

Evaluation of the performance of MP4-based procedures for a wide range of thermochemical and kinetic properties



Li-Juan Yu, Wenchao Wan, Amir Karton*

School of Chemistry and Biochemistry, The University of Western Australia, Perth, WA 6009, Australia

ARTICLE INFO

Article history:

Received 26 August 2016

In final form 10 October 2016

Available online 15 October 2016

Keywords:

SCS-MP2

MP2.5

SCS-MP3

MP3.5

CCSD(T)

ABSTRACT

We evaluate the performance of standard and modified MP n procedures for a wide set of thermochemical and kinetic properties, including atomization energies, structural isomerization energies, conformational energies, and reaction barrier heights. The reference data are obtained at the CCSD(T)/CBS level by means of the W n thermochemical protocols. We find that none of the MP n -based procedures show acceptable performance for the challenging W4-11 and BH76 databases. For the other thermochemical/kinetic databases, the MP2.5 and MP3.5 procedures provide the most attractive accuracy-to-computational cost ratios. The MP2.5 procedure results in a weighted-total-root-mean-square deviation (WTRMSD) of 3.4 kJ/mol, whilst the computationally more expensive MP3.5 procedure results in a WTRMSD of 1.9 kJ/mol (the same WTRMSD obtained for the CCSD(T) method in conjunction with a triple-zeta basis set). We also assess the performance of the computationally economical CCSD(T)/CBS(MP2) method, which provides the best overall performance for all the considered databases, including W4-11 and BH76.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

1.1. MP2-based procedures

The Møller–Plesset perturbation theory (MP n) energy can be written in terms of the Hartree–Fock energy (E^{HF}) and the correlation energy components:

$$E^{MPn} = E^{HF} + E^{(2)} + \dots + E^{(n)} \quad (1)$$

Where $E^{(n)}$ is the n th-order correction to the correlation energy. The second-order correction ($E^{(2)}$) can be further partitioned into:

$$E^{(2)} = E_{SS} + E_{OS} \quad (2)$$

where E_{SS} and E_{OS} are the same- and opposite-spin contributions to the MP2 correlation energy, respectively. A number of MP2-based methods have been developed over the past decade [1], in which the various correlation contributions are scaled by empirically or theoretically motivated scaling factors in order to improve the performance of the method. The first modified MP2-based procedure (SCS-MP2) was proposed by Grimme [2,3]. In this procedure the same- and opposite-spin components of the MP2 correlation energy are scaled separately by coefficients of $c_{SS} = 1/3$ and $c_{OS} = 6/5$, respectively. This significantly improves the performance of MP2

for thermochemistry [2,4]. For example, for the set of 51 reactions against which the SCS-MP2 method was initially benchmarked, this simple scaling procedure reduces the root-mean-square deviation (RMSD) by 50%, from 19.3 (MP2) to 9.6 kJ mol⁻¹ (SCS-MP2).

More recently, the SCS-MP2 procedure was benchmarked against an extensive set of reaction energies including isomerization, Diels–Alder, and ozonolyses reactions, which are subsets of the GMTKN30 database [5,6]. Against these benchmark sets, the SCS-MP2 method obtains a weighted total mean absolute deviation (WTMAD) of 7.5 kJ mol⁻¹, while MP2 results in a WTMAD which is roughly twice as large (namely, 15.1 kJ mol⁻¹) [1,5]. However, for the subsets of basic properties of the GMTKN30 database, which includes atomization energies, electron affinities, ionization potentials, proton affinities, and barrier heights [5,6], the WTMAD of SCS-MP2 (21.3) is only 11% smaller than that of MP2 (23.8 kJ mol⁻¹) [1]. We note in passing that Grimme also showed that the performance for both SCS-MP2 and MP2 can be significantly improved for heats of formation of large main group compounds by using an empirical higher-level-correction scheme [7]. Similarly, for the set of C₈H₈ isomerization energies, SCS-MP2 represents an appreciable improvement over MP2 with RMSDs of 7.9 and 13.4 kJ mol⁻¹, respectively [8]. Izgorodina et al. evaluated the performance of MP2-based methods for the prediction of a range of C–X (X = H, C, O, and F) bond dissociation energies (BDEs) [9]. They found that the SCS-MP2 method results in a mean absolute deviation (MAD) of 8.7 kJ mol⁻¹, which provides a significant

* Corresponding author.

E-mail address: amir.karton@uwa.edu.au (A. Karton).

improvement over MP2 for which an MAD of 19.6 kJ mol^{-1} is obtained. The SCS-MP2 procedure was also found to give excellent performance for the dispersion-driven isomerization of n-octane to hexamethylethane [10,11]. The SCS-MP2 isomerization energy (-5.6 kJ mol^{-1}) is in good agreement with the experimental value of -7.9 kJ mol^{-1} , while the MP2 method predicts an isomerization energy of $-18.1 \text{ kJ mol}^{-1}$ [10].

We have recently assessed the performance of the SCS-MP2 method for isomerization reactions involving migration of a double bond, which breaks the conjugated π -system [12,13]. For the set of 60 diene isomerization energies (the so-called DIE60 database), SCS-MP2 attains a stellar performance with an RMSD of merely 1.0 kJ mol^{-1} , while MP2 results in a larger RMSD of 2.7 kJ mol^{-1} [12]. Similar results were obtained for isomerization reactions in enones (i.e., the EIE22 database) [13]. In addition, for the relative energies of the 52 conformers of melatonin, SCS-MP2 yields superb performance with an RMSD of 0.9 kJ mol^{-1} , while MP2 results in an RMSD of 3.8 kJ mol^{-1} [14].

What about the performance of the SCS-MP2 method for reaction barrier heights? For the set of 26 pericyclic reactions in the BHPERI database [5,6], the SCS-MP2 method gives an RMSD of 5.9 kJ mol^{-1} . This represents a significant improvement over the MP2 method, which results in a whopping RMSD of 33.5 kJ mol^{-1} [15]. However, for other benchmark sets of reaction barrier heights, less dramatic improvements are obtained. For the set of 20 cycloreversion barrier heights in the CRBH20 database [16], the SCS-MP2 and MP2 methods result in RMSDs of 17.2 and 23.7 kJ mol^{-1} , respectively. For the set of 27 reactions in the water-catalyzed proton-transfer (WCPT) tautomerization database [17], SCS-MP2 and MP2 provide similar performance with MADs of 9.0 and 9.9 kJ mol^{-1} , respectively. It should be pointed out that for this database MP2 systematically underestimates the reaction barrier heights, while SCS-MP2 systematically overestimates them. Similarly, for the set of reaction barrier heights for multiple proton exchanges in water clusters in the BHPE13 database [18] SCS-MP2 and MP2 show similar performance with MADs of 15.5 and 14.6 kJ mol^{-1} , respectively. Where again, MP2 systematically underestimates the reaction barrier heights and SCS-MP2 systematically overestimates them.

Shortly after the development of the SCS-MP2 method, Head-Gordon et al. proposed the SOS-MP2 method [19]. SOS-MP2 completely neglects the same-spin contribution to the MP2 correlation energy (with $c_{ss} = 0$), while the opposite-spin contribution is scaled by $c_{os} = 1.3$. The SOS-MP2 method generally shows similar performance to the SCS-MP2 procedure for thermochemical properties [1,7,9,19] at an appreciable reduction of the computational cost (the SOS-MP2 method exhibits asymptotic CPU time scaling proportional to n^4 , while SCS-MP2 and MP2 exhibit asymptotic CPU time scaling proportional to n^5 , where n is the number of electrons in the system). In this context, the work of Fink on the S2-MP2 method must also be mentioned. The S2-MP2 method uses same- and opposite-spin scaling factors of $c_{ss} = 0.75$ and $c_{os} = 1.15$, respectively, which are motivated theoretically by Feenberg's scaling approach [20]. It has been shown that the S2-MP2 procedure gives poorer performance for atomization energies (with an MAD of 49.4 kJ mol^{-1}) than the MP2 procedure (MAD = 36.4 kJ mol^{-1}). However, for the reaction energies subset in the GMTKN30 database, S2-MP2 yields a WTMA of 10.9 kJ mol^{-1} which lies midway between SCS-MP2 (7.5) and MP2 (15.1 kJ mol^{-1}) [1,6].

1.2. MP3-based procedures

The first modified MP3-based procedure (SCS-MP3) was developed by Grimme shortly after the development of the SCS-MP2 procedure [21]. In this procedure a scaled third-order correction is added to the SCS-MP2 energy:

$$E^{\text{SCS-MP3}} = E^{\text{SCS-MP2}} + 0.25 \times E^{(3)} \quad (3)$$

Where the 3rd-order correction to the correlation energy ($E^{(3)}$) is scaled by a coefficient of 0.25. The SCS-MP3 procedure was initially assessed against a benchmark set of 32 isogyric reaction energies, 11 atomization energies, and 11 stretched geometries. In all cases, SCS-MP3 outperforms SCS-MP2, MP2, and MP3. More specifically, for the set of isogyric reaction energies the following MADs were obtained: 5.4 (SCS-MP3), 7.9 (SCS-MP2), 17.2 (MP2), and 20.5 kJ mol^{-1} (MP3). For the set of atomization energies the following MADs were obtained: 10.9 (SCS-MP3), 23.8 (SCS-MP2), 32.2 (MP3), and 33.1 kJ mol^{-1} (MP2). However, for a number of benchmark sets the SCS-MP3 and SCS-MP2 methods have been found to give similar performance. These include the following databases (RMSDs for SCS-MP3 and SCS-MP2 are given in parenthesis, respectively): the set of C_8H_8 isomerization energies (7.1 and 7.9) [8], the DIE60 database (0.9 and 1.0) [12], and the EIE22 database (1.0 and 0.8 kJ mol^{-1}) [13].

The SCS-MP3 method was also assessed for a number of datasets of reaction barrier heights. However, it shows similar or even poorer performance compared to SCS-MP2. The assessed benchmark sets include (RMSDs for SCS-MP3 and SCS-MP2 are given in parenthesis, respectively): the CRBH20 database (17.8 and 17.2) [16], the BHPERI database (10.0 and 5.9) [15], and the BHPE13 database (24.8 and 16.4 kJ mol^{-1}) [18].

Another MP3-based procedure is the so-called MP2.5 method – a simple arithmetic average of MP2 and MP3 – proposed by Hobza et al. for weak interactions [22]:

$$E^{\text{MP2.5}} = 0.5 \times (E^{\text{MP2}} + E^{\text{MP3}}) \quad (4)$$

MP2.5 has been initially tested for the 22 H-bonded, dispersion-controlled, and mixed non-covalent complexes in the S22 database [23]. It was shown that the performance of MP2.5 is superior to the spin-component scaled MP2-based methods, e.g. SCS-MP2 and SCS(MI)-MP2 [22,24]. The MP2.5 procedure results in an RMSD of merely 1.0 kJ mol^{-1} for the S22 database (cf. RMSDs of 5.1 for MP2, 3.4 for MP3, and 3.1 for SCS(MI)-MP2). The MP2.5 procedure also provides excellent performance for the database of barrier heights involving multiple proton exchanges in water clusters, with an overall RMSD of merely 2.1 kJ mol^{-1} (cf. RMSDs of 16.4 kJ mol^{-1} for SCS-MP2 and 24.8 kJ mol^{-1} for SCS-MP3) [18]. This outstanding performance is due to the fact that MP2 systematically underestimates the barrier heights, while MP3 systematically overestimates them by similar amounts. However, for other databases of barrier heights, e.g., the CRBH20 [16] and BHPERI databases [15], MP2.5 provides no significant improvement over SCS-MP2 and SCS-MP3.

