

## Convergence of the CCSD(T) Correction Term for the Stacked Complex Methyl Adenine–Methyl Thymine: Comparison with Lower-Cost Alternatives

M. Pitoňák,<sup>†,§</sup> T. Janowski,<sup>||</sup> P. Neogrády,<sup>§</sup> P. Pulay,<sup>||</sup> and P. Hobza<sup>\*,†,‡</sup>

*Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, v.v.i., Flemingovo Nám. 2, 166 10 Praha 6, Czech Republic, Department of Physical Chemistry, Palacký University, 771 46 Olomouc, Czech Republic, Department of Physical and Theoretical Chemistry, Faculty of Natural Sciences, Comenius University, Mlynská Dolina, 842 15 Bratislava 4, Slovak Republic, and Department of Chemistry and Biochemistry, Fulbright College of Arts and Sciences, University of Arkansas, Fayetteville, Arkansas 72701*

Received March 17, 2009

**Abstract:** We have performed large-scale calculations for the interaction energy of the stacked methyl adenine–methyl thymine complex at the CCSD(T)/aug-ccpVXZ (X = D,T) levels. The results can serve as benchmarks for the evaluation of two methods, MP2.5, introduced recently, and the widely used  $\Delta$ CCSD(T) correction defined as the difference between the CCSD(T) and MP2 energies. Our results confirm that the  $\Delta$ CCSD(T) correction converges much faster toward the complete basis set (CBS) limit than toward the MP2 or CCSD(T) energies. This justifies approximating the CBS energy by adding the  $\Delta$ CCSD(T) correction calculated with a modest basis set to a large basis MP2 energy. The fast convergence of the  $\Delta$ CCSD(T) correction is not obvious, as the individual CCSD and (T) contributions converge less rapidly than their sum. The MP2.5 method performs very well for this system, with results very close to CCSD(T). It is conjectured that using a  $\Delta$ MP2.5 correction, defined analogously to  $\Delta$ CCSD(T), with large basis sets may yield more reliable nonbonded interaction energies than using  $\Delta$ CCSD(T) with a smaller basis set. This would result in important computational savings as the MP3 scales computationally much less steep than CCSD(T), although higher than SCS-MP2, a similar approximation.

### 1. Theoretical Background

The CCSD(T) method in the complete basis set (CBS) limit provides accurate stabilization energies for various structures of molecular complexes. This is a very demanding task, but it must be kept in mind that CCSD(T)/CBS is in fact the only ab initio quantum mechanical method which is consistently capable of delivering benchmark quality results for single-reference systems. Other methods either are too

expensive or, if less expensive, fail to provide the right answer (e.g., MP2/CBS strongly overestimates stacking stabilization energies although it describes H-bonding energies reasonably well<sup>1</sup>). The performance of yet another group of less expensive methods is enhanced by incorporating empirical parameters. For example, the SCS-MP2 method<sup>2</sup> and its variants (SOS-MP2,<sup>3</sup> SCS(MI)-MP2,<sup>4</sup> and modified SCS-MP2<sup>5</sup>) provide better results than the MP2 method (the overestimation of the stacking interactions is corrected) but utilize one or two empirical parameters.

CCSD(T)/CBS provides excellent results, but it is very time consuming due to its unfavorable  $N^7$  scaling with the size of the system. Recently, a much more economical but still accurate method, MP2.5 (along with a generalized variant, “scaled MP3”), was proposed.<sup>1</sup> Unlike the SCS-MP2

\* Corresponding author. Phone: +420220410311. Fax: +420220410320. E-mail: pavel.hobza@uochb.cas.cz.

<sup>†</sup> Academy of Sciences of the Czech Republic.

<sup>||</sup> University of Arkansas.

<sup>§</sup> Comenius University.

<sup>‡</sup> Palacký University.

method and its variants, MP2.5 has correct asymptotic scaling at large intermolecular distances and uses only a single parameter. The MP2.5 method, more expensive than MP2, computational scaling is  $O(N^6)$  vs  $O(N^5)$  but still can be applied to significantly larger systems than CCSD(T). MP2.5 was shown to be superior when compared to SCS-MP2 based methods.<sup>1</sup>

An alternative, and more economical, approach uses modified density functional theory (DFT). DFT, extended by empirical dispersion correction by Grimme<sup>6</sup> or Jurečka et al.,<sup>7</sup> provides good stabilization energies and geometries for molecular complexes and clusters, but two parameters per atom are included in the empirical dispersion energy formula (Grimme's formulation includes an extra global scaling parameter as well). Promising new exchange–correlation functionals of Zhao and Truhlar (the M05 and M06 families<sup>8</sup>) use a large number of fitted parameters. Preliminary results obtained in the Hobza laboratory show good performance of the M06 family. They perform well for both noncovalent interactions and also for IR and visible spectra of isolated molecules. Parameters in most of the procedures mentioned above must, however, be fitted against benchmark data, resulting mostly from the CBS extrapolation of CCSD(T) energies.

