On geometries of stacked and H−bonded nucleic acid base pairs determined at various DFT, MP2, and CCSD(T) levels up to the CCSD(T)/complete basis set limit level

3 **AUTHORS**, INCLUDING:

Iwona Dabkowska
27 PUBLICATIONS   713 CITATIONS

SEE PROFILE

Pavel Hobza
Academy of Sciences of the Czech Republic
318 PUBLICATIONS   18,109 CITATIONS

SEE PROFILE

On geometries of stacked and H-bonded nucleic acid base pairs determined at various DFT, MP2, and CCSD(T) levels up to the CCSD(T)/complete basis set limit level

Iwona Dąbkowska
Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems, Flemingovo nám.2, 166 10 Prague 6, Czech Republic and Department of Chemistry, University of Gdańsk, Sobieskiego 18, 80-952 Gdańsk, Poland

Petr Jurečka and Pavel Hobza[a)]
Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems, Flemingovo nám.2, 166 10 Prague 6, Czech Republic

The geometries and interaction energies of stacked and hydrogen-bonded uracil dimers and a stacked adenine···thymine pair were studied by means of high-level quantum chemical calculations. Specifically, standard as well as counterpoise-corrected optimizations were performed at second-order Møller–Plesset (MP2) and coupled cluster level of theory with single, double, and perturbative triple excitations [CCSD(T)] levels with various basis sets up to the complete basis set limit. The results can be summarized as follows: (i) standard geometry optimization with small basis set (e.g., 6-31G*) provides fairly reasonable intermolecular separation; (ii) geometry optimization with extended basis sets at the MP2 level underestimates the intermolecular distances compared to the reference CCSD(T) results, whereas the MP2/cc-pVTZ counterpoise-corrected optimization agrees well with the reference geometries and, therefore, is recommended as a next step for improving MP2/cc-pVTZ geometries; (iii) the stabilization energy of stacked nucleic acids base pairs depends considerably on the method used for geometry optimization, so the use of reliable geometries, such as counterpoise-corrected MP2/cc-pVTZ ones, is recommended; (iv) the density functional theory methods fail completely in locating the energy minima for stacked structures and when the geometries from MP2 calculations are used, the resulting stabilization energies are strongly underestimated; (v) the self-consistent charges–density functional tight binding method, with inclusion of the empirical dispersion energy, accurately reproduces interaction energies and geometries of dispersion-bonded (stacked) complexes; this method can thus be recommended for prescanning the potential energy surfaces of van der Waals complexes. © *2005 American Institute of Physics*. [DOI: 10.1063/1.1906205]

I. INTRODUCTION

The geometry of noncovalent complexes is its primary property[1] and all other characteristics, such as energy or one-electron properties, depend critically on the accuracy of the geometrical parameters. Experimental techniques provide experimental data only indirectly, via rotation constants, and it is well known that these constants can yield ambiguous results.[1] Each set of rotational constants can apply to a large number of different structures. Theoretical methods of quantum chemistry offer another option and yield reliable structural and geometrical information if high-level procedures are used.

The stabilization energy of noncovalent complexes is known to be highly dependent on the theoretical level. We have recently shown that accurate stabilization energies for the H-bonded and stacked structures of nucleic acid (NA) base pairs are only obtained if second-order Møller–Plesset

(MP2) calculations are performed and extrapolated to the complete basis set (CBS) limit.[2] CBS values are obtained by extrapolating the total energies of supersystem and subsystems.[3] The first acceptable basis set is the aug-cc-pVDZ one. Extrapolation is thus performed from the aug-cc-pVDZ and aug-cc-pVTZ energies. Note that bare DZ+P basis set is generally not acceptable since the relevant stabilization energies are underestimated, which mainly concerns those cases where dispersion energy plays a dominant role (e.g., stacked structures). The calculation of the coupled cluster level of theory with single, double, and perturbative triple excitations [CCSD(T)] CBS limit ($\Delta E^{\mathrm{CCSD(T)}}/\mathrm{CBS}$) for complexes as large as NA base pairs is clearly impractical, therefore we approximated their values by

$$\Delta E^{\mathrm{CCSD(T)}}/\mathrm{CBS} \cong \Delta E^{\mathrm{MP2}}/\mathrm{CBS}$$
$$+ \Delta E[\mathrm{CCSD(T)} - \mathrm{MP2}]/\mathrm{DZ+P}, \quad (1)$$

where $\Delta E^{\mathrm{MP2}}/\mathrm{CBS}$ is the MP2 CBS stabilization energy and the latter term of Eq. (1) gives the difference between the CCSD(T) and MP2 stabilization energies determined using a

a)Author to whom correspondence should be addressed. Electronic mail: pavel.hobza@uochb.cas.cz

medium basis set of the DZ+P quality. The evaluation of the $\Delta E^{\text{CCSD(T)}}$/CBS limit using Eq. (1) is based on the assumption that the difference between the CCSD(T) and MP2 interaction energies is far less dependent on the quality of the basis set than the interaction energies themselves.[4]

The geometrical data in the studies mentioned above were obtained by gradient optimization of the total MP2 (or resolution of identity—RI-MP2) energies and the CCSD(T)/CBS stabilization energy was determined for the stationary point found. First, the basis set superposition error (BSSE) was eliminated by using the Boys–Bernardi function counterpoise procedure.[5] Furthermore, the CBS limit was determined by extrapolating the aug-cc-pVDZ and aug-cc-pVTZ, or aug-cc-pVTZ and aug-cc-pVQZ energies, and, finally, the CCSD(T) correction term was evaluated and added. This procedure for evaluating the stabilization energy is based on the assumption that either none of three steps mentioned affects the equilibrium geometry or that the effects of these steps compensate for each other.

There is some evidence from the literature[6] that counterpoise correction during gradient optimization of small complexes resulted in larger intermolecular distances. Less is known about the role of the remaining two factors. It seems that the MP2 and CCSD(T) stabilization energies of H-bonded complexes (determined at the MP2 optimized geometry) are very similar, but this does not necessarily mean that the MP2 and CCSD(T) optimized geometries of these complexes are similar as well. Regarding the stacked structures of nucleic acids base pairs, the CCSD(T) correction term is much larger than that for H-bonded systems. Moreover, for more than 30 stacked NA base pairs investigated by us up to now, this term is systematically positive, i.e., the CCSD(T) stabilization energy is smaller than the MP2 energy. This could mean that stacked structures adopt larger intersystem distances when optimized at the $\Delta E^{\text{CCSD(T)}}$/CBS level.

