

Accurate Reaction Barrier Heights of Pericyclic Reactions: Surprisingly Large Deviations for the CBS-QB3 Composite Method and Their Consequences in DFT Benchmark Studies

Amir Karton*^[a] and Lars Goerigk*^[b]

Accurate barrier heights are obtained for the 26 pericyclic reactions in the BHPERI dataset by means of the high-level Wn -F12 thermochemical protocols. Very often, the complete basis set (CBS)-type composite methods are used in similar situations, but herein it is shown that they in fact result in surprisingly large errors with root mean square deviations (RMSDs) of about $2.5 \text{ kcal mol}^{-1}$. In comparison, other composite methods, particularly G4-type and estimated coupled cluster with singles, doubles, and quasiperturbative triple excitations [CCSD(T)/CBS] approaches, show deviations well below the chemical-accuracy threshold of 1 kcal mol^{-1} . With the exception of SCS-MP2 and the herein newly introduced MP3.5 approach, all other tested Møller-Plesset perturbative procedures give poor performance with RMSDs of up to $8.0 \text{ kcal mol}^{-1}$. The finding that CBS-type methods fail for barrier heights of these reactions is unexpected and it is particularly troublesome given that they are often used

to obtain reference values for benchmark studies. Significant differences are identified in the interpretation and final ranking of density functional theory (DFT) methods when using the original CBS-QB3 rather than the new Wn -F12 reference values for BHPERI. In particular, it is observed that the more accurate Wn -F12 benchmark results in lower statistical errors for those methods that are generally considered to be robust and accurate. Two examples are the PW6B95-D3(BJ) hybrid-meta-general-gradient approximation and the PWPB95-D3(BJ) double-hybrid functionals, which result in the lowest RMSDs of the entire DFT study (1.3 and $1.0 \text{ kcal mol}^{-1}$, respectively). These results indicate that CBS-QB3 should be applied with caution in computational modeling and benchmark studies involving related systems. © 2015 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.23837

Introduction

Accurate computational prediction of chemical reaction rates requires the determination of reaction barrier heights with methods that can routinely achieve “chemical accuracy,” usually defined as deviations within 1 kcal mol^{-1} from accurate reference data. For example, according to the Arrhenius equation, a change of only $1.4 \text{ kcal mol}^{-1}$ in the reaction barrier height corresponds to a change of one order of magnitude in the reaction rate at room temperature. A few computationally economical composite procedures that are applicable to relatively large systems (with up to 20–30 nonhydrogen atoms) have been developed over the past two decades.^[1] These include the Gaussian- n (Gn) methods (e.g., G4,^[2] G4(MP2),^[3] G4(MP2)-6X,^[4] and G3X-K^[5]) and the complete basis set (CBS) methods (e.g., CBS-APNO^[6] and CBS-QB3).^[7] These procedures use MP4 and/or MP2-based additivity schemes in order to approximate the CCSD(T) energy (coupled cluster with singles, doubles, and quasiperturbative triple excitations) in conjunction with a triple- ζ quality basis set. Generally speaking, these procedures have been found to produce gas-phase thermochemical and kinetic properties (such as bond dissociation energies, enthalpies of formation, reaction energies and barrier heights) with mean absolute deviations (MADs) from reliable experimental or theoretical data below the threshold of “chemical accuracy”.^[1–9] Due to their attractive accuracy-to-

computational cost ratio, these procedures gained tremendous popularity and are now widely used for obtaining reaction barrier heights for both chemical applications (see Reference [10] for recent applications of Gn methods and Reference [11] for CBS methods) and benchmarking of more approximate theoretical procedures.^[12–18] Given the broader applicability of these computationally economical procedures, compared for example with composite methods that aim at reproducing the CCSD(T) infinite basis-set-limit energy (e.g., Wn ,^[19] Wn -F12,^[20] and ccCA^[21]), it is important to identify cases for which these procedures do not perform well so that their application in such instances would be handled with caution.^[1a,b]

Woodward and Hoffmann defined pericyclic reactions as “reactions in which all first-order changes in bonding

[a] A. Karton

School of Chemistry and Biochemistry, The University of Western Australia, Perth, WA 6009, Australia
E-mail: amir.karton@uwa.edu.au

[b] L. Goerigk

School of Chemistry, The University of Melbourne, Parkville, VIC, 3010, Australia
E-mail: lars.goerigk@unimelb.edu.au

Contract grant sponsor: Australian Research Council (ARC) Discovery Early Career Researcher Awards; Contract grant numbers: DE140100311; DE140100550; Contract grant sponsor: Selby Scientific Foundation through the 2014 Selby Research Award (L.G.)

© 2015 Wiley Periodicals, Inc.

Table 1. Pericyclic reactions in the BHPERI database.

Set ^[a]	Reaction	Reaction type	Reactants
Set-I	1	Electrocyclic ring closing	Cyclobutene
	2	Electrocyclic ring closing	cis-1,3,5-Hexatriene
	3	Electrocyclic ring closing	ortho-Xylylene
	4	Sigmatropic shift	1,3-Pentadiene
	5	Sigmatropic shift	1,3-Cyclopentadiene
	6	Sigmatropic shift	1,5-Hexadiene
	7	Diels–Alder cycloaddition	1,3-Butadiene + Ethylene
	8	Diels–Alder cycloaddition	1,3-Cyclopentadiene + Ethylene
	9	Diels–Alder cycloaddition	2 × 1,3-Cyclopentadiene
	10	Cycloreversion	Cis-triscyclopropacyclohexane
Set-II	11	1,3-Dipolar cycloaddition	Nitrous oxide + Ethylene
	12	1,3-Dipolar cycloaddition	Hydrazoic acid + Ethylene
	13	1,3-Dipolar cycloaddition	Diazomethane + Ethylene
	14	1,3-Dipolar cycloaddition	Fulminic acid + Ethylene
	15	1,3-Dipolar cycloaddition	Formonitrile imine + Ethylene
	16	1,3-Dipolar cycloaddition	Formonitrile ylide + Ethylene
	17	1,3-Dipolar cycloaddition	Methylene nitrene + Ethylene
	18	1,3-Dipolar cycloaddition	Formazomethine imine + Ethylene
	19	1,3-Dipolar cycloaddition	Formazomethine ylide + Ethylene
Set-III	20	Diels–Alder cycloaddition	Cyclopentadiene + Ethylene
	21	Diels–Alder cycloaddition	Silole + Ethylene
	22	Diels–Alder cycloaddition	Furan + Ethylene
	23	Diels–Alder cycloaddition	Pyrrrole + Ethylene
	24	Diels–Alder cycloaddition	Phosphole + Ethylene ^[b]
	25	Diels–Alder cycloaddition	Phosphole + Ethylene ^[c]
	26	Diels–Alder cycloaddition	Thiophene + Ethylene

[a] For further details, see main text Ref. [15]. [b] Barrier to the exo product. [c] Barrier to the endo product.

relationships take place in concert on a closed curve.”^[22] Examples for this important reaction class include Diels–Alder cycloaddition, sigmatropic shift, electrocyclization, and cheletropic reactions.^[22] Over the past decade a number of datasets of barrier heights of pericyclic reactions have been constructed for the purpose of benchmarking the performance of Kohn–Sham density functional theory (DFT)^[23] and low-level *ab initio* procedures.^[12,13] Houk and coworkers have constructed two such datasets.^[12] The first dataset (hereinafter referred to as set-I) comprised 11 electrocyclic, cycloaddition, cycloreversions, and sigmatropic-shift reactions involving unsaturated hydrocarbons.^[12a] The second dataset (hereinafter referred to as set-II) included 18 1,3-dipolar cycloaddition reactions involving heteroatom-containing 1,3-dipoles, and ethylene and acetylene dipolarophiles.^[12b] The third dataset (hereinafter referred to as set-III) was constructed by Sastry and coworkers and included Diels–Alder cycloaddition reactions of seven dienes (butadiene and six five-membered heterocyclic rings) with ethylene.^[13] The reference data for sets I and II were originally obtained using the CBS-QB3 procedure^[7] and for set-III at the CCSD(T)/6-31G(d) level of theory. Martin and coworkers obtained W1 reaction barrier heights for eight reactions of set-I and found large deviations between the CBS-QB3 and W1 values.^[24,25] In particular, for these eight reactions the CBS-QB3 procedure attains an MAD of 1.7 kcal mol⁻¹ relative to W1, with the largest deviation (in absolute value) being as large as 3.2 kcal mol⁻¹. The general-purpose GMTKN30 database for main-group thermochemistry, kinetics, and noncovalent interactions combined 10 reaction barrier heights from set-I, nine from set-

II, and the seven from set-III into a representative dataset of 26 barrier heights of pericyclic reactions (known as the BHPERI dataset, Table 1).^[15,16] The BHPERI dataset adopted the eight W1 reaction barrier heights for set-I and CBS-QB3 barrier heights for all the rest of the reactions. In the present work we calculate W1-F12 and W2-F12 reference numbers for the reactions in the BHPERI dataset and use these benchmark values to evaluate the performance of a variety of composite [e.g., CBS-QB3, CBS-APNO, G4, and G4(MP2)] and standard *ab initio* (e.g., MP2, SCS-MP2, MP3, and MP4) procedures. We find that the popular CBS-QB3 and CBS-APNO procedures result in very large deviations from our benchmark *Wn*-F12 values, with root mean square deviations (RMSDs) of 2.4 and 2.6 kcal mol⁻¹, respectively. The considered *Gn* methods result in smaller deviations, in particular G4 shows good performance with an RMSD below the threshold of chemical accuracy (namely, 0.8 kcal mol⁻¹). Of the standard *ab initio* procedures SCS-MP2 and the herein newly introduced MP3.5 procedures emerge as the best performers with RMSDs of 1.4 and 0.6 kcal mol⁻¹, respectively.

