JCTC Journal of Chemical Theory and Computation

"Mindless" DFT Benchmarking

Martin Korth and Stefan Grimme*

*Organisch-Chemisches Institut, Universität Münster, Corrensstrasse 40,
D-48149 Münster, Germany*

Abstract: A diversity-oriented approach for the generation of thermochemical benchmark sets is presented. Test sets consisting of randomly generated "artificial molecules" (AMs) are proposed that rely on systematic constraints rather than uncontrolled chemical biases. In this way, the narrow structural space of chemical intuition is opened up and electronically difficult cases can be produced in an unforeseeable manner. For the calculation of chemically meaningful relative energies, AMs are systematically decomposed into small molecules (hydrides and diatomics). Two different example test sets containing eight-atom, single-reference, main group AMs with chemically very diverse and unusual structures are generated. Highly accurate all-electron, estimated CCSD(T)/complete basis set reference energies are also provided. They are used to benchmark the density functionals S-VWN, BP86, B-LYP, B97-D, PBE, TPSS, PBEh, BH-LYP, B3-PW91, B3-LYP, B2-PLYP, B2GP-PLYP, BMK, MPW1B95, M05, M05-2X, PW6B95, M06, M06-L, and M06-2X. In selected cases, an empirical dispersion correction (DFT-D) has been applied. Due to the composition of the sets, it is expected that a good performance indicates "robustness" in many different chemical applications. The results of a statistical analysis of the errors for the entire set with 165 entries (average reaction energy of 117 kcal/mol, dubbed as the MB08-165 set) perfectly fit to the "Jacob's ladder" metaphor for the ordering of density functionals according to their theoretical complexity. The mean absolute deviation (MAD) decreases very strongly from LDA (20 kcal/mol) to GGAs (MAD of about 10 kcal/mol) but then was less pronounced to hybrid-GGAs (MAD of about 6−8 kcal/mol). The best performance (MAD of 4.1−4.2 kcal/mol) is found for the (fifth-rung) double-hybrid functionals B2-PLYP-D and B2GP-PLYP-D, followed by the M06-2X meta-hybrid (MAD of 4.8 kcal/mol). The significance of the proposed approach for thermodynamic benchmarking is discussed and related to the observed performance ranking also regarding wave function based methods.

1. Introduction

Especially with the rise of the density functional theory (DFT) based zoo of electronic structure methods in the past decade, benchmarking has become an intrinsically important task in quantum chemistry.[1−7] With state-of-the-art functionals that approach the accuracy of high-level wave function theory techniques at least for a variety of "standard" applications,[8,9] it is now becoming more and more obvious that the evaluation of new or improved methods via bench-

marking suffers from two major problems. First, high-quality experimental reference data are not always at hand, and it is often impossible to produce the corresponding theoretical data as an alternative in reasonable time spans. The second—but sometimes overlooked—problem is the composition of the test sets that is often strongly biased by factors like availability, chemical intuition, and the professional interest in "good" results, all leading to a pronounced narrowness of the chemical space under consideration.

An example of the influence of such a bias is the bad performance of the PBE[10] density functional for common benchmark sets, which stands in opposition to its good

* Corresponding author phone: +49-251-83-33241; fax: +49-251-83-36515; e-mail: grimmes@uni-muenster.de.

performance for a variety of real-life problems[5] and to its popularity in solid-state physics. The reason for the discrepancy in this case is also due to the fact that the common tests rely heavily on atomization energies. This produces a strong bias toward an accurate description of the free atoms relative to molecules,[11,12] which is not representative for many chemical applications.

A closer examination of the composition of existing benchmark sets reveals another problem. For example in the so-called Gn thermochemistry sets,[2] the individual compounds are explicitly selected according to the high accuracy of the experimental reference data. This leads to an accumulation of very stable, easily accessible compounds and thus to a lot of entries without much diverse information regarding their electronic structure (e.g., a large number of alkane derivatives with very similar electronic properties). In recent years, the composition of these sets was questioned and new benchmarks with model systems and theoretical reference values came up.[3,5,9] Because their composition is also guided by chemical intuition, these sets are nevertheless limited to the known chemical space, which has been proven to be very narrow in comparison with the chemical universe of possible compounds.[13]

An obvious way to open up the chemical space of test sets is to make use of random procedures in their generation. On the other hand, a completely randomized ansatz is not only impractical but also counterproductive: many biases correspond to some partial knowledge about the system of interest and the neglect of this knowledge results in a greatly diminished efficiency for the construction of test sets. Therefore, the only possible way of dealing with biases in benchmarking is the preferably complete transformation of *unperceived biases into known constraints*. In this spirit, one can make use of random elements constrained by systematic and controllable specifications to avoid unsystematic and uncontrolled criteria for the construction of benchmark sets.

On the basis of this insight, we here make a first suggestion of how to address the "selection" problem in practice and propose a diversity-oriented benchmarking procedure that is inspired by the "mindless chemistry" approach of Bera et al. for the problem of isomer-minima search.[14] Our ansatz is to randomly generate molecular systems for the calculation of relative energies, which are then used for thermochemical benchmarking purposes. Our goal is to produce theoretically demanding test cases in an unforeseeable manner through an opening of the chemical space beyond chemical intuition. This work is also rooted in the strong belief that only robust electronic structure methods will be useful in practice. The term "robust" here means that reliable results are provided even in electronically complicated situations ("extrapolative power"). It can be expected that only robust quantum chemical methods will provide good results also in an opened chemical space. The interest in molecular systems and relative energies reflects the "chemical" orientation of our approach, but algorithms for different properties and systems can surely be set up analogously.

2. Theory of Diversity Oriented Benchmarking

The basic idea of our approach is the generation of what we will call an "artificial molecule" (AM). An AM is not a molecule in the classical sense but a randomly generated minimum on an energetic hypersurface for a random conglomerate of a (predefined) set of atoms. To make sure that the generated systems are of use for quantum chemical research, we developed the following procedure for the generation of AMs. In the first step, the benchmark specification parameters are chosen to determine the general conditions (i.e., the constraints to randomness) of the test set:

1 How many systems are generated?
2 How many atoms does one AM contain?
3 How are the atoms chosen from the periodic system of elements?
4 How are the atoms arranged in space?
5 What is the desired complexity of electronic structure (e.g., total charge, number of open shells, or multireference character)?
6 What reaction scheme is applied?