The estimated CCSD(T)/CBS energy is defined as

$$\text{CCSD(T)}_{\text{CBS}} \approx \text{MP2}_{\text{CBS}} + \Delta\text{CCSD(T)}_{\text{small basis set}} \quad (1)$$

where the first term is the MP2 interaction energy, covering most of the correlation effects. This term is ideally calculated at a level close to that of the basis set limit. The second term, called the CCSD(T) correction term, is determined as a difference between CCSD(T) and MP2 interaction energies calculated using a smaller, computationally tractable basis set:

$$\Delta\text{CCSD(T)}_{\text{small basis set}} = \text{CCSD(T)}_{\text{small basis set}} - \text{MP2}_{\text{small basis set}} \quad (2)$$

It describes correlation effects such as pair couplings that are omitted in the MP2 method.

Fairly reliable interaction energies close to CBS can be obtained by two- (or more) point extrapolation by schemes proposed, for instance, by Helgaker et al.<sup>9</sup>

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

$$E_X^{\text{corr.}} = E_{\text{CBS}}^{\text{corr.}} + B \cdot X^{-3} \quad (4)$$

where A, B, and  $\alpha$  are fitting parameters, or by Lee et al.<sup>10,11</sup> (using BSSE corrected,  $\Delta E_X^b$ , and uncorrected,  $\Delta E_X^n$ , interaction energies):

$$\delta_X = \Delta E_X^b - \Delta E_X^n \quad (5)$$

$$\epsilon_X = \Delta E_X^b + \Delta E_X^n \quad (6)$$

$$\Delta E_{\text{CBS}} = 1/2(\delta_X \epsilon_{X+1} - \delta_{X+1} \epsilon_X) / (\delta_X - \delta_{X+1}) \quad (7)$$

based on systematically improved Dunning's aug-cc-pVXZ basis sets (e.g., aug-cc-pVDZ and aug-cc-pVTZ or, prefer-

ably, aug-cc-pVTZ and aug-cc-pVQZ). The Min et al. extrapolation can utilize two or more arbitrary basis sets.<sup>11</sup> These methods were developed in Kim's laboratory, and they will be referenced as Kim's extrapolation methods from now on.

The determination of the  $\Delta\text{CCSD(T)}$  correction term, needed in eq 1, is computationally expensive despite the fact that it is usually calculated in a small or medium size basis set. It is known that this term converges faster than MP2 interaction energy itself. About 300 CCSD(T)/CBS stabilization energies collected in the S22,<sup>12</sup> S26–07,<sup>13</sup> JCSH2005,<sup>12</sup> and BEGDB<sup>14</sup> databases were determined in the above-mentioned way, using the Helgaker's extrapolation. One very important observation follows from these also referred to as the “benchmark” data. The  $\Delta\text{CCSD(T)}$  correction term is almost negligible for H-bonded complexes while being systematically repulsive (up to 3.5 kcal/mol or more) for stacked complexes. Therefore, MP2/CBS results for stacked structures are strongly overbinding, with relative errors of as much as 50%. The question now remains whether the assumption of fast convergence, which is of key importance for the accurate evaluation of the CCSD(T)/CBS interaction energies for extended complexes, is fulfilled not only for model complexes (see, e.g. ref 12) but also for larger real life examples. Evidently, the only way to test this assumption is to carry out a “brute-force” attempt and perform the very time-consuming CCSD(T) calculations in extended basis sets for the systems of interest.

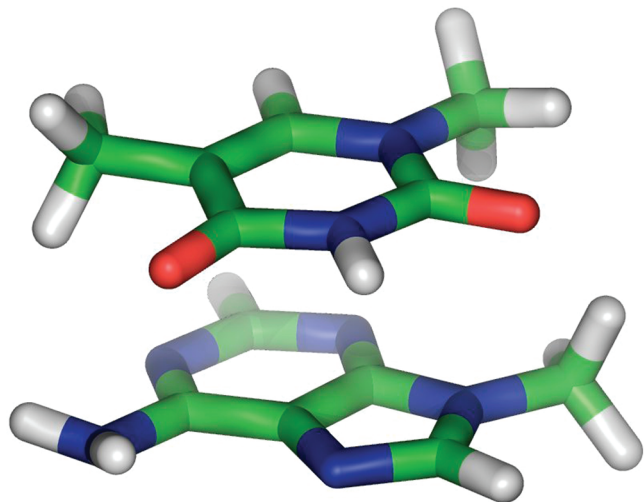
The benzene dimer is the most extensively investigated model of  $\pi$  interaction at high levels of correlation and large basis sets (up to aug-cc-pVQZ).<sup>15–17,19</sup> The total  $\Delta\text{CCSD(T)}$  term for the stacked dimer exhibits fast and monotonic convergence. Interestingly, its two components  $\Delta\text{CCSD}$  and  $\Delta\text{(T)}$  converge individually and less rapidly indicating a cancellation between these terms. Similar calculations on the uracil dimer<sup>19</sup> at the aug-cc-pVTZ level confirm this conclusion. However, the  $\Delta\text{CCSD(T)}$  correction is rather modest in these systems (below 2 kcal/mol), and it is desirable to check it for systems where it is larger.