We proceed to investigate the performance of a large variety of contemporary DFT and double-hybrid DFT (DHDFT)^[26,27] methods using our new *Wn*-F12 reaction barrier heights as a reference. Given its computational efficiency, DFT has become the workhorse in routine quantum-chemical applications. Although DFT itself is an exact theory, we have to rely on approximations to the “true” exchange-correlation functional and an unfortunate trend in density-functional-approximation (DFA) research is the large “zoo” of such approximations,^[28]

which at times can not only be intimidating to the general DFT users, but it can also confuse method developers. Therefore, designing new DFAs goes hand in hand with the construction of reliable benchmark sets to assess their accuracy before they can be used in subsequent applications.^[8f] While in the 1990s, the main interest of benchmark studies lay on properties such as heats of formation (or total atomization energies), electron affinities or ionization potentials,^[29] the focus shifted towards more “chemically inspired” problems in the 2000s, especially on reaction energies, barrier heights and noncovalent-interaction energies.^[8a–d,8f–i,25,30] The realization that not all DFAs are necessarily equally applicable to different kinds of chemical problems also sparked the development of compilations of benchmark databases to better gauge the robustness of a specific method. Such compilations were particularly promoted by Truhlar and coworkers^[31] and most recently by Goerigk and Grimme, who published the aforementioned GMTKN30 database.^[15,16] GMTKN30 is a compilation of 30 benchmark sets, including the herein discussed BHPERI set, and it has allowed thorough assessments of by now more than 50 different DFAs covering the entire range of Perdew’s Jacob’s Ladder^[32] classification of density functionals.^[17,27,33]

Thus, benchmark studies have their own important place in method development and assessment. Furthermore, it speaks for itself that the accuracy of the reference method is of utmost importance, as reference numbers of varying quality provide different outcomes and allow for completely different functional recommendations. Given that improved hardware architectures now allow the application of more “elaborate” methods, it is always worthwhile to reassess older, established methods against those more accurate ones. This is particularly important when said established methods are considered to be “gold standard” and are frequently used for obtaining reference data in benchmark studies. We recently demonstrated this point for relative energies in biologically relevant tetrapeptide conformations, where we showed that the popular MP2 method combined with a split-valence triple- ζ basis set was inadequate for correctly predicting their relative energies.^[34] We found that when this level of theory is used for benchmarking DFAs, it favors hybrid DFT functionals over double-hybrids, which is counterintuitive considering that the latter belong to a higher rung in the Jacob’s Ladder scheme. We additionally showed that when accurate CCSD(T)/CBS values are used as the reference the expected functional order is restored.

The herein presented Wn -F12 results for the BHPERI set allow us to carry out a conceptually similar analysis. We do not aim to recommend particular DFAs for the calculation of barrier heights of pericyclic reactions, but instead we aim to show how functional trends change when switching from less accurate (e.g., CBS-QB3) to more accurate (e.g., Wn -F12) benchmark values. Given that benchmark studies will always have to be carried out for newly developed methods, our DFA assessment serves as a reminder to emphasize the importance of the accuracy of reference data in such crucial investigations.

Computational Methods

To obtain reliable reference barrier heights for the reactions in the BHPERI dataset, calculations have been carried out using the high-level, *ab initio* Wn -F12 procedures ($n = 1$ and 2) with the Molpro 2012.1 program suite.^[35] $W1$ -F12 and $W2$ -F12 theories^[20] (and their earlier versions $W1$ and $W2.2$)^[19,36] represent layered extrapolations to the relativistic, all-electron CCSD(T)/CBS (coupled cluster with singles, doubles, and quasiperturbative triple excitations basis-set-limit energy). These composite theories can achieve “sub-chemical accuracy” for atomization reactions. For example, $W1$ -F12 and $W2$ -F12 theories are associated with RMSDs of 0.74 and 0.42 kcal mol⁻¹ for a set of 140 very accurate atomization energies obtained at the full configuration interaction (FCI) infinite basis-set limit.^[1,8f,19,20,36] Nevertheless, we point out that for systems containing only first-row elements (and H) $W1$ -F12 shows better performance. Specifically, for the 97 first-row atomization energies in the $W4$ –11 dataset,^[8f] $W1$ -F12 attains an RMSD of 0.45 kcal mol⁻¹ relative to reference atomization energies at the FCI infinite basis-set limit.^[20] Therefore in the present work we obtain reaction barrier heights with $W1$ -F12 for systems containing only first-row elements and with $W2$ -F12 for systems containing second-row elements.

$W1$ -F12 and $W2$ -F12 theories combine explicitly correlated F12 techniques^[37] with basis-set extrapolations in order to approximate the CCSD(T) infinite basis-set-limit energy. For the sake of making the article self-contained, we will briefly outline the various steps in these theories (for further details see Refs. [9] and [20]). In $W2$ -F12 the Hartree–Fock (HF) component is calculated with the VQZ-F12 basis set (where VnZ -F12 denotes the cc-pVnZ-F12 basis set of Peterson et al.^[38], which was specifically developed for explicitly correlated calculations). Note that the complementary auxiliary basis set singles correction is included in the self-consistent-field (SCF) energy.^[39–41] The valence CCSD-F12 correlation energy is extrapolated from the VTZ-F12 and VQZ-F12 basis sets, using the $E(L) = E_\infty + A/L^\alpha$ two-point extrapolation formula, with $\alpha = 5.94$. Optimal values for the geminal Slater exponents (β) used in conjunction with the VnZ -F12 basis sets were taken from Ref. [42]. The quasiperturbative triples, (T), corrections are obtained from standard CCSD(T)/VTZ-F12 calculations (i.e., without inclusion of F12 terms in the CCSD(T) procedure) and scaled by the factor $f = 0.987 \times E^{\text{MP2-F12}}/E^{\text{MP2}}$. This approach has been shown to accelerate the basis set convergence.^[20,43] In all of the explicitly correlated coupled cluster calculations the diagonal, fixed-amplitude 3C(FIX) ansatz^[40,44–46] and the CCSD-F12b approximation are employed.^[41,43] The CCSD innershell contribution is calculated with the core-valence weighted correlation-consistent aug-cc-pwCVTZ basis set of Peterson and Dunning,^[47] whilst the (T) innershell contribution is calculated with the cc-pwCVTZ(no f) basis set (where cc-pwCVTZ(no f) indicates the cc-pwCVTZ basis set without the f functions). The main difference between $W1$ -F12 and $W2$ -F12 theories is that $W1$ -F12 uses smaller basis sets for extrapolating the HF, CCSD-F12, and (T) contributions. Specifically, the HF and CCSD-F12 contributions are extrapolated from the VDZ-F12 and VTZ-F12 basis-sets using the two-point extrapolation formula

with $\alpha = 5.00$ and 3.38 , respectively.^[9,20] The (T) valence correlation energy is obtained in the same way as in the original W1 theory,^[19,20] that is, extrapolated from the A'VDZ and A'VTZ basis sets with $\alpha = 3.22$ (where A'VnZ indicates the combination of the standard correlation-consistent cc-pVnZ basis sets on H,^[48] the aug-cc-pVnZ basis sets on first-row elements,^[49] and the aug-cc-pV(n+d)Z basis sets on second-row elements).^[50]

As W1-F12 and W2-F12 theories represent layered extrapolations to the all-electron CCSD(T) basis-set-limit energy, it is of interest to estimate whether the contributions from post-CCSD(T) excitations are likely to be significant. The percentage of the total atomization energy accounted for by parenthetical connected triple excitations, $\%TAE_e[(T)]$, has been shown to be a reliable energy-based diagnostic for the importance of non-dynamical correlation effects. It has been suggested that $\%TAE_e[(T)] < 2\%$ indicates systems that are dominated by dynamical correlation, while $2\% < \%TAE_e[(T)] < 5\%$ indicates systems that are characterized by mild nondynamical correlation.^[8f,36] Supporting Information Table S1 gathers the $\%TAE_e[(T)]$ values for the transition structures (TSs) in the BHPERI dataset. The $\%TAE_e[(T)]$ values for these species lie in the range 1.9–4.1. In particular, the $\%TAE_e[(T)]$ values for reactions 11, 12, 14, and 15 lie between 3.1 and 4.1 and the $\%TAE_e[(T)]$ values for the rest of the TSs are < 2.9 . These values suggest that these TSs are dominated by mild nondynamical correlation effects and that our bottom-of-the-well CCSD(T)/CBS benchmark reaction barrier heights should be well below ~ 1 kcal mol⁻¹ from the barrier heights at the FCI basis-set limit.^[8f,20]

