

Describing Noncovalent Interactions beyond the Common Approximations: How Accurate Is the “Gold Standard,” CCSD(T) at the Complete Basis Set Limit?

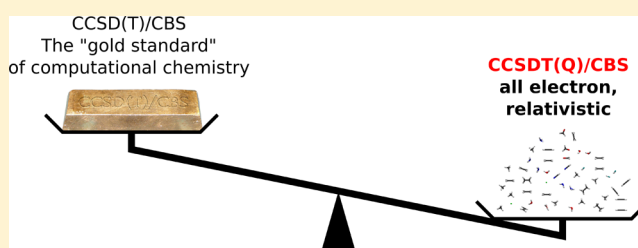
Jan Řezáč^{*,†} and Pavel Hobza^{†,‡}

[†]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, 166 10 Prague, Czech Republic

[‡]Regional centre of Advanced Technologies and Materials, Department of Physical Chemistry, Palacky University, 771 46 Olomouc, Czech Republic

S Supporting Information

ABSTRACT: We have quantified the effects of approximations usually made even in accurate CCSD(T)/CBS calculations of noncovalent interactions, often considered as the “gold standard” of computational chemistry. We have investigated the effect of excitation series truncation, frozen core approximation, and relativistic effects in a set of 24 model complexes. The final CCSD(T) results at the complete basis set limit with corrections to these approximations are the most accurate estimate of the true interaction energies in noncovalent complexes available. The average error due to these



approximations was found to be about 1.5% of the interaction energy.

A theoretical and computational description of noncovalent interactions is not an easy task because fine quantum mechanical effects must be properly evaluated if accurate and reliable results are sought. Only the recent development of computational methods, their implementation, and hardware make it possible to calculate noncovalent interactions in small molecules routinely with an estimated accuracy of units of percent. It is generally agreed that at such a level the correlation energy should include triple excitations and a very large basis must be used.¹ The most efficient way to satisfy these requirements is the composite CCSD(T)/CBS approach where accurate extrapolation to the complete basis set limit (CBS) is made at a lower (MP2) level and higher-order terms are calculated using coupled clusters with singles, doubles, and perturbative triples, CCSD(T), in a basis set as large as possible.² Due to its advantages and widespread use, this method has become the “gold standard” in the field. It has been used to produce a rather large body of benchmark data which now is widely used for the validation of other methods and parametrization of empirical models.^{3,4} Nevertheless, the accuracy of the CCSD(T)/CBS approach itself has not been studied sufficiently. Some information is available on the convergence of the results with a basis set size, but this makes it possible only to estimate the error of CCSD(T)/CBS using given combination of basis sets with respect to a more accurate CCSD(T)/CBS calculation.⁵ The other approximations involved in these calculations are usually neglected and not discussed at all. While some studies go beyond this level and report more accurate calculations,^{6–8} a systematic investigation of the effect of these approximations has not been carried out yet. In this work, we focus solely on the interaction energy

because it is the quantity used for comparison of computational methods in the field of noncovalent interactions. We are working on accurate calculations of vibrations, which will allow us to pass to the experimentally observable interaction enthalpy.

We should now identify the most important approximations used: (1) The most obvious one is the incompleteness of the basis set. The CBS limit can be only approached but never reached, and the computational cost of the calculation grows steeply when the size of the basis set is increased. A more accurate estimate of the CBS limit can be obtained by extrapolation; for accurate results, however, calculations in large basis sets are still a necessary prerequisite. (2) The second approximation is the incomplete evaluation of higher excitations. The CCSD(T) method recovers a vast majority of the correlation energy, but the contribution of higher excitations cannot be neglected. Here, a full configuration interaction (FCI) calculation is theoretically possible, but in practice it can be realized in systems with only few electrons because of the extremely steep scaling of the computational complexity. (3) In most calculations, the frozen core approximation is used to reduce the computational cost. In this approach, only the valence electrons are correlated. The contribution of the core electrons (core–core and core–valence correlation) is small but should be considered if we seek the best accuracy possible. (4) In lighter elements, such as in all organic molecules, the relativistic effects are usually