The aim of this paper is to show the roles of (i) counterpoise-corrected gradient optimization, (ii) optimization when the CBS limit of the MP2 stabilization energy is considered, and (iii) optimization when correction to higher-order correlation effects are taken into account. Additionally, an attempt is made to estimate the accuracy of the geometry determined by routinely used gradient optimizations performed at the MP2 level with a medium basis set. Finally, our last goal is to assess the quality of the results obtained by the most commonly used DFT methodologies. The study is performed for three different NA base pairs: planar H-bonded and stacked uracil dimer, and stacked adenine···thymine structures.

II. STRATEGY OF CALCULATIONS

First, a standard optimization procedure was performed at the correlated level with a medium-sized atomic orbital (AO) basis set (MP2/6-31G*). This is a frequently used theoretical approach and most of extended molecular clusters are studied only at this level. In the second step, correlated calculations were conducted using an extended AO basis set (RI-MP2/cc-pVTZ). This level yielded reliable geometries

when adopted in our recent studies on NA base pairs[2] *in vacuo*, which showed that RI-MP2 and MP2 energies and interaction energies are practically identical. Therefore, we also use the abbreviation MP2 for the RI-MP2 calculations in the following text. All the calculations with a basis sets larger than 6-31G** were performed employing RI approximation. In the third step, the geometry of the complexes considered were optimized at the same level, but the counterpoise-corrected gradient optimization was used instead of standard gradient optimization. In the fourth and fifth steps, only the intersystem separation was optimized by interpolating the MP2/CBS and CCSD(T)/CBS interaction energies for different intermolecular distances. The interpolation was made by the quadratic fit of the data obtained from the counterpoise-corrected RIMP2/cc-pVTZ geometry and two other points for which intermolecular separation was increased or decreased by 0.2 Å.

In addition to correlated quantum chemical methods, the reliability of very popular density functional techniques was tested as well. The DFT provides reasonable data for H-bonded complexes,[2(c)] but its performance is much worse for stacked systems, where it tends to fail completely.[7] The DFT calculations were performed using various functionals for single-point calculations (taking geometry from correlated calculations) and full geometry optimizations. Additionally, the approximate self-consisted charges–density functional tight binding (SCC-DFTB) method, derived from a tight binding (TB) scheme by an expansion of the density fluctuations around a reference density up to the second order, was used.[8(a)] The method represents a general extension of so-called tight-binding methods for charge self-consistency. The speed of SCC-DFTB computations is comparable to semiempirical quantum chemical methods (for comparison, whereas B-LYP/cc-pVDZ calculation for cytosine dimer takes 6 min of CPU time on a standard 1-processor PC, the same calculation using SCC-DFTB takes 0.3 s only). SCC-DFT does not, however, cover the dispersion properly. Therefore, Elstner *et al.*[8(b)] decided to add an empirical dispersion energy term for large distances, and include a damping of this term (f_{dmp}) with the onset of overlap of the charge density [see Eq. (2)].

The empirical dispersion energy is given by

$$E_{\text{dis}} = - \sum_{ij} f_{\text{dmp}}(R_{ij}) \frac{C_6^{ij}}{R_{ij}^6}, \qquad (2)$$

where f_{dmp} is the dumping function,[9] which approaches 1 for large intermolecular distances and fades to 0 for intermolecular distances small compared to the sum of van der Waals radii, C_6^{ij} are dispersion coefficients derived from atomic polarizabilities, and R_{ij}^6 is distance of given atom pair. This approach is similar to that adopted in commonly used force fields except that repulsion term is substituted by the dumping function.

SCC-DFTB treatment augmented by empirical dispersion term[8(b)] yielded satisfactory results for stacked and H-bonded DNA base pairs as well as stacked complexes of various intercalators (ethidium, daunomycin, etc.) with DNA base pairs.[8(c)]
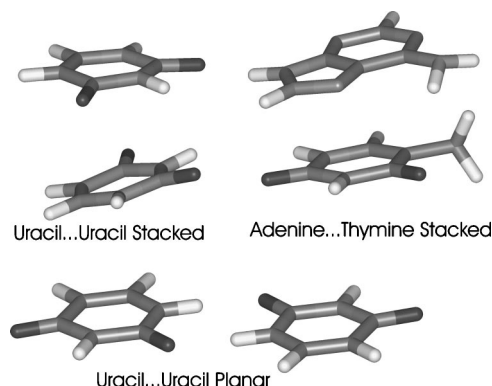
FIG. 1. The structures of stacked and H-bonded uracil dimers and adenine···thymine stacked base pair.

III. SYSTEMS STUDIED

The numerical evaluation of each step mentioned is time consuming, therefore, we selected the smallest NA base pair, namely, uracil dimer, as our model system. We consider the most stable stacked and planar H-bonded structures, which are depicted in Fig. 1.

Uracil dimer stacked structure is characterized by rather small value of the CCSD(T) correction term. This suggests that optimizing the geometry using $\Delta E^{\text{CCSD(T)}}$/CBS interaction energy could result in only small geometry changes (with respect to MP2/CBS geometry). Therefore, we have also included the stacked structure of the adenine···thymine pair (see Fig. 1) characterized by one of the largest values of the CCSD(T) correction term among NA base pairs. For stacked structures, the intermolecular distance defined as the distance between the centers of mass of both subsystems was considered. In the case of H-bonded uracil dimer, we defined intermolecular separation as the distance between heavy atoms involved in H bonds.