Among all the stacked DNA base pairs included in the S22 and JCSH2005 data sets, the largest  $\Delta\text{CCSD(T)}$  correction term, about 3.6 kcal/mol, was for the 9-methyl adenine (mA)–1-methyl thymine (mT) stacked complex. This complex is quite large (110 correlated electrons), so the previous calculations were done with the 6-31G\*(0.25) basis set only.

In this work, we are trying to approach the CCSD(T)/CBS for the stacked complex of mA–mT by applying the aug-cc-pVDZ and aug-cc-pVTZ basis sets and extrapolating the results to the complete basis set limit, using standard extrapolation techniques. Besides these benchmark calculations, we also tested the performance of the MP2.5 procedure.

## 2. Methodology

The geometry of the stacked mA–mT complex (see Figure 1) was taken from the S22 data set where it was determined by counterpoise-corrected all-coordinates gradient optimization at the MP2/cc-pVTZ level. It has been shown previously<sup>20</sup> that these geometries are close to the CCSD(T) ones.



**Figure 1.** Structure of the methyl adenine–methyl thymine complex.

To investigate the performance of other “small” basis sets which were used for calculation of other extended (stacked) systems, the CCSD(T) calculations were performed also with 6-31G\*, 6-31G\*(0.25), 6-31G\*\*, 6-31G\*\*(0.25), 6-31G\*\*(0.25,0.15), and 6-31G\*\*(diff aDZ) as well as with 6-31+G\*, 6-31+G\*(0.25), 6-31+G\*\*, 6-31+G\*\*(0.25), and 6-31+G\*\*(0.25,0.15) (Table 1 provides a complete list; note that all calculations involve spherical harmonic basis sets). Exponents of the *d* (for heavy elements, e.g. C, N, O) and *p* (hydrogen) polarization functions, in the 0.25 and (0.25,0.15) basis sets, were changed from default values to more diffuse ones (0.8→0.25 and 1.1→0.15), following the recommendations of van Lenthe et al.<sup>21</sup> Similarly, in the 6-31G\*\*(diff aDZ) basis set, these exponents of the polarization functions were changed to the most diffuse ones from aug-cc-pVDZ (e.g., 0.8→0.151 and 1.1→0.141). All correlated calculations use the frozen core approximation (the 1s orbital of the C, N, and O atoms). The interaction energy was, in all cases, corrected for the basis set superposition error using Boys–Bernardi<sup>22</sup> counterpoise correction.

Calculations involving the largest aug-cc-pVXZ (*X* = D, T) basis sets were performed using the PQS program package<sup>23</sup> with a recently developed parallel CCSD(T) module<sup>24,25</sup> that utilizes a parallel I/O filesystem.<sup>26</sup> This code was designed for an efficient computation of CCSD(T) energies for very large systems. The linear dependency in the basis set was not removed for this part of the calculations, but a tight integral threshold was applied ( $10^{-14}$ ) in order to avoid numerical difficulties that may be expected in such an extended system, where the overlap matrix has small eigenvalues.

All other MP3 and CCSD(T) calculations were carried out using the MOLCAS package.<sup>27,28</sup> This program uses Cholesky decomposed (CD) two-electron integrals in the Hartree–Fock module with no local exchange screening.<sup>29</sup> The MP3 and CCSD(T) calculations used a new, highly parallelized closed-shell code<sup>30</sup> and also made use of the CD decomposition of two-electron integrals. For all calculations, a threshold of the order of  $10^{-5}$  was used for the two-electron integrals

decomposition; in our experience, this gives an accuracy of higher than 0.01 kcal/mol.

### 3. Results and Conclusions

The total CCSD(T) stabilization energy (absolute value of the interaction energy) of mA–mT is strongly dependent on the basis set. As shown in the Table 1, going from aug-cc-pVDZ (618 basis functions), where the CCSD(T) calculations are already out of the range of most CCSD(T) codes on standard computer architectures, to aug-cc-pVTZ (1311 basis functions), the stabilization energy increases by as much as ~1.5 kcal/mol.