1.3. MP4-based procedures

Two MP4-based procedures have been recently proposed, MP3.5 and MP4_{avg} [15,16,25]. The MP3.5 procedure is a simple arithmetic average of the MP3 and MP4 methods:

$$E^{\text{MP3.5}} = 0.5 \times (E^{\text{MP3}} + E^{\text{MP4}}) \quad (5)$$

The motivation for the MP3.5 method is similar to that for the MP2.5 procedure, namely, the oscillatory behavior that the MP_n series exhibits for a number of chemical properties. For example, it has been found that for some thermochemical properties (such as the isomerization energies in the DIE60 and EIE22 databases [12,13] and N–Br bond dissociation energies [26]) MP2 and MP4 tend to systematically overestimate the CCSD(T)/CBS reference values, whilst MP3 tends to systematically underestimate them. On the other hand, for some reaction barrier heights (e.g., in the BHPERI, WCPT27, and BHPE13 databases [15,17,18]) the reverse oscillatory trend is observed. Namely, the MP2 and MP4 methods

systematically underestimate the CCSD(T)/CBS reference values, whilst MP3 systematically overestimates them. The oscillatory behavior of the MP n series is normally characterized by smaller oscillations as n increases. We note that this oscillatory behavior has also been previously observed for weak interactions [1,22,27]. In the present work, we find that this oscillatory behavior is also true for other thermochemical and kinetic datasets, e.g. the atomization energies in the W4-11 dataset, the isomerization energies in the H $_x$ C $_y$ O $_z$ dataset, and the catalyzed reaction barrier heights in the IACBH8 dataset. Fig. 1 illustrates the oscillatory behavior for the MP n series ($n = 2, 3, \text{ and } 4$) for the isomerization energies in the H $_x$ C $_y$ O $_z$ and DIE60 databases, as well as for the reaction barrier heights in the BHPE13 and WCPT27 databases.

The MP $_{4\text{avg}}$ method is the average of the MP4(SDTQ) and MP4(SDQ) methods (where MP4(SDTQ) includes all single, double, triple, and quadruple excitations, while MP4(SDQ) excludes the triple excitations) [28,29]:

$$E^{\text{MP}_{4\text{avg}}} = 0.5 \times (E^{\text{MP}_4(\text{SDQ})} + E^{\text{MP}_4(\text{SDTQ})}) \quad (6)$$

The development of this method was motivated by the systematic tendency of MP4(SDTQ) to underestimate the reaction barrier heights in the CRBH20 database, and the tendency of MP4(SDQ) to systematically overestimate them by similar amounts [16]. Consequently, the MP $_{4\text{avg}}$ method results in a near-zero mean signed deviation (MSD) of -0.4 kJ mol^{-1} and an RMSD of 2.5 kJ mol^{-1} for this database. It was later found that MP $_{4\text{avg}}$ also gives relatively good performance for N–Br bond dissociation energies due to the tendency of MP4(SDTQ) and MP4(SDQ) to err on different sides of the CCSD(T)/CBS reference values [26].

The MP3.5 method has been found to give excellent performance for the barrier heights in the BHPERI database with an RMSD of merely 2.5 kJ mol^{-1} . However, for the barrier heights in the CRBH20 database, it results in a rather disappointing RMSD of 11.3 kJ mol^{-1} . The MP $_{4\text{avg}}$ procedure, on the other hand, has been found to give superb performance for the barrier heights in both of the BHPERI and CRBH20 databases with RMSDs of 2.5 and 2.1 kJ mol^{-1} , respectively [15,16].

Despite the above-mentioned evidence for the promising performance of these MP4-based procedures, their performance has not been extensively benchmarked. One of the main goals of the present work is to benchmark the performance of the MP4-based procedures for a wide range of thermochemical and kinetic properties, and to compare their performance to that of the lower-cost MP2- and MP3-based procedures.

2. Databases

We evaluate the performance of the modified MP n -based procedures ($n = 2, 3, \text{ and } 4$) for a wide range of thermochemical and kinetic properties for which CCSD(T)/CBS reference values are available (coupled cluster with single, double, and quasiperturbative triple excitations extrapolated to the complete basis set limit). For this purpose we use a suite of 11 databases including atomization energies, isomerization energies, conformational energies, and reaction barrier heights [8,12,13,17,18,30,31,32–36]. Table 1 gives an overview of the benchmark databases employed in this work. All the reference energies are obtained at the CCSD(T)/CBS level of theory with only valence electrons correlated (i.e., where needed zero-point vibrational energies, core-valence, and relativistic contributions are excluded from the reference values).

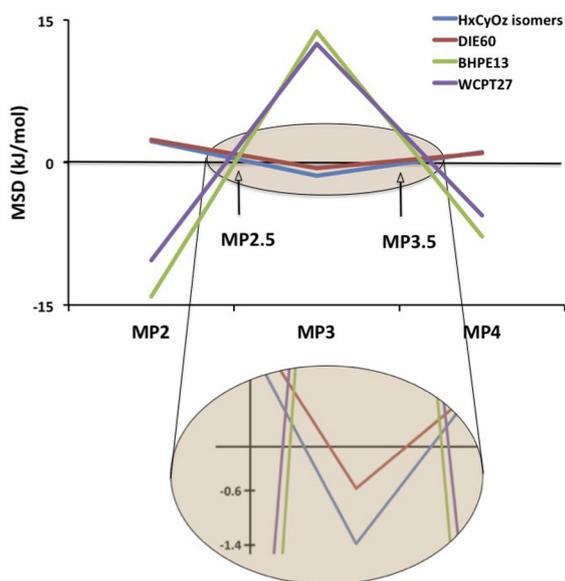


Fig. 1. Mean-signed deviations (MSDs, in kJ mol^{-1}) for the MP n methods for the isomerization energies in the H $_x$ C $_y$ O $_z$ and DIE60 datasets and the reaction barrier heights in the BHPE13 and WCPT27 datasets. Table S1 of the Supplementary data summarizes the MSDs for these (and other) databases.

Table 1

Overview of the benchmark databases used in the present work. All the reference values are obtained at the CCSD(T)/CBS level of theory by means of W_n or W_n -F12 theories ($n = 1, 2$)^a.

Database	Description	Ref. method	Ref.
W4-11	122 atomization energies of small molecules	W4 or higher	[30]
C ₈ H ₈ isomers	43 C ₈ H ₈ structural isomers	W1-F12	[8]
H $_x$ C $_y$ O $_z$ isomers	109 isomerization energies of H $_x$ C $_y$ O $_z$ isomers	W1-F12	[36]
DIE60	60 diene isomerization energies	W2-F12/W1-F12	[12]
EIE22	22 enecarbonyl isomerization energies	W1-F12	[13]
C $_n$ H $_{2n+2}$ conformers	18 conformational energies of n -butane, n -pentane, and n -hexane	W1 h	[32]
Tetrapeptide conformers	10 conformational energies of two tetrapeptides	Est. CCSD(T)/CBS	[33]
IACBH8	8 barrier heights of vinyl alcohol \rightarrow acetaldehyde reactions catalyzed by inorganic acids	W1-F12	[34]
BHPE13	13 barrier heights for proton exchange in small water, ammonia, and hydrogen fluoride clusters	W1-F12	[18]
WCPT27	27 barrier heights for a set of water-catalyzed proton transfer reactions	W2.2/W1	[17]
BH76	76 barrier heights of hydrogen transfer, heavy atom transfer, nucleophilic substitution, unimolecular and association reactions	W1	[35]

^a With the exception of the W4-11 database for which the reference values are obtained at the FCI/CBS level of theory from W4 (or higher) theory [30], and the tetrapeptide conformers database for which the reference values are taken as CCSD(T)/CBS \approx CCSD(T)/cc-pVDZ + MP2/CBS – MP2/cc-pVDZ [33].

2.1. Atomization reactions

2.1.1. The W4-11 dataset

The W4-11 database contains 140 atomization energies (with up to five heavy atoms) [30]. The species in the W4-11 database cover a broad spectrum of bonding situations and multireference character, and as such it is an ideal benchmark set for the validation of *ab initio* methods. We removed the three beryllium-containing compounds and the 15 pathologically multireference systems from the list [31]. All reference data in the W4-11 database are obtained at the FCI/CBS level of theory via the W4 (or higher) *ab initio* computational thermochemistry protocol [37,38].

2.2. Isomerization reactions

2.2.1. C₈H₈ structural isomers

This database includes 43 structural isomerization energies of C₈H₈ hydrocarbons. The C₈H₈ isomers involve a range of hydrocarbon functional groups, including linear and cyclic polyacetylene, polyene, and cumulene moieties, as well as aromatic, anti-aromatic, and strained rings [8]. Since many of these structural isomerization reactions involve breaking and forming of multiple bonds, this dataset proves to be a challenging test for many density functional theory (DFT) and standard *ab initio* methods. Reference isomerization energies are obtained by means of the W1-F12 thermochemical procedure [39].

2.2.2. H_xC_yO_z structural isomers

This database includes all structural isomers of a wide range of compounds with the molecular formula H_xC_yO_z, namely: H₂C₂O, H₄C₂O, H₆C₂O, H₂C₂O₂, H₄C₂O₂, H₆C₂O₂, H₂C₃O, H₄C₃O, H₆C₃O, H₈C₃O, H₂C₃O₂, H₄C₃O₂, H₆C₃O₂, and H₈C₃O₂ [36]. Overall, this database includes 109 accurate structural isomerization energies, which are calculated at the CCSD(T)/CBS level by means of the W2-F12 thermochemical protocol [39].

2.2.3. The DIE60 dataset

This database includes 60 isomerization reactions which involve a migration of one double bond that breaks the conjugated π -system [12]. The reactions in the DIE60 dataset are of the type [conjugated diene] \rightarrow [non-conjugated diene]. The dataset covers a broad spectrum of structures, including linear and branched (C_n-H_{2n-2}), and cyclic (C_nH_{2n-4}) dienes ($n = 5, 6, \text{ and } 7$). Reference isomerization energies are obtained by means of the high-level W2-F12 procedure for the C₅H₆, C₅H₈, C₆H₈, C₆H₁₀, and C₇H₁₀ dienes, and the W1-F12 procedure for the C₇H₁₂ dienes [39].

2.2.4. The EIE22 dataset

The EIE22 database is comprised of 22 prototypical isomerization reactions of the type [conjugated α, β -enecarbonyl] \rightarrow [non-conjugated β, γ -enecarbonyl] [16]. All the reactions from the EIE22 dataset involve a migration of one double bond, which breaks the conjugated π -system. This dataset can be viewed as an extension of the DIE60 dataset, which only contains hydrocarbons. The considered enecarbonyls involve a range of common functional groups (e.g. CH₃, NH₂, OCH₃, F, and CN). Accurate isomerization energies are obtained at the complete basis-set limit CCSD(T) level by means of the high-level W1-F12 thermochemical protocol [39].

2.3. Conformational energies

2.3.1. The C_nH_{2n+2} conformers dataset

This database of conformational C_nH_{2n+2} alkane isomers ($n = 4-6$) contains 18 conformational energies of *n*-butane, *n*-pentane, and

n-hexane [32]. Reference conformational energies are obtained by the W1h thermochemical procedure [40].

2.3.2. The tetrapeptide conformers dataset

This database includes 10 conformers of the ACE-ALA-GLY-ALA-NME and ACE-ALA-SER-ALA-NME tetrapeptide. ALA is alanine, GLY is glycine, and SER is serine. ACE and NME stand for acetyl and methylamide groups [41]. For each peptide, five conformers have been examined, i.e. parallel (β) and anti-parallel (β_a) β -sheets, right-handed (α_R) and left-handed (α_L) α -helices, and polyproline-II (PP-II) helix. Reference conformational energies are obtained via the CCSD(T)/CBS \approx CCSD(T)/cc-pVDZ + MP2/CBS - MP2/cc-pVDZ method, where the MP2/CBS energy is extrapolated to the complete basis-set limit from the cc-pVTZ and cc-pVQZ basis sets [33].

2.4. Reaction barrier heights

2.4.1. The IACBH8 dataset

This dataset contains reaction barrier heights for the vinyl alcohol \rightarrow acetaldehyde tautomerization reaction catalyzed by inorganic acids (HNO₃, H₂CO₃, H₃PO₄, HClO₄, and H₂SO₄), formic acid, and water [36]. The reference reaction barrier heights are obtained by means of the W1-F12 thermochemical protocol [39].