The second step is the generation of structures, in accordance with the answers to the first four questions. In the third step, one has to ensure that the intended number of molecules has the specified electronic structure according to the fifth question. The fourth step is the generation of reference data. This way, a vast amount of completely different benchmark sets can be systematically generated, characteristically depending on the countless possible answers to the above named questions. It has to be kept in mind that while the basic principle of our approach is completely general and unbiased, any generated benchmark set itself is necessarily constrained. We would like to emphasize that the restrictions we have applied for the generation of the example benchmark sets that we present in the following paragraphs (e.g., single reference cases and CCSD(T) reference values) are limitations of the generated test sets but not of the mindless benchmarking approach itself. While it would surely be a tedious task (that we found unreasonable to undertake for the benchmarking of current DFT methods), the use of multireference methods is in no way incompatible with our benchmarking approach. Along these lines, we have also excluded complicated spin states to keep the example benchmark set generation simple, not because our approach is generally limited in this direction (comparison of different density functionals also becomes problematic because they may yield different spin states as the lowest state). When speaking of the generation of electronically demanding test cases, it has to be kept in mind that alongside multireference character other reasons for electronic complexity (e.g., unusual bonding, small gaps, spin contamination) exist. In general, one cannot say that all multireference cases are necessarily electronically difficult (the multiplet of an atom being an example) and vice versa for single-reference molecules. The advantage of "mindless" benchmarking is the transparency of the process, which turns unperceived biases into known constraints, i.e. that it forces the originator of a benchmark set to explicitly determine the

"Mindless" DFT Benchmarking

J. Chem. Theory Comput., Vol. 5, No. 4, 2009 **995**

Table 1. Initial and Final Element Distributions (in percent) for the MB08-931 and MB08-ORG Benchmark Sets

	MB08-931			MB08-ORG		
element	initial	final	ratio[a]	initial	final	ratio[a]
H	69.2	64.8	0.94	61.8	58.2	0.94
Li	3.3	3.5	1.061	1.9	1.0	0.53
Be	3.3	3.6	1.091	0.5	0.0	0.00
B	3.3	4.7	1.42	1.0	1.8	1.80
C	3.3	3.9	1.18	15.5	17.6	1.14
N	3.3	3.8	1.15	3.9	4.3	1.10
O	3.3	2.9	0.88	3.9	4.5	1.15
F	3.3	3.2	0.97	1.0	1.3	1.30
Na	1.1	1.1	1.00	1.0	0.3	0.30
Mg	1.1	1.2	1.09	1.0	1.2	1.20
Al	1.1	2.0	1.82	1.0	1.5	1.50
Si	1.1	1.4	1.27	1.9	2.4	1.26
P	1.1	1.2	1.09	1.9	2.7	1.42
S	1.1	1.7	1.55	1.9	1.9	1.00
Cl	1.1	1.4	1.27	1.9	1.3	0.68

[a] Ratio of final to initial distribution.

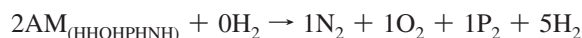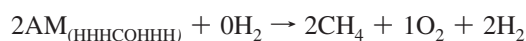boundary conditions to all dimensions of complexity found in the test set.

We have generated two different test sets for the illustration of our approach. For both sets we initially generated 300 random AMs. All systems contain exactly eight atoms, and only main group elements up to chlorine, excluding the noble gases helium and neon, are considered. The two test sets differ in their elemental composition; i.e., the atoms were chosen randomly with different probabilities for the different elements. For the first set, termed "MB08-931", the elemental probabilities were chosen so that the occurrence of the atoms Na to Cl is one-third of the occurrence of the atoms Li to F, which in turn is one-third of the occurrence of hydrogen (see Table 1). For the second set, termed "MB08-ORG", the element probabilities were chosen to roughly reproduce the element distribution found in organic compounds (see Table 1).[15]

The atoms were always placed randomly on the corners of a cube with an initial edge length of 2 b and then freely optimized on the PBE/TZVP[10,16] level. All systems were chosen to be uncharged, and open-shell calculations were restricted to (low-spin) doublet states.

Only systems that converged within a limit of 100 SCF and 100 geometry optimization steps and had a HOMO/LUMO gap of more than 0.5 eV (PBE/TZVP) were considered further. The number of open-shell AMs were chosen to be about 25% of all systems. To screen out multireference cases, an RCCSD(T)/cc-pVDZ[17] calculation was performed and systems with a T1-diagnostic larger than 0.02 or a D1-diagnostic larger than 0.10 were dropped (with both diagnostics rounded to two digits after the decimal point). Note that, when done automatically, checks for double entries and correct spin states have to be included in the benchmark set generation algorithm.

For both sets, a systematic decomposition into hydrides (for the group 1−4 elements) and homonuclear diatomic molecules (for the group 5−7 elements) was chosen as the reaction scheme. The only allowed reactant beside these products and the reactant AM is molecular hydrogen.

The stoichiometry is restricted to integer numbers. The first three reactions from the MB08-931 set are given as an example:

$$2AM_{(HHHCOHHH)} + 0H_2 \rightarrow 2CH_4 + 1O_2 + 2H_2$$

$$2AM_{(HHOHPHNH)} + 0H_2 \rightarrow 1N_2 + 1O_2 + 1P_2 + 5H_2$$

$$2AM_{(NLiBHCHHB)} + 8H_2 \rightarrow 4BH_3 + 2CH_4 + 2LiH + 1N_2$$

The above-described settings were chosen to restrict the test sets to small single-reference main group molecules, as these routinely applicable (black-box type) reference methods like CCSD(T) are known to yield results of high accuracy. A large number of initial structures was generated because we expected a lot of optimizations not to converge and aimed at a set size of about 100 entries. Of course, if the size of the test set is of importance, one can alter the scheme to generate random systems as long as enough systems have passed the following tests. Systems with a fixed number of eight atoms were chosen because a variable number of atoms seemed to be an unnecessary complication in this first study. Furthermore, eight-atom systems are just routinely manageable with high-level methods. Main group elements outweighing chlorine were also regarded as an unnecessary complication and the noble gases on the other side were viewed as too uninteresting for our purpose. Transition metals were excluded because even high-level coupled-cluster methods are known to have problems with the complex electronic situations in transition metal compounds, so the generation of a high-quality reference itself would be too demanding. The use of a cube for the initial starting geometries was believed as simple and sufficient after a few initial tests. An interesting alternative would be a "kick" procedure comparable to the one of Saunders,[18] which was also used by Bera et al.[14] for their isomer-minima search. The choice of the GGA for the optimization and initial screening is further to our disposal and might introduce a subtle bias toward single-reference cases. The SCF and structure optimization iteration limits were chosen to be quite high to avoid the introduction of additional biases through these limits. The goal of the applied threshold (gaps, diagnostics) was to completely screen out cases with very complex electronic structures, where CCSD(T) cannot be assumed to yield accurate reference values.