IV. METHODS

The geometries of the complexes studied were optimized using MP2, RI-MP2, and DFT methods with TZVPP 6-31G*, ([$5s3p2d1f/3s2p1d$]), and cc-pVTZ ([$4s3p2d1f/3s2p1d$]) basis sets and using the SCC-DFTB method with a standard (minimal) basis set ($1s1p/1s$). MP2/CBS interaction energies were determined using the Helgaker extrapolation scheme[3(b)] from both augmented and nonaugmented cc-pVDZ, cc-pVTZ, and cc-pVQZ energies

$$E_X^{\text{HF}} = E_{\text{CBS}}^{\text{HF}} + A \exp(-\alpha X), \tag{3}$$

$$E_X^{\text{corr}} = E_{\text{CBS}}^{\text{corr}} + BX^{-3}, \tag{4}$$

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 in the original work ($\alpha = 1.43$). So, for example, while extrapolating from the aug-cc-pVDZ and aug-cc-pVTZ point to CBS limit, we need to solve two sets of equations

$$\begin{cases} E_{\text{aDZ}}^{\text{HF}} = E_{\text{CBS}}^{\text{HF}} + A \exp(-1.43 \times 2) \\ E_{\text{aTZ}}^{\text{HF}} = E_{\text{CBS}}^{\text{HF}} + A \exp(-1.43 \times 3) \end{cases} \tag{5}$$

and

$$\begin{cases} E_{\text{aDZ}}^{\text{MP2cor}} = E_{\text{CBS}}^{\text{MP2cor}} + B \times 2^{-3} \\ E_{\text{aTZ}}^{\text{MP2cor}} = E_{\text{CBS}}^{\text{MP2cor}} + B \times 3^{-3}. \end{cases} \tag{6}$$

Finally, we obtain E_{CBS} as a sum of Hartree–Fock (HF) energy at its CBS limit ($E_{\text{CBS}}^{\text{HF}}$) and correlation energy extrapolated to its CBS limit ($E_{\text{CBS}}^{\text{MP2cor}}$):

$$E_{\text{CBS}} = E_{\text{CBS}}^{\text{HF}} + E_{\text{CBS}}^{\text{MP2cor}}. \tag{7}$$

This scheme was chosen because (i) it extrapolates HF and correlation energy separately and (ii) it uses the two-point form (extrapolates two successive basis sets results). The two-point extrapolation form is preferable because it was shown[3(b)] that inclusion of additional (lower quality basis set) points results in extrapolations of lower quality fit.

In this work, we extrapolated all interaction energy components including the BSSE and deformation energy. The CCSD(T) energies were evaluated using 6-31G*(0.25), 6-31G**(0.25,0.15), and aug-cc-pVDZ basis sets (numbers in parentheses refer to exponents of d- and p-polarization functions on heavy atoms and hydrogens, respectively). The CCSD(T)/CBS interaction energy was determined as the sum of the MP2/CBS interaction energy and the CCSD(T) correction term [the difference between MP2 and CCSD(T) interaction energies calculated with a smaller AO basis set]—see Eq. (1).

Counterpoise-corrected gradient optimization was performed using modified jobex script in the TURBOMOLE 5.6 package.

Throughout the study, GAUSSIAN 03,[10] TURBOMOLE 5.6,[11] and MOLPRO 2002.5 (Ref. 12) computational suites were utilized.

V. RESULTS AND DISCUSSION

A. Full gradient optimization

Basis set dependence of geometry. The geometry and energy characteristics of the uracil···uracil (UU) and adenine···thymine (AT) complexes are summarized in Table I.

For the stacked uracil···uracil system, the MP2/6-31G* optimization yields an intermolecular distance of 3.074 Å. The geometry of the complex changes when a larger basis set containing first and second polarization functions is adopted (MP2/cc-pVTZ). The extension of the basis set results in better description of the attractive dispersion energy and, consequently, shorter intermolecular distances should be found. Table I shows that the intermolecular distance decreases, but only slightly (by 0.018 Å). Progressing from standard geometry optimization to counterpoise-corrected gradient optimization should result, on the other hand, in increased intersystem distance. Evidently, (cf. Table I) the latter optimization substantially increases the intersystem distance by 0.111 Å. Analyzing the entries of Table I we observe that at the MP2/6-31G* level the errors due to underestimation of attraction forces and the lack of correction

TABLE I. The interaction energies (kcal/mol) of uracil$\cdots$uracil stacked, uracil$\cdots$uracil H-bonded, and adenine$\cdots$thymine stacked dimers. Numbering of structures 1, 2, and 3 stands for the geometries optimized at the MP2/6-31G*, RIMP2/cc-pVTZ, and RIMP2/cc-pVTZ-counterpoise-corrected levels of theory, respectively. E_{int} stands for interaction energy evaluated at the level at which optimization was done. The numbers in parentheses display total interaction energies after correction for deformation energies of monomers.

| | R^a | E_{int} | MP2 | | | $\Delta CCSD(T)^c$ | ΔE^d |
			aD^b	aT^b	CBS^b		
			uracil$\cdots$uracil stacked				
UU1	3.074	−6.14	−10.47	−11.55	−12.00	0.93	−11.07(−10.82)
UU2	3.056	−9.20	−9.92	−10.95	−11.39	1.13	−10.25(−9.80)
UU3	3.167	−9.10	−9.81	−10.63	−10.83	0.91	−9.93(−9.85)
			adenine$\cdots$thymine stacked				
AT1	3.106	−7.00	−13.54	−14.79	−15.32	2.70	−12.61(−12.11)
AT2	3.083	−12.10	−12.76	−13.81	−14.28	2.92	−11.36(−10.78)
AT3	3.162	−12.30	−13.32	−14.35	−14.77	2.56	−12.21(−12.01)
			uracil$\cdots$uracil H bonded				
UU1	2.855	−17.55	−18.75	−19.79	−20.22	−0.17	−20.39(−18.66)
UU2	2.760	−19.12	−18.62	−19.98	−20.52	0.01	−20.51(−18.61)
UU3	2.802	−18.74	−18.41	−19.60	−20.08	−0.07	−20.15(−18.69)

[a]Intersystem distance in Angstrom.
[b]aD, aT, CBS mean aug-cc-pVDZ, aug-cc-pVTZ, and complete basis set limit, respectively.
[c]CCSD(T)-MP2 energies determined in 6-31G*(0.25) basis set.
[d]$\Delta E^{CCSD(T)}$/CBS interaction energy evaluated as the sum of MP2/CBS and $\Delta CCSD(T)$ energies.

for BSSE partially cancel one another, which leads to a relatively reasonable intermolecular distance. The difference in intermolecular distances calculated by very fast MP2/6-31G* and time-consuming counterpoise-corrected RI-MP2/cc-pVTZ optimizations is smaller than 0.1 Å.