2.4.2. The BHPE13 dataset

This database contains 13 barrier heights for proton exchange in small (H₂O)_n clusters ($n = 2-6$), (NH₃)_n clusters ($n = 2-4$), and (HF)_n clusters ($n = 2-6$) [18]. The reference reaction barrier heights are obtained by means of the W1-F12 thermochemical protocol [39].

2.4.3. The WCPT27 dataset

This database contains 27 accurate reaction barrier heights for a set of nine proton-transfer tautomerization reactions that are either uncatalyzed, catalyzed by one water molecule, or catalyzed by two water molecules [17]. The reaction barrier heights are obtained by means of the high-level W2.2 [40] or W1 [42] thermochemical procedures.

2.4.4. The BH76 dataset

This database contains two subsets of reaction barrier heights: HTBH38 and NHTBH38. The HTBH38 set consists of 38 reaction barrier heights of hydrogen transfer reactions, whilst the NHTBH38 set consists of 12 heavy-atom transfer reactions (HATBH12), 16 nucleophilic substitution reactions (NSBH16), 10 unimolecular and association reactions (UABH10) [35]. The reference reaction barrier heights are obtained by means of the W1 thermochemical protocol [40].

3. Computational details

All the calculations were carried out using the Molpro 2012.1 program suite [43,44], except for the W4-11 and BH76 databases where all the calculations were performed with the Gaussian 09 program suite [45]. We consider Dunning's A'VnZ basis sets ($n = D, T, Q, \text{ and } 5$) for evaluating the basis set effects. The notation A'VnZ indicates the combination of the standard correlation-consistent cc-pVnZ basis sets on H [46], the aug-cc-pVnZ basis sets on first-row atoms [47], and the aug-cc-pV($n + d$)Z basis sets on second-row atoms [48]. For the DIE60, C_nH_{2n+2} conformers, and tetrapeptide conformers datasets, we only consider the cc-pVnZ basis sets ($n = D, T, \text{ and } Q$). The reference geometries and reference reaction energies and barrier heights for all the datasets are taken from the literature (see Table 1). The *ab initio* procedures consid-

ered in the present study are MP2, SCS-MP2 [2], MP2.5 [22], MP3, MP3.5 [15], SCS-MP3 [21], MP4(SDQ), MP4, MP4_{avg} [16], CCSD, SCS-CCSD [49], and CCSD(T). Unless otherwise mentioned, the performance of the MP2- and MP3-based methods is evaluated in conjunction with the A'VnZ or cc-pVnZ ($n = T$ and Q) basis sets, whereas the performance of the computationally more expensive MP4-based and coupled cluster procedures is evaluated in conjunction with the A'VTZ or cc-pVTZ basis set.

We also evaluate the performance of the cost-effective CCSD(T)/CBS(MP2) method. In this approach the CCSD(T)/CBS energy is estimated from the CCSD(T)/A'VDZ energy and an MP2-based basis-set-correction term ($\Delta MP2 = MP2/A'V\{T,Q\}Z - MP2/A'VDZ$, where the MP2/A'V{T,Q}Z energy is extrapolated to the basis-set limit with an extrapolation exponent of 3) [50]. This additivity scheme has been shown to give good performance for noncovalent interactions [33,51–53], reaction energies [12,54], and reaction barrier heights [15,16]. Here we will assess the performance of this cost-effective method for a more extensive set of thermochemical and kinetic data. As a point of reference for the performance of the MPn-based and CCSD(T)/CBS(MP2) methods we also include error statistics for the CCSD(T)/A'VnZ level of theory. The performance of all of these methods is evaluated against CCSD(T)/CBS reference values obtained from Wn theories (see Section 2 for further details).

4. Results and discussion

4.1. Performance of standard and modified *ab initio* procedures for thermochemistry

4.1.1. Performance for atomization energies (W4-11 database)

Table 2 gives an overview of the performance of the standard and modified *ab initio* methods for the atomization energies in the W4-11 dataset. As expected, the W4-11 database is an extremely challenging test for all the considered MPn-based methods, with RMSDs ranging between 13.3 (MP2.5) and 52.1 kJ mol⁻¹ (SCS-MP2) in conjunction with the A'V5Z basis set. In the following discussion we will use the results obtained with the A'V5Z basis set for the MPn-based methods, and the A'VQZ basis set for the coupled cluster methods.

Table 2
Statistical analysis for the performance of *ab initio* methods for the atomization energies in the W4-11 dataset (in kJ mol⁻¹).^a

Basis sets	Methods	RMSD	MAD	MSD	
A'V5Z	MP2	50.6	39.2	33.1	
	SCS-MP2	52.1	43.2	42.9	
	MP2.5	13.3	10.6	-2.8	
	MP3	46.5	38.8	-38.7	
	SCS-MP3	30.7	26.1	24.9	
	MP4(SDQ)	41.6	36.0	-36.0	
	MP4	16.7	12.5	11.0	
	MP3.5	16.5	14.0	-13.8	
	MP4 _{avg}	14.6	12.5	-12.5	
	A'VQZ	MP2	44.6	34.2	25.4
		SCS-MP2	43.1	34.8	33.9
		MP2.5	16.2	13.2	-9.7
MP3		52.7	44.8	-44.8	
SCS-MP3		22.4	18.5	16.3	
MP4(SDQ)		48.0	41.9	-41.9	
MP4		11.5	8.8	4.0	
MP3.5		23.3	20.4	-20.4	
MP4 _{avg}		21.5	19.0	-19.0	
CCSD		55.2	48.2	-48.2	
CCSD(T)		15.1	13.5	-13.5	
CBS(MP2) ^b		CCSD(T)	3.8	3.2	-0.3

^a RMSD = root mean square deviation, MAD = mean absolute deviation, MSD = mean signed deviation.

^b CCSD(T)/CBS(MP2) \approx CCSD(T)/A'VDZ + MP2/A'V{T,Q}Z - MP2/A'VDZ.

The MP2-based methods (MP2 and SCS-MP2) result in very large RMSDs of over 50 kJ mol⁻¹ in conjunction with the A'V5Z basis set, where in both cases the atomization energies are systematically overestimated. The MP3 method obtains a similarly high RMSD of 46.5 kJ mol⁻¹, but systematically underestimates the atomization energies. In this situation, MP2.5, which is an average of MP2 and MP3, provides significantly better performance with an RMSD of 13.3 kJ mol⁻¹ for the W4-11 dataset. The SCS-MP3 method also represents a significant improvement over both MP2 and MP3, but still attains a very large RMSD of 30.7 kJ mol⁻¹. The MP4 method results in an RMSD of 16.7 kJ mol⁻¹, and similarly to the MP2 method it tends to systematically overestimate the atomization energies, while MP4(SDQ) gives poor performance with an RMSD of 41.6 kJ mol⁻¹. The modified MP4-based procedures (MP3.5 and MP4_{avg}) show slightly improved performance, but tend to systematically underestimate the atomization energies.

The CCSD method obtains an RMSD of 55.2 kJ mol⁻¹ in conjunction with the A'VQZ basis set and, as expected, the RMSD of the CCSD(T) method is much lower (i.e., it is 15.1 kJ mol⁻¹). However, the cost-effective CCSD(T)/CBS(MP2) approach achieves a much lower RMSD of only 3.8 kJ mol⁻¹ and a near-zero MSD of 0.3 kJ mol⁻¹. This procedure significantly outperforms all the examined MPn-based procedures.

4.1.2. Performance for structural isomerization (C₈H₈ and H_xC_yO_z isomers databases)

Structural isomerization can be a challenging test for many DFT and *ab initio* methods since they often involve breaking and forming of several bonds. In this section we consider two structural isomerization databases, consisting of C₈H₈ and H_xC_yO_z isomers. Table 3 gives an overview of the performance of the MPn-based methods for the C₈H₈ isomers. We note that for the methods for which we have both A'VTZ and A'VQZ results (i.e. MP2, SCS-MP2, MP2.5, MP3, and SCS-MP3) the differences in the overall RMSDs (in absolute value) between these basis sets are smaller than 1.0 kJ mol⁻¹. In the following discussion, we will focus on the performance of the MP2- and MP3-based procedures in conjunction with the A'VQZ basis set, whilst the performance of the rest of the methods is evaluated with the A'VTZ basis set.

None of the MP2- and MP3-based procedures results in RMSDs below the threshold of chemical accuracy. The MP2-based methods exhibit relatively poor performance with RMSDs of 13.2 (MP2) and 8.1 kJ mol⁻¹ (SCS-MP2). The MP3-based procedures give

Table 3
Statistical analysis for the performance of *ab initio* methods for calculations of isomerization energies in the C₈H₈ isomers dataset (in kJ mol⁻¹).^a

Basis sets	Methods	RMSD	MAD	MSD	
A'VQZ	MP2	13.2	11.0	4.8	
	SCS-MP2	8.1	6.2	-3.2	
	MP2.5	8.1	6.7	5.3	
	MP3	8.9	6.0	5.8	
	SCS-MP3	6.9	4.4	-2.9	
	A'VTZ	MP2	12.7	10.6	4.2
SCS-MP2		8.7	6.5	-4.0	
MP2.5		7.2	5.9	4.4	
MP3		7.9	4.9	4.7	
SCS-MP3		7.6	4.7	-3.8	
MP4(SDQ)		9.1	7.8	-6.0	
MP4		6.0	4.4	-1.7	
MP3.5		3.0	2.3	1.5	
MP4 _{avg}		5.5	4.2	-3.8	
CCSD		6.4	4.8	-2.3	
SCS-CCSD		12.5	11.4	-11.4	
CCSD(T)		2.5	2.1	-1.5	
CBS(MP2)		CCSD(T)	2.5	2.2	-2.1

^a Footnotes a and b to Table 2 apply here.

slightly lower RMSDs, which range between 6.9 (SCS-MP3) and 8.9 kJ mol^{-1} (MP3). It is worth mentioning that the MP2.5 procedure does not improve the performance and results in practically the same RMSD as the SCS-MP2 procedure. The MP4(SDQ) method provides poor performance for the C_8H_8 isomers database with an RMSD of 9.1 kJ mol^{-1} , while the MP4 procedure shows better performance with an RMSD of 6.0 kJ mol^{-1} . The MP4_{avg} method gives slightly better performance with an RMSD of 5.5 kJ mol^{-1} . However, MP3.5 shows significantly better performance, attaining an RMSD of 3.0 kJ mol^{-1} , which is comparable to the performance of the CCSD(T) method, *vide infra*.

The CCSD and SCS-CCSD methods result in RMSDs of 6.4 and 12.5 kJ mol^{-1} , respectively. Thus, both methods show poor performance relative to the computationally more economical MP3.5 method. The CCSD(T) method shows the best performance with an RMSD of merely 2.5 kJ mol^{-1} . We note that the cost-effective approach CCSD(T)/CBS(MP2) provides very similar performance to the CCSD(T)/A'VTZ level of theory at a reduced computational cost.

We now turn our attention to the performance of *ab initio* procedures for the $\text{H}_x\text{C}_y\text{O}_z$ isomers database. Table 4 gives an overview of the performance of the MP n -based and coupled cluster methods. Similar to the results for the C_8H_8 database, the RMSDs for the A'VTZ basis set are within 1.0 kJ mol^{-1} from those for the A'VQZ basis set. Since we do not have the results for the coupled cluster methods with the A'VQZ basis set, we will concentrate on the performance of all the methods in conjunction with the A'VTZ basis set in the following text.

As shown in Table 4, in contrast to the results for the C_8H_8 database, all the examined methods perform reasonably well, with RMSDs ranging between 1.1 (CCSD(T)/CBS(MP2)) and 6.6 kJ mol^{-1} (MP2). The MP2 method results in an RMSD of 6.6 kJ mol^{-1} . This RMSD is decreased by 42% (to 3.8 kJ mol^{-1}) when the same- and opposite-spin components of the MP2 correlation energy are scaled in the SCS-MP2 procedure. The MP3-based methods (MP3, SCS-MP3, and MP2.5) show similar performance, with RMSDs between 3.8 and 4.0 kJ mol^{-1} .