The basis set converged T1-diagnostic is known to be a reliable indicator for the inherent importance of nondynamical electron correlation, and it is furthermore known that multireference effects are negligible for T1-diagnostic values of less than 0.02.[19] The convergence of the T1-diagnostic with respect to the basis set improvement is a measure for the coupling between the one- and N-electron basis sets, which is likely to be small for the single-reference molecules we aim at. In any case, as the T1-diagnostic decreases with larger one-particle basis sets (because more orbital relaxation is necessary for the more incomplete smaller basis sets), we can assume the T1-diagnostic already at the RCCSD/cc-pVDZ level to be a good threshold for the rejection of multireference cases. While the T1-diagnostic can be seen as an average measure over the whole molecule (for which contributions from a small problem area can be swamped

by that of the well described rest of the molecule), the D1-diagnostic is designed to yield a large value for systems with only a small problematic area[20] and is used as a complement for the T1-diagnostic here. We found that a cutoff of 0.10 for the D1-diagnostic leads to results that are consistent with the 0.02 value for the T1-diagnostic. We would like to mention that the T1- and D1-diagnostics are not foolproof and that it could be a possible refinement for our screening approach to use the percentage of atomization energy accounted for by the (T) component as a predictor for higher-order correlation effects, as suggested by Karton et al. (which nevertheless seems to lead to quantitatively comparable results in almost all cases).[21]

As a general reaction scheme, the decomposition of the AMs to small (mainly closed-shell) molecules was chosen. As already mentioned, we think that this is closer to everyday chemistry than the commonly employed atomizations which furthermore introduce a bias for the correct description of the molecule compared to the (mostly) open-shell atoms. However, we have included reference values for the atomization of the AMs in the Supporting Information. While in the present form our approach cannot deal with barrier heights, future benchmarks will include charged compounds with an odd number of electrons to allow investigations of, for example, self-interaction errors in DFT.

To summarize, both sets consist of reactions with small single-reference main group molecules, roughly three-quarters of them being closed-shell cases. The first set was generated with a general chemistry motivated "931" element distribution. The second one resembles element distribution in organic molecules more closely. Despite being single-reference main group molecules, the generated AMs show a large structural diversity with interesting bonding features, and many of them are chemically very unusual (see Figures 1 and 2; example AMs are discussed in section 4.1).

3. Computational Details

After the initial DFT geometry optimizations at the PBE/TZVP level[10,16] and RCCSD/cc-pVDZ[17] single point calculations for the T1- and D1-diagnostics, the reference reaction energy values were produced for the remaining systems as follows.

The RCCSD(T) complete basis set (CBS) limit correlation energy was extrapolated according to the method of Halkier et al.[22] using cc-pVTZ and cc-pVQZ[17] data points and added to the extrapolated SCF energy.[23] Core correlation/polarization effects were estimated by using the relative energy differences of valence RCCSD(T)/cc-pVTZ and all-electron RCCSD(T)/cc-pCVTZ[17] calculations as a correction to the valence RCCSD(T)/CBS results. Our final reaction energies thus correspond to all-electron CCSD(T)/CBS estimates.

These values serve as reference data for DFT single-point calculations with the S-VWN,[24,25] B97-D,[26] B-P86,[27] B-LYP,[28,29] PBE,[10] TPSS,[30] PBEh[31] (also known as PBE0 or PBE1PBE), BH-LYP,[32] B3-PW91,[33] B3-LYP,[33,34] B2-PLYP,[35] B2GP-PLYP-D,[36] BMK,[37] MPW1B95,[38] PW6B95,[39] M05,[9,40] M05-2X,[9,41] M06,[9,42] M06-L,[9,43] and M06-2X[9,42] density functionals using TZVPP[16] and QZVP[44] Gaussian AO basis sets. In standard notation these are written
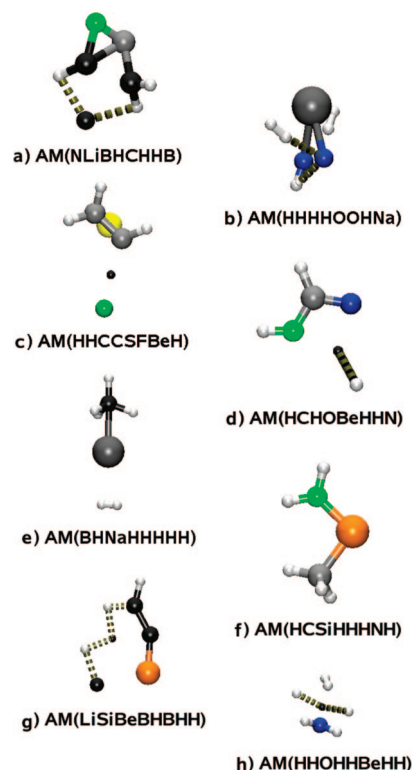

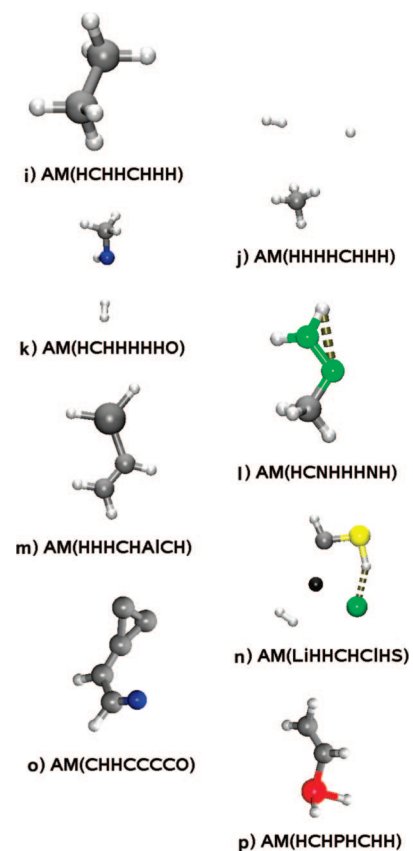**Figure 1.** Examples for artificial molecules (AMs) from the MB08-931 benchmark set.

Figure 2. Examples for artificial molecules (AMs) from the MB08-ORG benchmark set.

as (5s2p1d)/[3s2p1d] and (7s3p2d1f)/[4s3p2d1f] for hydrogen, (11s3p)/[5s3p] and (15s6p2d1f)/[6s4p2d1f] for lithium, (11s4p)/[5s3p] and (15s7p2d1f)/[7s4p2d1f] for beryllium,

"Mindless" DFT Benchmarking

J. Chem. Theory Comput., Vol. 5, No. 4, 2009 **997**

(11s6p2d1f)/[5s3p2d1f] and (15s8p3d2f1g)/[7s4p3d2f1g] for boron to fluorine, (14s8p1d)/[5s4p1d] and (20s12p4d2f)/[9s5p4d2f] for sodium, (14s8p1d)/[5s4p1d] and (20s12p5d2f)/[9s5p5d2f] for magnesium, and (14s9p2d1f)/[5s5p2d1f] and (20s14p4d2f1g)/[9s6p4d2f1g] for aluminum to chlorine.