Received: January 23, 2013

Published: April 4, 2013

Table 1. Accurate CCSD(T)/CBS Interaction Energies (in kcal/mol) and the Corrections Arising from Core Correlation, Relativistic Effects, and Inclusion of Quadruple Excitations (in kcal/mol and Percent of the Interaction Energy) in the A24 Set of Noncovalent Complexes, and the Total Corrected Interaction Energy, and Its Deviation (in %) from the CCSD(T)/CBS Value

	system		CCSD(T)/CBS	core correlation (%)	relativity (%)	Δ CCSDT(Q) (%)	total (%)
1	water...ammonia	C_s	-6.493	-0.041 (0.63)	0.009 (0.14)	0.001 (0.01)	-6.524 (0.48)
2	water dimer	C_s	-5.006	-0.032 (0.64)	0.013 (0.25)	0.012 (0.23)	-5.014 (0.15)
3	HCN dimer	C_s	-4.745	-0.018 (0.38)	0.007 (0.15)	0.007 (0.15)	-4.749 (0.08)
4	HF dimer	C_s	-4.581	-0.025 (0.54)	0.017 (0.36)	0.017 (0.38)	-4.572 (0.20)
5	ammonia dimer	C_{2h}	-3.137	-0.017 (0.55)	0.001 (0.04)	-0.004 (0.13)	-3.157 (0.64)
6	HF...methane	C_{3v}	-1.654	-0.020 (1.22)	0.002 (0.14)	-0.006 (0.37)	-1.679 (1.46)
7	ammonia...methane	C_{3v}	-0.765	-0.008 (0.98)	0.000 (0.03)	-0.006 (0.80)	-0.779 (1.81)
8	water...methane	C_s	-0.663	-0.006 (0.90)	0.000 (0.04)	-0.002 (0.38)	-0.672 (1.24)
9	formaldehyde dimer	C_s	-4.554	-0.004 (0.08)	0.009 (0.20)	0.075 (1.65)	-4.474 (1.76)
10	water...ethene	C_s	-2.557	-0.014 (0.56)	0.000 (0.00)	-0.007 (0.27)	-2.578 (0.83)
11	formaldehyde...ethene	C_s	-1.621	-0.006 (0.36)	0.000 (0.02)	-0.002 (0.10)	-1.629 (0.49)
12	ethyne dimer	C_{2v}	-1.524	-0.005 (0.31)	-0.003 (0.22)	-0.005 (0.34)	-1.537 (0.87)
13	ammonia...ethene	C_s	-1.374	-0.007 (0.50)	-0.001 (0.06)	-0.008 (0.59)	-1.389 (1.15)
14	ethene dimer	C_{2v}	-1.090	-0.003 (0.28)	-0.001 (0.08)	-0.016 (1.47)	-1.110 (1.83)
15	methane...ethene	C_s	-0.502	-0.003 (0.69)	0.000 (0.07)	-0.007 (1.47)	-0.514 (2.23)
16	borane...methane	C_s	-1.485	-0.004 (0.28)	-0.001 (0.05)	-0.028 (1.90)	-1.518 (2.22)
17	methane...ethane	C_s	-0.827	0.000 (0.02)	-0.001 (0.07)	-0.009 (1.07)	-0.837 (1.16)
18	methane...ethane	C_s	-0.607	0.000 (0.05)	-0.001 (0.09)	-0.007 (1.22)	-0.615 (1.26)
19	methane dimer	D_{3d}	-0.533	0.001 (0.15)	0.000 (0.08)	-0.006 (1.16)	-0.538 (1.09)
20	Ar...methane	C_{3v}	-0.405	-0.002 (0.37)	-0.002 (0.42)	0.000 (0.01)	-0.408 (0.80)
21	Ar...ethene	C_{2v}	-0.364	-0.002 (0.65)	-0.002 (0.68)	-0.001 (0.16)	-0.370 (1.48)
22	ethene...ethyne	C_{2v}	0.821	-0.011 (1.33)	0.001 (0.07)	-0.027 (3.31)	0.784 (4.57)
23	ethene dimer	D_{2h}	0.934	-0.012 (1.31)	0.001 (0.07)	-0.025 (2.70)	0.897 (3.94)
24	ethyne dimer	D_{2h}	1.115	-0.011 (0.95)	0.001 (0.10)	-0.031 (2.74)	1.075 (3.59)
	average			0.57	0.14	0.94	1.47

neglected, although their magnitude is comparable to the other approximations discussed here.

