

On the Optimization of the DF-SAPT-DFT Interaction Energies of Hydrogen-bonded Systems

JIŘÍ CZERNEK

Department of Bioanalogous and Special Polymers
 Institute of Macromolecular Chemistry
 Heyrovsky Square 2, 16206 Prague
 THE CZECH REPUBLIC
 czernek@imc.cas.cz <http://www.imc.cas.cz>

Abstract: The DF-SAPT-DFT interaction energies of 15 hydrogen-bonded systems were confronted with their supermolecular counterparts obtained by combining the RI-MP2/CBS and CCSD(T)/aug-cc-pVDZ results. The former were assumed to approach the CBS limit by the power law, whose exponent was estimated by fitting the two sets of interaction energies in the least-squares sense.

Key-Words: optimization, *ab initio*, DFT, interaction energy, hydrogen bonds, SAPT.

1 Introduction

The quantitative description of the properties of hydrogen-bonded complexes is vital for our understanding of many phenomena involving physical, chemical and biological processes in molecules and clusters. Thus, considerable effort has been spent on studying the characteristics of hydrogen-bonded systems experimentally [1] and also by applying a variety of theoretical techniques [2]. Generally, an application of the wavefunction-based quantum chemical methods which take into account electron correlation effects (beyond-Hartree-Fock, beyond-HF, methods) [3] is necessary to accurately describe the structure and energetics of hydrogen bonds, because an alternative treatment employing the standard techniques of the density-functional theory (DFT) [4] has been shown to suffer from the improper description of the London dispersion contribution to the intermolecular interactions (see [5] for the most recent discussion), and the HF method does not cover the dispersion forces at all. However, the application of beyond-HF methods together with sufficiently large basis sets is currently possible only for relatively small systems, their size being dependent on the level of sophistication of a given technique and not exceeding ca. 30 atoms in the case of highly accurate coupled-cluster [3] (CC) calculations. Fortunately, due to the recent progress (see review article [6]) in combining the cost-effective DFT description of monomer properties with the symmetry-adapted perturbation theory (SAPT) of intermonomer interactions, an important alternative strategy for modeling hydrogen-bonded systems emerged. In particular, the density-fitting (DF) approximation within this approach [7] makes it possible to study the interactions in sizeable (containing 50 atoms or more)

systems with the reliability of the results comparable to that of the CC methods [8], [9]. In this paper, two sophisticated approaches for an assessment of intermolecular interaction energies are combined. The first one is the supermolecular calculation of the sum of 1) the MP2 [3] (second-order Møller-Plesset perturbation theory) results extrapolated to the complete basis set (CBS) limit and of 2) the correction term, which approximates the higher-order correlation contributions (see Methods for details and reference [10] for a thorough discussion). This strategy has been successfully employed to obtain accurate interaction energies of relatively large systems including, for example, base pairs of nucleic acids [11], [12], [13] or complexes of benzene with hexafluorobenzene [14] and nitrobenzene [15]. The second approach involves the DF-SAPT-DFT calculations with two small augmented correlation-consistent basis sets (*cf.* Methods), which can be performed for even larger complexes. The set of the DF-SAPT-DFT data is subsequently fitted using a simple one-parameter form (Equation (4) below) to the supermolecular results for 15 hydrogen-bonded dimers. Importantly, the fit is shown to be robust and the resulting expression works reasonably well for a different kind of hydrogen-bonded systems not included in the original data set, thus capturing the convergence of the DF-SAPT-DFT results towards the most accurate data.

2 Methods

The training set consists of the systems 1 – 15, which can be subdivided into three structural classes. Thus, the water homodimer 1, the water-methanol dimer 2, the methanol homodimer 3 and the methanol-phenol dimer

4 all feature one strong O–H...O hydrogen bond. The next group contains the five formamide dimers first studied by Vargas et al. [16] and later by Frey and Leutwyler [17] and denoted as FA1 – FA5 (here 5 – 9 accordingly) together with the structurally similar *N*-methylacetamide dimer 10 (see ref. [16] for details). The last group comprises fluorine-containing systems, *i.e.*, the hydrogen fluoride homodimer 11, the HF–water dimer 12, the HF–hydrogen cyanide dimer 13, the HF–methanol dimer 14, and the C_{2h} -symmetric arrangement of the fluorobenzene homodimer 15 (the structure designated A in ref. [18]). The testing system is the coplanar complex of uracil and 4,6-diamino-1,3,5-triazine, which features three hydrogen bonds and is the model structure important in the description of hydrogen-bonded polymers [19]. It will be further referred to as 16.