All DFT treatments of open-shell species were performed unrestricted. For the local and semilocal functionals and for the second-order perturbation correction of the double-hybrids, the RI approximation for two-electron integrals[45,46] has been used. For all functionals except S-VWN, PBEh, TPSSh, BH-LYP, B3-PW91, BMK, MPW1B95, M05, M05-2X, PW6B95, M06, M06-L, and M06-2X, the DFT-D[26] correction for London dispersion energy was applied (added suffix "-D"). Its impact is small to moderate for the tested functionals (for a more detailed discussion, see below) mainly because relatively small eight-atom systems have been considered. All coupled-cluster calculations were done using Molpro 2006.1,[47] for the MPW1B95 and B3-PW91 calculations, we employed Gaussian 03,[48] and the PW6B95, M05, M05-02X, M06, M06-L, and M06-2X calculations were carried out with NWChem 5.1.[49] For all other calculations we used Turbomole 5.10.[50,51]

4. Results and Discussion

4.1. Benchmark Set Generation. For the MB08-931 example set, 162 (out of 300) geometry optimizations converged, but four of them had a HOMO/LUMO gap smaller than 0.5 eV. From the remaining 75 closed-shell and 83 open-shell systems, all closed-shell and the first 25 randomly generated open-shell systems were chosen. Twelve of these systems were identical to others, four had substantial multireference character, and for one system the RCCSD(T)/cc-pVQZ calculation did not converge. As a result, the first set contains 83 entries, with 21 (25%) open-shell cases, ranging in size from 11 to 44 electrons.

For the MB08-ORG example set, 176 (out of 300) geometry optimizations converged, but for three of them the RCCSD(T)/cc-pVDZ calculations did not converge, and three of them had a HOMO/LUMO gap smaller than 0.5 eV. From the remaining 82 closed-shell and 88 open-shell systems, all closed-shell and the first 27 randomly generated open-shell systems were chosen. Eleven of these systems were identical to others, six had substantial multireference character, and for eight systems the all-electron RCCSD(T)/cc-pCVTZ calculation did not converge. As a result, the second set contains 84 entries, with 19 (23%) open-shell cases, ranging in size from 12 to 50 electrons.

Two AMs of the organic set were also generated within the MB08-931 set (MB08-931 entries 12 and 49), so that the merging of MB08-931 and MB08-ORG leads to a benchmark set with 165 entries that is dubbed in the following (and for future reference) MB08-165. This set contains 39 open-shell systems of which nine have a significant amount of spin contamination (>0.03 deviation from the expected $\langle \hat{S}^2 \rangle$ value of 0.75 at the HF-level).

Not surprisingly, after the generation and selection process, the final element distributions turned out to be different from the one used as guideline. The initial and resulting distribu-tions are both given in Table 1. For the MB08-931 set no element is clearly underrepresented (less than 75% of the guideline distribution), while only aluminum and sulfur are clearly overrepresented (more than 150% of the guideline distribution). For the organic set, Li, Na, and Cl atoms are underrepresented and beryllium was completely screened out, while only boron is clearly overrepresented. The resulting distributions thus reflect the finite size of the sets, but another factor that also seems to be of influence is the element-specific different complexity of possible interactions. For instance, beryllium is found in seven out of the 300 initially generated "organic" structures, whereof three could not be optimized and the other four are skipped as open-shell systems, so that none is found between the finally selected ones. It is important to notice that the resulting distributions nevertheless nicely follow our initial intentions for the composition of the two example benchmarks as "931" and "ORG" sets.

Figures 1 and 2 show eight systems as examples from each test set. Both sets contain a number of chemically reasonable systems like e, f, and l, but while the organic set includes some "conventional molecules" like i (ethane) and k (methanol/H_2), nearly 30% of the 931 set are made up by fragmented systems like h and j. A handful of systems in both sets are quite complex like d and o, and a good part of both sets consists of chemically very unusual systems, like a, g, and n. The coordinates of all systems can be found in the Supporting Information.

4.2. Benchmark Calculations. *Overview and Discussion of the Reference Values.* For the given single-reference main group systems, already our valence CCSD(T)/CBS reference data can be assumed to be of high accuracy. In addition, because DFT functionals include core-correlation, we added an estimate for these effects to our reference data. It is based on the relative energy difference between valence RCCSD(T)/cc-pVTZ and all-electron RCCSD(T)/cc-pCVTZ calculations. As expected, this difference is small (about 1 kcal/mol on average with a maximum of 5 kcal/mol, compared to mean absolute reaction energies of 117 kcal/mol). For a typical density functional, it has little impact on the MAD if valence or all-electron data are taken as reference (e.g., a difference of less than 0.5 kcal/mol for PBE-D/QZVP).

The reference reaction energies should be briefly discussed first. The 931 set contains reaction energies between −199 and 434 kcal/mol and the organic set contains reaction energies between −571 and 302 kcal/mol. Opposed to the commonly employed atomization benchmarks that exclusively consider endothermic reactions, our data contain almost evenly around zero distributed reaction energies (see Figure 3). This means that without any applied bias in this direction our algorithm has generated simultaneously very stable and very unstable AMs. This is a greatly appreciated feature for a general thermodynamic benchmark set for chemistry where endothermic as well as exothermic reactions are of interest.

Overview of the DFT Data. This paragraph gives a short overview of the tables and figures with DFT data that are discussed in detail in the next paragraph.
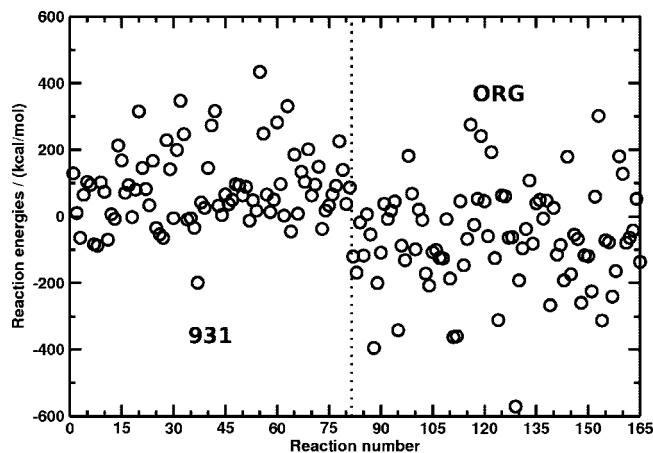
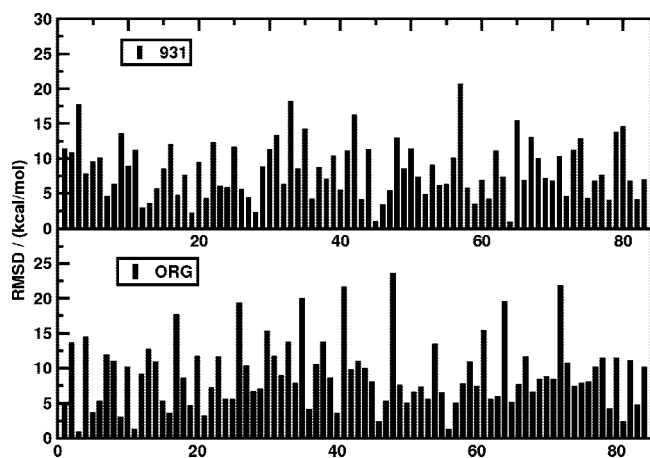**Figure 3.** Distribution of reaction energies of both benchmark sets.