We use our W1-F12 and W2-F12 benchmark barrier heights to evaluate the performance of a wide range of composite G_n-type procedures (namely, G4,^[2] G4(MP2),^[3] G4(MP2)–6X,^[4] G3,^[51] G3(MP2),^[52] G3B3,^[53] and G3(MP2)B3),^[53] and CBS-type methods (namely, CBS-APNO^[6] and CBS-QB3).^[7] The calculations for the G_n and CBS procedures were performed using the Gaussian 09 program suite.^[54] We also assess the performance of the CCSD(T) method using an additivity-based approach in which the CCSD(T)/A'V{T,Q}Z energy is estimated from the CCSD(T)/A'VDZ base-energy and an MP2-based basis-set-correction term ($\Delta MP2$). This cost-effective approach, which gives very good results for noncovalent interactions,^[34,55] has been recently shown to give excellent performance for reactions that involve double-bond migrations in conjugated dienes.^[8a,b] Here, we consider a more general version of this method in which the basis-set-correction term ($\Delta MP2$) is not restricted to second-order Møller–Plesset perturbation theory. Our additivity-based approach is based on the formula:

$$\text{CCSD(T)/CBS(MP}_n) \approx \text{CCSD(T)/A'VDZ} + \Delta \text{MP}_n \quad (1)$$

where ΔMP_n is given by:

$$\Delta \text{MP}_n = \text{MP}_n / \text{A'V}\{T, Q\}Z - \text{MP}_n / \text{A'VDZ} \quad (2)$$

In eq. (2), the MP_n/CBS energy ($n = 2, 3, 4$) is extrapolated to the basis-set limit from the A'V{T,Q}Z basis set pair using exponents of 5 and 3 for the Hartree–Fock and MP_n correlation contributions, respectively.^[56] We also evaluate the perform-

ance of a number of standard and modified Møller–Plesset procedures (e.g., MP2, MP3, MP4, MP2.5,^[57] SCS-MP2,^[58] and SCS-MP3^[59]). Similarly to the way to the MP2.5 procedure was defined as the average of MP2 and MP3, we define MP3.5 as the average of MP3 and MP4. We find that MP3.5 gives excellent performance that outperforms all of the considered Møller–Plesset-type procedures for the calculation of the reaction barrier heights in the BHPERI dataset (see Results and discussion).

Because BHPERI is part of the GMTKN30 database, it has already been investigated with a large number of density functionals.^[17,27] In most of those cases, Grimme's DFT-D3 dispersion correction with "zero-damping"^[60] has been applied and it has been established that London-dispersion plays a decisive role on the barrier heights given the respective system sizes.^[17] However, since the publication of GMTKN30, the updated DFT-D3 variant with Becke–Johnson^[61] (BJ) damping has been shown to be even more accurate.^[62] For our purpose, we therefore take the raw electronic energies from the previous studies and correct them with the newer DFT-D3(BJ) variant using Grimme's standalone program *dftd3*.^[63] The only exception are most of the Minnesota functionals, as they are not compatible with DFT-D3(BJ).^[17] Additionally, we also run calculations with the B2K-PLYP^[25,64] double-hybrid functional that was specifically designed for the description of barrier heights. These calculations are carried out with ORCA 3.0.2.^[65] The DFT-D3(BJ) parameters had to be newly determined for B2K-PLYP and they are listed in the SI. In total, we analyze data for 45 density functionals covering the four highest rungs of Jacob's ladder. They comprise 14 functionals belonging to the general-gradient approximation (GGA; rung 2), three meta-GGAs (rung3), 22 global and range-separated hybrids (rung 4), and six double-hybrids (rung 5). The whole list of tested functionals is shown in the SI. All the DFT calculations are carried out in conjunction with the large def2-QZVP^[66] quadruple- ζ basis set to ensure nearly complete basis-set convergence of the results to better gauge the functionals' "true" performance. Herein, we assess the DFAs against the original and our new W_n-F12 reference values for BHPERI.

Finally, we note that for a rigorous comparison with the DFT data, secondary effects that are not included in the DFT calculations, such as relativity, and zero-point vibrational energy corrections, are excluded from the W_n-F12 reference values. The W_n-F12 calculations were carried out using the reference geometries from the GMTKN30 database.^[15,16] However, as they are different from the level of theory prescribed in the W_n-F12 protocol for the geometry optimization (B3LYP-D3(BJ)/A'VTZ), we assess the effect of changing the reference geometry on the W1-F12 reaction barrier heights (Supporting Information Table S2); in this context also note that dispersion-corrections have been shown to be beneficial in DFT geometry optimizations of organic molecules.^[62,67] Inspection of Supporting Information Table S2 reveals that the reference geometries affect mainly the SCF and valence CCSD components in W1-F12, and the effect on the core–valence and valence (T) components is fairly small (< 0.2 kcal mol⁻¹). Specifically, the SCF component is affected by up to 0.9 kcal mol⁻¹ (in

Table 2. Statistical analysis for the performance of composite and *ab initio* procedures for the calculation of the barrier heights in the BHPERI database (kcal mol⁻¹).^[a]

Method	RMSD	MAD	MSD	LND	LPD
CBS-QB3	2.4	2.1	-2.1	-4.7 (9)	n/a
CBS-APNO	2.6	2.0	-2.3	-5.4 (9)	n/a
G4	0.8	0.6	-0.2	-2.9 (9)	0.9 (16)
G4(MP2)	1.3	1.0	-1.0	-3.6 (9)	0.4 (17)
G4(MP2)-6X	1.2	1.0	-0.9	-3.6 (9)	0.4 (17)
G3	1.4	1.1	-1.0	-3.7 (9)	0.8 (13)
G3B3	1.1	0.8	-0.7	-2.9 (9)	0.7 (5)
G3(MP2)	1.9	1.6	-1.6	-4.2 (9)	0.4 (5)
G3(MP2)B3	1.7	1.4	-1.3	-3.6 (9)	0.3 (5)
CCSD(T)/CBS(MP2) ^[b]	0.6	0.6	-0.6	-1.1 (14)	0.2 (5)
CCSD(T)/CBS(MP3) ^[c]	0.3	0.2	0.2	-0.3 (11)	0.6 (21)
CCSD(T)/CBS(MP4) ^[d]	0.5	0.4	-0.4	-1.5 (10)	0.3 (9)
MP2 ^[e]	8.0	7.6	-7.6	-12.2 (9)	n/a
MP3 ^[e]	4.2	3.9	3.9	n/a	6.6 (9)
MP4 ^[e]	3.5	3.4	-3.4	-5.2 (9)	n/a
MP2.5 ^[e]	2.2	2.0	-1.8	-3.0 (16)	1.4 (11)
MP3.5 ^[e]	0.6	0.5	0.3	-0.7 (16)	1.6 (11)
SCS-MP2 ^[e]	1.4	1.1	-0.7	-2.5 (10)	2.7 (11)
SCS-MP3 ^[e]	2.4	2.2	2.2	n/a	5.1 (11)
CCSD(T)/6-31G(d)	1.9	1.5	1.3	-1.0 (18)	5.0 (5)

[a] RMSD = root mean square deviation, MAD = mean absolute deviation, MSD = mean signed deviation, LND = largest negative deviation, LPD = largest positive deviation (the reaction numbers associated with the LND and LPD are given in parenthesis). [b] Obtained with eq. (1) with MP_n = MP2. [c] Obtained with eq. (1) with MP_n = MP3. [d] Obtained with eq. (1) with MP_n = MP4. [e] Extrapolated to the basis-set limit from the AV{T,Q}Z basis set pair with an extrapolation exponents of 5 and 3 for the HF and MP_n correlation contributions, respectively.

absolute value) (reactions **8** and **6**) and the CCSD component is affected by up to 2.1 kcal mol⁻¹ (reaction **6**). However, in nearly all cases the effects on the SCF and CCSD components have opposite signs (i.e., they tend to cancel each other). As a result the final W1-F12 reaction barrier heights obtained with the two geometries differ by less than 0.3 kcal mol⁻¹ in all cases (with the exception of reaction **6** for which a difference of 1.3 kcal mol⁻¹ is observed, see SI for further details). Nevertheless, as all the previous DFT and *ab initio* benchmark studies have used the reference geometries from the GMTKN30 database, we choose to use these geometries in order to facilitate a direct comparison with the previous studies. It is unlikely that the final trends discussed in the next sections would be affected by this.

Results and Discussion

Performance of CBS-type procedures

Table 2 gives an overview of the performance of economical composite procedures for the reaction barrier heights in the BHPERI dataset. We start our discussion with the CBS-type methods CBS-QB3 and CBS-APNO because of their particular relevance in this study. Other composite methods are discussed in the following section. The CBS-QB3 method, which was previously used for obtaining most of the reaction barrier heights in the BHPERI dataset, gives poor performance with an RMSD of 2.4 kcal mol⁻¹ from our benchmark W_n-F12 values. The computationally more expensive CBS-APNO method,

which usually shows better performance than CBS-QB3 for reaction energies,^[1a,8e,8i] results in even a higher RMSD of 2.6 kcal mol⁻¹. Both of these procedures systematically underestimate the reaction barrier heights, as evident from their mean signed deviations (MSDs) having the same absolute magnitudes as their MADs. In both cases the largest underestimation is obtained for the cycloaddition involving two 1,3-cyclopentadiene rings (reaction **9**, Table 1). In particular, CBS-QB3 underestimates this barrier by 4.7 kcal mol⁻¹ and CBS-APNO underestimates it by as much as 5.4 kcal mol⁻¹. Table 3 gives the RMSDs for the subsets in the BHPERI dataset (see also Table 1). Inspection of Table 3 reveals that the CBS methods fail most notably for the cycloaddition reactions in set-I and the Diels–Alder cycloaddition reactions in set-III. Specifically, the RMSDs for the CBS methods vary between 3.1 and 4.1 kcal mol⁻¹ for these two subsets of cycloaddition reactions. We note that the RMSDs for the electrocyclic, sigmatropic shift and 1,3-dipolar cycloaddition subsets are smaller (namely, they vary between 1.4 and 2.1 kcal mol⁻¹, Table 3).