All the structures 1 – 16 were subjected to the full geometrical optimization at the MP2(Frozen Core)/aug-cc-pVDZ level and the resulting stationary points were verified to be minima by calculating the harmonic vibrational frequencies (all real for each structure). The default algorithms and settings of Gaussian 03 [20] suite of quantum chemical programs were used. The coordinates of the structures can be obtained from the author upon request. The supermolecular calculations of the MP2 interaction energies of the complexes 1 – 16 were performed in the resolution of the identity (RI) integral approximation [21] using the TURBOMOLE V5-7-1 program package [22]. The frozen-core approximation and the augmented correlation-consistent polarized-valence X -tuple zeta basis sets, $X = 2$ (DZ), 3 (TZ) and 4 (QZ), where X is the cardinal number associated with each basis set, were applied in the RI-MP2 framework [23]. The complete basis set (CBS) estimates of the MP2 interaction energies were obtained using the mixed Gaussian/exponential form [24]:

$$E(X) = E_{\text{CBS}} + b \exp[-(X-1)] + c \exp[-(X-1)^2] \quad (1)$$

The parameters b and c and, most importantly, the basis set limit value of the RI-MP2 energy, E_{CBS} , are determined uniquely from the three energies. Thus, the CBS-extrapolated MP2 interaction energies, $\Delta E_{\text{MP2}}^{\infty}$, were obtained as the difference of the E_{CBS} of the dimer and of the E_{CBS} data for the monomers, which were corrected for the basis set superposition error (BSSE) employing the scheme of Boys and Bernardi [25]. The correction term approximating the contribution to the correlation energy not covered by the MP2 treatment, ΔE_{corr} , was obtained by taking the difference between the result provided by the coupled-cluster method with singles, doubles and noniterative triples excitations

[CCSD(T)] and by the canonical MP2 calculation, both values computed with the standard aug-cc-pVDZ basis set [26] using the default algorithms of the Molpro program package [27], [28]. As in the case of the RI-MP2 calculations, the frozen-core approximation and the counterpoise correction for the reduction of the BSSE [25] were adopted. The sum of the $\Delta E_{\text{MP2}}^{\infty}$ and ΔE_{corr} terms calculated as described above is often a satisfactory estimate of the CCSD(T) interaction energy at the basis set limit, $\Delta E_{\text{CCSD(T)}}^{\infty}$.

The DF-SAPT-DFT method was used as implemented in the Molpro 2006.1 code [28]. The PBE0AC asymptotically corrected exchange-correlation functional [7] was applied and two sets of results were obtained, the first one by applying the aug-cc-pVDZ atomic orbitals basis set [26], the cc-pVTZ auxiliary JK-fitting basis set [29] and the aug-cc-pVDZ MP2-fitting basis set [30], and the second utilizing, respectively, the analogous aug-cc-pVTZ, cc-pVQZ and aug-cc-pVTZ basis sets (references [26], [29] and [30] accordingly). The adiabatic local density approximation was adopted in the perturbation treatment of the second-order contributions, and the gradient-controlled shift procedure was employed [31] (the values of the differences between the vertical ionization potential and the energy of the highest molecular orbital for each system and the given combination of the basis sets can be obtained from the author upon request). Thus, the results calculated with the smaller and larger basis set will be denoted as $\Delta E_{\text{DFT}}^{\text{DZ}}$ and $\Delta E_{\text{DFT}}^{\text{TZ}}$, respectively. They were optimized to best-fit the supermolecular $\Delta E_{\text{CCSD(T)}}^{\infty}$ values as follows. Assuming that the DF-SAPT-DFT results approach their CBS limit, $\Delta E_{\text{DFT}}^{\infty}$, by the power law:

$$\Delta E_{\text{DFT}}^{XZ} = \Delta E_{\text{DFT}}^{\infty} + AX^{-\alpha} \quad (2)$$

and considering those obtained for the cardinal numbers $X = 2$ and 3 (*i.e.*, the $\Delta E_{\text{DFT}}^{\text{DZ}}$ and $\Delta E_{\text{DFT}}^{\text{TZ}}$ data), we eliminate the linear parameter A to get

$$\Delta E_{\text{DFT}}^{\infty} = \frac{\Delta E_{\text{DFT}}^{\text{DZ}} 2^{\alpha} - \Delta E_{\text{DFT}}^{\text{TZ}} 3^{\alpha}}{2^{\alpha} - 3^{\alpha}} \quad (3)$$

Then the optimal value of the parameter α in the least-squares sense is given by the solution to the minimization problem

$$\min_{\alpha} f: f = \sum_{i=1}^{15} \left(\frac{\Delta E_{\text{DFT}}^{\text{DZ}}(i) 2^{\alpha} - \Delta E_{\text{DFT}}^{\text{TZ}}(i) 3^{\alpha}}{2^{\alpha} - 3^{\alpha}} - \Delta E_{\text{CCSD(T)}}^{\infty}(i) \right)^2 \quad (4)$$

with i going over the energy-data for the complexes 1 – 15. The error bars for each i point representing the 95% confidence intervals on the parameter a were generated by the standard procedure [32], [33] on the assumption of Student's t distribution of errors employing the residuals and the covariance matrix as provided by the commercial NAG® FORTRAN Numerical Library routines E04FCF and E04YCF, respectively.

3 Results and Discussion

3.1 The Supermolecular Results

Table 1 summarizes the values of the interaction energies, which were obtained for the complexes 1 – 16 by the variational supermolecular approaches as described in Methods section. It should be mentioned that while the ΔE_{corr} values might seem to be small (maximum absolute value 1.91 kJ/mol for the complex 4, maximum relative value 6.7% in the case of the formamide dimer designated FA5 in ref. [17] and 9 in this work), they differ in sign and are expected to bring the MP2/CBS data to an almost-perfect agreement with the most accurate theoretical results. For example, a series of the CCSD(T)/aug-cc-pVXZ, $X = 2, 3$, and 4, interaction energies was calculated for the hydrogen fluoride homodimer 11, which were extrapolated using Equation (1) to obtain an alternative estimate of the CCSD(T)/CBS value. Resulting $\Delta E = -19.80$ kJ/mol lies closer to the $\Delta E_{\text{CCSD(T)}}$ than to the ΔE_{MP2} interaction energy thanks to the correct sign of the (small) ΔE_{corr} .

complex	$\Delta E_{\text{MP2}}^{\infty}$	ΔE_{corr}	$\Delta E_{\text{CCSD(T)}}^{\infty}$
1	-21.33	+0.21	-21.12
2	-21.37	+0.11	-21.26
3	-25.16	+0.44	-24.73
4	-37.30	+1.91	-35.38
5	-68.80	-0.36	-69.15
6	-45.62	-1.10	-46.72
7	-35.13	-0.06	-35.18
8	-30.86	-0.14	-31.00
9	-23.35	-1.70	-25.06
10	-75.48	+0.02	-75.46
11	-19.38	-0.12	-19.50
12	-38.11	+0.44	-37.66
13	-32.03	+1.38	-30.65
14	-44.31	+0.82	-43.48
15	-10.76	+0.02	-10.75
16	-88.44	+0.36	-88.08

Table 1. The interaction energies (in kJ/mol) obtained from the supermolecular calculations. See the text for details.