Figure 4. Rmsd averaged over 11 density functionals with the QZVP basis (see the text) for all reactions of both benchmark sets.

Figure 4 shows the root-mean-square deviation (rmsd) over 11 density functionals (S-VWN, PBE-D, B97-D, TPSS-D, PBEh, B3-LYP-D, M05, M05-2X, BMK, MPW1B95, and B2PLYP-D with the QZVP basis set) for the different entries of the two benchmark sets. This analysis is used to spot and check exceptionally problematic cases in the test sets, but it can clearly be seen that both sets are free from sizable outliers.

The mean absolute deviation (MAD), mean deviation (MD), rmsd, and the error spread (Δ, difference of maximum and minimum deviations) for representative density functionals (and for comparison also for HF and some coupled-cluster approximations) for both sets and the combined set are presented in Table 2. Low-order perturbation theory (i.e., MP2) has also been considered, but as expected, this and related methods suffer from spin-contamination in some of the open-shell systems. This makes a comparison with the DFT methods difficult, and therefore, we omit these methods in the present work. The MADs for DFT are also shown graphically in Figure 5. These statistical data are used for the evaluation of the different functionals and to find indicators for the example benchmark set quality. Additional information can be extracted from the histograms in Figure 7, where the error distribution for the functionals with the

lowest overall MAD (M06, MPW1B95, BMK, M06-2X, PW6B95, and B2-PLYP-D) is shown. To summarize the results, Table 3 shows DFT and WFT methods in order of their accuracy for the combined MB08-165 set. A complete listing of all results can be found in the Supporting Information.

Discussion of the DFT Data. Turning to Table 2, we first note the very similar performance of all tested density functionals for the two different test sets. It can therefore be concluded that a specific element distribution does not have a substantial impact on the difficulty of the AMs. This clearly supports previous experience that the main distinction for modern electronic structure methods is between main group chemistry and transition metal compounds. Methods that perform well for organic chemistry are usually also applicable to general main-group systems.

In Figure 6 we plotted the relative MAD per element in order to investigate possible element-specific characteristics of the observed errors. These values were obtained by dividing the absolute reaction energy errors between the atoms of the AM according to their number, taking the average over all reactions and 11 functionals, weighting this value with the element occurrence in the set, and dividing it by the corresponding value for hydrogen (which is set to unity in this way). Values larger (smaller) than unity thus indicate higher (smaller) "difficulty" for the corresponding element. While not much can be concluded for the MB08-931 set, it looks as if oxygen is rather problematic (presumably because of the open-shell reaction product O_2) in the MB08-ORG set. A trend can be observed for the second-row where the errors increase with atomic number. A possible reason why sodium is a particularly easy case in both sets could be that it only occurs in simple ionic structures. We expect that a similar error analysis would be extremely helpful for semiempirical approaches where often atom-specific parameters are used. The finding that DFT does not produce very pronounced elemental error distributions is in agreement with its fundamental "ab initio"-like character.

The second conclusion that can be drawn from Table 2 is that albeit basis set incompleteness effects are less severe for DFT than for WFT methods, a thorough analysis should be made at the QZVP one-electron basis set level. Note that the CCSD(T)/cc-pVDZ level of theory (which is quite often used in the literature) performs a lot worse than any tested DFT approach (except LDA)! On the other hand, the extrapolated CCSD(T)/cc-pV(DT)Z data are comparable even to CCSD(T)/cc-pVQZ, which seems to be an important result for future generation of similar reference data for larger systems.

In some cases the now well-established DFT-D method[26] to account for nonlocal London dispersion effects has been applied, which is indicated by the suffix "-D" after the functional name. The MAD values with and without the dispersion correction for selected functionals are given in Tables 2 and 3 for comparison. Because of the relatively small size of the systems studied, dispersion effects are moderate. Except for PBE, application of this correction always leads to a lowering of the MAD ranging from 0.7 kcal/mol (TPSS) over 1.7 kcal/mol (B2-PLYP) to 2.2 kcal/

Table 2. Statistical Performance Indicators (in kcal/mol) for the MB08-931, MB08-ORG, and Both (MB08-165) Benchmark Sets

| | | MB08-931 QZVP | | MB08-ORG QZVP | | MB08-165 | | | | | | | |
| | | | | | | TZVPP | | | | QZVP | | | |
entry	method	MAD	rmsd	MAD	rmsd	MAD	MD	rmsd	$\Delta_{Min-Max}$	MAD	MD	rmsd	$\Delta_{Min-Max}$
						DFT							
1	S-VWN	22.4	28.7	17.9	24.4	19.4	14.1	25.5	134.9	20.3	16.2	26.8	138.5
2	Slater−Dirac	16.2	20.9	14.2	18.6	14.9	7.6	19.4	112.5	15.1	9.9	19.8	95.9
3	PBE-D	10.0	12.4	9.6	12.0	9.6(9.2) [a]	2.8	12.4	86.0	9.8	4.9	12.2	69.6
4	B97-D	9.7	12.3	10.9	14.6	11.5	−5.7	15.2	92.3	10.3	−3.5	13.5	87.1
5	TPSS-D	9.2	11.8	10.7	13.7	10.4(11.1) [a]	−2.7	14.1	79.5	10.0	−0.7	12.8	66.3
6	PBEh	8.0	10.0	9.1	12.0	9.3	−2.9	12.7	86.9	8.6	−0.5	11.1	69.6
7	B3-LYP-D	6.2	8.0	7.0	9.4	7.8(10.0) [a]	−4.8	10.5	55.4	6.6	−2.7	8.8	49.1
8	M05	6.4	8.0	7.5	9.6	8.0	−3.7	10.1	58.7	7.0	−1.3	8.9	57.6
9	M05-2X	7.8	11.3	5.7	8.2	6.3	1.2	8.9	70.0	6.8	3.5	9.9	72.2
10	BMK	5.1	6.3	5.9	8.0	6.5	−2.8	8.8	54.4	5.5	−1.1	7.3	50.9
11	MPW1B95	5.6	6.9	5.6	7.3	6.1	−1.9	8.2	54.7	5.6	0.1	7.1	40.8
12	M06	6.3	8.1	5.6	7.0	7.2	−4.3	9.6	48.5	6.0	−2.2	7.6	40.4
13	M06-2X	5.0	7.1	4.6	6.0	4.6	2.1	6.2	44.6	4.8	3.9	6.6	33.1
14	PW6B95	4.5	5.8	5.1	6.5	5.5	−2.8	7.7	48.6	4.8	−0.7	6.1	33.9
15	B2-PLYP-D	3.7	4.8	4.5	6.1	6.2(7.9) [a]	−5.6	8.4	38.7	4.1	−2.6	5.5	29.7
						WFT							
16	HF	24.0	29.6	28.0	37.0	26.6	−23.0	33.6	133.2	26.1	−22.4	33.7	158.7

entry	method	MAD	MD	rmsd	$\Delta_{Min-Max}$
17	CCSD/cc-pVQZ	5.6	−5.2	7.7	34.3
18	CCSD(T)/cc-pVDZ	14.2	−13.9	18.2	78.3
19	CCSD(T)/cc-pVTZ	5.5	−5.3	7.0	24.6
20	CCSD(T)/cc-pVQZ	2.6	−2.5	3.4	11.8
21	CCSD(T)/cc-pV(DT)Z	2.8	−2.4	3.8	17.8