A rough measure, of whether a basis set is saturated for calculation of the interaction energy, is a comparison of the BSSE corrected and uncorrected values. In aug-cc-pVTZ, these two values at the CCSD(T) level differ by ~4.4 kcal/mol. Furthermore, even for such a large basis set as the aug-cc-pVQZ, the corrected and uncorrected stabilization energies at the MP2 level still differ by 2 kcal/mol. It is very important to use diffuse basis functions. For instance, the CCSD(T) stabilization energy obtained using the cc-pVTZ-f basis set (the standard Dunning’s cc-pVTZ basis set without *f* functions) is ~1.8 kcal/mol lower than at the aug-cc-pVDZ level, i.e., a basis set of the same size. This is also apparent by comparing the results obtained, for instance, with the 6-31+G\*\* and 6-31+G\*\*(0.25,0.15) basis sets. They differ by 3.75 kcal/mol.

Further analysis of results in Tables 1 and 2 clearly indicates where this basis set dependence comes from. The HF results (see Table 1) are repulsive and show almost no basis set dependence, amounting to 7.31 kcal/mol in the least diffuse basis set (6-31G\*) and to 7.05 kcal/mol in the most diffuse and extended aug-cc-pVQZ basis set. The effect of the BSSE correction is significant even at this level of theory, being most pronounced in the smallest but very diffuse 6-31G\*\*(diff aDZ) basis set at ~10 kcal/mol (in absolute value), and leads to artificial stabilization of the complex. Let us skip the MP2 method for a while and focus on the effect of correlation beyond MP2, i.e.,  $\Delta$ CCSD(T) shown in Table 2.

The basis set dependence of the  $\Delta$ CCSD(T) correction term is significantly smaller than the changes in the total CCSD(T) stabilization energy. The dependence of the  $\Delta$ CCSD(T) on the diffuseness of basis set is reversed for the 6-31G\*\*, 6-31+G\*, and 6-31+G\*\* family of basis sets but increases in going from the aug-cc-pVDZ to the aug-cc-pVTZ basis. Surprisingly, not even the ordering of the BSSE corrected and uncorrected values of  $\Delta$ CCSD(T) is uniform. This situation is similar to that of the stacked structure of the uracil dimer.<sup>19</sup> The maximum scatter of the  $\Delta$ CCSD(T) values is much smaller compared to the total CCSD(T) values ~1.3 vs ~8.6 kcal/mol or ~30 vs 66% of the total combination. What differs tremendously is the absolute value of the difference between the BSSE corrected and uncorrected values. The largest  $\Delta$ CCSD(T) difference is ~0.8 kcal/mol for the 6-31+G\*\*(0.25,0.15) basis set, while the largest difference for the total CCSD(T) values is more than 27 kcal/mol. For the highest quality  $\Delta$ CCSD(T) and CCSD(T) values (aug-cc-pVTZ), these differences are

**Table 1.** Basis Set Dependence of HF, MP2, and CCSD(T) Total Stabilization Energies for the mA–mT Complex<sup>a,b</sup>

basis set	no. of AOs	HF	MP2	CCSD(T)
6-31G* <sup>c</sup>	324	−7.31 (−2.59)	7.65 (17.49)	4.24 (14.15)
6-31G*(0.25)	324	−6.99 (0.66)	12.89 (30.21)	9.32 (26.85)
6-31G**	369	−7.20 (−2.45)	8.10 (17.89)	4.60 (14.41)
6-31G**(0.25)	369	−6.91 (0.69)	13.07 (30.12)	9.48 (26.67)
6-31G**(0.25,0.15)	369	−6.98 (1.69)	13.63 (34.00)	10.28 (30.88)
6-31G**(diff aDZ)	369	−7.22 (2.86)	13.23 (40.56)	10.23 (37.92)
6-31+G*	408	−7.10 (−4.73)	10.61 (21.42)	7.05 (18.48)
6-31+G*(0.25)	408	−6.59 (−1.99)	14.80 (31.35)	11.46 (28.67)
6-31+G**	453	−7.17 (−4.73)	11.09 (22.06)	7.41 (19.05)
6-31+G**(0.25)	453	−6.72 (−2.35)	14.06 (30.15)	10.63 (27.31)
6-31+G**(0.25,0.15)	453	−6.77 (−0.98)	14.36 (34.75)	11.16 (32.33)
cc-pVTZ-f	618	−7.14 (−5.02)	14.20 (21.19)	9.82 (17.04)
aug-cc-pVDZ	618	−7.11 (−4.43)	15.77 (27.00)	11.61 (23.31)
aug-cc-pVTZ	1311	−7.03 (−6.43)	17.31 (21.92)	13.06 (17.44)
aug-cc-pVQZ	2370	−7.05 (−6.89)	17.79 (19.70)	—

<sup>a</sup> Total stabilization energies in kcal/mol; negative values are repulsive. <sup>b</sup> Values in parentheses are stabilization energies without BSSE correction. <sup>c</sup> Spherical harmonics (5d, 7f,...) are used in all basis sets.