Similar observations can be made concerning the AT-stacked complex. The MP2/6-31G* optimization yields intersystem separation of 3.106 Å and this distance decreases when the MP2/TZVPP optimization is performed (cf. Table I). The MP2/cc-pVTZ counterpoise-corrected gradient optimization leads to an increase of the intersystem distance (3.162 Å). All these distances agree well with their counterparts from UU-stacked dimer, demonstrating that intermolecular separation in stacked structures depends only a little on the nature of the pair.

The geometry dependence on the basis set is qualitatively the same for the UU H-bonded complex. The structure of the H-bonded pair (cf. Fig. 1) corresponds to the global minimum at the potential energy surface of the uracil dimer[13] at the HF/6-31G** level. Progressing to the MP2/6-31G* level, the intersystem distance decreases from 2.917 to 2.855 Å (cf. Table I) and expanding the basis set to cc-pVTZ causes only a small geometry change of 0.095 Å. By performing counterpoise-corrected optimization at the same level of theory, one obtains an intermolecular distance increase of 0.04 Å, but this increase is significantly smaller than that for both of the stacked complexes.

Basis set dependence of the stabilization energy. The stabilization energy depends strongly on the basis set used. For UU-stacked dimer, the MP2/6-31G* value (6.14 kcal/mol) is considerably smaller than the MP2/cc-pVTZ value (9.20 kcal/mol). This increase is understandable and is due to much better description of the correlation (dis-

persion) energy. Extrapolation to the CBS limit further increases the MP2 stabilization energy so that 12.00 and 11.39 kcal/mol are obtained for 6-31G* and MP2/cc-pVTZ counterpoise-corrected geometries. Both stabilization energies are significantly larger than that determined at the MP2/cc-pVTZ level (even though the cc-pVTZ is already an extended basis set, which contains f functions on heavy atoms and d functions on hydrogens) and it again confirms the need to extrapolate the stabilization energies to the CBS limit.

Similarly, in the case of AT-stacked dimer, the MP2/6-31G* and MP2/TZVPP stabilization energies differ considerably (by 5.10 kcal/mol) and this increase reflects the slow convergence of the intersystem correlation energy for stacked structures. Extrapolating to the CBS limit additionally enlarges the stabilization energy by more than 2 kcal/mol, again underlining the necessity of using an extended AO basis set for obtaining reliable stabilization energies.

However, by investigating the stabilization energies for UU H-bonded pair, we found that they are less dependent on the calculation level than in the case of stacked structures. Progressing from MP2/6-31G* to MP2/cc-pVTZ yields a 9% stabilization energy increase, while these values for stacked UU and AT complexes reach 50% and 73%, respectively. Evidently, H-bonded structures are described well at the lower theoretical level, while the description of stacked clusters requires higher levels. This conclusion is confirmed by comparing the interpolated MP2/CBS value (see the following paragraph) with the MP2/cc-pVTZ value (20.09 and 19.12 kcal/mol, respectively)—they are rather close. Furthermore, the interpolated CCSD(T)/CBS stabilization energy (20.20 kcal/mol) is slightly larger than the MP2/CBS energy, which is opposite the case of both stacked clusters.

TABLE II. Point energies by point optimization of uracil···uracil stacked and uracil···uracil H-bonded pairs. All distances are in angstrom, all energy values in kcal/mol.

Method	Basis[a]	UU stacked				UU H bonded			
		Energy for given separation[b]				Energy for given separation[b]			
		−0.2	0.0	+0.2	Interpolation[c]	−0.2	0.0	+0.2	Interpolation[c]
MP2	DZ	−11.56	−11.77	−10.54	(3.097;−11.86)	−19.49	−21.74	−20.55	(2.833;−21.77)
	TZ	−11.95	−12.16	−10.99	(3.099;−12.25)	−19.45	−21.07	−19.40	(2.800;−21.08)
	QZ	−11.44	−11.68	−10.57	(3.104;−11.76)	−19.07	−20.81	−19.17	(2.805;−20.80)
	CBS	−11.41	−11.63	−10.53	(3.101;−11.72)	−19.03	−20.80	−19.16	(2.805;−20.81)
	aDZ	−16.04	−15.58	−13.82	(2.999;−16.05)	−19.53	−21.72	−20.33	(2.824;−21.75)
	aTZ	−13.17	−12.18	−11.02	(3.065;−13.33)	−19.73	−21.53	−19.92	(2.807;−21.54)
	aQZ	−11.83	−11.98	−10.16	(3.083;−12.15)	−19.08	−20.89	−19.31	(2.808;−20.88)
	CBS	−11.03	−11.28	−9.31	(3.094;−11.27)	−18.73	−20.51	−18.93	(2.807;−20.52)
MP2	DZ	−3.84	−5.93	−6.19	(3.293;−6.30)	−12.65	−16.13	−15.85	(2.887;−16.46)
	TZ	−8.11	−9.10	−8.56	(3.197;−9.12)	−16.41	−18.74	−17.56	(2.834;−18.80)
	QZ	−9.74	−10.30	−9.46	(3.147;−10.29)	−17.64	−19.75	−18.38	(2.823;−19.78)
	CBS	−10.90	−11.16	−10.11	(3.107;−11.21)	−18.49	−20.43	−18.92	(2.814;−20.43)
CP corrected	aDZ	−8.75	−9.81	−9.81	(3.198;−9.82)	−15.26	−18.41	−17.69	(2.864;−18.58)
	aTZ	−10.08	−10.63	−9.76	(3.145;−10.64)	−17.28	−19.60	−18.36	(2.832;−19.65)
	aQZ	−10.51	−10.91	−9.28	(3.107;−11.00)	−17.96	−20.07	−18.67	(2.822;−20.09)
	CBS	−10.82	−11.11	−9.17	(3.094;−11.27)	−18.43	−20.37	−18.87	(2.815;−20.39)
CCSD (T)	DZ	−10.29	−10.87	−9.87	(3.141;−10.88)	−19.15	−21.80	−20.83	(2.848;−21.89)
	TZ	−10.68	−11.25	−10.32	(3.144;−11.26)	−19.11	−21.13	−19.68	(2.818;−21.15)
	QZ	−10.17	−10.77	−9.91	(3.150;−10.79)	−18.73	−20.87	−19.45	(2.822;−20.88)
	CBS	−10.15	−10.73	−9.87	(3.148;−10.74)	−18.69	−20.86	−19.44	(2.822;−20.89)