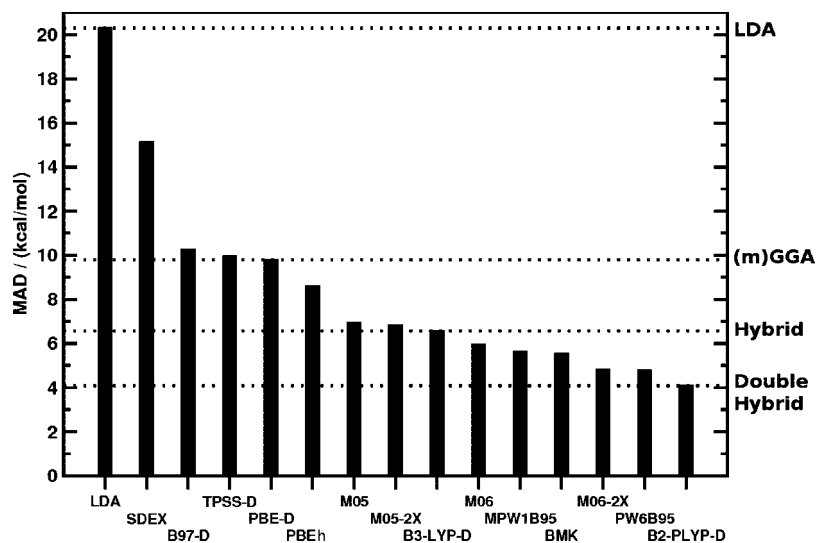[a] Values without DFT-D[26] dispersion correction in parentheses.

Figure 5. MAD of different density functionals for the MB08-165 set.

mol (B3-LYP). This is consistent with previous experience about the implicit account of dispersion effects by the electronic part of the functionals.[52]

Comparing the performance of the different functionals (Table 2 and Figure 5), one finds that the observed accuracy fits nicely to the "Jacob's ladder" metaphor of Perdew et al.[53] for the ordering of density functionals based upon the information of electron density they use. Suited on the first rung (local density approximation, LDA), the S-VWN functional yields quite bad results, with MADs of 22.4 (MB08-931 set), 17.9 (MB08-ORG set), and 20.3 kcal/mol

(MB08-165 set), respectively. Note that exchange-only LDA (Slater−Dirac exchange) produces significantly better results than S-VWN.

The biggest improvement is found when climbing up to the second rung (generalized gradient approximation, GGA). Our examples PBE-D (MADs of 10.0, 9.6 and 9.8 kcal/mol) and B97-D (9.7, 10.9 and 10.3 kcal/mol) yield similar results, although we note a slightly better performance of the nonempirical (PBE) compared to the empirically parametrized functional. The next step up the ladder to the third rung (meta-GGA) with TPSS-D (9.2,
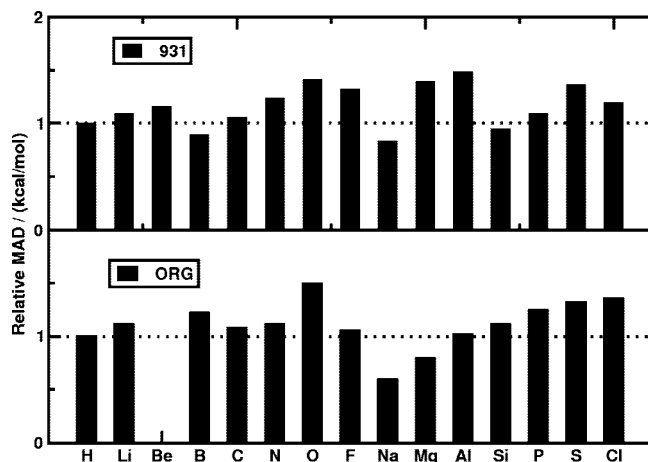
Figure 6. Relative MAD as distributed over the elements (for details, see the text) for the MB08-931 and the MB08-ORG set.

Table 3. Performance (statistical descriptors in kcal/mol) of DFT and WFT[a] Methods for the Combined MB08-165 Benchmark Set

method	rmsd	MAD
CCSD(T)/cc-pVQZ	3.38	2.61
CCSD(T)/cc-pV(DT)Z	3.80	2.82
B2-PLYP-D/QZVP	5.49	4.09
PW6B95/QZVP	6.14	4.78
M06-2X/QZVP[a]	6.59	4.82
B2-PLYP/QZVP	6.63	5.12
CCSD(T)/cc-pVTZ	7.01	5.45
MPW1B95/QZVP	7.13	5.64
BMK/QZVP	7.26	5.53
M06/QZVP[a]	7.60	5.96
CCSD/cc-pVQZ	7.69	5.64
B3-LYP-D/QZVP	8.75	6.58
M05/QZVP	8.88	6.95
B3-PW91/QZVP	9.31	7.04
M05-2X/QZVP	9.89	6.83
B3-LYP/QZVP	10.61	8.19
TPSSh/QZVP	10.98	8.38
PBEh/QZVP	11.09	8.62
B-P86-D/QZVP	11.50	9.27
B-LYP-D/QZVP	12.14	9.23
PBE-D/QZVP	12.24	9.81
TPSS-D/QZVP	12.85	9.97
B97-D/QZVP	13.54	10.27
BH-LYP/QZVP	16.66	12.03
M06-L/QZVP[a]	17.17	11.13
CCSD(T)/cc-pVDZ	18.24	14.25
S-VWN/QZVP	26.78	20.31
HF/QZVP	33.68	26.07

[a] Only valence electrons have been correlated. [b] Two (four for M06-L) entries have been omitted because SCF convergence is lacking.

10.7, and 10.0 kcal/mol) shows no advancement at all, which is in agreement with the results of two recent studies[5,12] on "real" molecules.