In this work, we quantify the effect of all these approximations on noncovalent interactions using the best means currently possible. We do so for an extended set of model noncovalent complexes, which allows us to draw more general conclusions using statistics. The goal is to quantify the accuracy of the CCSD(T)/CBS calculations in comparison to more realistic results. Understanding the accuracy of the method commonly used as a benchmark is necessary for a proper handling of the benchmark data. It should be noted that we have recently found that CCSD[T] might describe noncovalent interactions better than CCSD(T) when compared to higher-order methods.⁹ This work brings new data relevant to this topic, but here we will focus only on the more widespread CCSD(T) method; the full comparison will be discussed in more detail in a separate work.

For this study, we have built a set of 24 noncovalent complexes featuring interactions of a different nature, covering hydrogen bonds, mixed electrostatics/dispersion, and dispersion-dominated interactions (including π - π stacking). To make accurate calculations possible, only small model systems can be used. The complexes, listed in Table 1 and shown in Figure 1, contain up to four second-group element atoms and take advantage of a symmetry to reduce the cost of the calculations. We do not consider rare gas dimers because they are too simple to serve as a model for molecular complexes.

To evaluate the error associated with the approximation discussed above, we use the methods described here (technical details are provided later). (1) The starting point is a CCSD(T) calculation extrapolated to the CBS limit from the largest basis

sets possible (up to quintuple-zeta) where all the other approximations are employed. This is also used as a reference for the less accurate CCSD(T)/CBS estimates used for larger systems. (2) To evaluate the contribution of higher-order excitations, we use the best method applicable to complexes of this size, coupled clusters with perturbative quadruples—CCSDT(Q). It had been shown to produce results very close to full CCSDTQ⁹ and recover a major part of the correlation missing in CCSD(T).¹⁰ (3) To evaluate the effect of the frozen core approximation, the CCSD(T) calculations are repeated with all electrons correlated. (4) Relativistic effects are calculated using the scalar fourth-order Douglas-Kroll-Hess Hamiltonian¹¹ at the CCSD(T) level with all electrons correlated.

The results presented in this work form the first comprehensive data set of interaction energies calculated with such a high accuracy. We hope that it will serve as a valuable tool for a future benchmarking of the most accurate computational methods. To make the use of the data set more convenient, we have made the data available in the online database at www.begdb.com¹² as the A24 set.

METHODS

Geometries. The complexes have been optimized at a very high level so that the geometry itself can serve as a benchmark. We have used a composite CCSD(T)/CBS scheme identical to the setup used previously for interaction energies: HF energy is calculated in the aug-cc-pVQZ basis set; MP2 correlation energy is extrapolated¹³ from aug-cc-pVTZ and aug-cc-pVQZ basis sets and the higher-order contribution Δ CCSD(T), defined as $E^{\text{CCSD(T)}} - E^{\text{MP2}}$, is calculated in the aug-cc-pVDZ