3.2 The DF-SAPT-DFT and Optimized Results

The data resulting from the DF-SAPT-DFT calculations and from the optimization process are collected in Table 2. Thus, the $\Delta E_{\text{DFT}}^{\text{DZ}}$ and $\Delta E_{\text{DFT}}^{\text{TZ}}$ interaction energies are shown, which were employed (together with the set of the $\Delta E_{\text{CCSD(T)}}^{\infty}$ values) to solve the problem as formulated by Equation (4). The value of $a = 1.90$ was obtained by an application of the nonlinear fitting procedure and used to get the $\Delta E_{\text{DFT}}^{\infty}$ results from Equation (3). The residuals of the fit, *i.e.*, the $\Delta(i) = \Delta E_{\text{DFT}}^{\infty}(i) - \Delta E_{\text{CCSD(T)}}^{\infty}(i)$ differences for integer $i = 1, 2, \dots, 15$ are also given in Table 2 together with the Δ value for the complex 16 not included in the parameter estimation. A highly accurate fit of the DF-SAPT-DFT data to their supermolecular counterparts was obtained (the maximum and average absolute errors amount to 1.43 and 0.56 kJ/mol, respectively, and the average relative error is just 1.6%). Significantly, the model described by Equation (4) was applied to a wide range of hydrogen-bonded systems, with various atoms involved, and with the interaction energies spanning the interval from ca. 10 to almost 80 kJ/mol. The model is robust: from an inspection of Figure 1 it is apparent that no outliers are present in the data. Moreover, the $\Delta E_{\text{DFT}}^{\infty}$ result extrapolated for the dimer 16, while lying outside the parametrized range of the interaction energies, is in a fairly good agreement with the $\Delta E_{\text{CCSD(T)}}^{\infty}$ value (the absolute error of about one half of kcal/mol).

complex	$\Delta E_{\text{DFT}}^{\text{DZ}}$	$\Delta E_{\text{DFT}}^{\text{TZ}}$	$\Delta E_{\text{DFT}}^{\infty}$	Δ
1	-18.22	-19.87	-21.30	-0.18
2	-18.18	-19.86	-21.31	-0.05
3	-20.37	-22.35	-24.04	+0.69
4	-28.82	-31.72	-34.22	+1.16
5	-61.39	-65.56	-69.16	-0.01
6	-41.86	-44.73	-47.21	-0.49
7	-29.66	-35.25	-34.47	+0.71
8	-27.72	-29.67	-31.34	-0.34
9	-21.97	-23.65	-25.11	-0.05
10	-68.57	-73.04	-76.09	-1.43
11	-16.35	-17.75	-18.96	+0.54
12	-32.87	-35.54	-37.84	-0.18
13	-26.75	-29.16	-31.23	-0.58
14	-37.27	-40.21	-42.74	+0.74
15	-7.87	-8.75	-9.51	+1.24
16	-76.70	-81.60	-85.80	+2.28

Table 2. The DF-SAPT-DFT-based interaction energies and the $\Delta E_{\text{DFT}}^{\infty} - \Delta E_{\text{CCSD(T)}}^{\infty}$ differences, Δ (all values are in kJ/mol). See the text for details.

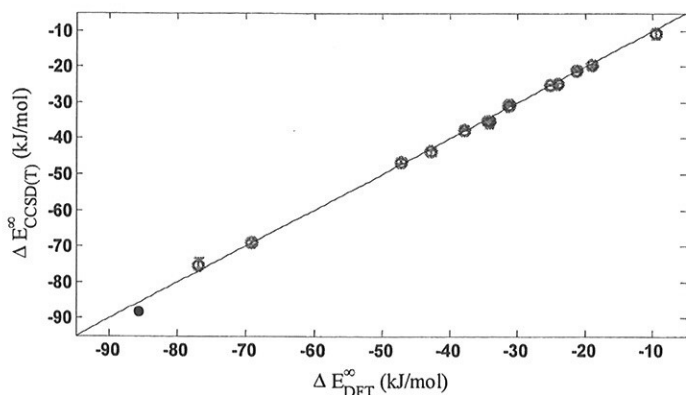


Figure 1. The supermolecular vs. optimized results (open circles, with the error bars shown in red; the filled circle represents the point not included in the parameter estimation, and the line would correspond to a perfect fit). See the text for details.

4 Conclusion

The DF-SAPT-DFT interaction energies of the training set of 15 hydrogen-bonded complexes of various structural types, which covered the range of values from ca. 10 to almost 80 kJ/mol, were confronted with their $\Delta E_{\text{CCSD(T)}}^{\infty}$ supermolecular counterparts estimated by combining the RI-MP2/CBS and CCSD(T)/aug-cc-pVDZ results. Thus, the DF-SAPT-DFT data calculated with two smaller basis sets were assumed to approach the CBS limit by the power law and successfully fitted to the model (Equation (4)) to minimize the differences from the $\Delta E_{\text{CCSD(T)}}^{\infty}$ results. The testing complex **16**, not included in the parameter estimation, was described by the model well within the chemical accuracy (*cf.* the discussion of this term in ref. [34]).