The inclusion of nonlocal information from occupied orbitals by adding Fock-exchange on the fourth rung (hyper-
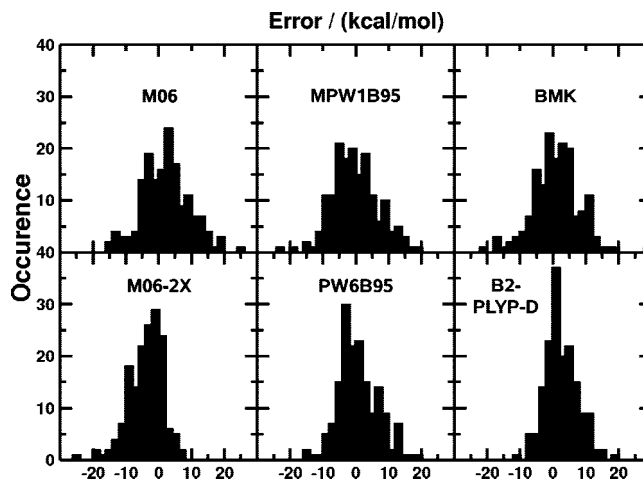

Figure 7. Error distribution for the density functionals with the lowest MADs. The binning interval is 2 kcal/mol.

GGA or hybrids) with PBEh (8.0, 9.1 and 8.6 kcal/mol) and B3-LYP-D (6.2, 7.0, and 6.6 kcal/mol) represents again a big step toward higher accuracy. However, less is gained from the GGA level to hybrids than from LDA to GGAs, i.e., it becomes more and more difficult to climb up the ladder.

Turning to the performance of the hybrid-meta-GGAs (which are mostly represented by highly parametrized forms), one finds a mixed picture. For the M05 and M06 family of functionals, it was possible to incorporate dispersion effects through flexibility and parametrization, and therefore no dispersion correction was applied for these functionals. Compared to conventional hybrids like B3-LYP-D, the MAD increases even slightly for M05 and M05-2X. BMK and MPW1B95 perform a bit better and the best results in this class are found for the PW6B95 (4.5, 5.1, and 4.8 kcal/mol) and M06-2X (5.0, 4.6, and 4.8 kcal/mol) functionals. However, in the SCF calculations with the M06 functionals, we noted in several cases convergence problems, and a few AMs therefore had to be omitted. This behavior is known in the literature.[54] Note also how well MAD, rmsd, and maximum errors provide the same picture about performance, which is a clear indication for the statistical quality of the sets.

The B2-PLYP-D and B2GP-PLYP-D double-hybrid functionals include nonlocal information about correlation, which is absent in any of its competitors, via the second-order perturbation correction. They can therefore be classified as true (but empirical) fifth-rung density functionals. This view is indeed supported by the very good performance of B2-PLYP-D for all benchmark sets with MADs of 3.7, 4.5, and 4.1 kcal/mol, which is again an appreciable step into the right direction. B2GP-PLYP-D that contains more nonlocal information compared to B2-PLYP-D (65 vs 53% Fock-exchange and 36 vs 27% MP2-type correlation) shows indistinguishably good performance. Note that also the maximum errors for the double-hybrids are lower than for any other tested functional (28−30 vs 33.1 kcal/mol for the best other, which is M06-2X).

While the conventional DFT approaches are shown to have reached the basis set limit already at the TZVPP level, the

"Mindless" DFT Benchmarking

J. Chem. Theory Comput., Vol. 5, No. 4, 2009 **1001**

double-hybrids substantially benefit from a further basis set enlargement because of the included perturbation correction.

Additional information about the performance of the functionals with the lowest MADs (M06, MPW1B95, BMK, M06-2X, PW6B95, and B2-PLYP-D) can be taken from Figure 7, which displays error distributions in a 5 kcal/mol interval. The hybrid-GGAs M06, MPW1B95, and BMK show a behavior very similar to each other, while the M06-2X and the PW6B95 functionals seem to be slightly superior. Nevertheless, they cannot reach the accuracy of B2-PLYP-D, which shows the sharpest and most well-behaved Gaussian-shape error distribution.

This view is further supported by Table 3, where DFT and WFT methods are shown in order of their accuracy for the combined MB08-165 set. As additional functionals, BH-LYP, B3-PW91, and M06-L are included here for comparison. On the basis of the MAD, we tried to identify groups of methods with similar performance. Perusing Table 3, one finds that B2-PLYP-D outperforms the highly parametrized meta-hybrid GGAs, which in turn are even better than CCSD(T) with a sizable AO basis. In the next group, we find three other meta-hybrids that are better than B3-LYP, even when it is corrected for missing dispersion effects. Two almost nonempirical hybrids (TPSSh and PBEh) and uncorrected B3-LYP follow, which are only slightly better than typical GGAs. Here, the old B-P86 seems to be most robust, although we note only a small (probably statistically insignificant) difference between the best and worst (B-P86 and B97-D) GGAs. This seems to be important in many solid-state or surface-science applications, where semilocal functionals are dominant.

5. Concluding Remarks

We present a diversity-oriented approach for the preferably unbiased generation of thermochemical benchmark data. We have addressed the "selection" problem through the random generation of "artificial molecules" (AMs) in accordance with systematic and controllable generation specifications. The most appealing feature of our approach in comparison to existing test sets is the opening of chemical space, which produces electronically demanding cases in an unforeseeable manner. It is hoped that these sets can provide a clear, unbiased, statistically significant, and comprehensive picture about the performance of electronic structure methods at a minimum of computational effort.

For the generation of two example sets, we have chosen several strict constraints (that could have been released or further "randomized") to keep the examples simple and to ensure a high accuracy of the reference data. We ended up with two test sets with chemically very diverse and unusual structures, despite the single-reference main group nature of all generated AMs. For the evaluation of the benchmark sets, we performed DFT calculations, applying several different density functionals, and compared them to highly accurate CCSD(T)/CBS(all electron) estimates. We have found a systematic improvement along "Jacob's ladder" for most of the tested functionals, which is a promising first result for our "mindless" approach and supports the significance of Perdew's metaphor.

Concerning individual density functionals and wave function based methods, the following main conclusions can be drawn:

1. Even for these relatively small systems, B3-LYP is not better than other simple hybrids. Only when the relatively large effect of the dispersion correction (about 2 kcal/mol lowering of the MAD) is included does it outperform the modern and highly parametrized M05 meta-hybrid GGAs that seem to represent relatively inaccurate points in the M0X-parameter space (this also holds for M05-L). However, the performance of B3-LYP for larger molecules is known to deteriorate substantially, and since more accurate and robust alternatives are available, we cannot recommend its application in general.

2. Although no atomizations have been considered, the PBE family of functionals (including TPSS and hybrid versions) performs slightly worse than B3-LYP. This is contrast to some recent applications in, for example, organic chemistry where they outperform B3-LYP and is likely a result of the relatively small size of the systems. As known from previous work on van der Waals complexes, the PBE(h) functionals simulate to some extent medium-range correlation effects related to short-range dispersion[55] that seem to be of some importance even in eight-atom systems.