Moving to the MP4-based methods, MP4(SDQ) and MP4_{avg} show similar performance to the MP3-based methods with RMSDs of 3.5 and 3.6 kJ mol^{-1} , respectively. In contrast, MP4 attains a

Table 4
Statistical analysis for the performance of *ab initio* methods for the isomerization energies in the $\text{H}_x\text{C}_y\text{O}_z$ dataset (in kJ mol^{-1}).^a

Basis sets	Methods	RMSD	MAD	MSD	
A'VQZ	MP2	7.6	5.9	2.7	
	SCS-MP2	3.7	2.9	1.0	
	MP2.5	5.3	3.9	1.0	
	MP3	4.9	4.1	-0.7	
	SCS-MP3	2.9	2.3	0.2	
	MP4(SDQ)	3.9	3.1	1.8	
	MP4	3.9	3.2	1.8	
	MP3.5	2.0	1.5	0.6	
	MP4_{avg}	3.3	2.6	1.8	
	A'VTZ	MP2	6.6	5.3	2.2
		SCS-MP2	3.8	3.0	0.4
		MP2.5	4.0	3.0	0.4
		MP3	3.8	3.1	-1.4
SCS-MP3		3.8	3.0	0.4	
MP4(SDQ)		3.5	2.8	1.0	
MP4		4.7	3.8	1.1	
MP3.5		1.6	1.3	-0.2	
MP4_{avg}		3.6	2.8	1.0	
CCSD		2.3	1.8	-0.1	
SCS-CCSD		5.0	3.8	-3.5	
CCSD(T)		2.4	1.7	-0.9	
CBS(MP2)		CCSD(T)	1.1	0.9	-0.9

^a Footnotes a and b to Table 2 apply here.

somewhat larger RMSD of 4.7 kJ mol^{-1} . However, the MP3.5 procedure again results in an exceptionally low RMSD of 1.5 kJ mol^{-1} . The excellent performance of MP3.5 is also demonstrated by a near-zero MSD of -0.2 kJ mol^{-1} . Indeed this method significantly outperforms the computationally more expensive coupled cluster procedures CCSD, SCS-CCSD, and CCSD(T)/A'VTZ as well as the all the considered MP n -based methods (Table 4). Finally we note that the CCSD(T)/CBS(MP2) method provides superb performance with an RMSD of merely 1.1 kJ mol^{-1} .

4.1.3. Performance for isomerization reactions involving double-bond migration (DIE60 and EIE22 databases)

In this subsection we consider the DIE60 and EIE22 databases of isomerization reactions that involve a migration of one double bond that breaks the π -conjugated system. With the exception of the newly proposed MP4-based methods (MP3.5 and MP4_{avg}), all the examined methods have been previously assessed for these two databases [12,13]. However, for the sake of making the article self-contained, we will describe the performance of all the examined methods.

Table 5 gives an overview of the performance of the *ab initio* methods for the DIE60 dataset. We start by noting that practically all the *ab initio* procedures converge relatively smoothly and rapidly to the basis set limit. For example, for the methods for which we have both cc-pVQZ and cc-pVTZ results (MP2, SCS-MP2, MP2.5, MP3, and SCS-MP3), the differences in the overall RMSDs (in absolute value) between these two basis sets are equal to or smaller than 0.1 kJ mol^{-1} , whilst the differences in the overall RMSDs between the cc-pVDZ and cc-pVQZ basis sets are equal to (or smaller than) 0.7 kJ mol^{-1} (see Table S5, Supplementary data). Thus, for these reactions even the cc-pVDZ basis set gives results that are close to the basis set limit. Since we do not have the results for the coupled cluster methods with the cc-pVQZ basis set, we will concentrate on the results with the cc-pVTZ basis set in the following text.

The MP2 procedure systematically overestimates the isomerization energies and results in an RMSD of 2.9 kJ mol^{-1} . This RMSD is reduced to 1.1 kJ mol^{-1} when the same- and opposite-spin components of the MP2 correlation energy are scaled in the SCS-MP2 procedure. Inclusion of higher-order correlation corrections in procedures such as MP2.5, MP3, SCS-MP3, MP4(SDQ), and MP4 results in similar performance, with RMSDs of 1.0–1.8 kJ mol^{-1} . It is note-

Table 5
Statistical analysis for the performance of *ab initio* methods for the isomerization energies in the DIE60 dataset (in kJ mol^{-1}).^a

Basis sets	Methods	RMSD	MAD	MSD
cc-pVQZ	MP2	2.8	2.4	2.4
	SCS-MP2	1.0	0.6	0.4
	MP2.5	1.1	0.9	0.8
	MP3	1.1	1.1	-0.7
	SCS-MP3	0.9	0.7	-0.4
	MP4	1.3	1.0	1.0
cc-pVTZ	MP2	2.9	2.4	2.4
	SCS-MP2	1.1	0.8	0.5
	MP2.5	1.1	0.9	0.9
	MP3	1.1	0.9	-0.6
	SCS-MP3	1.0	0.7	-0.3
	MP4(SDQ)	1.8	1.5	-1.3
	MP4	1.3	1.0	1.0
	MP3.5	0.5	0.4	0.2
	MP4_{avg}	0.8	0.6	-0.1
	CCSD	1.9	1.6	-1.5
	SCS-CCSD	2.2	2.0	-2.0
	CCSD(T)	0.6	0.5	-0.2
CBS(MP2) ^b	CCSD(T)	0.4	0.3	-0.3

^a Footnote a to Table 2 applies here.

^b $\text{CCSD(T)/CBS(MP2)} \approx \text{CCSD(T)/cc-pVDZ} + \text{MP2/cc-pV[T,Q]Z} - \text{MP2/cc-pVDZ}$.

worthy that the two newly proposed MP3.5 and MP4_{avg} methods also show very good performance with RMSDs of 0.5 and 0.8 kJ mol⁻¹, respectively. The CCSD method attains a somewhat disappointing RMSD of 1.9 kJ mol⁻¹ considering its computational cost, and its spin-component scaled variant SCS-CCSD gives an even larger RMSD (2.2 kJ mol⁻¹). The CCSD(T) method attains an RMSD of merely 0.6 kJ mol⁻¹, while the CCSD(T)/CBS(MP2) method performs slightly better with an RMSD of 0.4 kJ mol⁻¹.

We now turn our attention to the EIE22 database. In contrast to the DIE60 dataset which only involves hydrocarbons, the performance of the *ab initio* methods for the EIE22 dataset is evaluated in conjunction with the A'VTZ basis sets since this database involves polar bonds. Table 6 lists the RMSDs, MADs, and MSDs for the EIE22 dataset for the examined methods. Practically all the tested methods converge fairly rapidly to the basis-set limit. For instance, for the methods for which we have both A'VTZ and A'VQZ results (MP2, SCS-MP2, MP2.5, MP3, and SCS-MP3), the differences in the RMSDs (in absolute value) between these two basis sets are smaller than or equal to 0.1 kJ mol⁻¹. Here we focus on the performance of all the methods with the A'VTZ basis set.

The MP2 method attains an RMSD of 2.1 kJ mol⁻¹, while the RMSD is reduced to only 0.9 kJ mol⁻¹ using the SCS-MP2 method. The MP3- and MP4-based procedures result in RMSDs ranging from 0.5 (MP3.5 and MP4_{avg}) to 1.9 kJ mol⁻¹ (MP4(SDQ)). In particular, the MP2.5 procedure provides superb performance and obtains an RMSD of 0.7 kJ mol⁻¹. The modified MP4-based procedures (MP3.5 and MP4_{avg}) give slightly better performance at a much higher computational cost. The CCSD method performs similarly to MP2, with an RMSD of 2.1 kJ mol⁻¹. The SCS-CCSD method systematically underestimates the isomerization energies in the EIE22 database (as evident from MSD ≈ -1 × MAD = -2.5 kJ mol⁻¹) and results in an RMSD of 2.7 kJ mol⁻¹. The CCSD(T) method gives the best performance with an RMSD of merely 0.4 kJ mol⁻¹, while the CCSD(T)/CBS(MP2) approach attains a slightly higher RMSD of 0.6 kJ mol⁻¹.

4.1.4. Performance for conformational energies in C_nH_{2n+2} and tetrapeptide systems

From the electronic structure point of view, conformational energies are generally easier to calculate than the isomerization energies discussed in Sections 4.1.2 and 4.1.3 since they do not involve breaking and forming of covalent bonds. In this section we will consider two datasets of conformers, namely, dispersion-

Table 6
Statistical analysis for the performance of *ab initio* methods for calculations of isomerization energies in the EIE22 dataset (in kJ mol⁻¹).^a

Basis sets	Methods	RMSD	MAD	MSD	
A'VQZ	MP2	2.1	1.9	1.8	
	SCS-MP2	0.8	0.7	0.2	
	MP2.5	0.7	0.6	0.4	
	MP3	1.5	1.2	-1.1	
	SCS-MP3	1.2	0.9	-0.5	
	A'VTZ	MP2	2.1	1.9	1.9
		SCS-MP2	0.9	0.8	0.3
		MP2.5	0.7	0.6	0.4
		MP3	1.5	1.3	-1.0
		SCS-MP3	1.3	1.0	-0.4
MP4(SDQ)		1.9	1.6	-1.5	
MP4		1.4	1.3	1.3	
MP3.5		0.5	0.4	0.1	
MP4 _{avg}		0.5	0.4	-0.1	
CCSD		2.1	1.8	-1.7	
SCS-CCSD	2.7	2.5	-2.4		
CCSD(T)	0.4	0.3	-0.1		
CBS(MP2)	CCSD(T)	0.6	0.5	-0.5	

^a Footnotes a and b to Table 2 apply here.

driven conformational energies in C_nH_{2n+2} (n = 4–6) hydrocarbons [32] and conformational energies in tetrapeptides [33].

Table 7 gives an overview of the performance of the standard and modified *ab initio* procedures for the C_nH_{2n+2} conformational energies. The performance of the MP2-based methods (MP2 and SCS-MP2) is assessed in conjunction with the cc-pVTZ and cc-pVQZ basis sets, whilst the performance of the rest of the methods is evaluated in conjunction with the cc-pVTZ basis set. In the following discussion we will use the results obtained with the cc-pVTZ basis set for all the examined methods.

The MP2 and SCS-MP2 methods result in similar RMSDs of 0.62 and 0.68 kJ mol⁻¹ respectively, while the MP3 and SCS-MP3 methods show slightly poorer performance with RMSDs of 0.75 and 0.99 kJ mol⁻¹, respectively. Since MP2 tends to underestimate the conformational energies and MP3 tends to overestimate them by similar amounts, the MP2.5 method shows excellent performance with an RMSD of merely 0.19 kJ mol⁻¹. Similarly, the MP3.5 method shows significantly better performance (RMSD = 0.15 kJ mol⁻¹) than the MP4 method (RMSD = 0.49 kJ mol⁻¹). Again, this is due to the fact that MP3 tends to overestimate the conformational energies, while MP4 tends to underestimate them. The MP4_{avg} method provides similar performance to MP3.5 and gives a near-zero RMSD of 0.09 kJ mol⁻¹. The superb performance of the MP3.5 and MP4_{avg} procedures is also demonstrated by the near-zero MSDs, suggesting that they are free of systematic bias.

The CCSD method represents poor performance relative to the MP4-based procedures, attaining an RMSD of 0.69 kJ mol⁻¹. The SCS-CCSD procedure results in a similar RMSD. Finally, we note that the CCSD(T)/CBS(MP2) method provides similar performance to the MP3.5 and MP4_{avg} procedures with an RMSD of 0.14 kJ mol⁻¹.