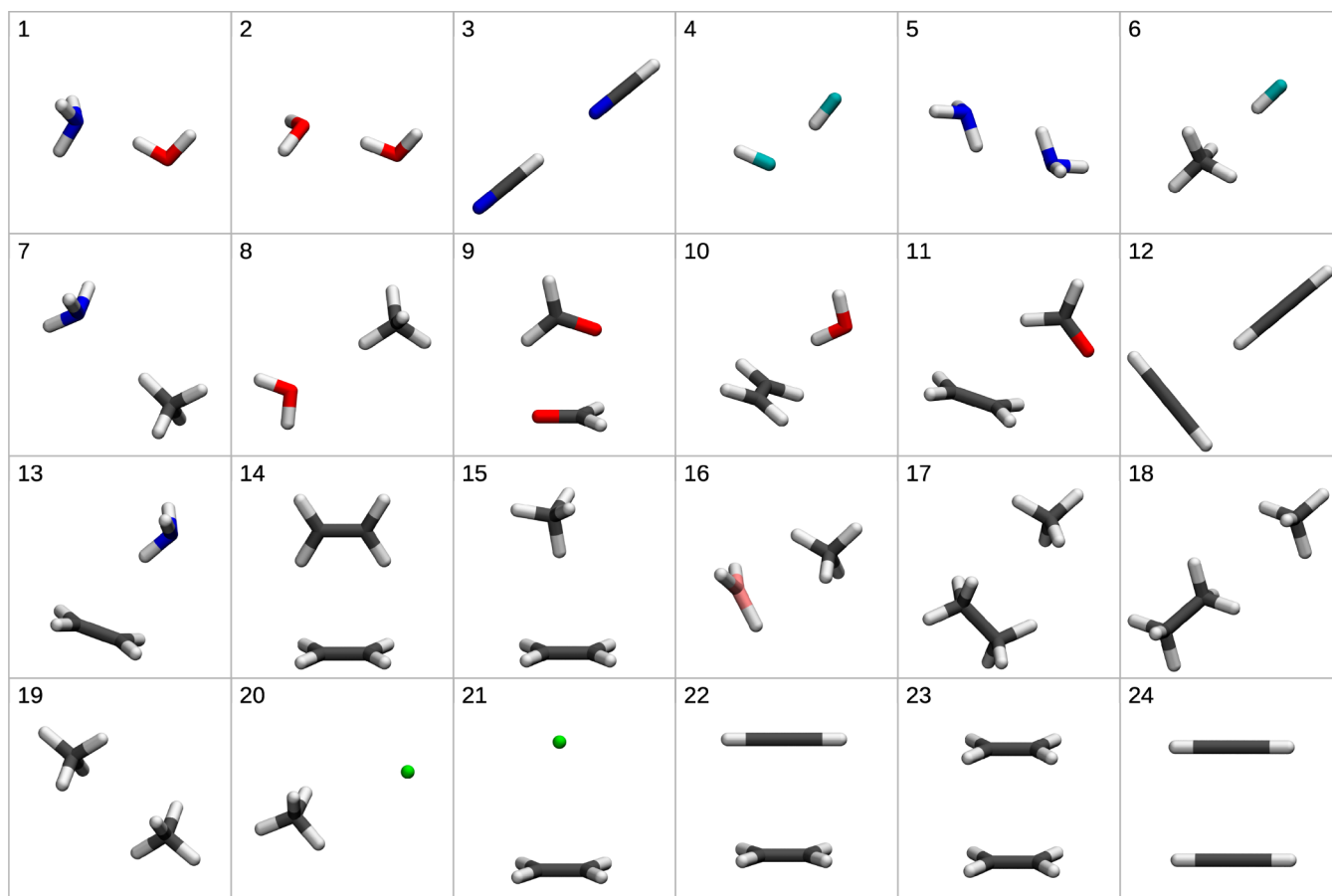


Figure 1. Structure of the 24 noncovalent complexes investigated in this study (the A24 set): hydrogen bonds (1–5), mixed interactions (6–15), and dispersion dominated (16–24).

basis set. In all calculations, counterpoise correction is used to eliminate the basis set superposition error (BSSE). We have tested the accuracy of this scheme on complexes 2 and 19 (models for hydrogen bond and dispersion interaction). Intermolecular distance was compared to one interpolated from CCSD(T)/CBS calculations extrapolated from aug-cc-pVTZ and aug-cc-pVQZ basis sets. The difference is very small (0.4 and 0.003% of the distance) and affects the interaction energies only negligibly (by less than 1 cal/mol).