3. Concerning the meta-hybrid GGAs, the M06 parametrization seems to be a clear step forward, although we also note the good performance of the less parametrized PW6B95 functional. However, meta-GGAs or meta-hybrid-GGAs do not in general represent improvements compared to their GGA counterparts, as seen for the PBE(h)/TPSS(h) pairs.

4. The best functionals contain a relatively large fraction of nonlocal Fock-exchange (about 40−60%) and adding more seems to improve the results (the only exception to this rule is BH-LYP). This is consistent with our choice that only main-group systems have been considered. By construction the sets do not contain cases with strong correlation effects (for which semilocal functionals should work better) but seem to be dominated by self-interaction error including cases (which are better described with "high-X" functionals). This bias is not rooted in the basic ansatz of "mindless" benchmarking but a result of the choice of the necessary (single) reference CCSD(T) method as a basis for evaluation.

5. The dispersion-corrected double-hybrid functionals, despite having only three (two "electronic") adjustable parameters (not counting the global, "black-box"-type dispersion correction), outperform all other tested DFT approaches by a significant margin and yield results in between CCSD(T)/cc-pVTZ and CCSD(T)/cc-pVQZ quality. The fact that two different parametrizations (B2-PLYP and B2GP-PLYP) provide similar rmsd values indicates the robustness of the general idea. This could not be expected, because standard MP2 often breaks down in electronically complicated situations. This supports the previous claim that the Kohn−Sham orbitals greatly improve the behavior of the perturbation treatment used in the double-hybrids.

6. The more approximate coupled-cluster treatments provide results for the MB08-165 set, in full agreement with common knowledge about their behavior: the (T)-correction is of utmost importance (e.g., CCSD/cc-pVQZ is not better

than the best meta-hybrids), single DZ (or even TZ) AO basis set computations are a waste of computer time, and two-point extrapolations provide results that improve like one step in cardinal number (i.e., DT-extrapolation yields results comparable to a single QZ calculation). This further supports the significance of the proposed test set.

All in all, it is hoped that this work will help to pave the way for improved electronic structure methods in chemistry. Future work in this promising direction will explore in particular larger AMs than the eight-atom systems considered herein.

Supporting Information Available: Cartesian coordinates of all AMs and relative energies at various CCSD(T) levels and for all described density functionals. This material is available free of charge via the Internet at http://pubs.acs.org

References

(1) Cioslowski, J., Ed. *Quantum Mechanical Prediction of Thermochemical Data*; Kluwer: Dordrecht, 2001; 1 ff.

(2) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764–7776.

(3) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 8996–8999.

(4) Zhao, Y.; Gonzalez-Garcia, N.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 2012–2018.

(5) Grimme, S.; Steinmetz, M.; Korth, M. *J. Org. Chem.* **2007**, *72*, 2118–2126.

(6) Jurecka, P.; Sponer, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

(7) Johnson, E. R.; Mori-Sanchez, P.; Cohen, A. J.; Yang, W. *J. Chem. Phys.* **2008**, *129*, 204112.

(8) Schwabe, T.; Grimme, S. *Acc. Chem. Res.* **2008**, *41*, 569–579.

(9) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157–167.

(10) Perdew, J. P.; Burke, K.; Enzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(11) Csonka, G. I.; Ruzsinszky, A.; Tao, J.; Perdew, J. P. *Int. J. Quantum Chem.* **2005**, *101*, 506–511.

(12) Grimme, S. *J. Phys. Chem. A* **2005**, *109*, 3067–3077.

(13) Fink, T.; Reymond, J.-L. *J. Chem. Inf. Model.* **2007**, *47*, 342–352.

(14) Bera, P. B.; Sattelmeyer, K. W.; Saunders, M.; Schaefer, H. F., III; Schleyer, P. v. R. *J. Phys. Chem. A* **2006**, *110*, 4287–4290.

(15) Clayden, J.; Greeves, N.; Warren, S.; Wothers, P. *Organic Chemistry*; Oxford University Press: Oxford, 2001; p 12.

(16) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829–5835.

(17) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(18) Saunders, M. *J. Comput. Chem.* **2004**, *25*, 621–626.

(19) Lee, T. J.; Taylor, P. R. *Int. J. Quantum Chem.* **1989**, *S23*, 199–207.

(20) Lee, T. J. *Chem. Phys. Lett.* **2003**, *372*, 362–367.

(21) Karton, A.; Rabinovich, E.; Martin, J. M. L.; Ruscic, B. *J. Chem. Phys.* **2006**, *125*, 144108.

(22) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243–252.

(23) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Olsen, J. *Chem. Phys. Lett.* **1999**, *302*, 437–446.

(24) Slater, J. C. *Phys. Rev.* **1951**, *81*, 385–390.

(25) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200–1211.

(26) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

(27) Perdew, J. P. *Phys. Rev. B* **2000**, *33*, 8822–8824.

(28) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(29) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

(30) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.

(31) Adamo, C.; Barone, V. J. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

(32) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

(33) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(34) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

(35) Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108.

(36) Karton, A.; Tarnopolsky, A.; Lamere, J.-F.; Schatz, G. C.; Martin, J. M. L. *J. Phys. Chem. A* **2008**, *112*, 12868.

(37) Boese, A. D.; Martin, J. M. L. *J. Chem. Phys.* **2004**, *121*, 3405–3416.

(38) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6908–6918.

(39) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 5656–5667.

(40) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Phys.* **2005**, *123*, 161103.

(41) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364–382.

(42) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

(43) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.

(44) Weigend, F.; Furche, F.; Ahlrichs, R. *J. Chem. Phys.* **2003**, *119*, 12753–12762.

(45) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *242*, 652.

(46) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119–124.

(47) Werner, H. J.; et al. MOLPRO version 2006.1, a package of ab initio programs; University College Cardiff Consultants, Ltd.: Cardiff, 2006. See also http://www.molpro.net (accessed Feb 3, 2009).

(48) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone,

V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R. Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J. Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghava-chari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision E.01*; Gaussian, Inc.: Wallingford CT, 2004.

(49) Bylaska, E. J.; et al. *NWChem, A Computational Chemistry Package for Parallel Computers, Version 5.1*; Pacific Northwest National Laboratory: Richland, 2007.

(50) Ahlrichs, R.; et al. *TURBOMOLE V5.10*; University of Karlsruhe: Karlsruhe, 2008. See also: http://www.turbomole. com (accessed Feb 3, 2009).

(51) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kömel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169.

(52) Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397–3406.

(53) Perdew, J. P.; Ruzsinszky, A.; Tao, J.; Staroverov, V. N.; Scuseria, G. E.; Csonka, G. *J. Chem. Phys.* **2005**, *123*, 062201.

(54) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 1849–1868.

(55) Grimme, S. *Angew. Chem., Int. Ed.* **2006**, *45*, 4460−4464.