These findings come as a surprise because such large deviations for reaction barrier heights are rarely found for the CBS-type composite procedures. Given the widespread use of these procedures for benchmarking DFT and for quantum chemical modeling it is important to highlight cases for which they do not perform well. How our findings influence the evaluation of

Table 3. Root mean square deviations (RMSDs) of composite and *ab initio* procedures for the subsets in the BHPERI database of barrier heights (kcal mol⁻¹).

Method	Set-I			Set-II [d]	Set-III [e]
	[a]	[b]	[c]		
CBS-QB3	2.0	1.4	3.4	1.5	3.1
CBS-APNO	2.1	2.1	4.1	1.6	3.5
G4	0.5	0.2	1.7	0.5	0.8
G4(MP2)	1.2	0.7	2.3	0.3	1.5
G4(MP2)-6X	1.1	0.7	2.3	0.3	1.5
G3	0.9	0.6	2.3	1.1	1.6
G3B3	0.9	0.6	1.8	0.4	1.4
G3(MP2)	1.6	1.1	2.8	1.2	2.2
G3(MP2)B3	1.6	1.0	2.5	0.6	2.1
CCSD(T)/CBS(MP2) ^[f]	0.4	0.4	0.4	0.7	0.7
CCSD(T)/CBS(MP3) ^[g]	0.2	0.4	0.5	0.2	0.3
CCSD(T)/CBS(MP4) ^[h]	0.5	0.5	0.3	0.4	0.3
MP2 ^[i]	5.2	7.2	10.4	6.8	9.4
MP3 ^[i]	3.0	2.9	4.9	4.0	4.8
MP4 ^[i]	2.3	2.8	4.1	3.5	3.9
MP2.5 ^[i]	1.5	2.2	2.8	2.2	2.3
MP3.5 ^[i]	0.6	0.3	0.4	0.8	0.5
SCS-MP2 ^[i]	0.8	1.0	1.3	1.7	0.9
SCS-MP3 ^[i]	2.0	1.9	2.8	2.5	2.7
CCSD(T)/6-31G(d)	1.1	3.6	2.0	0.7	1.9

[a] Electrocyclic ring closing reactions (reactions **1–3**, Table 1). [b] Sigmatropic shift reactions (reactions **4–6**). [c] Cycloaddition reactions (reactions **7–10**). [d] 1,3-dipolar cycloaddition reactions (reactions **11–19**). [e] Diels–Alder cycloaddition reactions (reactions **20–26**). [f] Obtained with eq. (1) with MP_n = MP2. [g] Obtained with eq. (1) with MP_n = MP3. [h] Obtained with eq. (1) with MP_n = MP4. [i] Extrapolated to the basis-set limit from the AV{T,Q}Z basis set pair with an extrapolation exponents of 5 and 3 for the HF and MP_n correlation contributions, respectively.

approaches, such as DFAs, is discussed later. However, first we analyze the performance of other popular *ab initio* methods for these reaction types.

Performance of other composite procedures

We begin this section with an evaluation of the popular G_n -type methods: G4, G4(MP2), G4(MP2)–6X, G3, G3(MP2), G3B3, and G3(MP2)B3. The G4-type procedures show significantly better performance than the CBS-type procedures. For the entire BHPERI database G4 attains an RMSD of 0.8 kcal mol⁻¹ (Table 2). Furthermore, the G4 procedure gives excellent performance for four out of the five subsets of the BHPERI dataset (Table 3) with RMSDs below the “chemical accuracy” threshold, namely 0.2 (sigmatropic shifts), 0.5 (1,3-dipolar cycloadditions), 0.5 (electrocyclic reactions), and 0.8 kcal mol⁻¹ (Diels–Alder cycloadditions). However, for the challenging subset of cycloaddition reactions in set-I, G4 attains an RMSD of 1.7 kcal mol⁻¹. The G4(MP2) procedure, which is computationally more economical than G4, attains an overall RMSD of 1.3 kcal mol⁻¹ for the entire BHPERI dataset, and the following RMSDs are obtained for the various subsets: 0.3 (1,3-dipolar cycloadditions), 0.7 (sigmatropic shifts), 1.2 (electrocyclic reactions), 1.5 (Diels–Alder cycloadditions), and 2.3 (cycloaddition reactions in set-I) kcal mol⁻¹. The G4(MP2)–6X procedure, which has the same computational cost as G4(MP2), shows very similar performance to G4(MP2) (Tables 2 and 3). All in all, the G4-type procedures are clearly superior to the CBS procedures, for example, for the entire BHPERI dataset the RMSD of G4 (0.8 kcal mol⁻¹) is smaller by nearly 70% than that of CBS-QB3 (2.4 kcal mol⁻¹). Furthermore, for the various subsets in Table 3 the RMSDs of G4 are smaller than those for CBS-QB3 by about 50–80%. Regarding the performance of the G3-type procedures, we note that the G3B3 and G3(MP2)B3 systematically give better performance than G3 and G3(MP2), respectively. However, there seems to be no merit in using the G3-type procedures over the more recent G4-type procedures, as the latter have a similar computational cost and in general give better performance (Tables 2 and 3).

We now turn our attention to the performance of the CCSD(T)/CBS method approximated by eq. (1). Calculating the ΔMP_n basis-set-correction term in eq. (1) at the MP2 level (i.e., the CCSD(T)/CBS(MP2) method) results in excellent performance with an RMSD of only 0.6 kcal mol⁻¹ for the entire BHPERI dataset. We note that this cost-effective method outperforms all of the G_n -type composite methods—including the G4 procedure, which gives an overall RMSD of 0.8 kcal mol⁻¹. Inspection of Table 3 reveals that the CCSD(T)/CBS(MP2) method yields RMSDs that are below the threshold of “chemical accuracy” for all the subsets of the BHPERI dataset. Namely, the RMSDs are: 0.4 (all subsets of Set-I) and 0.7 (Set-II and Set-III) kcal mol⁻¹. We also note that the largest deviation (underestimation) is 1.1 kcal mol⁻¹. For comparison the largest deviation for the G4 methods (again an underestimation) is about 2.5 times higher (namely, 2.9 kcal mol⁻¹).

We examine the effect of calculating the ΔMP_n basis-set-correction term eq. (1) using higher-orders of perturbation

theory. Using $MP_n = MP3$ (i.e., the CCSD(T)/CBS(MP3) method) leads to an overall RMSD of 0.3 kcal mol⁻¹ for the entire BHPERI dataset. This RMSD is smaller by a factor of two compared with the RMSD of 0.6 kcal mol⁻¹ obtained for the CCSD(T)/CBS(MP2) method. Note also that the largest deviation for the CCSD(T)/CBS(MP3) method is 0.6 kcal mol⁻¹ (Table 2), and the RMSDs for all the subsets of the BHPERI dataset are well below the threshold of “chemical accuracy.” Namely, they are: 0.2 (1,3-dipolar cycloaddition and electrocyclic reactions), 0.3 (Diels–Alder cycloadditions), 0.4 (sigmatropic shifts), and 0.5 (cycloaddition reactions of Set-I) kcal mol⁻¹ (Table 3). These results are consistent with previous findings for the barrier heights for multiple proton exchange reactions in water, ammonia, and hydrogen fluoride clusters.^[8d] In particular, it was found that the performance of methods based on eq. (1) using Pople-type basis sets was significantly improved for $MP_n = MP3$ compared to $MP_n = MP2$ (see Refs. [8]c–d for further details). Finally, we note that using $MP_n = MP4$ in eq. (1), generally leads to similar performance as $MP_n = MP3$. Specifically, the RMSD for the entire BHPERI dataset increases from 0.3 [CCSD(T)/CBS(MP3)] to 0.5 [CCSD(T)/CBS(MP4)] kcal mol⁻¹, and the RMSDs for nearly all the subsets are slightly higher for the latter method (apart from the cycloaddition reactions of Set-I, Table 3).