Acknowledgements

This research has been supported by the Academy of Sciences of the Czech Republic (Grant IAA400500602) and by the Ministry of Education, Youths and Sports of the Czech Republic (Project 2B08021). Time allocation in the Czech Academic Supercomputer Centre is gratefully acknowledged.

References:

[1] D. M. P. Mingos, *Supramolecular Assembly via Hydrogen Bonds*, Springer, 2004.
 [2] A. Cavalli, P. Carloni, M. Recanatini, Target-related applications of first principles quantum chemical methods in drug design, *Chemical Reviews*, Vol. 106, No. 9, 2006, pp. 3497-3519.
 [3] C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*, Wiley, 2004.

[4] W. Koch, M. C. Holthausen, *A Chemist's Guide to Density Functional Theory*, Wiley, 2000.

[5] A. Krishtal, K. Vanommeslaeghe, A. Olsz, T. Veszprémi, C. V. Alsenoy, P. Geerlings, Accurate interaction energies at density functional theory level by means of an efficient dispersion correction, *Journal of Chemical Physics*, Vol. 130, No. 17, 2009, Art. No. 174101.

[6] K. Szalewicz, K. Patkowski, B. Jezierski, Intermolecular Interactions via perturbation theory: From diatoms to biomolecules, *Structure and Bonding*, Vol. 116, 2005, pp. 43-117.

[7] A. Heßelmann, G. Jansen, M. Schütz, Density-functional theory-symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies, *Journal of Chemical Physics*, Vol. 122, No. 1, 2005, Art. No. 014103.

[8] A. Heßelmann, G. Jansen, M. Schütz, Interaction energy contributions of H-bonded and stacked structures of the AT and GC DNA base pairs from the combined density functional theory and symmetry-adapted perturbation theory approach, *Journal of the American Chemical Society*, Vol. 128, No. 36, 2006, pp. 11730-11731.

[9] A. Fiethen, G. Jansen, A. Heßelmann, M. Schütz, Stacking energies for average B-DNA structures from the combined density functional theory and symmetry-adapted perturbation theory approach, *Journal of the American Chemical Society*, Vol. 130, No. 6, 2008, pp. 1802-1803.

[10] T. Janowski, P. Pulay, High accuracy benchmark calculations on the benzene dimer potential energy surface, *Chemical Physics Letters*, Vol. 447, No. 1-3, 2007, pp. 27-32.

[11] P. Jurečka, J. Šponer, J. Černý, P. Hobza, Benchmark database of accurate (MP2 and CCSD(T) basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs, *Physical Chemistry Chemical Physics*, Vol. 8, No. 117, 2006, pp. 1985-1993.

[12] J. A. Frey, A. Müller, M. Losada, S. Leutwyler, Isomers of the uracil dimer: An ab initio benchmark study, *Journal of Physical Chemistry B*, Vol. 111, No. 13, 2007, pp. 3534-3542.

[13] M. L. Leininger, I. M. B. Nielsen, M. E. Colvin, C. L. Janssen, Accurate structures and binding energies for stacked uracil dimers, *Journal of Physical Chemistry A*, Vol. 106, No. 15, 2002, pp. 3850-3854.

[14] S. Tsuzuki, T. Uchimaru, M. Mikami, Intermolecular interaction between hexafluorobenzene and benzene: Ab initio calculations including CCSD(T) level electron correlation correction, *Journal of Physical Chemistry A*, Vol. 110, No. 5, 2006, pp. 2027-2033.