We now turn our attention to the performance of the *ab initio* methods for the tetrapeptide conformers (Table 8). Due to the large

Table 7
Statistical analysis for the performance of *ab initio* methods for conformational energies in the C_nH_{2n+2} hydrocarbons (in kJ mol⁻¹) (n = 4–6).^a

Basis sets	Methods	RMSD	MAD	MSD
cc-pVQZ	MP2	0.50	0.39	-0.39
	SCS-MP2	0.83	0.72	0.72
cc-pVTZ	MP2	0.62	0.53	-0.53
	SCS-MP2	0.68	0.57	0.57
	MP2.5	0.19	0.14	0.06
	MP3	0.75	0.66	0.66
	SCS-MP3	0.99	0.86	0.86
	MP4(SDQ)	0.49	0.41	0.41
	MP4	0.49	0.46	-0.46
	MP3.5	0.15	0.11	0.10
	MP4 _{avg}	0.09	0.07	-0.02
	CCSD	0.69	0.59	0.59
CBS(MP2)	SCS-CCSD	0.66	0.61	-0.61
	CCSD(T)	0.24	0.23	-0.23
	CCSD(T)	0.14	0.12	-0.12

^a Footnote a to Table 2 and footnote b to Table 5 apply here.

Table 8
Statistical analysis for the performance of *ab initio* methods for calculations of conformational energies in the tetrapeptides (in kJ mol⁻¹).^a

Basis sets	Methods	RMSD	MAD	MSD
cc-pVQZ	MP2 ^b	2.9	2.4	-1.4
	cc-pVTZ	MP2 ^b	6.1	4.5
cc-pVTZ	SCS-MP2	2.2	1.6	-1.6
	MP2.5	2.7	2.1	-1.8
	MP3	1.1	1.0	0.7
	SCS-MP3	1.0	0.7	-0.4

^a Footnote a to Table 2 and footnote b to Table 5 apply here.

^b The results are taken from Ref. [33].

sizes of the tetrapeptides, we were only able to perform the MP2- and MP3-based calculations in conjunction with the cc-pVTZ basis set. Table S8 of the Supplementary data gives the results for all the methods in conjunction with the cc-pVDZ basis set. However, as pointed out in reference [33], this basis set is too small for adequately describing the H-bonds in these systems, and these results will not be discussed here. Suffice to say that all the considered methods perform poorly in conjunction with the cc-pVDZ basis set with RMSDs ranging between 8.9 (MP3) and 14.7 kJ mol⁻¹ (MP4).

Except for the MP2 method, which exhibits a particularly strong basis set dependence [33], all the methods perform reasonably well in conjunction with the cc-pVTZ basis set, with RMSDs well below the threshold of chemical accuracy. In particular, SCS-MP2 results in an RMSD of 2.2 kJ mol⁻¹, whilst MP2.5 provides slightly deteriorated performance with an RMSD of 2.7 kJ mol⁻¹. MP3 and SCS-MP3 provide excellent performance with RMSDs of 1.1 and 1.0 kJ mol⁻¹, respectively, where MP3 tends to overestimate the conformational energies and SCS-MP3 tends to underestimate them.

4.2. Performance of standard and modified *ab initio* procedures for reaction barrier heights

Now we turn our attention to the performance of standard and modified *ab initio* methods for reaction barrier heights. We have previously found that the MP4_{avg} and CCSD(T)/CBS(MP2) methods give very good performance for the CRBH20 dataset [16]. In particular, in conjunction with the A'VTZ basis set these methods attain RMSDs of 2.1 and 1.7 kJ mol⁻¹, respectively, compared to RMSDs of 11.3 (MP3.5), 6.1 (MP4), and 5.0 kJ mol⁻¹ (CCSD(T)) in conjunction with the same basis set [16]. We note however, that for the reaction barrier heights in the BHPERI database, MP3.5 was found to significantly outperform all the other MP n -based procedures with an overall RMSD of 2.5 kJ mol⁻¹ [15]. It is therefore of interest to evaluate the performance of the modified MP4-based procedures for a wider range of reaction barrier heights.

4.2.1. Performance for barrier heights of reactions catalyzed by inorganic acids (IACBH8 database)

Table 9 gathers the RMSDs, MADs, and MSDs for the IACBH8 dataset. We note that for the methods for which we have both A'VTZ and A'VQZ results (MP2, SCS-MP2, MP2.5, MP3, and

Table 9
Statistical analysis for the performance of *ab initio* methods for the reaction barrier heights in the IACBH8 dataset (in kJ mol⁻¹).^a

Basis sets	Methods	RMSD	MAD	MSD	
A'VQZ	MP2	7.9	7.3	-7.3	
	SCS-MP2	11.9	11.4	11.4	
	MP2.5	4.7	4.7	4.7	
	MP3	16.7	16.6	6.6	
	SCS-MP3	17.6	17.4	17.4	
	MP2	9.6	9.0	-9.0	
A'VTZ	SCS-MP2	10.0	9.4	9.4	
	MP2.5	3.2	3.0	3.0	
	MP3	15.1	14.9	14.9	
	SCS-MP3	15.6	15.4	15.4	
	MP4(SDQ)	11.3	11.3	11.3	
	MP4	8.4	8.0	-8.0	
	MP3.5	3.6	3.5	3.5	
	MP4 _{avg}	2.0	1.8	1.7	
	CCSD	13.4	13.3	13.3	
	SCS-CCSD	5.4	5.3	5.3	
	CCSD(T)	1.3	1.2	-1.1	
	CBS(MP2)	CCSD(T)	2.1	2.0	2.0

^a Footnotes a and b to Table 2 apply here.

SCS-MP3) the differences in the overall RMSDs between these two basis sets are smaller than (or equal to) 2.0 kJ mol⁻¹ (in absolute value).

Inspection of Table 9 reveals that the IACBH8 database is a challenging test for nearly all the standard and modified MP n procedures. Only three of the nine procedures result in RMSDs below the threshold of chemical accuracy, i.e. MP2.5 (3.2), MP3.5 (3.6), and MP4_{avg} (2.0 kJ mol⁻¹). For the other six methods the RMSDs range from 8.4 (MP4) to 15.6 kJ mol⁻¹ (SCS-MP3). The CCSD procedure attains a large RMSD of 13.4 kJ mol⁻¹, while its spin-component scaled variant SCS-CCSD performs better, with an RMSD of 5.4 kJ mol⁻¹. This RMSD is reduced to merely 1.3 kJ mol⁻¹ for the CCSD(T) procedure, whilst the CCSD(T)/CBS(MP2) approach attains a slightly higher RMSD of 2.1 kJ mol⁻¹.

4.2.2. Performance for barrier heights of proton-exchange reactions in NH₃, H₂O, and HF clusters (BHPE13 database)

Table 10 lists the error statistics for the BHPE13 database, which is divided into three subsets of barrier heights, namely the NH₃, H₂O, and HF clusters. We start by noting that for the methods for which we have both A'VTZ and A'VQZ results (MP2, SCS-MP2, MP2.5, MP3, and SCS-MP3) the differences in the overall RMSDs of the whole BHPE13 database between these two basis sets are smaller than 3.2 kJ mol⁻¹. For nearly all the MP n -based methods the A'VTZ basis set performs slightly better than the A'VQZ basis set (relative to the W1-F12 reference values). In the following discussion, we will use the A'VTZ basis set to assess the performance of all the considered methods for the entire BHPE13 dataset and its three subsets. Inspection of Table 10 reveals that only five of the 13 tested methods result in RMSDs below the chemical accuracy threshold, i.e. MP2.5 (2.4), MP3.5 (4.0), SCS-CCSD (3.2), CCSD(T)/CBS(MP2) (3.4), and CCSD(T) (4.1 kJ mol⁻¹). For the other eight methods the RMSDs range between 4.7 (MP4_{avg}) and 18.8 kJ mol⁻¹ (SCS-MP3).

We now move to the performance of the standard and modified *ab initio* procedures for the barrier heights in the three subsets of the BHPE13 database. We begin by noting that the MP2 and SCS-MP2 procedures perform poorly for all the subsets. Specifically, MP2, which systematically underestimates the barrier heights (as evident from MSD = -1 × MAD for all the three subsets), results in RMSDs of 10.5, 18.1, and 13.5 kJ mol⁻¹ for the barrier heights in the NH₃, H₂O, and HF subsets, respectively. Similarly, SCS-MP2 attains RMSDs of 12.9 (NH₃ and H₂O subsets) and 9.7 kJ mol⁻¹ (HF subset). The MP3 and SCS-MP3 methods show even poorer performance with RMSDs varying between 14.9 (MP3 for the NH₃ subset) and 21.3 kJ mol⁻¹ (SCS-MP3 for the H₂O subset), where the barrier heights are systematically overestimated. In this situation, the MP2.5 procedure shows good performance with RMSDs of 2.3 (NH₃), 1.8 (H₂O), and 3.0 kJ mol⁻¹ (HF subset). Similarly, the MP3.5 method shows good performance for the H₂O and HF subsets with RMSDs of 3.7 and 2.2 kJ mol⁻¹, respectively, while for the NH₃ subset it attains a larger RMSD of 6.1 kJ mol⁻¹.

With regards to the computationally more expensive coupled cluster procedures, the CCSD method performs worse than MP3.5 and MP4_{avg}, obtaining much larger RMSDs of 19.7 (NH₃ subset), 20.2 (H₂O subset), and 12.6 kJ mol⁻¹ (HF subset). However, its spin-component scaled variant SCS-CCSD performs very well, with RMSDs ranging between 0.9 (H₂O subset) and 4.5 kJ mol⁻¹ (HF subset). The CCSD(T) method shows good performance for the NH₃ and H₂O subsets (RMSDs = 1.7 and 2.6 kJ mol⁻¹, respectively). However, for the HF subset, CCSD(T) attains a somewhat disappointing RMSD of 5.9 kJ mol⁻¹. The CCSD(T)/CBS(MP2) method, on the other hand, shows good performance for all the subsets with RMSDs of 2.5 (NH₃ subset), 4.4 (H₂O subset), and 2.6 kJ mol⁻¹ (HF subset).

Table 10
Statistical analysis for the performance of *ab initio* methods for the reaction barrier heights in the BHPE13 dataset and its three subsets (in kJ mol^{-1}).^a

