enthalpy should be compared with the weighted average of stabilization enthalpies of all nonnegligibly populated structures rather than with the energy of the most stable (in terms of $\Delta H_T°$) but barely present structure (mA···mT Hoogsten, 4.0%; mG···mC Watson−Crick, 28.2%). Computing the stabilization energies of 12 mA···mT and 12 mG···mC structures at the present level of theory is, however, impractical. Furthermore, because ab initio calculations based on the rigid rotor—harmonic oscillator—ideal gas approximation do not provide reliable values of $\Delta G°$ for such a high temperatures as 323 or 381 K, the populations of different complexes cannot be based on quantum chemical calculations only, and, instead, molecular dynamics (MD) calculations should be adopted. We have estimated the population of various structures on the basis of results from our previous MD study.[6] Populations (weighting factors) were taken from MD/Q results, and previous low-level ab initio interaction energies were scaled. H-bonding and stacked energies were scaled by the ratio of $\Delta H_T°$ from Table 1 and the respective low-level ab initio value from ref 6, while the T-shaped structures were scaled by a factor obtained from the average of the factors found for H-bonding and stacking. Resulting interaction enthalpies, $\Delta H_{323K}° = -11.3$ kcal/mol (mA···mT) and $\Delta H_{381K}° = -18.0$ kcal/mol (mG···mC), are in good agreement with experiment ($\Delta H_{323K}° = -13.0$ kcal/mol (mA···mT) and $\Delta H_{381K}° = -21.0$ kcal/mol (mG···mC)). Here we would like to point out that although the present theoretical values are considerably larger than those published previously, the resulting weighted average is still smaller than the experimental value. This implies that the actual stabilization energies of methylated A···T and methylated G···C pairs should be even larger.

Gas-phase calculations and experiments clearly support the fact that H-bonded structures are either more stable than the stacked ones (G···C and mG···mC) or comparably stable (A···T and mA···mT). When passing from the gas phase to the water, the situation is changed and the stacked structure became clearly dominant.[25]

(21) Gould, I. R.; Kollman, P. A. *J. Am. Chem. Soc.* **1994**, *116*, 2493.
(22) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590.
(23) Šponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers* **2001**, *61*, 3.
(24) Kabeláč, M.; Kratochvíl, M.; Šponer, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **2000**, *1077*, 17.
(25) Kabeláč, M.; Hobza, P. *Chem. Eur. J.* **2001**, *2067*, 7.

Possible Sources of Error. We first consider the error of the single-point calculations. Provided that the basis set dependence of the ($\Delta E^{CCSD(T)} - \Delta E^{MP2}$) correction term is negligible, the extrapolated MP2 interaction energies corrected by the CCSD(T) correction should be very similar to the CBS extrapolated CCSD(T) interaction energies. For 16 small closed-shell molecules Klopper et al.[26] showed that the maximum error of the CCSD(T) interaction energy extrapolated using cc-pVTZ and cc-pVQZ basis sets is below 1 kcal/mol and the mean error is well below this limit. We believe that the same holds for our interaction energies estimated by the T→Q extrapolation of the aug-cc-pVTZ and aug-cc-pVQZ results (recalling that the augmented basis sets describe intermolecular interaction considerably better than the nonaugmented ones).

Considerably larger errors may originate from the geometry optimization performed without inclusion of the BSSE. Although the MP2/TZVPP method yields satisfactory results[9] for H-bonded complexes, the question arises whether a comparable accuracy will also be reached for stacked structures. It must be kept in mind that the BSSE for these structures is still rather large (e.g. 4.6 kcal/mol in the case of mG···mC). The energy minimum on the BSSE-corrected PES should be deeper than that on the BSSE-uncorrected PES with a posteriori correction for BSSE. This will bring desirable improvement (enlargement) of the present stabilization energies and, consequently, also closer agreement with experimental data.

The third source of error might originate from using the populations of individual dimer structures based on MD/Q/empirical potential calculations. The larger is the difference in stabilization energies of individual structures in the mixture (H-bonded, stacked, and T-shaped), the larger error is introduced via inaccurate weighting factors. This means the mG···mC results will be affected more than the mA···mT ones, in full agreement with our observation.

It is presently not possible to estimate the relative contribution of errors mentioned, but we believe that the most important one originates from the use of BSSE-uncorrected geometry optimization, especially for the stacked structures.

(26) Bak, K. L.; Halkier, A.; Jørgensen, P.; Olsen, J.; Helgaker, T.; Klopper, W. *J. Mol. Struct.* **2001**, *567*, 375.

Conclusions

(i) The convergency of the MP2 interaction energy for the present complexes is rather slow, and it is found to be essential to include the extrapolation to the complete basis set limit. Working at the MP2/aug-cc-pVDZ level results in an error of about 10% of the MP2/CBS interaction energy.

(ii) The D→T extrapolations yield interaction energies comparable to those obtained from the considerably more demanding T→Q extrapolation that concerns in particular the Truhlar-type extrapolation.

(iii) The aug-cc-pVQZ interaction energies are already close to the CBS interaction energies.

(iv) Inclusion of the CCSD(T) corrections is vital for obtaining reliable relative values between planar H-bonding and stacking interactions.

(v) Final stabilization energies are very large, much larger than published up to now. On the basis of comparison with experimental data, we conclude that they represent the lower bound of the true stabilization energies. On the basis of error analysis, we expect that the present H-bonding energies are close to the true values while stacked energies are still about 10% too low.

(vi) The stacking energies for the mG···mC pair are considerably lower than the respective H-bonding energies, but they are still larger than the mA···mT H-bonding energies. This conclusion can change our view on the importance of specific H-bonding interactions and nonspecific stacking interactions in nature.

(vii) Present stabilization energies for H-bonding and stacking energies represent the most accurate and reliable values and can be considered as new reference data.

Acknowledgment. This project, LN00A032 (Center for Complex Molecular Systems and Biomolecules), was supported by the Ministry of Education of the Czech Republic. We thank Prof. Klaus Müller-Dethlefs for reading and improving the manuscript.

JA036611J