On this potential, optimization was carried out using a numerical gradient. Tight convergence criteria were used to ensure a high quality of the geometries (gradient norm of 0.006 kcal/mol/Å). Optimizations were started from nonsymmetric geometries in order to avoid optimization to a saddle point. After each optimization, the structure was symmetrized, and when the symmetrization led to a non-negligible change of geometry, the optimization was repeated. The symmetry of the optimized complexes is reported in Table 1.

The three complexes intended as a model of π - π stacking (complexes 22–24) are an exception. At this computational level, the stacked structures are not minima and optimize to geometries featuring CH- π interactions (such as complexes 11, 12, and 14). Because the π - π interactions are important and should be included in such a set, we have built stacked geometries from optimized monomers, placing them at a fixed distance of 3.5 Å (the sum of the van der Waals radii of carbon atoms) in the most symmetric arrangement (see Figure 1). In these cases, the overall interaction is repulsive due to an unfavorable orientation of the molecular quadrupoles, but the

contribution of the correlation energy is attractive and of a magnitude similar to the other systems studied.

CCSD(T) Calculations. The CCSD(T) calculations were performed in the aug-cc-pVXZ (X = T, Q, 5) basis sets.¹⁴ The complete basis set was estimated by fitting the three CCSD(T) correlation energies with the function

$$E_{\text{corr},N} = E_{\text{corr,CBS}} + aN^b \quad (1)$$

where N is the cardinal number of the basis set and a and b are the fitted parameters. This three-point fit is more general than the more common two-point fit¹³ where the exponent b is fixed. The free exponent adapts to each system individually which should lead to more accurate results (although the difference is small). The Hartree–Fock energy is taken from the aug-cc-pV5Z calculation and is not extrapolated.

The core correlation has been calculated as the difference between all-electron and frozen-core CCSD(T) correlation energies extrapolated¹³ from polarized-core basis sets aug-cc-pCVTZ and aug-cc-pCVQZ. Here, the quintuple-zeta calculations (aug-cc-pCV5Z) were not feasible.

CCSDT(Q) Calculations. The all-electron CCSDT(Q) calculations have been performed with the 6-31G**(0.25, 0.15) basis set, a modified 6-31G** where the exponent of the polarization function on hydrogen is set to 0.15 and in other elements decreased to 0.25. This basis set was found to yield excellent results for noncovalent interactions,¹⁵ while it is small enough to make these calculations possible.

Relativistic Effects. The relativistic effects were evaluated using fourth-order Douglas–Kroll–Hess (DKH) Hamiltonian¹¹ in an all-electron CCSD(T) calculation using the aug-cc-pCVQZ-DK basis set. The fourth order DKH method provides accurate, parameter-free treatment of the relativistic effects.

Computational Details. All the calculations presented here use tight convergence criteria (10^{-10} au) in both SCF and CC procedures. Counterpoise correction is applied in order to remove the basis set superposition error. All the CCSD(T) calculations were performed in MOLPRO 2010.¹⁶ The CCSDT(Q) calculations were carried out in the MRCC program^{17,18} coupled to CFOUR.¹⁹

■ RESULTS AND DISCUSSION

The results are summarized in Table 1. The CCSD(T)/CBS interaction energies range from -6.5 kcal/mol (the strongest hydrogen bond) down to -0.4 kcal/mol (in the complexes of an argon atom with hydrocarbons); positive values were obtained in the π – π stacks where an attractive geometry is not available. It must be noted that the correlation part of the interaction energy is still negative and amounts to -2.576 , -2.936 , and -2.356 kcal/mol in the systems 22, 23, and 24.

As expected, the inclusion of the core correlation in the all-electron calculations stabilizes all of the noncovalent complexes. The effect is the strongest in hydrogen bonds, where there are very close contacts between the molecules. The average (mean signed) error introduced by the frozen-core approximation is -0.011 kcal/mol; in relative measures (the average of these errors relative to the CCSD(T)/CBS interaction energy), it amounts to 0.57% (in this work, we preferentially use the average relative errors, because the interaction energies in the set vary significantly; the use of error measures in absolute units would overly emphasize the strongest interactions such as hydrogen bonds).