Datasets	Basis sets	Methods	RMSD	MAD	MSD		
NH ₃ subset	A'VQZ	MP2	10.8	10.6	-10.6		
		SCS-MP2	12.9	12.2	12.2		
		MP2.5	2.1	2.0	2.0		
		MP3	14.7	14.6	14.6		
		SCS-MP3	19.2	18.5	18.5		
	A'VTZ	MP2	10.5	10.3	-10.3		
		SCS-MP2	12.9	12.4	12.4		
		MP2.5	2.3	2.3	2.3		
		MP3	14.9	14.8	14.8		
		SCS-MP3	19.2	18.6	18.6		
		MP4(SDQ)	18.4	18.0	18.0		
		MP4	2.9	2.7	-2.7		
		MP3.5	6.1	6.0	6.0		
		MP4 _{avg}	7.8	7.6	7.6		
		CCSD	19.7	19.3	19.3		
		SCS-CCSD	3.2	3.2	3.2		
		CCSD(T)	1.7	1.6	1.6		
		CBS(MP2)	CCSD(T)	2.5	2.3	2.3	
		H ₂ O subset	A'VQZ	MP2	15.0	14.5	-14.5
				SCS-MP2	16.5	15.3	15.3
MP2.5	2.2			2.1	2.1		
MP3	18.9			18.7	18.7		
SCS-MP3	24.8			23.6	23.6		
A'VTZ	MP2		18.1	17.4	-17.4		
	SCS-MP2		12.9	12.0	12.0		
	MP2.5		1.8	1.6	-0.7		
	MP3		16.1	16.0	16.0		
	SCS-MP3		21.3	20.3	20.3		
	MP4(SDQ)		18.0	17.5	17.5		
	MP4		8.9	8.7	-8.7		
	MP3.5		3.7	3.6	3.6		
	MP4 _{avg}		4.6	4.4	4.4		
	CCSD		20.2	19.6	19.6		
	SCS-CCSD		0.9	0.9	-0.7		
	CCSD(T)		2.6	2.3	-2.3		
	CBS(MP2)		CCSD(T)	4.4	4.2	4.2	
	HF subset		A'VQZ	MP2	9.1	8.9	-8.9
				SCS-MP2	14.7	13.9	13.9
MP2.5		3.2		3.0	3.0		
MP3		15.0		14.9	14.9		
SCS-MP3		20.5		19.8	19.8		
A'VTZ		MP2	13.5	13.0	-13.0		
		SCS-MP2	9.7	9.3	9.3		
		MP2.5	3.0	2.6	-1.0		
		MP3	11.3	11.0	11.0		
		SCS-MP3	15.7	15.3	15.3		
		MP4(SDQ)	10.8	10.7	10.7		
		MP4	10.4	10.1	-10.1		
		MP3.5	2.2	1.8	0.5		
		MP4 _{avg}	0.7	0.6	0.3		
		CCSD	12.6	12.6	12.6		
		SCS-CCSD	4.5	4.0	-3.8		
		CCSD(T)	5.9	5.2	-5.2		
		CBS(MP2)	CCSD(T)	2.6	2.6	2.6	
		Overall	A'VQZ	MP2	12.0	11.4	-11.4
				SCS-MP2	15.0	14.0	14.0
MP2.5	2.6			2.4	2.4		
MP3	16.5			16.3	16.3		
SCS-MP3	22.0			21.1	21.0		
A'VTZ	MP2		14.9	14.1	-14.1		
	SCS-MP2		11.8	11.0	11.0		
	MP2.5		2.4	2.2	-0.1		
	MP3		14.2	13.8	13.8		
	SCS-MP3		18.8	18.0	18.0		
	MP4(SDQ)		15.7	15.0	15.0		
	MP4		8.6	7.8	-7.8		
	MP3.5		4.0	3.5	3.0		
	MP4 _{avg}		4.7	3.7	3.6		
	CCSD		17.5	16.8	16.8		
	SCS-CCSD		3.2	2.6	-1.0		
	CCSD(T)		4.1	3.3	-2.5		
	CBS(MP2)		CCSD(T)	3.4	3.2	3.2	

^a Footnotes a and b to Table 2 apply here.

4.2.3. Performance for barrier heights of water-catalyzed proton-transfer reactions (WCPT27 database)

Table 11 gives error statistics for the overall WCPT27 dataset as well as its three subsets: uncatalyzed reactions (denoted as uncat), reactions catalyzed by one water molecule (denoted as 1H₂O), and reactions catalyzed by two water molecules (denoted as 2H₂O). Firstly, we note that for the methods for which we have both A'VTZ and A'VQZ results (MP2, SCS-MP2, MP2.5, MP3, and SCS-MP3) the differences in the RMSDs of the whole WCPT27 database between these two basis sets are less than 0.9 kJ mol^{-1} . In the following text, we will use the results of the A'VTZ basis set to assess the performance of the computational methods for the WCPT27 dataset and its three subsets.

The MP2 method attains RMSDs of 7.2, 11.2, and 13.7 kJ mol^{-1} for the uncat, 1H₂O, and 2H₂O subsets, respectively; while the SCS-MP2 method performs better with RMSDs of 5.8, 8.7, and 11.7 kJ mol^{-1} , respectively. The MP3 method systematically overestimates the barrier heights for the three subsets and obtains RMSDs of 14.9 (uncat), 11.4 (1H₂O), and 12.0 kJ mol^{-1} (2H₂O). The SCS-MP3 procedure also shows poor performance with RMSDs ranging between 10.8 (uncat) and 17.9 kJ mol^{-1} (2H₂O). For the overall WCPT27 database, the RMSDs of MP3 and SCS-MP3 are 12.8 and 14.6 kJ mol^{-1} , respectively, where both procedures overestimate the barrier heights, as evident from MSD = MAD. Since MP2 tends to underestimate the barrier heights and MP3 tends to overestimate them, MP2.5 shows very good performance and attains RMSDs of 4.0 (uncat), 1.8 (1H₂O), 2.1 (2H₂O), and 2.8 kJ mol^{-1} (overall WCPT27 dataset).

We now turn our attention to the performance of the MP4-based methods. The MP4 method gives RMSDs of 5.2 (uncat), 5.9 (1H₂O), 7.0 (2H₂O), and 6.1 kJ mol^{-1} (overall WCPT27 dataset). The modified MP4-based methods, MP3.5 and MP4_{avg}, show somewhat better performance for the subsets as well as the whole WCPT27 dataset. Specifically, MP3.5 gives RMSDs of 5.1 (uncat), 3.0 (1H₂O), 2.9 (2H₂O), and 3.8 kJ mol^{-1} (overall WCPT27 dataset). MP4_{avg} provides similar performance with RMSDs of 4.7 (uncat), 4.0 (1H₂O), 4.7 (2H₂O), and 4.5 kJ mol^{-1} (overall WCPT27 dataset).

With regards to the coupled cluster procedures, CCSD obtains large RMSDs ranging between 14.3 (uncat) and 17.0 kJ mol^{-1} (2H₂O). Scaling the same and opposite-spin components, however, significantly improves the performance. The SCS-CCSD method gives RMSDs ranging between 3.8 (uncat) and 4.0 kJ mol^{-1} (1H₂O). The CCSD(T) method shows excellent performance for the three subsets, with RMSDs equal to (or smaller than) 2.2 kJ mol^{-1} . The CCSD(T)/CBS(MP2) method provides outstanding performance for the whole database as well as its three subsets and achieves RMSDs of 1.6 (uncat), 0.8 (1H₂O), 2.6 (2H₂O), and 1.8 kJ mol^{-1} (overall WCPT27 dataset).

4.2.4. Performance for barrier heights of hydrogen-transfer and nonhydrogen-transfer reactions (BH76 database)

We also evaluate the performance of standard and modified *ab initio* methods for the reaction barrier heights in the BH76 database. Table 12 gives an overview of the performance of the examined methods in conjunction with the A'VQZ basis sets for the BH76 database and its two subsets (HTBH38 and NHTBH38) [35]. Table S12 (Supplementary data) gives an overview of the basis set convergence for the BH76 database with the A'VnZ basis sets ($n = \text{D}$ and T). We note that practically all the MPn-based methods converge rapidly to the basis set limit. For example, the differences in RMSDs for the overall BH76 database between the A'VTZ and A'VQZ basis sets for the MPn-based methods (except for MP2, for which the difference is 1.0 kJ mol^{-1}) are equal to or smaller than 0.4 kJ mol^{-1} (in absolute value).

Inspection of Table 12 reveals that the whole BH76 database is a challenging test for all the considered MPn-based methods, with

Table 11

Statistical analysis for the performance of *ab initio* methods for the reaction barrier heights in the WCPT27 dataset and its three subsets (in kJ mol^{-1}).^a

Datasets	Basis sets	Methods	RMSD	MAD	MSD		
Uncat ^b	A'VQZ	MP2	7.7	7.5	-7.5		
		SCS-MP2	5.3	4.7	4.7		
		MP2.5	3.8	3.5	3.5		
		MP3	14.7	14.5	14.5		
		SCS-MP3	10.4	10.2	10.2		
	A'VTZ	MP2	7.2	7.1	-7.1		
		SCS-MP2	5.8	5.0	5.0		
		MP2.5	4.0	3.8	3.8		
		MP3	14.9	14.7	14.7		
		SCS-MP3	10.8	10.4	10.4		
		MP4(SDQ)	13.1	12.6	12.6		
		MP4	5.2	4.7	-4.7		
		MP3.5	5.1	5.0	5.0		
		MP4 _{avg}	4.7	4.0	3.9		
		CCSD	14.3	13.8	13.8		
	CBS(MP2)	SCS-CCSD	3.8	3.2	3.2		
		CCSD(T)	0.8	0.7	0.0		
		CCSD(T)	1.6	1.4	-1.4		
		1H2O cat ^c	A'VQZ	MP2	10.8	10.4	-10.4
				SCS-MP2	9.2	9.1	9.1
MP2.5	1.7			1.4	0.5		
MP3	11.8			11.5	11.5		
SCS-MP3	14.8			14.5	14.5		
A'VTZ	MP2		11.2	10.8	-10.8		
	SCS-MP2		8.7	8.4	8.4		
	MP2.5		1.8	1.4	0.1		
	MP3		11.4	11.1	11.1		
	SCS-MP3		14.2	13.9	13.9		
	MP4(SDQ)		13.2	12.9	12.9		
	MP4		5.9	5.5	-5.5		
	MP3.5		3.0	2.8	2.8		
	MP4 _{avg}		4.0	3.7	3.7		
	CCSD		14.5	14.1	14.1		
CBS(MP2)	SCS-CCSD	4.0	3.3	2.8			
	CCSD(T)	1.5	1.3	-0.8			
	CCSD(T)	0.8	0.7	0.6			
	2H2O cat ^d	A'VQZ	MP2	12.2	11.7	-11.7	
			SCS-MP2	13.4	13.3	13.3	
MP2.5			1.9	1.6	0.7		
MP3			13.3	13.0	13.0		
SCS-MP3			19.6	19.4	19.4		
A'VTZ		MP2	13.7	13.2	-13.2		
		SCS-MP2	11.7	11.5	11.5		
		MP2.5	2.1	1.7	-0.8		
		MP3	12.0	11.7	11.7		
		SCS-MP3	17.9	17.7	17.7		
		MP4(SDQ)	15.6	15.3	15.3		
		MP4	7.0	6.5	-6.5		
		MP3.5	2.9	2.6	2.6		
		MP4 _{avg}	4.7	4.4	4.4		
		CCSD	17.0	16.7	16.7		
CBS(MP2)	SCS-CCSD	3.9	3.3	2.7			
	CCSD(T)	2.2	1.8	-1.4			
	CCSD(T)	2.6	2.5	2.5			
	Overall ^e	A'VQZ	MP2	10.4	9.9	-9.9	
			SCS-MP2	9.9	9.0	9.0	
MP2.5			2.7	2.2	1.6		
MP3			13.3	13.0	13.0		
SCS-MP3			15.4	14.7	14.4		
A'VTZ		MP2	11.0	10.3	-10.3		
		SCS-MP2	9.0	8.3	8.3		
		MP2.5	2.8	2.3	1.1		
		MP3	12.8	12.5	12.5		
		SCS-MP3	14.6	14.0	14.0		
		MP4(SDQ)	14.0	13.6	13.6		
		MP4	6.1	5.6	-5.6		
		MP3.5	3.8	3.4	3.4		
		MP4 _{avg}	4.5	4.0	4.0		
		CCSD	15.3	14.9	14.9		
CBS(MP2)	SCS-CCSD	3.9	3.3	2.9			
	CCSD(T)	1.6	1.3	-0.8			
	CCSD(T)	1.8	1.5	0.6			

Table 12

Statistical analysis for the performance of *ab initio* methods for the reaction barrier heights in the BH76 dataset and its two subsets HTBH38 and NHTBH38 (in kJ mol^{-1}).^a

Datasets	Basis sets	Methods	RMSD	MAD	MSD		
HTBH38 subset	A'VQZ	MP2	18.5	14.1	5.1		
		SCS-MP2	24.3	21.0	21.0		
		MP2.5	14.9	10.8	9.9		
		MP3	17.5	14.7	14.7		
		SCS-MP3	25.5	23.4	23.4		
		MP4(SDQ)	15.3	12.9	12.9		
		MP4	9.1	6.4	2.7		
		MP3.5	11.8	8.7	8.7		
		MP4 _{avg}	10.7	7.8	7.8		
		CCSD	10.9	9.6	9.3		
		CCSD(T)	1.2	1.0	0.2		
		CBS(MP2)	CCSD(T)	1.5	1.2	0.5	
		NHTBH38 subset	A'VQZ	MP2	10.7	8.3	4.8
				SCS-MP2	15.2	12.8	11.8
				MP2.5	12.1	11.0	10.6
MP3	19.5			16.3	16.3		
SCS-MP3	17.0			15.2	14.6		
MP4(SDQ)	12.7			10.5	10.5		
MP4	6.5			4.5	-1.6		
MP3.5	8.8			7.4	7.4		
MP4 _{avg}	6.3			5.1	4.4		
CCSD	12.1			10.0	10.0		
CCSD(T)	0.8			0.6	0.0		
CBS(MP2)	CCSD(T)			1.6	1.3	0.0	
BH76 set	A'VQZ			MP2	15.1	11.2	5.0
				SCS-MP2	20.3	16.9	16.4
				MP2.5	13.6	10.9	10.2
		MP3	18.5	15.5	15.5		
		SCS-MP3	21.6	19.3	19.0		
		MP4(SDQ)	14.0	11.7	11.7		
		MP4	7.9	5.5	0.6		
		MP3.5	10.4	8.0	8.0		
		MP4 _{avg}	8.8	6.5	6.1		
		CCSD	11.5	9.8	9.6		
		CCSD(T)	1.0	0.8	0.1		
		CBS(MP2)	CCSD(T)	1.6	1.2	0.3	

^a Footnotes *a* and *b* to Table 2 apply here.

