

Stability of Hydrocarbons of the Polyhedrane Family: Convergence of ab Initio Calculations and Corresponding Assessment of DFT Main Approximations

J. C. Sancho-García*

Departamento de Química Física, University of Alicante, E-03080 Alicante, Spain

ABSTRACT: Highly accurate coupled-cluster (CC) calculations with large basis sets have been performed to study the binding energy of the $(CH)_{12}$, $(CH)_{16}$, $(CH)_{20}$, and $(CH)_{24}$ polyhedral hydrocarbons in two, cage-like and planar, forms. We also considered the effect of other minor contributions: core-correlation, relativistic corrections, and extrapolations to the limit of the full CC expansion. Thus, chemically accurate values could be obtained for these complicated systems. These nearly exact results are used to evaluate next the performance of main approximations (i.e., pure, hybrid, and double-hybrid methods) within density functional theory (DFT) in a systematic fashion. Some commonly used functionals, including the B3LYP model, are affected by large errors, and only those having reduced self-interaction error (SIE), which includes the last family of conjectured expressions (double hybrids), are able to achieve reasonable low deviations of 1-2 kcal/mol especially when an estimate for dispersion interactions is also added.

1. INTRODUCTION

The stability of the polyhedrane family of hydrocarbons $(CH)_{2n}$ (n = 6, 8, 10, 12; see Figures 1 and 2) has recently attracted much attention¹⁻⁶ due to the subtle interplay found between intramolecular short-, medium-, and long-range (in the weak overlap region) physical effects when increasing the size of the system. The field of computational organic chemistry^{7,8} is expected to face a compromise between accuracy and computational cost in these cases; thus, the assessment of efficient yet accurate methods for this kind of systems is a topic of ongoing research. To do so appropriately, benchmarking always needs highly accurate and/or experimental results for the systematic validation of less costly methods. Furthermore, we note that two of the (CH)₁₂ possible isomers are also part of the DC9 subset of the stringent GMTKN24 general benchmark database,9 which was recently developed to facilitate a nonbiased comparison between different theoretical methods, intrinsically proving the current interest on these molecules as complicated and thus challenging systems.

In this context and as it has been done before, ^{4,5} we carefully calculate the binding energies (BE) of polyhedranes relative to the number of constituing acetylene units

$$\Delta E(BE) = \frac{1}{n} E[(CH)_{2n}] - E(C_2H_2)$$
 (1)

Before any attempt to assess the accuracy of an approximate model (i.e., DFT) to deal with these and related systems, we will calculate these energies with the current standard for covalent and noncovalent interactions of medium-size molecules: 10,11 high-level coupled-cluster (CC) theory with singles, doubles, and perturbatively estimated triple excitations, the CCSD(T) method. Note the extraordinary effort needed for a method formally scaling as N^7 , where N can be related to the size of the systems, to the whole (CH) $_{2n}$ family of compounds shown in Figures 1 and 2. These calculations are 2 orders of magnitudes

more demanding than those obtained before, based on spin-scaled variants of the Møller—Plesset perturbation theory up to second order (MP2), and subsequently used to benchmark a large number of density functionals. Additionally, the latter authors have recently corrected their benchmark values due to some inadvertent overestimation of the original numbers. Thus, at this stage, our objective is 2-fold: (i) to provide first highly accurate reference values for binding energies of the members of the (CH)_{2n} family and (ii) to further use these numbers to carefully benchmark the results of density functional theory (DFT).

2. TECHNICAL DETAILS

The ab initio final energy for each system was calculated as follows

$$E_{\text{CCSD(T)}}^{\text{CBS}} \approx E_{\text{MP2}}^{\text{cc-pVQZ//cc-pVTZ}} + \left(E_{\text{CCSD(T)//MP2}}^{\text{cc-pVTZ}} - E_{\text{MP2}}^{\text{cc-pVTZ}}\right)$$
(2)

where the first term tries to saturate the correlation effects at the second order of the Møller–Plesset perturbation theory with respect to the space expanded by the basis sets, the slowest converging step, while the second term, denoted for the sake of simplicity $\Delta \text{CCSD}(T)$ in the following, adds the remaining correlation effects which have commonly a less marked dependence with respect to the basis sets. Note that these terms are single-point corrections at the sufficiently converged MP2/ccpVTZ geometries. This strategy intends to estimate the correlation effects at the complete basis sets (CBS) limit. The family of cc-pVnZ basis sets (n = D, T, Q) has been consequently used along the study. All single-point ab initio calculations were performed keeping frozen the core orbitals, a minor effect that