Performance of perturbative wave-function procedures

We turn now to the performance of standard and spin-component-scaled Møller–Plesset perturbation theory for the reaction barrier heights in the BHPERI dataset. We note to begin that MP2, MP3, and MP4 perform poorly with RMSDs for the entire BHPERI dataset of 8.0, 4.2, and 3.5 kcal mol⁻¹, respectively (Table 2). MP2 and MP4 systematically underestimate the barrier heights ($MSD = -1 \times MAD$), whereas MP3 systematically overestimates them ($MSD = MAD$). This is consistent with the MP2, MP3, and MP4 results obtained for a range of tautomerization reactions,^[8a,b] and barrier heights in proton-transfer reactions involving multiple proton exchanges.^[8c,d] Furthermore, as the TSs of pericyclic reactions may partly be viewed as dispersion-bound complexes, these trends are also consistent with the tendency of MP2 and MP4 to overbind dispersion-bound complexes and MP3 to underbind them.^[57,58,68] In this situation, MP2.5, which is an average of MP2 and MP3, outperforms both MP2 and MP3. However, MP2.5 still results in a rather large RMSD of 2.2 kcal mol⁻¹ for the entire BHPERI dataset. We note that MP2.5 has been developed for the description of binding energies in noncovalently bound dimers, and our assessment goes beyond its originally intended use. The MP3.5 procedure, which is similarly defined here as the average of MP3 and MP4, outperforms MP2.5, MP3, and MP4 with an RMSD of merely 0.6 kcal mol⁻¹. We note that MP3.5 gives RMSDs below 1 kcal mol⁻¹ for all the subsets of the BHPERI dataset (specifically, RMSD = 0.3–0.8 kcal mol⁻¹, Table 3). Scaling the same-spin and opposite-spin components of the MP2 correlation energy in the SCS-MP2 procedure largely eliminates the bias of MP2 to underestimate the reaction barrier heights, that is, the MSD improves from –7.6

Table 4. Overview of the basis set convergence of standard wave-function procedures for the calculation of the barrier heights in the BHPERI database (Δ RMSDs, kcal mol⁻¹).^[a]

Method	A'VDZ ^[b]	A'VTZ ^[c]	A'VQZ ^[d]
MP2	2.1	1.0	0.4
MP3	-2.7	-1.2	-0.5
MP4	2.5	1.1	0.4
MP2.5	2.4	1.1	0.4
MP3.5	2.0	0.5	-0.1
SCS-MP2	2.3	0.9	0.3
SCS-MP3	-1.1	-1.0	-0.4

[a] The tabulated values are the differences (Δ RMSDs) between the RMSDs obtained with the A'V{T,Q}Z basis set pair and the A'VnZ basis set ($n = D, T, Q$). [b] Δ RMSD = RMSD(A'VDZ) - RMSD(A'V{T,Q}Z). [c] Δ RMSD = RMSD(A'VTZ) - RMSD(A'V{T,Q}Z). [d] Δ RMSD = RMSD(A'VQZ) - RMSD(A'V{T,Q}Z).

(MP2) to -0.7 (SCS-MP2) kcal mol⁻¹. The overall RMSD for the SCS-MP2 procedure is 1.4 kcal mol⁻¹. Therefore, the new MP3.5 procedure is the only MPn-based procedure that attains an RMSD below the threshold of "chemical accuracy" (Table 2).

The performance of the standard and modified MPn procedures in Tables 2 and 3 is evaluated at the basis-set limit (i.e., at the MPn/A'V{T,Q}Z level of theory). It is also of interest to evaluate the performance of these procedures using smaller basis sets. Table 4 gives an overview of the basis-set convergence for the MPn-based procedures. The tabulated values give the difference between the RMSDs obtained with A'V{T,Q}Z basis set pair (reported in Table 2) and those obtained with smaller basis sets, Δ RMSD = RMSD(A'VnZ) - RMSD(A'V{T,Q}Z), $n = D, T$, and Q . Thus, a positive Δ RMSD indicates that the use of the smaller basis set results in an overall deterioration in the performance of the method, whereas a negative value indicates an improvement in performance with the smaller basis set. Inspection of Table 4 reveals that the MP2, MP4, MP2.5, MP3.5, and SCS-MP2 procedures converge fairly smoothly to the basis-set limit. Using the A'VQZ basis set results in RMSDs that are larger by up to 0.4 kcal mol⁻¹ than those obtained at the basis-set limit. For these methods, the use of the A'VTZ basis set results in RMSDs that are larger by about 1 kcal mol⁻¹ than those obtained at the basis-set limit (with the exception of MP3.5 which seems to converge faster to the basis-set limit with a Δ RMSD of 0.5 kcal mol⁻¹). The A'VDZ basis set results in Δ RMSDs > 2 kcal mol⁻¹ and is therefore not recommended. Finally, we note that MP3 and SCS-MP3 have negative Δ RMSDs, that is their performance deteriorates with the size of the basis set due to error cancellation between basis set incompleteness errors and "high-level corrections" (CCSD(T) - MP3 or CCSD(T) - SCS-MP3) differences.

On the quality of the reference values in a DFT benchmark study

The main purpose of this section is to demonstrate in what way reference values of different quality can affect a DFA benchmark study. Supporting Information Table S3 shows the statistical values for all 45 tested density functionals over the entire BHPERI set based on the old (mostly) CBS-QB3 and the

updated *Wn*-F12 reference values. An analysis of the individual RMSDs reveals that for the new reference values 18 of the tested methods — 11 of which are GGA functionals—show a chemically significant increase, that is, Δ RMSD > 1 kcal mol⁻¹. Conversely, 11 methods show a significant improvement of up to 2 kcal mol⁻¹. For the remaining methods we observe absolute changes below the 1 kcal mol⁻¹ threshold of chemical accuracy. In summary, 64% of the evaluated DFAs show significant changes between the old and new sets of reference values.

To further analyze the importance of this finding, we average the statistical values for each rung of Jacob's Ladder. For instance, an averaged RMSD for GGA functionals means that we obtained 22 separate RMSDs over the BHPERI set for all tested GGAs and then formed the average over those values. Average RMSDs, MSDs and error ranges are shown in Table 5. First of all, we notice that all averaged MSDs are shifted down by 1.6 kcal mol⁻¹ when going from the old to the new *Wn*-F12 benchmark—a value that again is chemically significant. MSDs can be used to identify specific trends, such as a systematic over- or under-estimation of the reaction barrier heights, or a rather balanced description if MSDs are close to zero. A change in reference values can therefore also influence the way those trends are perceived. For the GGA functionals this appears not to be a problem because already with respect to the old reference they tend to strongly underestimate barrier heights (averaged MSD = -4.4 kcal mol⁻¹). This effect is simply more pronounced for the *Wn*-F12 benchmark (averaged MSD = -6.0 kcal mol⁻¹). Contrary to that, we see a different picture for the remaining DFAs, particularly for the hybrids and double-hybrids. In the case of the original reference values, the first show on average an overestimation of about 1.1 kcal mol⁻¹, while the latter seem to be very balanced with an averaged MSD close to zero. This picture is reversed for the new benchmark, with the hybrids being closer to zero and the double-hybrids having an averaged MSD of -1.5 kcal mol⁻¹. Thus, it is obvious that the quality of reference values directly influences our general perception of a method.

The overall accuracy in benchmark studies is usually gauged with MADs or, as herein, with RMSDs. The averaged RMSDs are given in Table 5. Both sets of reference data order the functionals according to the Jacob's Ladder scheme, with GGA functionals yielding the largest errors and double-hybrids the smallest. However, a significant difference between using the old and new reference values can be seen for GGA functionals

Table 5. Root mean square deviations (RMSDs), mean signed deviations (MSDs) and error ranges, averaged for the various rungs of Jacob's Ladder (kcal mol⁻¹).

Functional class	Averaged RMSD		Averaged MSD		Averaged error range	
	Old	New	Old	New	Old	New
GGA	5.3	6.5	-4.4	-6.0	10.7	9.9
meta-GGA	3.7	3.8	-1.0	-2.6	8.5	6.3
hybrid-GGA	3.3	2.9	1.1	-0.5	10.3	8.3
double-hybrid	1.6	1.9	0.1	-1.5	5.2	3.7

Table 6. The three best and worst functionals for their respective rung on Jacob's Ladder based on the root mean square deviations shown in parentheses (kcal mol⁻¹). The results are based on both the CBS-QB3/W1 ("Old") and *Wn*-F12 ("New") reference values.^[a]

		Best		Worst	
		Old	New	Old	New
GGA	1.	B97-D ^[71] (3.3)	mPW-LYP ^[72,73] (3.8)	O-PBE ^[74,75] (9.8)	O-PBE (11.4)
	2.	rPW86-PBE ^[75,76] (3.3)	rPW86-PBE (4.1)	PBESOL ^[77] (8.6)	PBESOL (10.1)
	3.	B-LYP ^[73,78] (3.6)	B-LYP/B97-D (4.3)	O-LYP ^[73,74] (6.0)	O-LYP/B-PBE ^[75,78] (7.5)
Hybrid	1.	TPSS0 ^[79] (2.0)	PW6B95 ^[70] (1.3)	BH-LYP ^[80] (6.2)	TPSSH ^[81] (5.9)
	2.	M05-2X ^[69] (2.0)	B3-LYP ^[82] (1.5)	LC- ω PBE ^[83] (5.1)	BH-LYP (4.4)
	3.	B1B95 ^[84] /B3-LYP (2.3)	M05-2X (1.7)	M05 ^[85] (5.0)	M06-HF ^[86] (4.0)
Double-Hybrid	1.	DSD-BLYP ^[87] (1.4)	PWPB95 ^[16] (1.0)	PWPB95 (2.3)	DSD-PBEP86 ^[88] (2.7)
	2.	B2GP-PLYP ^[25] (1.4)	B2K-PLYP ^[64] (1.5)	B2-PLYP ^[26] (1.7)	DSD-BLYP (2.3)
	3.	B2K-PLYP (1.4)	B2GP-PLYP (1.8)	DSD-PBEP86 (1.5)	B2-PLYP (2.0)

The numbers were obtained with respect to the original CBS-QB3/W1 ("Old") and *Wn*-F12 ("New") reference values. [a] All functionals are dispersion-corrected with Grimme's DFT-D3(BJ), except for M052X and M06HF, which were corrected with DFT-D3 and "zero-damping." The D3 suffix was omitted in the table for better clarity.

that turn out to be even less accurate than originally thought (averaged RMSD = 5.3 kcal mol⁻¹ for CBS-QB3/W1 vs. 6.5 kcal mol⁻¹ for *Wn*-F12). Interestingly, the changes are much smaller for the three higher rungs. Inspection of the individual results reveals that this is mainly due to two reasons. First, the reaction barrier heights of some DFT methods lie in between the CBS-QB3 and *Wn*-F12 barriers, with similar absolute deviations from both. For instance, the double-hybrid B2K-PLYP-D3(BJ)^[25,64] gives a barrier height of 19.6 kcal mol⁻¹ for reaction **24** (see Table 1), which overestimates the CBS-QB3 barrier by 1.4 kcal mol⁻¹, but underestimates the *W1*-F12 result by 1.8 kcal mol⁻¹. A second reason is that two methods "swap places," meaning that an average over both of them gives rise to very similar statistics. For instance, with respect to CBS-QB3, the hybrid M052X-D3^[69] slightly overestimates the barrier of reaction **20** by 0.4 kcal mol⁻¹, while the hybrid PW6B95-D3(BJ)^[70] overestimates it by 2.4 kcal mol. The order of those two functionals and the absolute magnitude of their deviations reverse when going to *W1*-F12, with M05-2X-D3 underestimating the barrier by 2.7 kcal mol⁻¹, while PW6B95-D3(BJ) deviates only by -0.7 kcal mol⁻¹.