[a]The abbreviations DZ, TZ, QZ, aDZ, aTZ, and aQZ stand for cc-pVDZ, cc-pVTZ, cc-pVQZ, aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ basis sets.
[b]For "0.0" geometries, we utilized MP2/cc-pVTZ counterpoise-corrected geometries, where the intermolecular distances equal to 3.167 and 2.802 Å for UU stacked and UU H-bonded pairs, respectively.
[c]The point displayed in the table is the minimum of the quadratic function obtained by interpolating the stabilization energies for the reference geometry and two points around it at 0.2 Å increment.

Previously, we extensively used[7(b),14] a modified 6-31G* basis set where a standard exponent of d-polarization functions on heavy atoms ($\alpha=0.8$) was replaced by a lower value of 0.25 (Ref. 15) and this basis set yielded surprisingly accurate stabilization energies even for stacked NA base pairs. The MP2/6-31G*(0.25) stabilization energy for the UU-stacked complex for the MP2/6-31G* and MP2/cc-pVTZ geometries amounts to −8.43 and −9.20 kcal/mol (cf. Table III), respectively, which is close to higher-level energies. This agreement justifies the use of this basis set for stacked (but also H-bonded) structures. However, it must be mentioned, on the basis of our earlier studies,[14] that the 6-31G*(0.25) basis set is not suitable for geometry optimization.

The effect of the geometry on the stabilization energy. The CCSD(T)/CBS stabilization energies evaluated for MP2/6-31G*, MP2/cc-pVTZ, and MP2/cc-pVTZ-counterpoise-corrected geometries are presented in Table I. Comparing their values for MP2/6-31G* and MP2/cc-pVTZ geometries one can see a very small change for UU H-bonded system (by 0.12 kcal/mol), while the changes for stacked pairs are unignorable (about 0.8 kcal/mol). Analo-gously, while passing from counterpoise-uncorrected to counterpoise-corrected geometries, the energy change for the planar system is smaller than for the stacked structures.

This difference shows that the CCSD(T)/CBS interaction energy of the stacked system is highly sensitive to the geometry. The necessity of using reliable geometries as possible is thus clearly demonstrated. However, one should also take into account the cost of the calculations in respect to the gain in quality.

B. Point-by-point one-dimensional optimization

Full gradient optimization with large basis sets such as cc-pVQZ is currently impractical for such big systems. Therefore, we optimized only one coordinate—the intersystem distance. This data is collected in Table II and displayed in Fig. 2.

For both uracil complexes, the optimal intermolecular distance and energy are interpolated using three single-point energy calculations: one on the counterpoise-corrected MP2/cc-pVTZ minimum geometry, one with the intermolecular distance increased by 0.2 Å, and one with the distance de-
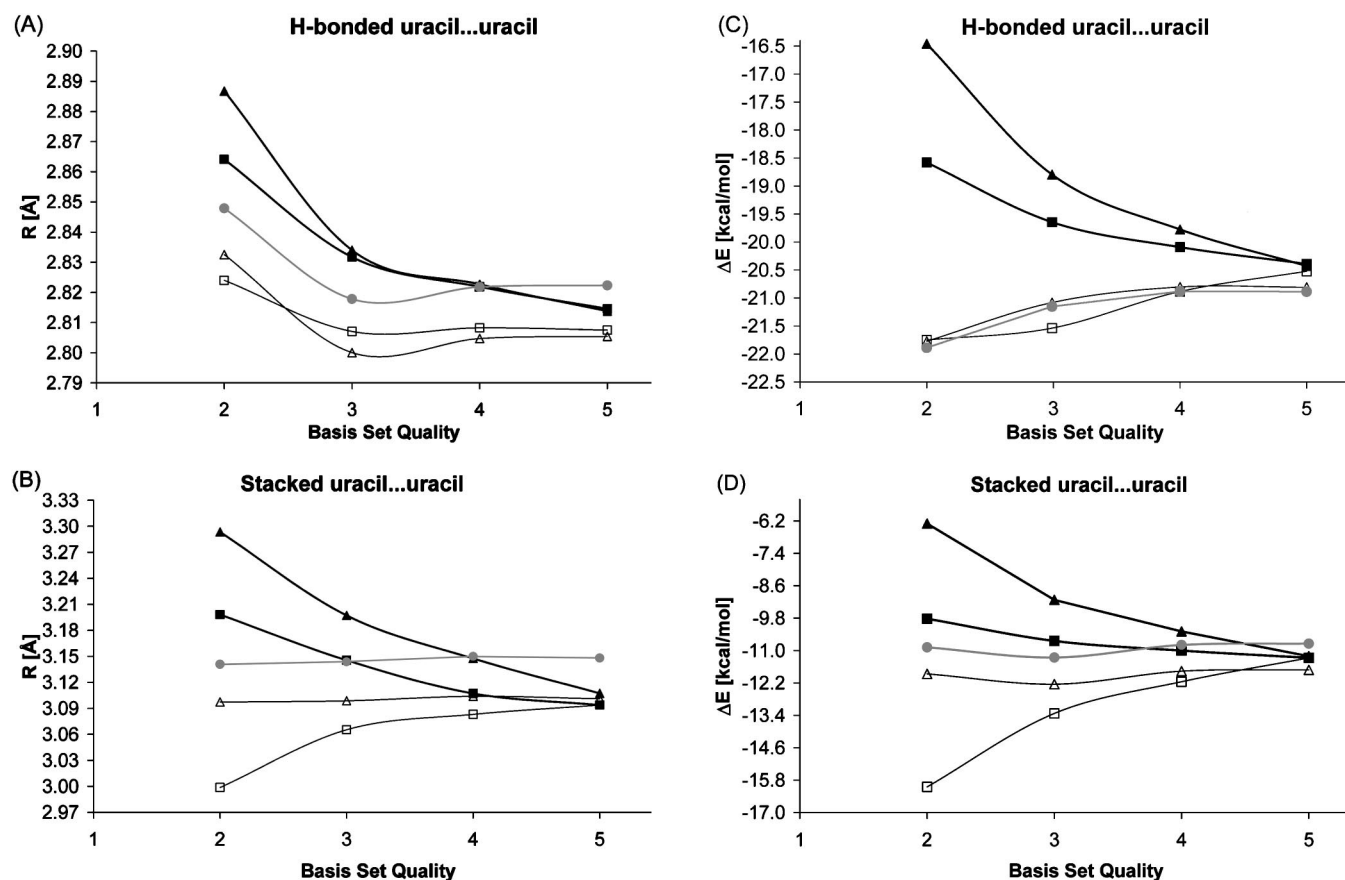
FIG. 2. The basis set dependence of MP2 intermolecular distances and interaction energies of H-bonded [plots (A) and (C)] and stacked [plots (B) and (D)] uracil dimer. Solid symbols represent counterpoise-corrected values, open symbol represents uncorrected values. Square symbols correspond to MP2/aug-cc-pVXZ, triangles to MP2/cc-pVXZ, and gray circles to CCSD(T)/cc-vVXZ values, where X=D,T,Q. X-axis marks refer to the highest angular momentum quantum number in the basis set given $(D, T, \ldots \sim 2, 3, \ldots)$, and 5 is for TZ→QZ extrapolation to CBS.