Received: March 23, 2011 Published: July 21, 2011

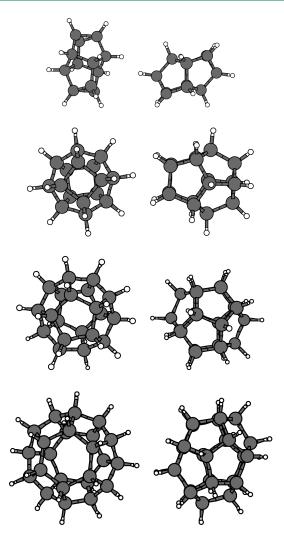


Figure 1. Representation of the molecular geometries of cage-like $(CH)_{2n}$ (n = 6, 8, 10, 12) compounds 1a-4a (from top to bottom) in two views. Figure created with XMakemol (see http://www.nongnu.org/xmakemol/).

will however be further investigated at the MP2 level, and without the counterpoise correction for basis sets superposition error, which facilitates comparison with previous results.

The 2.8.0 version of the ORCA quantum-chemical package¹² was consistently used for all calculations reported here. Concerning the use of DFT-based methods, note that (i) we significantly increased all the default thresholds (i.e., integration grid or convergence of self-consistent resolution of the Kohn—Sham equations) to reduce numerical errors and (ii) the geometries of the compounds are fully optimized with the cc-pVTZ basis set, which seems enough to ensure nearly converged results with respect to the one-particle space. Additionally, the resolution-of-the-identity (RI) and chain-of-spheres (COSX) techniques^{13,14} were employed if needed to alleviate the computational cost of the more demanding steps. The auxiliary basis functions were taken from the corresponding basis set library.

3. POLYHEDRANE FAMILY (CH)_{2N}: AB INITIO RESULTS

We will investigate first the convergence of the MP2 values with respect to the basis sets: Increasing the level from cc-

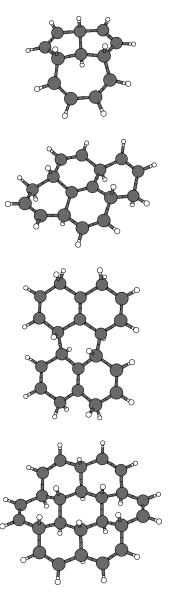


Figure 2. Representation of the molecular geometries of planar $(CH)_{2n}$ (n = 6, 8, 10, 12) compounds 1b-4b (from top to bottom). Figure created with XMakemol (see http://www.nongnu.org/xmakemol/).

pVDZ to cc-pVTZ reduces the values of $\Delta E(BE)$ by 2.0–2.5 (1.5-1.8) kcal/mol for the 1a-4a (1b-4b) compounds along the $(CH)_{2n}$ series, whereas going from cc-pVTZ to ccpVQZ leaves the values almost unaltered (maximum deviation less than 0.1 kcal/mol for both cage-like and planar forms). Thus, the first term of eq 2 can be considered as practically converged at this stage. Before analysis of the complete CCSD(T)/CBS results for the entire $(CH)_{2n}$ family, let us focus first on further envisioned computational savings. We see two strategies to reduce the cost of the more demanding $\Delta CCSD(T)$ contribution: (i) to employ the cc-pVDZ basis sets, which reduces the number of two-electron integrals, and (ii) to employ the CCSD level, which reduces the formal scaling of the procedure by an order of magnitude. However, this last strategy is discouraged due to csome compensation of errors between the convergence of CCSD and (T) contributions.¹⁵ We will thus investigate next the efficiency of the first possibility for the $(CH)_{12}$ (1a) case: $\Delta E(BE)$ only

Table 1. Binding Energies, $\Delta E(BE)$ in kcal/mol, of the $(CH)_{2n}$ (n = 6, 8, 10, 12) Family of Compounds Calculated at the Approximate CCSD(T)/CBS Level

	$(CH)_{12}$		(C	(CH) ₁₆		$(CH)_{20}$		(CH) ₂₄	
compound	1a	1b	2a	2b	3a	3b	4a	4b	
$\Delta E(BE)$ -	-47.36	-42.28	-51.85	-45.86	-56.38	-48.81	-54.93	-49.60	

varies (in kcal/mol) from $-47.27~(\Delta CCSD(T) = E^{\rm cc-pVTZ}_{\rm CCSD(T)//MP2} - E^{\rm cc-pVTZ}_{\rm MP2})$ to $-47.36~(\Delta CCSD(T) = E^{\rm cc-pVDZ}_{\rm CCSD(T)//MP2} - E^{\rm cc-pVDZ}_{\rm MP2})$, which completely validates the use of this approximation in the following. Therefore, the possible impact of further basis sets extension is not expected to significantly influence the conclusions reached up to here.