Finally, Table 5 also shows averaged error ranges, which can give further insight into a method's robustness and reliability. Also here, both sets of reference values provide the smallest error ranges for double-hybrids. For all four functional classes the new, higher-quality reference values ultimately yield smaller error distributions than the old ones. The largest effect for changing the reference can be seen for the hybrid functionals. While with respect to CBS-QB3/W1 they have an average error range close to that of the GGAs (10.3 vs. 10.7 kcal mol⁻¹), the gap between hybrids and GGAs is larger for the updated reference values (8.3 vs. 9.9 kcal mol⁻¹). Thus the interpretation about the robustness of hybrids relatively to GGAs changes. As expected from the discussion in the previous sections, the new reference values have the largest impact on sets II and III, as eight of the ten reactions in set-I already had *W1* numbers as the reference, which are practically the same as *W1*-F12. However, overall the same conclusions can be drawn as for the entire BHPERI set (see Supporting Information Tables S6 and S7). Noteworthy is the influence of the

new reference values on set-III for which we notice a large decrease in all MSDs of 3.1 kcal mol⁻¹!

Usually a benchmark study ends with recommending a method that particularly works well for the investigated problem. The BHPERI set of 26 barrier heights had been designed to be part of the large GMTKN30 database, which had the aim to test methods for the robustness and to recommend only those that seemed to be generally applicable over an entire range of chemical problems. In the spirit of GMTKN30 and because our aim is not to identify one particular functional that works well for pericyclic reactions, we therefore try to limit such recommendations to a minimum. Nevertheless, we can demonstrate how the quality of the reference data does influence the process of identifying promising methods. For that purpose we contemplate not only the best method, but rather the functional ranking, which means we demonstrate which functionals can be considered the three best and the three worst in their respective class. Such rankings are shown for GGAs, hybrids and double-hybrids in Table 6; meta-GGAs are left out of this discussion, as we have only analyzed a total of three. As Table 6 shows, the effect of the quality of the reference values has only a small influence on the GGA functionals and the three best and three worst methods are very similar for both sets of reference numbers. Contrary to that, some surprising changes are observed for the hybrids and double-hybrids. According to the old CBS-QB3/W1 benchmark, TPSS0-D3(BJ)^[79] is the best hybrid, followed by the more established M052X-D3, B1B95-D3(BJ)^[84] and B3LYP-D3(BJ)^[82] approaches. TPSS0 is similar to the TPSSH hybrid,^[81] only with 25% of Fock exchange instead of 10%. This recommendation seems rather exotic, as TPSS0 has not been established as particularly robust^[17] nor has it appeared very often in the literature. Indeed, when the new reference values are applied, TPSS0 disappears from the list of the best three functionals and moreover the related TPSSH-D3(BJ) is now identified as the worst hybrid for BHPERI. M052X-D3 and B3LYP-D3(BJ) remain in the list of the top three hybrids, however the best hybrid with respect to the new reference values turns out to be PW6B95-D3(BJ). We note that PW6B95 has been found to be one of the most accurate and robust hybrid functionals for

main-group, transition-metal, and polypeptide chemistry, both for energetics and geometries.^{[8e,h],[17,25,62,64,89–92]} In this sense, it is encouraging to see that using reference values with a better quality, also the statistics of known accurate DFAs seem to improve.

In passing, we also note that functionals that are more likely to be free from the self-interaction error (SIE) appear in the list of the worst three functionals. Examples are range-separated methods, such as LC- ω PBE-D3(BJ),^[83] or functionals with relatively large fractions of Fock-exchange (BH-LYP-D3(BJ)^[80] and M06HF-D3).^[86] It is therefore conceivable that also SIE-unrelated problems contribute to the DFAs' deviations from the reference values, in contrast to other common barrier-height test sets.^[17,30a,b]

We end our discussion with an analysis of double-hybrid functionals. In this context it is important to keep in mind that the energetic differences among those are much smaller than for the other classes of functionals and that our comparison here is done at a very high level. In other words, usually any of the six tested double-hybrids can be used in related applications with reasonable reliability. Similar to what has been seen for hybrids, two functionals belong to the best three approaches for both sets of reference values, namely the general-purpose functional B2GP-PLYP-D3(BJ)^[25] and the kinetics functional B2K-PLYP-D3(BJ), which in contrast to the hybrids is the only specific functional for kinetics that stands out in this study. These two methods yield the same RMSD for BHPERI as the DSD-BLYP-D3(BJ)^[87] functional (1.4 kcal mol⁻¹) with respect to the original reference. However, when reanalyzing BHPERI with the new reference values, the gap between the functionals increases and moreover PWPB95-D3(BJ)^[16] moves to the top of the list with the best RMSDs of the entire study of 45 DFAs (1.0 kcal mol⁻¹). Interestingly, the same method is the worst of the six tested double-hybrids when benchmarked against the old numbers (RMSD = 2.3 kcal mol⁻¹). However, PWPB95 has generally been shown to be accurate and reliable and the new analysis based on the updated reference values goes perfectly in line with those previous findings.^[16,17,27,34,91] As with PW6B95 for the hybrids, we therefore see again that better reference values give better statistical values for methods that are already been known to be robust and accurate.

The last statement is further strengthened by the fact that a similar connection between quality of the benchmark and accuracy of reliable DFAs has also been found for the relative stability of tetrapeptide conformations.^[34] Thus, we can state that although we have only dealt with one specific benchmark set, the demonstrated influence of the reference values on the interpretation of DFA applicability is representative for general benchmarking problems.

Conclusions

We obtain accurate electronic barrier heights for the 26 reactions in the BHPERI dataset, which includes electrocyclic ring-closing, sigmatropic-shift, and 1,3-dipolar and Diels–Alder cycloaddition reactions. Originally, most of the reaction barrier heights in BHPERI had been obtained with the popular com-

posite CBS-QB3 method. Herein, we report new benchmark values for all 26 reactions by means of the high-level, *ab initio* W1-F12 and W2-F12 thermochemical protocols. These benchmark reaction barrier heights allow us to evaluate the performance of more approximate theoretical procedures for the calculation of barrier heights of pericyclic reactions. We initially assess the composite thermochemical procedures CBS-QB3 and CBS-APNO, followed by the *Gn*-type approaches and estimated CCSD(T)/CBS(MP n) procedures. We find a surprisingly large deviation for CBS-QB3 with an overall RMSD of 2.4 kcal mol⁻¹ from our benchmark *Wn*-F12 values. The computationally more expensive CBS-APNO procedure results in a slightly larger RMSD of 2.6 kcal mol⁻¹, indicating a general difficulty of the CBS-type approach for these reaction barrier heights. Contrary to that, the remaining composite procedures fare better, and particularly the *G4*-based methods give RMSDs that are close or within the 1 kcal mol⁻¹ threshold of “chemical accuracy.” The same is also true for the CCSD(T)/CBS(MP n) procedures, which are based on MP n theory at the CBS limit corrected by the systematic differences between MP n and CCSD(T) correlation energies.

We round up our analysis of wave-function methods by assessing various Møller–Plesset perturbative approaches. While MP2, MP3 and MP4 result in large RMSDs ranging from 3.5 to 8.0 kcal mol⁻¹, SCS-MP2 results in an acceptable RMSD of 1.4 kcal mol⁻¹, and the herein newly introduced MP3.5 procedure results in an RMSD of only 0.6 kcal mol⁻¹. Note that the related SCS-MP3 and MP2.5 methods yield significantly larger RMSDs of 2.4 and 2.2 kcal mol⁻¹, respectively.

The finding that CBS-QB3 fails for barrier heights of pericyclic reactions comes as a surprise, as this method generally gives reliable thermochemical and kinetic data. We therefore advise caution when applying this method to similar problems and recommend the use of the other procedures outlined above. The importance of obtaining accurate thermochemical and kinetic data is particularly important in the framework of DFT benchmarking, as demonstrated in an analysis of 45 density functionals covering the four highest rungs of the Jacob's Ladder classification. In 64% of the tested cases significant differences are observed when the *Wn*-F12 values were used as a reference instead of the CBS-QB3 values. In particular, the RMSDs and MSDs change by more than 1 kcal mol⁻¹ in these cases. The error ranges for most density-functional-theory approximations were also smaller when the new reference numbers were used. Although both sets of benchmark values reproduce the correct functional order according to Jacob's Ladder, they provide different functional rankings. Furthermore, in some cases the CBS-QB3 reference values favor functionals that had not been routinely applied before (e.g., TPSS0), while the *Wn*-F12 reference values favor methods that had been shown to be robust and accurate for general applications in previous benchmark studies (e.g., PW6B95 and PWPB95). In other words, the errors for density functionals that have been previously shown to give better general performance decrease when the more accurate benchmark values are used.