**Table 2.** Basis Set Dependence of  $\Delta$ MP3 ( $\Delta$ MP2.5) and  $\Delta$ CCSD(T) Correction to the Interaction Energy (in kcal/mol)<sup>a</sup>

basis set	$\Delta$ MP3	$\Delta$ CCSD	$\Delta$ CCSD(T)	$c_{\text{opt}}^b$	$\Delta$ MP2.5 error <sup>c</sup>
6-31G*	5.93 (6.77)	5.33 (6.10)	3.41 (3.34)	0.58	0.45
6-31G*(0.25)	7.60 (8.45)	6.30 (7.02)	3.57 (3.36)	0.47	−0.23
6-31G**	6.11 (6.98)	5.49 (6.31)	3.50 (3.48)	0.57	0.44
6-31G**(0.25)	7.67 (8.58)	6.38 (7.17)	3.59 (3.45)	0.47	−0.25
6-31G**(0.25,0.15)	7.67 (8.48)	6.29 (7.10)	3.36 (3.13)	0.44	−0.48
6-31G**(diff aDZ)	7.12 (7.92)	5.90 (6.97)	2.99 (2.64)	0.42	−0.57
6-31+G*	6.95 (7.41)	6.01 (6.71)	3.56 (2.94)	0.51	0.09
6-31+G*(0.25)	7.79 (7.82)	6.34 (6.66)	3.33 (2.68)	0.43	−0.56
6-31+G**	7.15 (7.63)	6.20 (6.94)	3.68 (3.01)	0.52	0.10
6-31+G**(0.25)	7.87 (7.98)	6.44 (6.83)	3.43 (2.84)	0.44	−0.50
6-31+G**(0.25,0.15)	7.81 (7.80)	6.33 (6.71)	3.20 (2.42)	0.41	−0.71
cc-pVTZ-f	8.67 (9.01)	7.34 (7.78)	4.38 (4.15)	0.50	0.05
aug-cc-pVDZ	8.94 (9.43)	7.48 (8.20)	4.16 (3.69)	0.47	−0.31
aug-cc-pVTZ	9.08 (9.46)	7.93 (8.45)	4.25 (4.47)	0.47	−0.29

<sup>a</sup> Values in parentheses are contributions without correction for the BSSE. <sup>b</sup>  $c_{\text{opt}} = \Delta\text{CCSD(T)}/\Delta\text{MP3}$ . <sup>c</sup>  $\Delta\text{MP2.5 error} = \Delta\text{CCSD(T)} - 0.5^* \Delta\text{MP3}$ .

~0.2 vs ~4.4 kcal/mol. The difference between the  $\Delta\text{CCSD(T)}$  calculated in the better performing small basis set, 6-31G\*\*(0.25), and the “reference”, aug-cc-pVTZ, is ~0.7 kcal/mol, while for the better performing “medium” basis set, 6-31+G\*\*, it is ~0.6 kcal/mol. This represents the limit of accuracy of the “estimated” CCSD(T)/CBS calculated according to the eq 1 using the small basis set for the  $\Delta\text{CCSD(T)}$  correction. Another option is to stay with the “plain” CCSD(T) values, which would, however, lead to much larger errors, ~2.8 kcal/mol for better performing small basis set 6-31G\*\*(0.25,0.15) and 1.6 kcal/mol for better performing “medium” basis 6-31+G\*\*(0.25,0.15).

The largest basis set dependence is obtained for the MP2 correlation energy. Going from the aug-cc-pVTZ to aug-cc-pVQZ, the stabilization energy still increases by almost 0.5 kcal/mol, while the difference between the BSSE corrected and uncorrected values decreases from ~4.6 to ~1.9 kcal/mol. For comparison, going from aug-cc-pVDZ to aug-cc-pVTZ, the  $\Delta\text{CCSD(T)}$  correction term increases only by ~0.1 kcal/mol and the difference between the BSSE corrected and uncorrected values reduces from ~−0.5 to ~0.2 kcal/mol. Considering the slow convergence of MP2 and the rather fast one of  $\Delta\text{CCSD(T)}$ , the composite scheme eq 1 should be, in general, successful.