RMSDs ranging between 7.9 (MP4) and 21.6 kJ mol^{-1} (SCS-MP3). The CCSD method attains an RMSD of 11.5 kJ mol^{-1} , whilst CCSD (T) gives superb performance with an RMSD of merely 1.0 kJ mol^{-1} . The cost-effective CCSD(T)/CBS(MP2) approach performs very well with an RMSD of 1.6 kJ mol^{-1} .

The HTBH38 subset consists of 38 barrier heights for hydrogen-transfer reactions. All the MP n -based methods provide poor performance for this subset with RMSDs ranging between 9.1 (MP4) and 25.5 kJ mol^{-1} (SCS-MP3). Interestingly, nearly all the examined MP n -based procedures systematically overestimate the barrier heights for the HTBH38 subset, as evident from $\text{MSD} \approx \text{MAD}$. The CCSD procedure also results in a large RMSD of 10.9 kJ mol^{-1} , however, CCSD(T), as expected, provides excellent performance with an RMSD of 1.2 kJ mol^{-1} . The CCSD(T)/CBS(MP2) method gives similar performance, resulting in an RMSD of 1.5 kJ mol^{-1} .

The subset of non-hydrogen transfer barrier heights, NHTBH38, contains three types of barrier heights, namely for heavy atom transfer (HATBH12), nucleophilic substitution (NSBH16), and unimolecular/association (UABH10) reactions. The MP n -based procedures give RMSDs ranging from 6.3 (MP4_{avg}) to 19.5 kJ mol^{-1} (MP3). Once again nearly all the tested MP n -based methods systematically overestimate the barrier heights for the NHTBH38 subset, with $\text{MSD} \approx \text{MAD}$, except for MP4. The MP4 procedure shows very similar performance to MP4_{avg}, obtaining an RMSD of 6.5 kJ mol^{-1} . The CCSD procedure gives poor performance, however, the CCSD(T) method results in an exceptionally low RMSD of 0.8 kJ mol^{-1} , which increases to 1.6 kJ mol^{-1} for the cost-effective CCSD(T)/CBS(MP2) procedure.

^a Footnotes *a* and *b* to Table 2 apply here.

^b Subset of uncatalyzed reactions.

^c Subset of reactions catalyzed by one water molecule.

^d Subset of reactions catalyzed by two water molecules.

^e Including all the uncatalyzed and catalyzed reaction barrier heights.

4.3. Overview of the performance of the modified MP n procedures for thermochemical and kinetic properties

In order to compare the overall performance of the various MP n -based procedures over a range of databases it is useful to define a weighted total root mean square deviation (WTRMSD), which combines the RMSDs for all the databases into a single RMSD for each method. We note that the W4-11 and BH76 databases have been excluded from this statistical analysis since many of the MP n -based procedures show poor performance for these datasets (see above). We will use a similar approach to that used by Goerigk and Grimme [5,55]. Namely, the WTRMSD is taken as:

$$WTRMSD = \frac{1}{523.1} \times \sum_{i=1}^9 N_i \times \frac{RMSD_i^{MP2}}{RMSD_i^{MP4}} \times RMSD_i \quad (7)$$

Since some of the chemical properties (e.g., reaction barrier heights) are more challenging than others (e.g., conformational and isomerization energies), this equation takes into account the relative difficulty of each database as the ratio between the RMSDs for the MP2 and MP4 methods. However, we note that removing this ratio from Eq. (7) has a very small effect on the final WTRMSDs, in particular they change by less than 0.2 kJ mol⁻¹. The WTRMSDs (without the challenging W4-11 and BH76 datasets) are summarized in Fig. 2, whilst the WTRMSDs with the W4-11 and BH76 datasets are listed in Table S13 (Supplementary data).

Of the MP2- and MP3-based procedures, the MP2.5 method puts in the best overall performance with a WTRMSD of 3.4 kJ mol⁻¹, and even outperforms the MP4 and CCSD methods. The MP3 and CCSD methods result in similar performance with WTRMSDs of 5.1 and 5.0 kJ mol⁻¹, respectively. We note that similar observations have been previously found for weak interactions [22,56–58]. The other MP2- and MP3-based methods (MP2, SCS-MP2, MP3, and SCS-MP3) show poor overall performance with WTRMSDs ranging between 4.7 (SCS-MP2) and 7.1 kJ mol⁻¹ (MP2). Fig. 2 shows that there is a significant improvement in performance of the MP4-based methods in the order: MP4(SDQ) → MP4 → MP4_{avg} → MP3.5. The MP4(SDQ) method attains a WTRMSD of 5.6 kJ mol⁻¹. The MP4 method attains a lower WTRMSD of 4.2 kJ mol⁻¹, which is smaller than that for the CCSD and SCS-CCSD procedures. The modified MP4-based methods (MP4_{avg} and MP3.5) show better performance than MP4 at no

additional computational cost. The MP4_{avg} procedure results in a WTRMSD of 3.1 kJ mol⁻¹, however it does not represent a significant improvement over the computationally more economical MP2.5 method (with a WTRMSD of 3.4 kJ mol⁻¹). The MP3.5 procedure shows stellar performance with a WTRMSD of 1.9 kJ mol⁻¹, which is the same as that for the CCSD(T)/A'VTZ method.

We note in passing that optimizing the empirical coefficients of the MP3.5 and MP4_{avg} methods (Eqs. (5) and (6)) to minimize the RMSD for each of the databases leads to optimal scaling coefficients which are close to 0.5 for most of the databases. The optimized coefficients for the MP3.5 and MP4_{avg} methods are summarized in Table S14 of the Supplementary data.

It is also of interest to compare the performance of the spin-component scaled methods (SCS-MP2, SCS-MP3, and SCS-CCSD). For a number of thermochemical databases, the SCS-MP2 and SCS-MP3 procedures have been found to give similar performance [8,12,13]. Here we also find that the three procedures show similar overall performance with WTRMSDs ranging between 4.7 (SCS-MP2) and 5.4 kJ mol⁻¹ (SCS-MP3). Finally, a comment is due on the computational cost of these methods. Table 13 gives the relative central processing unit (CPU) times used by the MP n methods for naphthalene (C₁₈H₁₂, in D_{2h} symmetry). All the calculations were carried out in conjunction with the cc-pVTZ basis set. We can see that for a highly symmetric, medium-sized system such as naphthalene the MP2 and MP3 calculations take a fraction of the CPU time compared to the MP4 and CCSD(T) calculations. Therefore, methods such as SCS-MP2, SCS-MP3, and MP2.5 provide significant computational savings relative to the MP4 and CCSD(T)

Table 13
Computational resources used for single-point-energy MP n and CCSD(T) calculations for naphthalene.^{a,b}

Level of theory	CPU hours ^c
MP2/VTZ	0.4
MP3/VTZ	0.6
MP4/VTZ	20.6
CCSD(T)/VTZ	31.4

^a All the calculations ran on 1 core of dual Intel Xeon E5-2670v2 systems (3.1 GHz, 256 GB RAM).

^b Scratch disk usage was similar in all cases (i.e., 60–70 GB of scratch disk was used).

^c Central processing unit (CPU) times in hours.

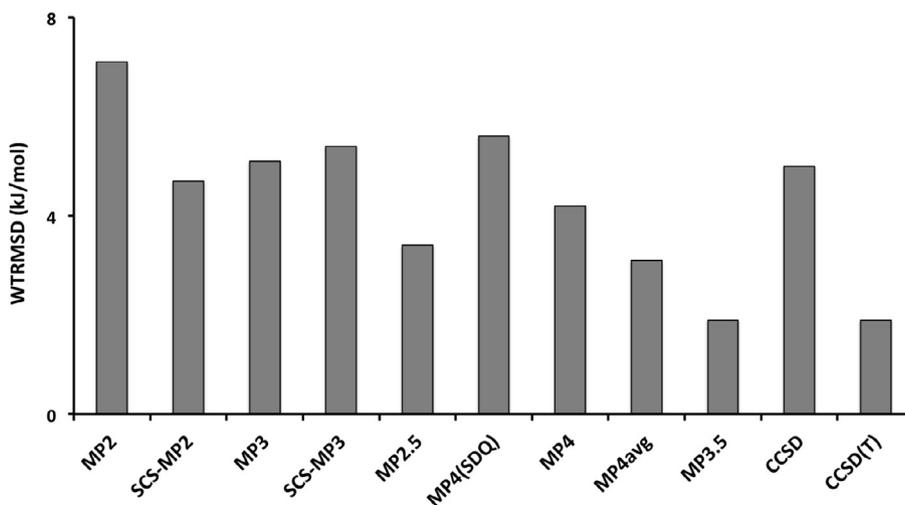


Fig. 2. Weighted total root mean square deviations (WTRMSDs, in kJ mol⁻¹) for the considered methods. The challenging W4-11 and BH76 datasets are excluded from these statistics (see main text). The values for the tetrapeptide conformers dataset are based on data obtained with the cc-pVDZ basis set, while the values for all the other databases are based on data obtained with the cc-pVTZ or A'VTZ basis sets.

methods, and are recommended in cases where they show good performance. Nevertheless, the MP4 calculation requires 66% of the CPU time required for the CCSD(T) calculation. Therefore, methods such as MP3.5 and MP4_{avg} still provide reduced computational cost relative to CCSD(T) and are recommended in cases where they give similar performance to the CCSD(T) method.

5. Conclusions

We have assessed the performance of standard and modified *ab initio* methods for a wide range of thermochemical and kinetic properties including atomization energies, isomerization energies, conformational energies, barrier heights for hydrogen transfer and non-hydrogen transfer reactions. The reference values were obtained at the CCSD(T)/CBS level of theory by means of the *Wn* and *Wn-F12* ($n = 1$ and 2) thermochemical protocols (except for the W4-11 dataset for which the reference values were obtained at the FCI/CBS level). With regards to the performance for thermochemical properties, we make the following observations:

- None of the MP n -based methods give acceptable performance for the extremely challenging W4-11 dataset. The best performing MP n -based methods result in RMSDs of 13.3 (MP2.5), 14.6 (MP4_{avg}), 16.5 (MP3.5), and 16.7 kJ mol⁻¹ (MP4). Whilst the worst performing methods (MP2, SCS-MP2, and MP3) result in RMSDs > 45.0 kJ mol⁻¹. The cost-effective approach CCSD(T)/CBS(MP2) method, on the other hand, shows good performance with an RMSD of 3.8 kJ mol⁻¹.
- MP2.5 in conjunction with the A'VTZ or cc-pVTZ basis set performs very well for isomerization and conformational energies and obtains RMSDs of 4.0 (H_xC_yO_z isomers), 1.1 (DIE60 isomers), 0.7 (EIE22 isomers), 0.19 (C_nH_{2n+2} conformers), and 2.7 kJ mol⁻¹ (tetrapeptide conformers).
- The modified MP4-based method, MP3.5, provides superb performance for all of the tested thermochemical properties (apart from atomization energies). The following RMSDs are obtained with the A'VTZ basis set: 3.0 (C₈H₈ isomers), 1.6 (H_xC_yO_z isomers), 0.5 (DIE60 and EIE22 isomers), and 0.15 kJ mol⁻¹ (C_nH_{2n+2} conformers). The MP4_{avg} method gives slightly higher RMSDs for the abovementioned databases.
- With no exceptions, the cost-effective CCSD(T)/CBS(MP2) approach shows excellent performance for all the considered datasets, resulting in RMSDs between 0.14 (C_nH_{2n+2} conformers) and 3.8 kJ mol⁻¹ (W4-11 dataset).