Finally, we conclude that dispersion-corrected double-hybrid density functionals belong to the most accurate methods in our study, outperforming (SCS)-MP n methods, thus confirming previous studies.^[17,34] Moreover, for the barrier heights in the BHPERI dataset, double-hybrids are more accurate than the CBS-QB3 method. This is in itself an important finding, particularly as CBS-QB3 has been used before to assess the performance of double-hybrids for this dataset.^[16,17,27] Of the double-hybrid functionals, PWPB95-D3(BJ) and B2K-PLYP-D3(BJ) give the best performance with RMSDs of only 1.0 and 1.5 kcal mol⁻¹, respectively.

Acknowledgments

The authors gratefully acknowledge the generous allocation of computing time from the National Computational Infrastructure (NCI) National Facility within the National Computational Merit Allocation Scheme (projects dv9 and fk5), and the support of iVEC through the use of advanced computing resources located at iVEC@UWA.

Keywords: reaction barrier heights · pericyclic reactions · CBS-QB3 · W1-F12 theory · density functional theory benchmarking

How to cite this article: A. Karton, L. Goerigk *J. Comput. Chem.* **2015**, DOI: 10.1002/jcc.23837

 Additional Supporting Information may be found in the online version of this article.

- [1] (a) K. A. Peterson, D. Feller, D. A. Dixon, *Theor. Chem. Acc.* **2012**, *131*, 1079; (b) T. Helgaker, W. Klopper, D. P. Tew, *Mol. Phys.* **2008**, *106*, 2107; (c) L. A. Curtiss, P. C. Redfern, K. Raghavachari, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 810. (d) J. M. L. Martin, *Annu. Rep. Comput. Chem.* **2005**, *1*, 31.
- [2] L. A. Curtiss, P. C. Redfern, K. Raghavachari, *J. Chem. Phys.* **2007**, *126*, 084108.
- [3] L. A. Curtiss, P. C. Redfern, K. Raghavachari, *J. Chem. Phys.* **2007**, *127*, 124105.
- [4] B. Chan, J. Deng, L. Radom, *J. Chem. Theory Comput.* **2011**, *7*, 112.
- [5] G. da Silva, *Chem. Phys. Lett.* **2013**, *558*, 109.
- [6] J. W. Ochterski, G. A. Petersson, J. A. Montgomery, Jr., *J. Chem. Phys.* **1996**, *104*, 2598.
- [7] (a) J. A. Montgomery, Jr., M. J. Frisch, J. W. Ochterski, G. A. Petersson, *J. Chem. Phys.* **1999**, *110*, 2822; (b) J. A. Montgomery, Jr., M. J. Frisch, J. W. Ochterski, G. A. Petersson, *J. Chem. Phys.* **2000**, *112*, 6532.
- [8] (a) A. Karton, L.-J. Yu, *Chem. Phys.* **2014**, *441*, 166; (b) L.-J. Yu, F. Sarrami, A. Karton, R. J. O'Reilly, *Mol. Phys.*, in press, DOI:10.1080/00268976.2014.986238; (c) A. Karton, R. J. O'Reilly, L. Radom, *J. Phys. Chem. A* **2012**, *116*, 4211; (d) A. Karton, R. J. O'Reilly, B. Chan, L. Radom, *J. Chem. Theory Comput.* **2012**, *8*, 3128; (e) A. Karton, J. M. L. Martin, *Mol. Phys.* **2012**, *110*, 2477; (f) A. Karton, S. Daon, J. M. L. Martin, *Chem. Phys. Lett.* **2011**, *510*, 165; (g) L. A. Curtiss, P. C. Redfern, K. Raghavachari, *Chem. Phys. Lett.* **2010**, *499*, 168; (h) J. Zheng, Y. Zhao, D. G. Truhlar, *J. Chem. Theory Comput.* **2009**, *5*, 808; (i) A. Karton, D. Gruzman, J. M. L. Martin, *J. Phys. Chem. A* **2009**, *113*, 8434; (j) S. Parthiban, G. de Oliveira, J. M. L. Martin *J. Phys. Chem. A* **2001**, *105*, 895.
- [9] A. Karton, L.-J. Yu, M. K. Kesharwani, J. M. L. Martin, *Theor. Chem. Acc.* **2014**, *133*, 1483.
- [10] (a) S. So, U. Wille, G. da Silva, *Environ. Sci. Technol.* **2014**, *48*, 6694; (b) P. Ghesquière, D. Talbi, A. Karton, *Chem. Phys. Lett.* **2014**, *595*, 13; (c) S. Liu, S. Srinivasan, M. C. Grady, M. Soroush, A. M. Rappe, *Int. J. Quantum Chem.* **2014**, *114*, 345; (d) B. Sun, M. L. McKee, *J. Phys. Chem. A* **2013**, *117*, 9076; (e) B. Sun, M. L. McKee, *Inorg. Chem.* **2013**, *52*, 5962; (f) G. da Silva, *J. Phys. Chem. A* **2012**, *116*, 5317; (g) R. S. Assary, L. A. Curtiss, *Energy Fuels* **2012**, *26*, 1344; (h) G. da Silva, *J. Phys. Chem. A* **2011**, *115*, 291; (i) G. da Silva, A. J. Trevitt, *Phys. Chem. Chem. Phys.* **2011**, *13*, 8940.
- [11] (a) A. C. Davis, S. M. Sarathy, *J. Phys. Chem. A* **2013**, *117*, 7670; (b) H. K. Chakravarty, R. X. Fernandes, *J. Phys. Chem. A* **2013**, *117*, 5028; (c) Y. Sakai, H. Ando, T. Oguchi, Y. Murakami, *Chem. Phys. Lett.* **2013**, *556*, 29; (d) T. L. Nguyen, J. Park, K. Lee, K. Song, J. R. Barker, *J. Phys. Chem. A* **2011**, *115*, 4894; (e) X. Song, C. A. Parish, *J. Phys. Chem. A* **2011**, *115*, 2882; (f) A. C. Davis, J. S. Francisco, *J. Phys. Chem. A* **2011**, *115*, 2966; (g) R. Asatryan, G. da Silva, J. W. Bozzelli, *J. Phys. Chem. A* **2010**, *114*, 8302; (h) S. Sharma, S. Raman, W. H. Green, *J. Phys. Chem. A* **2010**, *114*, 5689; (i) G. Black, J. M. Simmie, *J. Comp. Chem.* **2010**, *31*, 1236.
- [12] (a) V. Guner, K. S. Khuong, A. G. Leach, P. S. Lee, M. D. Bartberger, K. N. Houk, *J. Phys. Chem. A* **2003**, *107*, 11445; (b) D. H. Ess, K. N. Houk, *J. Phys. Chem. A* **2005**, *109*, 9542.
- [13] T. C. Dinadayalane, R. Vijaya, A. Smitha, G. N. Sastry, *J. Phys. Chem. A* **2002**, *106*, 1627.
- [14] S. Grimme, C. Mück-Lichtenfeld, E.-U. Würthwein, A. W. Ehlers, T. P. M. Goumans, K. Lammertsma, *J. Phys. Chem. A* **2006**, *110*, 2583.
- [15] L. Goerigk, S. Grimme, *J. Chem. Theory Comput.* **2010**, *6*, 107.
- [16] L. Goerigk, S. Grimme, *J. Chem. Theory Comput.* **2011**, *7*, 291.
- [17] L. Goerigk, S. Grimme, *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670.
- [18] M. Korth, W. Thiel, *J. Chem. Theory Comput.* **2011**, *7*, 2929.
- [19] (a) J. M. L. Martin, G. de Oliveira, *J. Chem. Phys.* **1999**, *111*, 1843; (b) S. Parthiban, J. M. L. Martin, *J. Chem. Phys.* **2001**, *114*, 6014.
- [20] A. Karton, J. M. L. Martin, *J. Chem. Phys.* **2012**, *136*, 124114.
- [21] (a) N. J. DeYonker, T. R. Cundari, A. K. Wilson, *J. Chem. Phys.* **2006**, *124*, 114104; (b) N. J. DeYonker, T. Grimes, S. Yockel, A. Dinescu, B. Mintz, T. R. Cundari, A. K. Wilson, *J. Chem. Phys.* **2006**, *125*, 104111.
- [22] (a) R. B. Woodward, R. Hoffmann, *Angew. Chem. Int. Ed. Engl.* **1969**, *8*, 781; (b) K. N. Houk, Y. Li, J. D. Evanseck, *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 682.
- [23] W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133.
- [24] A. D. Boese, J. M. L. Martin, *J. Chem. Phys.* **2004**, *121*, 3405.
- [25] A. Karton, A. Tarnopolsky, J.-F. Lamere, G. C. Schatz, J. M. L. Martin, *J. Phys. Chem. A* **2008**, *112*, 12868.
- [26] S. Grimme, *J. Chem. Phys.* **2006**, *124*, 034108.
- [27] L. Goerigk, S. Grimme, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 576.
- [28] R. Peverati, D. G. Truhlar, *Phil. Trans. R. Soc. A* **2013**, *372*, 20120476.
- [29] (a) J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, L. A. Curtiss, *J. Chem. Phys.* **1989**, *90*, 5622; (b) L. A. Curtiss, K. Raghavachari, P. C. Redfern, J. A. Pople, *J. Chem. Phys.* **1997**, *106*, 1063.
- [30] (a) Y. Zhao, B. J. Lynch, D. G. Truhlar, *J. Phys. Chem. A* **2004**, *108*, 2715; (b) Y. Zhao, N. Gonzalez-Garcia, D. G. Truhlar, *J. Phys. Chem. A* **2005**, *109*, 2012; (c) P. Jurecka, J. Sponer, J. Cerny, P. Hobza, *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985; (d) S. Grimme, M. Steinmetz, M. Korth, *J. Org. Chem.* **2007**, *72*, 2118.
- [31] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215.
- [32] J. P. Perdew, K. Schmidt, in *Density Functional Theory and Its Applications to Materials*, AIP Conference Proceedings; V. E. Van Doren, C. Van Alsenoy, P. Geerlings, Eds.; American Institute of Physics: Melville, **2001**; Vol. 577, pp 1–20.
- [33] (a) W. Hujo, S. Grimme, *J. Chem. Theory Comput.* **2011**, *7*, 3866; (b) L. Goerigk, *J. Chem. Theory Comput.* **2014**, *10*, 968.
- [34] L. Goerigk, A. Karton, J. M. L. Martin, L. Radom, *Phys. Chem. Chem. Phys.* **2013**, *15*, 7028.
- [35] (a) H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklaß, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, *MOLPRO is a package of ab initio programs*, Available at: <http://www.molpro.net>; (b) H.-J. Werner, P. J. Knowles, G. Knizia,