Table 2 includes the  $\Delta\text{CCSD}$  corrections. Comparing them with the  $\Delta\text{CCSD(T)}$  values, it is clear that there is significant cancellation between the  $\Delta\text{CCSD}$  and  $\Delta(T)$  contributions.

Table 2 also shows the MP3 correction term,  $\Delta\text{MP3}$ . Comparing with  $\Delta\text{CCSD(T)}$ ,  $\Delta\text{MP3}$  is too repulsive by 42–59%, see the “ $c_{\text{opt}}$ ” column in the table. Scaling the  $\Delta\text{MP3}$  correction by 1/2, i.e., doing MP2.5, results are in absolute errors of ~0.7–~0.05 kcal/mol relative to the  $\Delta\text{CCSD(T)}$  value, depending on the basis set. The error of MP2.5 is indicated by the difference between the optimum scaling coefficient and 0.5 for the particular structure and basis set. Some trends of “ $c_{\text{opt}}$ ”, at least for the nucleic acid base pairs, have already been discussed previously.<sup>1</sup> The optimum scaling coefficient is closer to 0.4 for small and diffuse basis sets and closer to (or larger than) 0.5 for less diffuse basis sets. Interestingly, the basis set dependence of the higher-order correlation terms, e.g.,  $\Delta\text{MP3}$  and  $\Delta\text{CCSD(T)}$ , for mA–mT is strong enough to make the MP2.5 method in larger basis sets (aug-cc-pVDZ or aug-cc-pVTZ) more accurate than the estimated CCSD(T)/CBS value, calculated using the exact  $\Delta\text{CCSD(T)}$  correction obtained by small/medium basis sets (e.g., 6-31G\*/6-31G\*\*- or 6-31+G\*/6-31+G\*\*-type). The error of MP2.5 is determined by the absolute value of the  $\Delta\text{MP3}$  term. According to our experience from the S22 test set,<sup>1</sup> which is quite



**Table 3.** Comparison of the MP3, CCSD, and (T) Wall Clock Timings<sup>a</sup>

basis set	setup	PP/CPP	MP3 <sup>b</sup>	CCSD <sup>b</sup>	(T) <sup>b</sup>
6-31G**(0.25,0.15)	6 x A, 4 x B	14/4	0.19	5.93	—
6-31G**(0.25,0.15)	12 x C	24/4	0.05	1.23	5.23
6-31+G**(0.25,0.15)	12 x C	24/4	0.10	2.47	15.75
cc-pVTZ-f	1 x B	1/8	3.78	—	—
cc-pVTZ-f	7 x B	7/8	—	31.50	488.00

<sup>a</sup> The following symbols were used for node architectures: “A” Intel Core2 Quad (4 cores), 2.40 GHz, “B” Intel Xeon E5345 (8 cores), 2.33 GHz, and “C” 2x Quad-Core AMD Opteron 2354, 2.20 GHz. The column setup explains how the machines were utilized. “PP” stands for a number of MPI parallel processes, while “CPP” stands for a number of cores utilized by threaded-BLAS routines per one MPI process. In all the CCSD calculations listed below, 21 iterations were necessary for the convergence. <sup>b</sup> Results are in hours.

diverse in terms of intermolecular interaction types, the optimum value of the scaling coefficient lies between 0.4 and 0.6. This means that the absolute error of the  $\Delta$ MP2.5 correction is generally within 10% of the  $\Delta$ MP3 value. This error is comparable with or lower than the basis set effect on  $\Delta$ CCSD(T).

Timings shown in Table 3 illustrate computational savings of the MP3 method.

MP3 is significantly faster than CCSD despite the same asymptotic scaling of these methods with the system size. The computer time is saved by skipping numerous nonlinear terms from the CCSD equations. However, the major difference is that the CCSD method is iterative, while MP3 is equivalent to a single configuration interaction doubles (CID) iteration. This is especially beneficial in parallel runs. The savings of MP2.5 with respect to CCSD(T) result mainly from its lower asymptotic scaling,  $O(N^6)$  vs  $O(N^7)$ . This makes MP2.5, although it lacks a solid theoretical basis, an efficient alternative for CCSD(T) when the latter exceeds available computational resources. Another problem not yet fully recognized by the computational chemistry community is that the numerical accuracy of the (T) combination obtained using double precision arithmetic may be insufficient due to the very large number arithmetic operations in its evaluation. This is particularly the case if diffuse basis sets are utilized.

For the sake of completeness, we also report a few details involving our calculation for the biggest basis sets. The computation for the dimer in the aug-cc-pVTZ basis set was performed employing a 40 Xeon E5430 with 2.66 GHz nodes, each node with 8 processing cores and 16 GB of memory. The wall clock time for the most expensive part, the (T) correction, was 160 h. Calculation of energies of the monomers in the basis set of the dimer needed for counterpoise correction took less than 120 h on 10 nodes. The basis set dimension for this system is 1311, and the number of correlated occupied orbitals is 55.