Further investigation of these results, expectedly taken as benchmarks, is carried out next again for compound 1a, taken again as example. To do so, we invoke the technique called valence "focal-point analysis" (FPA) 16,17 to study again the dual convergence of both basis sets and methods toward nearly exact results. The asymptotic value is estimated as follows: (i) the binding energy provided by zeroth-order energy (RHF) is calculated with the cc-pVnZ ($n=\mathrm{D}$, T, Q) sequence and extrapolated to its CBS limit by a n^{-5} function; 18 (ii) the remaining correlation contributions to $\Delta E(\mathrm{BE})$ employing CC-based methods (δ_{CCSD} and $\delta_{\mathrm{CCSD}(\mathrm{T})}$, respectively) are included at the extrapolated values; and (iii) the CC series is extrapolated toward completeness (δ_{FCC}) thanks to the availability of a continued fraction approximation. 19 Among all the available extrapolation techniques for CC-based correlation energies, E_{c} , we choose the expression 20

$$E_{c}(\infty) = [E_{c}(n+1) - E_{c}(n)]F + E_{c}(n)$$
(3)

where F is a coefficient for extrapolation depending on n and n+1 and $E_{\rm c}(\infty)$ is the final extrapolated value. Application of the aforesaid scheme to $({\rm CH})_{12}$ (1a) brings a final value of -47.39 kcal/mol, in complete agreement with calculations based on eq 2; $\delta_{\rm CCSD}$, $\delta_{\rm CCSD(T)}$, and $\delta_{\rm FCC}$ contributing, respectively, with -8.36, -0.30, and -0.09 kcal/mol to the initial RHF/ccpV ∞ Z value of -38.63 kcal/mol. As a byproduct of this FPA result, we confirm again the hypothesis implicit in eq 2 about the negligible impact on further basis sets extension.

Additionally, one needs to complementarily address other effects which are normally overlooked when trying to reach errors in the subchemical accuracy range (best estimates deviating the most with respect to the exact value by 1 kcal/mol): core correlation effects and scalar relativistic corrections. The former is calculated here by the difference between MP2/cc-pCVTZ// MP2/cc-pVTZ and MP2/cc-pVTZ energies and amounts to less than 0.1 kcal/mol (in absolute value) for the whole $(CH)_{2n}$ set. The latter correction employs the Douglas-Kroll-Hess (DKH) Hamiltonian²¹ and is estimated as the difference between MP2-DKH/cc-pVTZ//MP2/cc-pVTZ and MP2/cc-pVTZ energies, which thus gives a value around 0.1 kcal/mol for all compounds. We have not evaluated the impact of other minor effects yet, i.e., the influence of the CCSDT-CCSD(T) correction, although they are expected to be negligibly small according to the FPA value obtained for $\delta_{ ext{FCC}}$ of the smallest compound. To summarize, our best-estimates (nonrelativistic) core-correlation corrected CCSD(T)/CBS binding energies for compounds 1a-4a (cage-like) and 1b-4b (planar) are gathered in Table 1

Table 2. Mean Unsigned Error, MUE in kcal/mol, for Binding Energies of the $(CH)_{2n}$ (n = 6, 8, 10, 12) Family of Compounds at Various MP2 and DFT Levels

method		MUE		
		1a-4a	1b-4b	
ab initio ^a	MP2	2.4	0.9	
	SCS-MP2	1.7	2.2	
DFT^b	BLYP	13.1	9.4	
	BLYP-D	8.0	5.2	
	B3LYP	7.4	5.0	
	B3LYP-D	2.8	1.3	
	BHHLYP	2.0	0.9	
	BHHLYP-D	0.7	0.2	
	B2-PLYP	3.6	2.7	
	B2-PLYP-D	1.2	0.8	
	B2GP-PLYP	1.0	0.9	
	B2GP-PLYP-D	0.7	0.5	

^a With the cc-pVQZ basis set at the corresponding cc-pVTZ-optimized geometry. ^b At the cc-pVTZ-optimized geometry.

and will be used next to benchmark the results obtained by other methods such as density functional theory (DFT) using mainly hybrid and double-hybrid models. Note that orbital-dependent density functionals are called to play a major role in next decades and further applications to chemically interesting systems are largely envisioned.