- F. R. Manby, M. Schütz, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, 2, 242.
- [36] A. Karton, E. Rabinovich, J. M. L. Martin, B. Ruscic, *J. Chem. Phys.* **2006**, 125, 144108.
- [37] C. Hättig, W. Klopper, A. Köhn, D. P. Tew, *Chem. Rev.* **2012**, 112, 4.
- [38] K. A. Peterson, T. B. Adler, H.-J. Werner, *J. Chem. Phys.* **2008**, 128, 084102.
- [39] J. Noga, S. Kedzuch, J. Šimunek, *J. Chem. Phys.* **2007**, 127, 034106.
- [40] G. Knizia, H.-J. Werner, *J. Chem. Phys.* **2008**, 128, 154103.
- [41] T. B. Adler, G. Knizia, H.-J. Werner, *J. Chem. Phys.* **2007**, 127, 221106.
- [42] J. G. Hill, K. A. Peterson, G. Knizia, H.-J. Werner, *J. Chem. Phys.* **2009**, 131, 194105.
- [43] G. Knizia, T. B. Adler, H.-J. Werner, *J. Chem. Phys.* **2009**, 130, 054104.
- [44] S. Ten-no, J. Noga, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, 2, 114.
- [45] S. Ten-no, *Chem. Phys. Lett.* **2004**, 398, 56.
- [46] H.-J. Werner, T. B. Adler, F. R. Manby, *J. Chem. Phys.* **2007**, 126, 164102.
- [47] K. A. Peterson, T. H. Dunning, *J. Chem. Phys.* **2002**, 117, 10548.
- [48] T. H. Dunning, *J. Chem. Phys.* **1989**, 90, 1007.
- [49] R. A. Kendall, T. H. Dunning, Jr., R. J. Harrison, *J. Chem. Phys.* **1992**, 96, 6796.
- [50] T. H. Dunning, K. A. Peterson, A. K. Wilson, *J. Chem. Phys.* **2001**, 114, 9244.
- [51] L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov, J. A. Pople, *J. Chem. Phys.* **1998**, 109, 7764.
- [52] L. A. Curtiss, P. C. Redfern, K. Raghavachari, V. Rassolov, J. A. Pople, *J. Chem. Phys.* **1999**, 110, 4703.
- [53] A. G. Baboul, L. A. Curtiss, P. C. Redfern, K. Raghavachari, *J. Chem. Phys.* **1999**, 110, 7650.
- [54] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian 09, Revision D.01; Gaussian, Inc.: Wallingford CT, **2009**.
- [55] (a) D. G. Liakos, F. Neese, *J. Phys. Chem. A* **2012**, 116, 4801; (b) P. Jurečka, P. Hobza, *Chem. Phys. Lett.* **2002**, 365, 89; (c) W. Klopper, H. P. Lüthi, *Mol. Phys.* **1999**, 96, 559; (d) J. M. L. Martin, *Theor. Chem. Acc.* **1997**, 97, 227.
- [56] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, A. K. Wilson, *Chem. Phys. Lett.* **1998**, 286, 243.
- [57] M. Pitonak, P. Neogady, J. Cerny, S. Grimme, P. Hobza, *Chem. Phys. Chem.* **2009**, 10, 282.
- [58] (a) S. Grimme, *J. Chem. Phys.* **2003**, 118, 9095; (b) S. Grimme, L. Goerigk, R. F. Fink, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, 2, 886.
- [59] S. Grimme, *J. Comput. Chem.* **2003**, 24, 1529.
- [60] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, 132, 154104.
- [61] (a) A. D. Becke, E. R. Johnson, *J. Chem. Phys.* **2005**, 123, 024101; (b) E. R. Johnson, A. D. Becke, *J. Chem. Phys.* **2005**, 123, 154101; (c) E. R. Johnson, A. D. Becke, *J. Chem. Phys.* **2006**, 124, 174104.
- [62] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, 32, 1456.
- [63] S. Grimme, DFT-D3 program, Grimme group website, Available at: <http://www.thch.uni-bonn.de/tc/index.php?section=downloads&subsection=DFT-D3&lang=english>.
- [64] A. Tarnopolsky, A. Karton, R. Sertchook, D. Vuzman, J. M. L. Martin, *J. Phys. Chem. A* **2008**, 112, 3.
- [65] F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, 2, 73.
- [66] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, 7, 3297.
- [67] S. Grimme, M. Steinmetz, *Phys. Chem. Chem. Phys.* **2013**, 15, 16031.
- [68] (a) P. Pyykkö, *Chem. Rev.* **1997**, 97, 597; (b) N. Runeberg, P. Pyykkö, *Int. J. Quant. Chem.* **1998**, 66, 131; (c) S. M. Cybulski, M. L. Lytle, *J. Chem. Phys.* **2007**, 127, 141102.
- [69] Y. Zhao, N. E. Schultz, D. G. Truhlar, *J. Chem. Theory Comput.* **2006**, 2, 364.
- [70] Y. Zhao, D. G. Truhlar, *J. Phys. Chem. A* **2005**, 109, 5656.
- [71] S. Grimme, *J. Comput. Chem.* **2006**, 27, 1787.
- [72] C. Adamo, V. Barone, *J. Chem. Phys.* **1998**, 108, 664.
- [73] (a) C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, 37, 785; (b) B. Miehlich, A. Savin, H. Stoll, H. Preuss, *Chem. Phys. Lett.* **1989**, 157, 200.
- [74] N. C. Handy, A. J. Cohen, *Mol. Phys.* **2001**, 99, 403.
- [75] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, 77, 3865.
- [76] E. D. Murray, K. Lee, D. C. Langreth, *J. Chem. Theory Comput.* **2009**, 5, 2754.
- [77] J. P. Perdew, A. Ruzsinsky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, K. Burke, *Phys. Rev. Lett.* **2008**, 100, 136406.
- [78] A. D. Becke, *Phys. Rev. A* **1988**, 38, 3098.
- [79] S. Grimme, *J. Phys. Chem. A* **2005**, 109, 3067.
- [80] A. D. Becke, *J. Chem. Phys.* **1993**, 98, 1372.
- [81] V. N. Staroverov, G. E. Scuseria, J. Tao, J. P. Perdew, *J. Chem. Phys.* **2003**, 119, 12129.
- [82] (a) A. D. Becke, *J. Chem. Phys.* **1993**, 98, 5648; (b) P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, 98, 11623.
- [83] O. A. Vydrov, G. E. Scuseria, *J. Chem. Phys.* **2006**, 125, 234109.
- [84] A. D. Becke, *J. Chem. Phys.* **1996**, 104, 1040.
- [85] Y. Zhao, N. E. Schultz, D. G. Truhlar, *J. Chem. Phys.* **2005**, 123, 161103.
- [86] Y. Zhao, D. G. Truhlar, *J. Phys. Chem. A* **2007**, 110, 13126.
- [87] S. Kozuch, D. Gruzman, J. M. L. Martin, *J. Phys. Chem. C* **2010**, 114, 20801.
- [88] S. Kozuch, J. M. L. Martin, *Phys. Chem. Chem. Phys.* **2011**, 12, 20104.
- [89] M. M. Quintal, A. Karton, M. A. Iron, A. D. Boese, J. M. L. Martin, *J. Phys. Chem. A* **2006**, 110, 709.
- [90] M. Korth, S. Grimme, *J. Chem. Theory Comput.* **2009**, 5, 993.
- [91] M. Steinmetz, S. Grimme, *ChemistryOpen* **2013**, 2, 115.
- [92] (a) L. Goerigk, J. R. Reimers, *J. Chem. Theory Comput.* **2013**, 9, 3240; (b) L. Goerigk, C. A. Collyer, J. R. Reimers, *J. Phys. Chem. B* **2014**, 118, 14612.

Received: 3 November 2014

Revised: 17 December 2014

Accepted: 21 December 2014

Published online on 00 Month 2014