With regards to the performance of the standard and modified *ab initio* methods for reaction barrier heights, we draw the following conclusions:

- The BH76 database is a challenging test for all the MP n -based methods. The RMSDs for this database range between 7.9 (MP4) and 21.6 kJ mol⁻¹ (SCS-MP3).
- The MP2.5 method in conjunction with the A'VTZ basis set performs very well for the reaction barrier heights of proton-transfer or proton-exchange reactions, with RMSDs of 3.2 (IACBH8), 2.4 (BHPE13), and 2.8 kJ mol⁻¹ (WCPT27). The MP3.5 and MP4_{avg} methods also show good performances for these databases, albeit at a higher computational cost.
- MP3 and its spin-components scaled procedure SCS-MP3 in conjunction with the A'VTZ basis set give poor performance for the proton-transfer and proton-exchange reactions, with RMSDs ranging from 12.8 (MP3 for the WCPT27 dataset) to 18.8 kJ mol⁻¹ (SCS-MP3 for the BHPE13 dataset). For the BH76 database, MP3 and SCS-MP3 obtain RMSDs of 18.5 and 21.6 kJ mol⁻¹, respectively with the A'VQZ basis set.

- The cost-effective CCSD(T)/CBS(MP2) method shows outstanding performance for all the datasets with RMSDs of 2.1 (IACBH8), 3.4 (BHPE13), 1.8 (WCPT27), and 1.6 kJ mol⁻¹ (BH76).

With regards to the overall performance of the considered methods for thermochemistry and kinetics (excluding the W4-11 and BH76 datasets), we draw the following conclusions:

- The MP2.5 method provides excellent overall performance (with WTRMSD = 3.4 kJ mol⁻¹) and even outperforms the MP4 and CCSD methods (with WTRMSDs = 4.2 and 5.0 kJ mol⁻¹, respectively).
- The other MP2- and MP3-based methods (MP2, SCS-MP2, MP3, and SCS-MP3) show poorer overall performance with WTRMSDs ranging between 4.7 (SCS-MP2) and 7.1 kJ mol⁻¹ (MP2).
- The modified MP4-based methods (MP3.5 and MP4_{avg}) show better performance than MP4 at no additional computational cost. In particular, the MP3.5 procedure shows stellar performance with a WTRMSD of 1.9 kJ mol⁻¹, which is the same WTRMSD obtained for the CCSD(T) method in conjunction with the same basis set.

Acknowledgments

We gratefully acknowledge the generous allocation of computing time from the National Computational Infrastructure (NCI) National Facility, and system administration support provided by the Faculty of Science at the University of Western Australia (UWA) to the Linux cluster of the Karton group, the provision of an Australian Postgraduate Award (to L.-J.Y.), and an Australian Research Council (ARC) Discovery Early Career Researcher Award (to A.K., project number: DE140100311). We would like to thank the reviewers of the current manuscript for their valuable suggestions.

Appendix A. Supplementary data

The MSD values of MP n -based methods ($n = 2, 3,$ and 4) for the tested datasets (Table S1); overview of basis set convergence for the W4-11 database (Table S2); statistical analysis for the performance of the considered methods in conjunction with the cc-pVDZ or A'VDZ basis set for the C₈H₈ isomers (Table S3), HxCyOz isomers (Table S4), DIE60 isomers (Table S5), EIE60 isomers (Table S6), C_n-H_{2n+2} conformers (Table S7), C_nH_{2n+2} conformers (Table S8), IACBH8 reaction barrier heights (Table S9), BHPE13 reaction barrier heights (Table S10), and WCPT27 reaction barrier heights (Table S11); statistical analysis for the performance of the tested *ab initio* methods in conjunction with the A'VnZ basis sets ($n = D$ and T) for the BH76 database (Table S12); WTRMSDs for all the 11 databases (Table S13); optimal coefficients for the MP3.5 and MP4_{avg} methods for each of the databases (Table S14); and full references for Gaussian 09 (Ref. [45]) and Molpro (Ref. [44]).

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.chemphys.2016.10.009>.

References

- [1] S. Grimme, L. Goerigk, R.F. Fink, *WIREs Comput. Mol. Sci.* 2 (2012) 886.
- [2] S. Grimme, *J. Chem. Phys.* 118 (2003) 9095.
- [3] Á. Szabados, *J. Chem. Phys.* 125 (2006) 214105.
- [4] F. Neese, T. Schwabe, S. Kossmann, B. Schirmer, S. Grimme, *J. Chem. Theory Comput.* 5 (2009) 3060.
- [5] L. Goerigk, S. Grimme, *Phys. Chem. Chem. Phys.* 13 (2011) 6670.
- [6] L. Goerigk, S. Grimme, *J. Chem. Theory Comput.* 7 (2011) 291.

- [7] S. Grimme, *J. Phys. Chem. A* 109 (2005) 3067.
- [8] A. Karton, J.M. Martin, *Mol. Phys.* 110 (2012) 2477.
- [9] E.I. Izgorodina, M.L. Coote, L. Radom, *J. Phys. Chem. A* 109 (2005) 7558.
- [10] A. Karton, A. Tarnopolsky, J.-F. Lamere, G.C. Schatz, J.M. Martin, *J. Phys. Chem. A* 112 (2008) 12868.
- [11] S. Grimme, *Angew. Chem. Int. Ed.* 45 (2006) 4460.
- [12] L.-J. Yu, A. Karton, *Chem. Phys.* 441 (2014) 166.
- [13] L.-J. Yu, F. Sarrami, A. Karton, R.J. O'Reilly, *Mol. Phys.* 113 (2015) 1284.
- [14] U.R. Fogueri, S. Kozuch, A. Karton, J.M. Martin, *J. Phys. Chem. A* 117 (2013) 2269.
- [15] A. Karton, L. Goerigk, *J. Comput. Chem.* 36 (2015) 622.
- [16] L.-J. Yu, F. Sarrami, R.J. O'Reilly, A. Karton, *Chem. Phys.* 458 (2015) 1.
- [17] A. Karton, R.J. O'Reilly, L. Radom, *J. Phys. Chem. A* 116 (2012) 4211.
- [18] A. Karton, R.J. O'Reilly, B. Chan, L. Radom, *J. Chem. Theory Comput.* 8 (2012) 3128.
- [19] Y. Jung, R.C. Lochan, A.D. Dutoi, M. Head-Gordon, *J. Chem. Phys.* 121 (2004) 9793.
- [20] R.F. Fink, *J. Chem. Phys.* 133 (2010) 174113.
- [21] S. Grimme, *J. Comput. Chem.* 24 (2003) 1529.
- [22] M. Pitoňák, P. Neogradý, J. Černý, S. Grimme, P. Hobza, *Chem. Phys. Chem.* 10 (2009) 282.
- [23] P. Jurečka, J. Šponer, J. Černý, P. Hobza, *Phys. Chem. Chem. Phys.* 8 (2006) 1985.
- [24] R.A. Distasio Jr, M. Head-Gordon, *Mol. Phys.* 105 (2007) 1073.
- [25] L.-J. Yu, F. Sarrami, R.J. O'Reilly, A. Karton, *Mol. Phys.* 114 (2016) 21.
- [26] R.J. O'Reilly, A. Karton, *Int. J. Quantum Chem.* 116 (2016) 52.
- [27] D. Cremer, *WIREs Comput. Mol. Sci.* 1 (2011) 509.
- [28] R. Krishnan, J.A. Pople, *Int. J. Quantum Chem.* 14 (1978) 91.
- [29] R. Krishnan, M. Frisch, J. Pople, *J. Chem. Phys.* 72 (1980) 4244.
- [30] A. Karton, S. Daon, J.M. Martin, *Chem. Phys. Lett.* 510 (2011) 165.
- [31] The following systems were removed from the W4–11 dataset: BeCl₂, BeF₂, Be₂, B₂, C₂(1 Σ^+), BN(1 Σ^+), OF, F₂O, FOO, FOOF, Cl₂O, ClOO, OCIO, O₃, S₃, S₄, cis-HOOO, and trans-HOOO.
- [32] D. Gruzman, A. Karton, J.M. Martin, *J. Phys. Chem. A* 113 (2009) 11974.
- [33] L. Goerigk, A. Karton, J.M. Martin, L. Radom, *Phys. Chem. Chem. Phys.* 15 (2013) 7028.
- [34] A. Karton, *Chem. Phys. Lett.* 592 (2014) 330.
- [35] Y. Zhao, N. González-García, D.G. Truhlar, *J. Phys. Chem. A* 109 (2005) 2012.
- [36] A. Karton, D. Talbi, *Chem. Phys.* 436 (2014) 22.
- [37] A. Karton, E. Rabinovich, J.M. Martin, B. Ruscic, *J. Chem. Phys.* 125 (2006) 144108.
- [38] A. Karton, *WIREs Comput. Mol. Sci.* 6 (2016) 292.
- [39] A. Karton, J.M. Martin, *J. Chem. Phys.* 136 (2012) 124114.
- [40] J.M. Martin, S. Parthiban, *Quantum-Mechanical Prediction of Thermochemical Data*, Springer, 2001, p. 31.
- [41] J. Jiang, Y. Wu, Z.-X. Wang, C. Wu, *J. Chem. Theory Comput.* 6 (2010) 1199.
- [42] J.M. Martin, G. de Oliveira, *J. Chem. Phys.* 111 (1999) 1843.
- [43] H.J. Werner, P.J. Knowles, G. Knizia, F.R. Manby, M. Schütz, *WIREs Comput. Mol. Sci.* 2 (2012) 242.
- [44] H. Werner, P. Knowles, G. Knizia, F. Manby, M. Schütz, *MOLPRO*, version 2012.1, a package of ab initio programs, see <http://www.molpro.net>, 2012.
- [45] M. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. Petersson, et al. *Gaussian 09, Revision D.01*; Gaussian Inc: Wallingford CT, 2009.
- [46] T.H. Dunning Jr, *J. Chem. Phys.* 90 (1989) 1007.
- [47] R.A. Kendall, T.H. Dunning Jr, R.J. Harrison, *J. Chem. Phys.* 96 (1992) 6796.
- [48] T.H. Dunning Jr, K.A. Peterson, A.K. Wilson, *J. Chem. Phys.* 114 (2001) 9244.
- [49] T. Takatani, E.G. Hohenstein, C.D. Sherrill, *J. Chem. Phys.* 128 (2008) 124111.
- [50] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, A.K. Wilson, *Chem. Phys. Lett.* 286 (1998) 243.
- [51] D.G. Liakos, F. Neese, *J. Phys. Chem. A* 116 (2012) 4801.
- [52] P. Jurečka, P. Hobza, *Chem. Phys. Lett.* 365 (2002) 89.
- [53] W. Klopper, H.P. Luthi, *Mol. Phys.* 96 (1999) 559.
- [54] J. Friedrich, *J. Chem. Theory Comput.* 11 (2015) 3596.
- [55] L. Goerigk, S. Grimme, *J. Chem. Theory Comput.* 6 (2010) 107.
- [56] T. Van Mourik, A.K. Wilson, T.H. Dunning Jr, *Mol. Phys.* 96 (1999) 529.
- [57] S. Tsuzuki, T. Uchimaru, K. Matsumura, M. Mikami, K. Tanabe, *Chem. Phys. Lett.* 319 (2000) 547.
- [58] A. Rappe, E. Bernstein, *J. Phys. Chem. A* 104 (2000) 6117.