4. POLYHEDRANE FAMILY (CH)_{2N}: DFT RESULTS

In this context, we present in Table 2 the results of the systematic analysis performed; the mean unsigned error (MUE) will be used to judge the quality of the approximations. Before analyzing the DFT-based results, we would like to underline first that spin-component-scaled MP2 models (as SCS-MP2³⁰) are still an approximation to the CCSD(T) values yet certainly an improvement with respect to unscaled MP2 upon inspection (vide infra) of the corresponding values; thus, previous claims^{4,5} should also benefit from the current CCSD(T)/CBS values.

We start with the discussion with systems 1a-4a. Table 2 also shows that whereas BLYP and B3LYP behave inaccurately, a MUE of 13.1 and 7.4 kcal/mol is obtained, respectively, a higher value of exact-like exchange (BHHLYP) clearly improves the results and reduces the corresponding MUE to 2.0 kcal/mol. However, further steps in this direction (i.e., increasing the value of the exact-like exchange weight in an hybrid-like model) is not recommended, and in fact, the HF-LYP model, with full exactlike exchange together with a coupled correlation potential in a self-consistent fashion, drops an MUE of 10.2 kcal/mol. On the other hand, the double hybrids selected (B2-PLYP and B2GP-PLYP) are undoubdtedly a major improvement with respect to the golden standard (B3LYP) since the corresponding MUE is at least halved. Concerning compounds 1b-4b, simply speaking, the trends found are the same as before. Dispersion-uncorrected MUE values are 9.4, 5.0, 0.9, and 8.6 kcal/mol for BLYP, B3LYP, BHHLYP, and HF-LYP models, lower than for cage-like forms; note that double-hybrid forms behave very respectably again.

There is still room for further improvement if one couples the dispersion energy (-D) to any of the calculated DFT values.³¹ To do that, at essentially no extra computational cost, we resort to a

Table 3. Mean Unsigned Error, MUE in kcal/mol, for Isomerization Energies between 1a-4a and 1b-4b Members of the $(CH)_{2n}$ (n=6,8,10,12) Family of Compounds at Various MP2 and DFT Levels

	method	MUE
ab initio ^a	MP2	12.2
	SCS-MP2	3.8
DFT^b	BLYP	32.9
	BLYP-D	23.8
	B3LYP	20.6
	B3LYP-D	12.5
	BHHLYP	9.4
	BHHLYP-D	7.0
	B2-PLYP	8.0
	B2-PLYP-D	3.7
	B2GP-PLYP	1.7
	B2GP-PLYP-D	1.5

^a With the cc-pVQZ basis set at the corresponding cc-pVTZ-optimized geometry. ^b At the cc-pVTZ-optimized geometry.

first-order function which explicitly depends on the well-known \mathbb{R}^{-6} decay of these interactions

$$E = s_6 \sum_{AB} f_d(R_{AB}) \frac{C_6^{AB}}{R_{AB}^6}$$
 (4)

where s_6 is a scaling parameter used to efficiently couple the exchange-correlation and dispersion energies, $f_{\rm d}(R_{\rm AB})$ is a damping function of the interatomic distance $(R_{\rm AB})$, and $C_6^{\rm AB}$ is a dispersion coefficient for the atomic pair AB. $^{32-34}$ Table 2 nicely shows how the MUE is dramatically altered in the right direction after addition of the dispersion term to the final values, independently of the model chosen.

If one averages now the MUE obtained for both sets of compounds, 1a-4a and 1b-4b, a range of hybrid (BHHLYPD) and double-hybrid (B2-PLYP-D, B2GP-PLYP-D, and B2GP-PLYP-D) methods are able to be obtained now, in some cases by far the target of "chemical accuracy" (a maximum error of ± 1 kcal/mol with respect to energy reference values). Note that even in the worst possible scenario any double hybrid would at least halve the MUE with respect to B3LYP, a somehow superseded method now.