creased by 0.2 Å. This approximation allows us to use basis sets as large as aug-cc-pVQZ. Unfortunately, it appears that the results obtained by interpolating three points on the potential energy surface cannot be directly compared with the intermolecular separations and energies from the full gradient optimization. The reason for this is that the potential energy surface is not perfectly quadratic, therefore the use of the quadratic interpolation function introduces some error. However, the relative effects of different basis sets (convergence behavior) and trends in the convergence rates should be captured reasonably well within the quadratic approximation.

The convergence of the MP2 intermolecular distances and interaction energies for both uracil dimers is shown in Fig. 2. Some curves do not converge to the CBS limit monotonously due to the different basis set dependence of the Hartree–Fock and correlation energies (both components are extrapolated separately). The extrapolated CBS values for both augmented and nonaugmented cc-pVXZ basis set series are very close to each other and can thus be considered as converged. In the following text, the counterpoise-uncorrected MP2/cc-pVXZ values extrapolated to their CBS limit (last points of open-triangles' curves) are referred to as MP2/CBS reference energies and distances. Therefore, the reference values are 2.805 Å and −20.81 kcal/mol for H bonded, and 3.101 Å and −11.72 kcal/mol for stacked uracil dimer.

In each case, both the basis set superposition error and the basis set incompleteness error are relatively large even for basis sets containing f functions. Furthermore, it seems that no generally valid conclusion can be derived regarding whether counterpoise corrected or uncorrected values behave better for every kind of complex and basis set series (augmented and nonaugmented). Nevertheless, it seems that counterpoise-uncorrected cc-pVXZ series with $X \geq T$ converge faster than the others, probably because BSSE is almost completely compensated for by the growing dispersion contribution. *Counterpoise-uncorrected cc-pVTZ geometries are, in all cases, already especially close to their complete basis set limits.* It should be emphasized that this does not mean that this level of theory is recommended for obtaining highly precise intermolecular separations, since MP2 method is known to overestimate the dispersion contribution of the intermolecular interaction energy (this issue will be addressed in the following paragraph). However, if we assume the small basis set dependence of ΔCCSD(T) term, it is evident that CCSD(T)/cc-pVTZ counterpoise-uncorrected geometries are superior to the counterpoise-corrected ones and they would be practically identical to their CCSD(T)/CBS limits.

The gray curves in Fig. 2 show the effect of higher-order correlation contributions [approximated by ΔCCSD(T) term] on the intermolecular separations and interaction energies of

TABLE III. The saturation of the basis set at the MP2 and CCSD(T) levels for planar and stacked UU dimers. For stacked system we used geometry optimized at RIMP2/cc-pVTZ UU2 and for planar system at the MP2/6-31G* level of theory. All quantities are in kcal/mol.

	6-31G*(0.25)	6-31G** (0.25,0.15)	aug-cc-pVDZ	aug-cc-pVTZ
Uracil···Uracil–stacked				
CCSD(T)	−8.01	−8.20	−9.09	
MP2	−8.94	−9.01	−10.47	−11.55
CCSD(T)-MP2	0.93	0.81	1.38	
Uracil···Uracil–H bonded				
CCDS(T)	−16.36	−16.67	−17.05	
MP2	−15.92	−16.17	−16.87	−17.63
CCSD(T)-MP2	−0.44	−0.50	−0.18	

uracil dimers in H-bonded and stacked geometries. As expected, in the case of the stacked complex [Figs. 2(B) and 2(D)] large ΔCCSD(T) correction (about 1 kcal/mol) causes elongation of the intermolecular distance. An increase of 0.046 Å (with respect to the reference MP2/CBS geometry—open-triangles curve) is unignorable, because it is comparable to geometry changes which occur when we extrapolate from cc-pVTZ values to the complete basis set limit. In the case of H-bonded dimer, the inclusion of ΔCCSD(T) term resulted in elongation of the H-bond lengths as well (0.016 Å, in respect to the reference geometry), although the ΔCCSD(T) correction to energy at the MP2 equilibrium distance is negative (−0.1 kcal/mol). This means that the CCSD(T) minimum is deeper and the intermolecular separation is larger compared to MP2. Though the higher-order correlation correction is comparable to the effects of the basis set size increase, its absolute value is small (note that the potential energy valley of the H-bonded structure is much steeper than for the dispersion-bonded, stacked structure). Adding these two values to the MP2/CBS limit we obtain approximate CCSD(T)/CBS intermolecular separations of 2.822 Å for H-bonded uracil dimer and of 3.148 Å for stacked uracil dimer.