The relativistic effects are the smallest of the contributions investigated, because the systems studied contain only light atoms. The relativistic effects range from -0.003 to 0.017 kcal/mol; the average relative error caused by omitting them is 0.14%.

The most important contribution missing in the CCSD(T)/CBS calculations is the higher-order excitations in the correlation treatment. In this case, we cannot eliminate the approximation completely, because such calculations are not tractable. As we need a reasonably large basis set to obtain reliable results, we extend our CC calculations by only one order to CCSDT(Q), where triple excitations are treated iteratively and quadruples perturbatively. As we have shown recently, it provides a substantial improvement over CCSDT and yields interaction energies very close to full CCSDTQ calculations.⁹ It has also been shown on the total energies of diatomic molecules that the CCSDT(Q) method recovers approximately 95% of the correlation energy missing in CCSDT, using CCSDTQP as a reference.¹⁰ This means that the CCSDT(Q) level provides a sufficiently accurate estimate of the higher-order effects in general. This contribution, Δ CCSDT(Q), is evaluated as a difference between the CCSDT(Q) and CCSD(T) correlation energies calculated in the 6-31G** (0.25, 0.15) basis set. The largest value of 0.075 kcal/mol is observed in the formaldehyde dimer, the strongest complex among the larger ones in the set. The largest relative differences are found in the stacked complexes, because the change in correlation energy is made relative to the overall

interaction energy, which contains the repulsive HF component. When the values are made relative to the correlation energy only, there is no systematic preference of the stacked complexes, and the contributions are practically random, regardless of the nature of the interaction. The average relative error associated with the neglect of the quadruple excitations is 0.94% (please note that this term is only an estimate because of the size of the basis set used).

When all of the corrections are combined, the resulting average relative error of 1.47% is slightly smaller than the sum of the individual average errors because partial error compensation occurs.

Until now, we have used a very accurate estimate of the CBS limit in our comparison. In most practical calculations, we cannot afford calculations in such a large basis set, and an additional error is introduced due to the inaccuracy of the CBS extrapolation from smaller basis sets. There are many options that would lead to different results; here we will test the affordable composite scheme that we have found to work best for noncovalent interactions. It is the method that we use for the production of our extended benchmark data sets such as S66^{4,20} and X40.²¹ Here, HF energy is calculated in the aug-cc-pVQZ basis set. MP2 correlation energy is extrapolated from aug-cc-pVTZ and aug-cc-pVQZ basis sets, and the Δ CCSD(T) term is calculated in aug-cc-pVTZ without diffuse functions on hydrogens (all calculations use frozen core approximation). When applied to the 24 complexes presented here, the average error compared to the accurate CBS limit (without further corrections) is 1.04%. This excellent accuracy is a result of favorable error compensation; other combinations of the basis sets of a comparable size yield larger errors (we will provide more details in a dedicated paper). Nevertheless, it is a good indicator of the accuracy of practically applicable methods, provided that a careful choice of basis sets has been made.

The same results can be compared to the most accurate total interaction energies including all the additional corrections. This would show more clearly how far the common benchmark calculations are from the reality. Surprisingly, the average error here is even lower (although negligibly), namely 0.94%. On the other hand, the agreement is not as even as indicated by the increase of the root-mean-square error from 0.017 kcal/mol in the case of uncorrected CBS reference to 0.022 kcal/mol. These results indicate that all the effects beyond CCSD(T)/CBS are so small that they are lost in the noise of the inaccurate CBS estimation and can be, like in this case, even compensated favorably.

■ CONCLUSIONS

We have reported accurate CCSD(T)/CBS interaction energies and further corrections toward an all-electron, relativistic CCSDT(Q) calculation for a set of 24 noncovalent complexes (introduced as the A24 data set).