In addition, concerning these nonrelativistic all-core DFT-based calculations, we would like to remark first that a large set of calculated key properties of organic systems is known to be mostly affected by the self-interaction error (SIE) of common density functionals³⁵ (i.e., the spurious self-interaction of an electron with itself). Generally speaking, functionals having minimal SIE should be used for theoretical studies of organic systems.³⁶ As commonly admitted in the field and far from being trivial,³⁷ the SIE is very often used interchangeably with the concept of delocalization error, which is maybe more intuitive: a functional with large SIE would overstabilize delocalized densities while giving too high energies for localized densities. Note that this effect would strongly manifest when computing the energy differences between cage-like (1a-4a) and planar (1b-4b) isomer forms.

In this respect, Table 3 presents the values of the MUE obtained for isomerization energies E(na) - E(nb) (n = 1-4) with respect to the reference values given again by eq 2. The

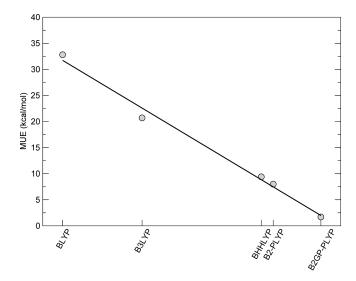


Figure 3. Evolution of MUE (in kcal/mol) for the isomerization energies of cage-like vs planar forms as a function of the exact-like exchange term introduced into the functional.

striking decrease of the MUE values with the weight of exact-like exchange introduced into the (double-)hybrid form is shown in Figure 3; we remind that the exact-like exchange is, by definition, SIE free, which helps to interpret the trend found with a correlation coefficient $r^2 = 0.991$. It is maybe also illustrative to compare these numbers with other modern methods belonging to the Minnesota family of functionals, ³⁸ as M06-L and M06-2X. The numbers taken from ref 4 lead to a MUE of 8.0 and 4.4 kcal/mol, respectively, which can be considered as competitive too. All in all, we would like to warn about the marked influence of SIE and/ or dispersion corrections here and some strategies to cope with the former among others: high weights of exact-like exchange (i.e., BHHLYP), double-hybrid models (i.e., B2-PLYP or B2GP-PLYP), or long-range-corrected functionals (i.e., CAM-B3LYP or LC-BLYP, the latter drops a MUE of 8.2 kcal/mol according to the values taken from ref 4). The latter effect can be efficiently and accurately taken into account by the correction given by eq 4 or related.39

5. CONCLUSIONS

We briefly remind here that our main purpose was to convincingly establish a set of benchmark values for the stability energy of two forms, cage-like and planar, of polyhedrane hydrocarbons, which have recently attracted much interest. In the pursuit of the greatest possible accuracy, we relied on CCSD(T) calculations at the estimated basis sets limit together with some minor corrections in an attempt to successfully reach the subchemical accuracy range. Having done that, further assessment of less costly methods, such as spin-component-scaled MP2 variants (SCS-MP2) as well as hybrid and double-hybrid functionals, was carried out. Since there exists in the literature a plethora of density functionals, our intention was not to extensively cover all of them; we selected examples from the main existing categories (pure, hybrid, and double-hybrid methods) without changing the exchange and correlation density kernels: BLYP, B3LYP, BHHLYP, B2-PLYP, and B2GP-PLYP. A dispersion correction (-D) was then added to each functional form to test its influence on the final results. One can clearly observe upon inspection of the results and independently of the isomer type (cage-like or

planar) how standard methods such as B3LYP keep unacceptably large errors and a higher weight of exact-like exchange in the composition of the hybrid functional definitively helps. Note that double hybrids normally have, by default, larger weights of exact-like exchange, and thus, they have also shown excellent behavior. This seems to indicate the large influence of SIE, especially for isomerization energies, although intramolecular dispersion interactions are also important mostly for binding energies. Overall, to be always on the safer side due to the possible existing interplay between SIE and intramolecular dispersion interactions, we believe that older functionals should be progressively discarded, for a routine use in the field of computational organic chemistry, and double hybrids might be routinely applied if the computational cost does not become a serious bottleneck.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jc.sancho@ua.es.