A question arises about the importance of higher-order correlation on intermolecular separations in structures where CCSD(T) correction term is of profound importance. In this study we determine this effect for an AT stacked pair, where the ΔCCSD(T) term is considerably larger than that of the UU dimer. Consequently, a larger difference between MP2/CBS and $\Delta E^{\text{CCSD(T)}}$/CBS geometries can be expected. Indeed, by optimizing the $\Delta E^{\text{CCSD(T)}}$/CBS interaction energy we obtain energy minimum at 3.204 Å, i.e., the intermolecular distance increases by 0.083 Å (in comparison to MP2/CBS limit, which in this case, was obtained by extrapolating from MP2/aug-cc-pVDZ and MP2/aug-cc-pVTZ values), which is considerably more than the analogical value for uracil dimer (0.047 Å).

The data presented in Fig. 2(A) demonstrates that the approximate CCSD(T)/CBS geometry of H-bonded uracil dimer (2.822 Å) (the last point on the gray curve) agrees well with the counterpoise-corrected MP2/cc-pVTZ geometry (2.834 Å) (the second point on solid-square curve). Regarding the stacked uracil dimer [Fig. 2(B)], the difference is slightly larger (3.148 Å vs 3.197 Å, respectively). However,

this difference is expected to decrease if larger basis sets are used to evaluate the ΔCCSD(T) correction (see Table III). In both cases, counterpoise-corrected MP2/cc-pVTZ optimization performs slightly better than counterpoise-uncorrected MP2/cc-pVTZ optimization. Furthermore, due to overestimation of the dispersion energy, the MP2 counterpoise-uncorrected method gives insufficient intermolecular separation, hence, it introduces artificial deformation of monomers. Thus, it can be concluded that counterpoise-corrected MP2 optimization can be recommended as a method of choice to further improve MP2/cc-pVTZ geometries. However, it seems that the good performance of counterpoise-corrected MP2/cc-pVTZ optimization with respect to CCSD(T)/CBS optimization is partially due to the fortunate cancellation of the basis set incompleteness error with effects of higher-order correlation. In other words, the error in counterpoise-corrected optimization is comparable to the difference between the MP2 and CCSD(T) results.

However, according to unpublished results from our laboratory, it seems that this only holds true for complexes above a certain size. For small complexes with less than ten heavy atoms, intermolecular distances determined at counterpoise-corrected MP2/cc-pVTZ level are substantially overestimated. This effect is probably connected to the higher flexibility of the wave function of large polyatomic molecules (basis set saturation), which allows better description of larger systems than small ones within the same atomic basis set.[16]

C. Basis set convergence of ΔCCSD(T) term

The ΔCCSD(T) correction term in Tables I and II is determined using a 6-31 G*(0.25) basis set. In our previous studies,[4] we investigated model (H bonded and stacked) complexes and showed that smaller basis sets such as 6-31 G*(0.25) or 6-31 G**(0.25,0.15) reproduce the CCSD(T)-MP2 energy difference determined with aug-cc-pVDZ basis set well. In this paper we calculated, for the first time, the CCSD(T) correction term for UU dimer using an aug-cc-pVDZ basis set.

Table III shows the CCSD(T) correction term determined using 6-31 G*(0.25), 6-31 G**(0.25,0.15), and aug-cc-pVDZ basis sets. The first one reproduces the reference [CCSD(T)-MP2] value slightly better. However, all numbers

TABLE IV. DFT interaction energies (in kcal/mol) determined with 6-31 +G^{**} basis set for the planar H-bonded (UU-PL) and stacked (UU-ST) uracil dimer. All values are BSSE corrected.

	Frozen geometry[a]		Full optimization	
Functional	UU-PL	UU-ST	UU-PL	UU-ST
PBE1PBE	−17.16	−3.75	−19.85	[b]
MPW1PW91	−16.33	−2.05	−18.87	[b]
PW91PW91	−17.08	−3.59	−20.80	[b]
B3LYP	−16.16	−0.82	−18.29	[b]
DFT-TB_no disp	−13.27	−4.88	−15.93	−4.39
DFT-TB_disp	−14.62	−9.85	−17.36	−9.22

[a]MP2/6-31G^* geometries.
[b]No minimum was detected and all structures converged to planar H-bonded ones.

are similar and show that the CCSD(T) correction term is repulsive for stacked structures. It must be mentioned here that up till now we have collected results for more than 30 stacked NA base pairs and the CCSD(T) correction term is *systematically repulsive*. The CCSD(T) correction term for the H-bonded structure is much smaller and also here the 6-31 $G^*(0.25)$ value is closer to the reference value.

D. DFT performance

Finally, we would like to comment on the performance of the DFT methods. Table IV summarizes the DFT interaction energies for stacked UU dimer.

When the MP2/6-31 G^* geometry is adopted, the resulting stabilization energies are considerably smaller than those obtained using the MP2/6-31 G^* procedure (−6.14 kcal/mol, cf. Table I). The smallest difference (however still higher than 2.5 kcal/mol) is obtained with functionals of Perdew, Burke, and Ernzerhof (PBE1PBE) and of Perdew and Wang (PW91PW91), while the largest difference results when the B3LYP functional is used. The situation becomes much worse when the geometry of a complex is optimized within the functional used. For all starting geometries, no stationary point was found for stacked structures. All optimizations result in planar H-bonded structures. This unambiguously means that the DFT potential energy surface does not exhibit any energy minimum corresponding to the stacked arrangement (in our previous paper we have shown that identical results were obtained when DFT calculations were performed with a much larger, cc-pVTZ basis set).[7(c)] This is certainly a very serious drawback of DFT and, moreover, varying a functional does not improve the situation. We would like to stress here that this dramatic failure is not limited to stacked structures of NA base pairs, but similar ones have been found for stacked (or better nonplanar) amino acids dimers.[17] The explanation for these failures is simple—the DFT techniques do not cover the London dispersion energy. The last two lines of Table III show the results for when the DFTB methods (with and without addition of dispersion energy) are applied. Evidently, the dispersion energy plays an important role and when covered (the last line) the respective results are better than those from MP2/6-31 G^* calculations and agree well with MP2/cc-

pVTZ results, even when full optimization is performed. These results again underline the importance of dispersion energy in biomolecular clusters and clearly show the limiting usage of methods not covering this energy contribution.