Among the corrections beyond CCSD(T)/CBS, the contribution of higher excitations to the correlation energy has the largest effect. Evaluated here as the difference between CCSDT(Q) and CCSD(T), it amounts on average to 0.94% of the interaction energy. The second largest difference is brought by passing from frozen core approximation to all-electron calculations, with the average difference being 0.57%. The relativistic effects are less important in the studied systems (which consist of light elements only) with an average difference of 0.14%. Overall, these three effects change the interactions by 1.47%; this number should be considered as the

error of an ideal CCSD(T)/CBS calculation with respect to more realistic values.

We have also shown that more efficient estimates of the CCSD(T)/CBS interaction energies could yield results in excellent agreement (around 1%) with the most accurate calculations. This confirms the high accuracy of the benchmark data sets, such as S66 and X40, calculated using such a setup.

■ ASSOCIATED CONTENT

■ Supporting Information

Optimized geometries of the complexes used in this work are provided. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rezac@uochb.cas.cz.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was part of the Research Project RVO: 61388963 of the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic. The support of Praemium Academiae, Academy of Sciences of the Czech Republic, awarded to P.H. in 2007 is also acknowledged. This work was also supported by the Czech Science Foundation [P208/12/G016] and the operational program Research and Development for Innovations of the European Social Fund (CZ 1.05/2.1.00/03/0058).

■ REFERENCES

- (1) Hobza, P. *Acc. Chem. Res.* **2012**, *45*, 663–672.
- (2) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Am. Chem. Soc.* **2002**, *124*, 104–112.
- (3) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.
- (4) Řezáč, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2011**, *7*, 2427–2438.
- (5) Marshall, M. S.; Burns, L. A.; Sherrill, C. D. *J. Chem. Phys.* **2011**, *135*, 194102–194102–10.
- (6) Pittner, J.; Hobza, P. *Chem. Phys. Lett.* **2004**, *390*, 496–499.
- (7) Jankowski, P.; McKellar, A. R. W.; Szalewicz, K. *Science* **2012**, *336*, 1147–1150.
- (8) Harding, M. E.; Klopper, W. *ChemPhysChem* **2012**, DOI: 10.1002/cphc.201200834.
- (9) Řezáč, J.; Šimová, L.; Hobza, P. *J. Chem. Theory Comput.* **2013**, *9*, 364–369.
- (10) Bomble, Y. J.; Stanton, J. F.; Kállay, M.; Gauss, J. *J. Chem. Phys.* **2005**, *123*, 054101–054101–8.
- (11) Reiher, M.; Wolf, A. *J. Chem. Phys.* **2004**, *121*, 10945–10956.
- (12) 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*, 1261–1270.
- (13) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243–252.
- (14) Woon, D. E.; Dunning, T. H. *J. Chem. Phys.* **1994**, *100*, 2975.
- (15) Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, *99*, 3247–3276.
- (16) Werner, H. J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Lindh, R.; Mitrushenkov, A.; Rauhut, G.; Shamasundar, K. R.; Adler, T. B.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Goll, E.; Hampel, C.; Hesselmann, A.; Hetzer, G.; Hrenar, T.; Jansen, G.; Köppl, C.; Liu, Y.; Lloyd, A. W.; Mata, R. A.; May, A. J.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; O'Neill, D. P.;

Palmieri, P.; Pflüger, K.; Pitzer, R.; Reiher, M.; Shiozaki, T.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T.; Wang, M.; Wolf, A. *MOLPRO*, version 2010.1; Cardiff University: Cardiff, U. K.; Universität Stuttgart: Stuttgart, Germany, 2010.

(17) Kállay, M. *MRCC*, a string-based many-body program. See also www.mrcc.hu (accessed Jan. 2013).

(18) Kállay, M.; Surján, P. R. *J. Chem. Phys.* **2001**, *115*, 2945–2954.

(19) Stanton, J. F.; Gauss, J.; Harding, M. E.; Szalay, P. G.; et al. *CFOUR*, a quantum chemical program package. For the current version, see www.cfour.de (accessed Jan. 2013).

(20) Řezáč, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2011**, *7*, 3466–3470.

(21) Řezáč, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2012**, *8*, 4285–4292.