To obtain new CCSD(T) benchmark values for the stabilization energy of mA–mT, we performed Helgaker and Kim's type of extrapolations, shown in Table 4.

For both types of extrapolations, two variants are presented. First, rows “aDZ→aTZ” in Table 4 correspond to extrapolations of the total CCSD(T) correlation energies

**Table 4.** Helgaker's and Kim's Extrapolation of the Total MP2 and CCSD(T) Stabilization Energies and  $\Delta$ CCSD(T) Correction Term (Repulsive) Calculated from Given Values<sup>e</sup>

extrapolation	MP2	CCSD(T)	$\Delta$ CCSD(T)
Helgaker aDZ→aTZ <sup>a</sup>	17.95	13.66	4.29
Helgaker aTZ→aQZ <sup>b</sup>	17.99	13.70	4.29
Kim aDZ→aTZ <sup>c</sup>	18.38	13.92	4.45
Kim aTZ→aQZ <sup>d</sup>	18.14	13.74	4.40

<sup>a</sup> HF/aQZ + corr. CCSD(T)/(aDZ→aTZ). <sup>b</sup> HF/aQZ + corr. (MP2/X +  $\Delta$ CCSD(T)/X – 1)/(aTZ→aQZ). <sup>c</sup> Total CCSD(T)/(D→T). <sup>d</sup> (Total MP2/X +  $\Delta$ CCSD(T)/X – 1)/(aTZ→aQZ). <sup>e</sup> Extrapolation values calculated from the values outlined in footnotes b–e. For details see text; aXZ stands for Dunning's aug-cc-pVXZ (X = D, T, Q) basis sets.

obtained by the aug-cc-pVDZ and aug-cc-pVTZ basis set. In Helgaker's extrapolation, the HF energy was taken from the aug-cc-pVQZ basis set, which was considered converged to the CBS limit. The values in the second row, labeled as “aTZ→aQZ”, were constructed by combining the MP2 correlation (or in case of Kim's extrapolation total) energies with the  $\Delta$ CCSD(T) obtained in the basis set of one cardinality lower, i.e., MP2/aug-cc-pVQZ +  $\Delta$ CCSD(T)/aug-cc-pVTZ, and analogously for the aug-cc-pVTZ and aug-cc-pVDZ pair.

Assuming that the “aTZ→aQZ” extrapolation for both types is closer to the real CBS limit, the total CCSD(T) stabilization energies from the Helgaker and Kim schemes differ by less than 0.05 kcal/mol, both being  $\sim$ 13.7 kcal/mol, while the  $\Delta$ CCSD(T) of the both correlation schemes agree within  $\sim$ 0.1 kcal/mol, both being  $\sim$ 4.30 kcal/mol.

## 4. Summary

(1) The  $\Delta$ CCSD(T) correction term determined from calculations using extended basis sets (aug-cc-pVDZ and aug-cc-pVTZ) is 4.3 kcal/mol, which is  $\sim$ 0.7 kcal/mol more than that determined using small basis sets (e.g., 3.6 kcal/mol in 6-31G\*(0.25)). This means that the CCSD(T)/CBS values in the S22, S26–07, and JCSH2005 data sets (where the CCSD(T) correction term is determined using small or medium basis sets) are reasonably reliable. However, it indicates that the CBS value of  $\Delta$ CCSD(T) increases by 10–20% (at most), if evaluated using large basis sets.

(2) The  $\Delta$ CCSD(T) correction converges more rapidly than the  $\Delta$ CCSD and  $\Delta$ (T) corrections separately, showing a partial cancellation between these terms.

(3) Helgaker- and Kim-types of extrapolations from extended basis sets (aug-cc-pVXZ, X = D, T, Q) are in good mutual agreement (within 0.1 kcal/mol), which supports the universality and robustness of both schemes.

(4) The MP2.5 method provides excellent values for stabilization energies at a much lower cost and, unlike the SCS-MP2 method<sup>2</sup> and its variants, yields the correct asymptotic behavior for dispersion. Its computational cost, although higher than MP2 based methods, is reasonable, and it is recommended for extended molecular complexes. Based on calculations of other systems, the MP2.5 methods accuracy is expected to be within 10% of the  $\Delta$ MP3 correction.