VI. CONCLUSIONS

(1) Standard geometry optimization with a small AO basis set (e.g., 6-31 G^*) provides relatively reasonable intermolecular separation which is due to the compensation of errors by the increase of the basis set, the counterpoise-corrected gradient optimization, and the inclusion of high-order correlation energy contributions.

(2) Geometry optimization with extended basis sets at the MP2 level underestimates the intermolecular distances compared to the reference CCSD(T) results. But MP2/cc-pVTZ counterpoise-corrected optimization agrees better with reference geometries and is, therefore, recommended as a next step for improving MP2/cc-pVTZ geometries.

(3) Higher-order correlation effects on the intermolecular distance are found to be comparable in magnitude to the basis set truncation effect for both H-bonded and stacked complexes. Because these effects are of the opposite sign, they almost cancel each other if counterpoise-corrected optimization at the MP2/cc-pVTZ level is performed, which is why this method performs relatively well for large dimers.

(4) The stabilization energies of stacked NA base pairs depend significantly on their intermolecular geometry. Use of reliable counterpoise-corrected MP2/cc-pVTZ geometries is recommended. If lower-level geometry is adopted, it results in an overestimation of the stabilization energy by about 10%.

(5) The CCSD(T) correction term for NA base pairs is determined for the first time using extended (aug-cc-pVDZ) basis set. Its value is close to that calculated with medium (6-31 G^*, 6-31 G^{**}) basis sets. This finding is of key importance as it justifies the existing procedure for determining the total stabilization energy [cf. Eq. (1)] of extended molecular clusters.

(6) The DFT methods fail completely in locating energy minima for stacked structures. When the geometries from MP2 calculations are used the resulting stabilization energies are strongly underestimated.

(7) The SCC-DFTB method, with inclusion of empirical dispersion energy (SCC-DFTB+D) reproduces well both the interaction energies and the geometries of dispersion-bonded (stacked) complexes. It can thus be recommended for prescanning of the potential energy surfaces of van der Waals and biomolecular complexes.

ACKNOWLEDGMENTS

[1] K. Müller-Dethlefs and P. Hobza, Chem. Rev. (Washington, D.C.) **100**, 143 (2000).

[2] P. Jurečka and P. Hobza, J. Am. Chem. Soc. **125**, 15608 (2003); P. Jurečka, J. Šponer, and P. Hobza, J. Phys. Chem. B **108**, 5466 (2004); J. Šponer, P. Jurečka, and P. Hobza, J. Am. Chem. Soc. **126**, 10142 (2004); I. Dąbkowska, H. Valdes, P. Jurečka, and P. Hobza, J. Phys. Chem. A **109**, 1131 (2005).

[3] (a) D. Feller, J. Chem. Phys. **96**, 6104 (1992); (b) A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, O. Jeppe, and A. K. Wilson, Chem. Phys. Lett. **286**, 243 (1998); (c) A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, and J. Olsen, *ibid.* **302**, 437 (1999); (d) D. G. Truhlar, *ibid.* **294**, 45 (1998).

[4] P. Jurečka and P. Hobza, Chem. Phys. Lett. **365**, 89 (2002); P. Hobza and J. Šponer, J. Am. Chem. Soc. **124**, 11802 (2002).

[5] S. F. Boys and F. Bernardi, Mol. Phys. **19**, 553 (1970).

[6] S. Simon, M. Duran, and J. J. Dannenberg, J. Chem. Phys. **105**, 11024 (1996); P. Hobza and Z. Havlas, Theor. Chem. Acc. **99**, 372 (1998); P. Hobza, O. Bludský, and S. Suhai, Phys. Chem. Chem. Phys. **1**, 3073 (1999); P. K. Wawrzyniak, J. Panek, Z. Latajka, and J. Lundell, J. Mol. Struct. **704**, 297 (2004).

[7] (a) P. Hobza, J. Šponer, and T. Reschel, J. Comput. Chem. **16**, 1315 (1995); (b) J. Šponer and P. Hobza, Collect. Czech. Chem. Commun. **68**, 2231 (2003); (c) J. Černý and P. Hobza, Phys. Chem. Chem. Phys. **7**, 1624

[8] (a) M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, Phys. Rev. B **58**, 7260 (1998); (b) M. Elstner, P. Hobza, T. Frauenheim, S. Suhai, and E. Kaxiras, J. Chem. Phys. **114**, 5149 (2001); (c) D. Reha, M. Kabeláč, F. Ryjacek, J. Šponer, J. E. Šponer, M. Elstner, S. Suhai, and P. Hobza, J. Am. Chem. Soc. **124**, 336 (2002).

[9] T. A. Halgren, J. Am. Chem. Soc. **114**, 7827 (1992).

[10] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 03, Revision A.1, Gaussian, Inc., Pittsburgh, PA, 2003.

[11] R. Ahlrichs, M. Bär, M. Häser, H. Horn, and C. Kölmel, Chem. Phys. Lett. **162**, 165 (1989).

[12] R. D. Amos, A. Bernhardsson, A. Berning *et al.*, MOLPRO, a package of *ab initio* programs designed by H.-J. Werner and P. J. Knowles, version 2002.1.

[13] J. Šponer and P. Hobza, J. Phys. Chem. A **104**, 4592 (2000).

[14] P. Hobza and J. Šponer, Chem. Rev. (Washington, D.C.) **99**, 3247 (1999).

[15] L. M. J. Kroon-Batenburg and F. B. van Duijneveldt, J. Mol. Struct. **121**, 185 (1985).

[16] P. Hobza and R. Zahradník, Chem. Rev. (Washington, D.C.) **88**, 871 (1988).

[17] J. Vondrášek, L. Bendová, V. Klusák, and P. Hobza, J. Am. Chem. Soc. **127**, 2615 (2005).