REFERENCES

- (1) Schreiner, P. R.; Fokin, A. A.; Pascal, R. A., Jr.; de Meijere, A. Org. Lett. 2006, 8, 3635.
 - (2) Schreiner, P. R. Angew. Chem., Int. Ed. 2007, 46, 4217.
- (3) Shamov, G. A.; Budzelaar, P. H. M.; Schreckenbach, G. J. Chem. Theory Comput. 2010, 6, 477.
- (4) Shamov, G. A.; Schreckenbach, G.; Budzelaar, P. H. M. J. Chem. Theory Comput. 2001, 6, 3442.
- (5) Shamov, G. A.; Schreckenbach, G.; Budzelaar, P. H. M. J. Chem. Theory Comput. 2011, 7, 804.
- (6) Swart, M.; Solá, M.; Bickelhaupt, M. J. Comput. Chem. 2011, 32, 1117.
- (7) Bachrach, S. Computational Organic Chemistry; Wiley: New York, 2007.
 - (8) Bachrach, S. Annu. Rep. Prog. Chem. Sect. B 2008, 104, 394.
 - (9) Goerigk, L.; Grimme, S. J. Chem. Theory Comput. 2010, 6, 107.
 - (10) Černý, J.; Hobza, P. Phys. Chem. Chem. Phys. 2007, 9, 5291.
- (11) Riley, K. E.; Pitoňák, M.; Jurečka, P.; Hobza, P. Chem. Rev. 2010, 110, 5023.
- (12) Neese, F. ORCA 2.8.0; University of Bonn: Bonn, Germany; http://www.thch.uni-bonn.de/tc/orca/
 - (13) Neese, F. J. Comput. Chem. 2003, 24, 1740.
- (14) Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Chem. Phys. 2009, 356, 98.
- (15) Pitoňák, M.; Riley, K. E.; Neogrady, P.; Hobza, P. Chem-PhysChem 2008, 9, 1636.
- (16) Császár, A. G.; Allen, W. D.; Schaefer, H. F., III J. Chem. Phys. 1998, 108, 9751.
 - (17) Sancho-Garca, J. C. Chem. Phys. Lett. 2009, 468, 138.
 - (18) Parthiban, S.; Martin, J. M. L. J. Chem. Phys. 2001, 114, 6014.
 - (19) Goodson, D. Z. J. Chem. Phys. 2002, 116, 6948.
 - (20) Schwenke, D. W. J. Chem. Phys. 2005, 122, 014107.
- (21) Neese, F.; Wolf, A.; Fleig, T.; Reiher, M.; Hess, B. A. J. Chem. Phys. 2005, 122, 204107.
 - (22) Becke, A. D. J. Chem. Phys. 1993, 98, 5648.
 - (23) Becke, A. D. J. Chem. Phys. 1993, 98, 1372.
 - (24) Grimme, S. J. Chem. Phys. 2006, 124, 034108.
- (25) Schwabe, T.; Grimme, S. Phys. Chem. Chem. Phys. 2006, 8, 4398.
- (26) Tarnopolsky, A.; Karton, A.; Sertchook, R.; Vuzman, D.; Martin, J. M. L. J. Phys. Chem. A 2008, 112, 3.
- (27) Karton, A.; Tarnopolsky, A.; Lamère, J. F.; Schatz, G. C.; Martin, J. M. L. J. Phys. Chem. A 2008, 112, 12868.
- (28) Zhang, Y.; Xu, X.; Goddard, W. A., III Proc. Natl. Acad. Sci. U.S.A. 2009, 106, 4963.
 - (29) Kummel, S.; Kronik, L. Rev. Mod. Phys. 2008, 3, 80.

- (30) Grimme, S. J. Chem. Phys. 2003, 118, 9095.
- (31) Burns, L. A.; Vázquez-Mayagoitia, A.; Sumpter, B. A.; Sherrill, C. D. J. Chem. Phys. **2011**, 134, 084107.
 - (32) Wu, Q.; Yang, W. J. Chem. Phys. 2002, 116, 515.
 - (33) Grimme, S. J. Comput. Chem. 2006, 27, 1787.
- (34) Jurečka, P.; Černý, J.; Hobza, P.; Salahub, D. R. J. Comput. Chem. 2007, 28, 555.
- (35) Mori-Sánchez, P.; Cohen, A. J.; Yang, W. J. Chem. Phys. 2006, 125, 201102.
- (36) Johnson, E. R.; Mori-Sánchez, P.; Cohen, A. J.; Yang, W. J. Chem. Phys. 2008, 129, 204112.
 - (37) Körzdörfer, T. J. Chem. Phys. 2011, 134, 094111.
 - (38) Zhao, Y.; Truhlar, D. G. Acc. Chem. Res. 2008, 41, 157.
- (39) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. J. Chem. Phys. **2010**, 132, 154104.