**Acknowledgment.** This work was supported by grants no. LC512 and MSM6198959216 from the Ministry of Education, Youth, and Sports (MŠMT) of the Czech Republic and was part of research project no. Z4 055 0506. It was also supported by the Slovak Research and Development Agency (contract no. APVV-20-018405) and by grant CHE 0515922 of the U.S. National Science Foundation. The authors wish to acknowledge the support of Praemium Academiae of the Academy of Sciences of the Czech Republic, awarded to P.H. in 2007. Generous computation time from the Star of Arkansas supercomputer, which was purchased in part by funds obtained under grant no. MRE 072265 of the U.S. National Science Foundation, is gratefully acknowledged.

### References

- (1) Pitoňák, M.; Neogrády, P.; Černý, J.; Grimme, S.; Hobza, P. *ChemPhysChem* **2009**, *10*, 282.
- (2) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095.
- (3) Jung, Y. S.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. *J. Chem. Phys.* **2004**, *121*, 9793.
- (4) Distasio, R. A.; Head-Gordon, M. *Mol. Phys.* **2007**, *105*, 1073.
- (5) Hill, J. G.; Platts, J. A. *J. Chem. Theory Comput.* **2007**, *3*, 80.
- (6) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463.
- (7) Jurečka, P.; Černý, J.; Hobza, P.; Salahub, D. R. *J. Comput. Chem.* **2007**, *28*, 555.
- (8) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.
- (9) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243.
- (10) Lee, E. C.; Kim, D.; Jurečka, P.; Tarakeshwar, P.; Hobza, P.; Kim, K. S. *J. Phys. Chem. A* **2007**, *111*, 3446.
- (11) Min, S. K.; Lee, E. C.; Lee, H. M.; Kim, D. Y.; Kim, D.; Kim, K. S. *J. Comput. Chem.* **2008**, *29*, 1208.
- (12) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.
- (13) Riley, K. E.; Hobza, P. *J. Phys. Chem. A* **2007**, *111*, 8257.
- (14) Řezáč, J.; Jurečka, P.; Riley, K. E.; Černý, J.; Valdes, H.; Pluháčková, K.; Berka, K.; Řezáč, T.; Pitoňák, M.; Vondrášek, J.; Hobza, P. *Collect. Czech. Chem. Commun.* **2008**, *73*, 1261.
- (15) Sinnokrot, M. O.; Sherrill, C. D. *J. Chem. Phys.* **2006**, *110*, 10656–10668.
- (16) Hill, J. G.; Platts, J. A.; Werner, H.-J. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4072–4078.
- (17) Janowski, T.; Pulay, P. *Chem. Phys. Lett.* **2007**, *447*, 27–32.
- (18) Pitoňák, M.; Neogrády, P.; Řezáč, J.; Jurečka, P.; Urban, M.; Hobza, P. *J. Chem. Theory Comput.* **2008**, *4*, 1829.
- (19) Pitoňák, M.; Riley, K. E.; Neogrády, P.; Hobza, P. *ChemPhysChem* **2008**, *9*, 1636.
- (20) Dabkowska, I.; Jurečka, P.; Hobza, P. *J. Chem. Phys.* **2005**, *122*, 204322.
- (21) van Lenthe, J. H.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Duijneveldt, F. B. *Weakly Bonded Systems. In Advances in Chemical Physics; Volume LXIX; Prigogine, I., Rice, S. A., Eds.; John Wiley & Sons Ltd.: 1987.*
- (22) Boys, S. F.; Bernardi, F. *Mol. Phys.* **2002**, *100*, 65.
- (23) *PQS version 3.2; Parallel Quantum Solutions: 2013 Green Acres Road, Fayetteville, Arkansas 72703.*
- (24) Janowski, T.; Ford, A. R.; Pulay, P. *J. Chem. Theory Comput.* **2007**, *3*, 1368.
- (25) Janowski, T.; Pulay, P. *J. Chem. Theory Comput.* **2008**, *4*, 1585.
- (26) Ford, A. R.; Janowski, T.; Pulay, P. *J. Comput. Chem.* **2007**, *28*, 1215.
- (27) Karlstrom, G.; Lindh, R.; Malmqvist, P.-Å.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P.-O.; Cossi, M.; Schimmelpfennig, B.; Neogrády, P.; Seijo, L. *Comput. Mater. Sci.* **2003**, *28*, 222.
- (28) Aquilante, F.; De Vico, L.; Ferre, N.; Malmqvist, P.-Å.; Neogrády, P.; Pedersen, T.; Pitoňák, M.; Reiner, M.; Roos, B.; Serrano-Andres, L.; Urban, M.; Veryazov, V.; Lindh, R. *J. Comput. Chem.* . in press.
- (29) Aquilante, F.; Pedersen, T. B.; Lindh, R. *J. Chem. Phys.* **2007**, *126*, 194106.
- (30) Neogrády, P.; Aquilante, F.; Noga, J.; Pitoňák, M.; Hobza, P.; Urban, M. manuscript in preparation.

CT900126Q