- [15] S. Tsuzuki, K. Honda, T. Uchimaru, M. Mikami, Intermolecular interaction of nitrobenzene-benzene and nitrobenzene dimer: Significant stabilization of slipped-parallel orientation by dispersion interaction, *Journal of Chemical Physics*, Vol. 125, No. 12, 2006, Art. No. 124304.
- [16] R. Vargas, J. Garza, R. A. Friesner, H. Stern, B. P. Hay, D. A. Dixon, Strength of the N-H...O=C and C-H...O=C bonds in formamide and *N*-methylacetamide dimers, *Journal of Physical Chemistry A*, Vol. 105, No. 20, 2001, pp. 4963-4968.
- [17] J. A. Frey, S. Leutwyler, An ab initio benchmark study of hydrogen bonded formamide dimers, *Journal of Physical Chemistry A*, Vol. 110, No. 45, 2001, pp. 12512-12518.
- [18] J. D. Dunitz, W. B. Schweizer, Molecular pair analysis: C-H...F interactions in the crystal structure of fluorobenzene? And related matters, *Chemistry - A European Journal*, Vol. 12, No. 26, 2006, pp. 6804-6815.
- [19] W. H. Binder, R. Zirbs, Supramolecular polymer networks with hydrogen bonds in the main- and side-chain, *Advances in Polymer Science*, Vol. 207, 2007, pp. 1-78.
- [20] M. J. Frisch et al., Gaussian 03, Revision C02. Gaussian, Inc., Wallingford CT, 2004.
- [21] F. Weigend, M. Häser, RI-MP2: first derivatives and global consistency, *Theoretical Chemistry Accounts*, Vol. 97, Nos. 1-4, 1997, 143-152.
- [22] R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, Electronic-structure calculations on workstation computers - the program system TURBOMOLE, Vol. 162, No. 3, 1989, pp. 165-169.
- [23] F. Weigend, M. Häser, H. Patzelt, R. Ahlrichs, RI-MP2: optimized auxiliary basis sets and demonstration of efficiency, *Chemical Physics Letters*, Vol. 294, No. 1-3, 1998, pp. 143-152.
- [24] K. A. Peterson, D. E. Woon, T. H. Dunning, Jr., Benchmark calculations with correlated wave-functions, *Journal of Chemical Physics*, Vol. 100, No. 10, 1994, 7410-7415.
- [25] S. F. Boys, F. Bernardi, Calculation of small molecular interactions by differences of separate total energies - some procedures with reduced errors, *Molecular Physics*, Vol. 19, No. 4, 1970, 553-557.
- [26] E. Woon, T. H. Dunning, Jr., Gaussian-basis sets for use in correlated molecular calculations, *Journal of Chemical Physics*, Vol. 103, No. 11, 1995, 4572-4585.
- [27] C. Hampel, K. A. Peterson, H.-J. Werner, A comparison of the efficiency and accuracy of the quadratic configuration-interaction (QCISD), coupled cluster (CCSD), and Brueckner coupled cluster (BCCD) methods, *Chemical Physics Letters*, Vol. 190, No. 1-2, 1992, pp. 1-12.
- [28] MOLPRO, version 2006.1, a package of *ab initio* programs, H.-J. Werner, P. J. Knowles, R. Lindh, *et. al.*, see <http://www.molpro.net>.
- [29] F. Weigend, A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency, *Physical Chemistry Chemical Physics*, Vol. 4, No. 18, 2002, pp. 4285-4291.
- [30] F. Weigend, A. Köhn, C. Hättig, Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations, *Journal of Chemical Physics*, Vol. 116, No. 8, 2002, pp. 3175-3183.
- [31] A. Heßelmann, G. Jansen, The helium dimer potential from a combined density functional theory and symmetr-adapted perturbation theory approach using an exact exchange-correlation potential, *Physical Chemistry Chemical Physics*, Vol. 5, No. 22, 2003, pp. 5010-5014.
- [32] Y. Bard, *Nonlinear Parameter Estimation*, Academic Press, 1974.
- [33] G. A. F. Seber, C. J. Wild, *Nonlinear Regression*, Wiley, 2003.
- [34] M. Pitoňák, K. E. Riley, P. Neogrady, P. Hobza, Highly accurate CCSD(T) and DFT-SAPT stabilization energies of H-bonded and stacked structures of the uracil dimer, *ChemPhysChem*, Vol. 9, No. 11, 2008, pp. 1636-1644.