# Successes and Failures of DFT Functionals in acid/base and redox reactions of organic and biochemical interest

*Pedro J. Silva*[1] *and Maria João Ramos*[2*]

[1]REQUIMTE, Fac. de Ciências da Saúde, Univ. Fernando Pessoa, Rua Carlos da Maia, 296, 4200-150 Porto-Portugal

[2*]REQUIMTE, Faculdade de Ciências do Porto, Rua do Campo Alegre, 687, 4169-007 Porto – Portugal

mjramos@fc.up.pt

**Abstract:** The performance of 18 different DFT functionals in the prediction of absolute and relative energies of organic and biochemical acid/base and redox reactions was evaluated, using MP2 extrapolated to the complete basis set limit and CCSD(T)/aug-cc-pVTZ energies as benchmark. Absolute reduction energies were predicted with relatively large average errors (2-4 kcal.mol$^{-1}$) except for the best functional, PBE0 (1.3±1.2 kcal.mol$^{-1}$). The DFT predictions of *relativ*e reduction energies afforded mean unsigned errors (2.1 kcal.mol$^{-1}$ for the best functional, PBE0) which, although relatively large, are smaller than those obtained with MP2 with comparable basis sets (3.2 kcal.mol$^{-1}$). This relatively poor performance was mostly due to significant underprediction of the reduction energy of a model disulfide bridge, and overprediction of that of a model quinone, as eliminating these reaction from the test set enables most of the tested hybrid-GGA functionals (B3LYP, B3PW91, B97-1, BHHLYP, PBE0, PBE1PW91, X3LYP), though not meta-GGA or meta-hybrid-GGA functionals, to achieve small mean unsigned errors (<1.5 kcal.mol$^{-1}$). In the acid/base reaction test set, several hybrid-GGA and meta-hybrid GGA functionals (BHHLYP, B97-1, B97-2, X3LYP, M06-2X, as well as the popular B3LYP) yielded small errors (0.8-1.5 kcal.mol$^{-1}$) in the computation of *relative* energies across the tested model reactions. For most of the proton transfers between different acid/base pairs (or electron transfer between redox pairs) tested we could find at least one functional with very high accuracy (error <0.2 kcal.mol$^{-1}$). The reactions involving electron transfer between quinone (or disulfide) and other redox groups stand out as the clearest example of the shortcomings of DFT methods, as the best functionals are most often wrong by 1-3 kcal.mol$^{-1}$.

**Keywords:** DFT; benchmarking; Acid/base reactions; redox potential

## Introduction

The rapid rate of increase of computing power and the development of fast algorithms and user-friendly software packages has turned computational chemistry from a small field restricted to specialists into a large applied field increasingly used by every branch of chemistry. The accurate computation of geometries and energies of molecular species can nowadays be performed using coupled-cluster methods, but their high computational cost and unfavorable scaling properties render them effectively out of reach for routine computations. Density-functional theory (DFT), on the other hand, affords much more favorable scaling and a lower computational overhead and can therefore be used for routine computations and for larger systems (>80 atoms). DFT suffers, however, from a serious drawback: the Hohenberg-Kohn theorem[1], which states the existence of a functional relating the ground-state electronic energy of a non-degenerate system with the electron density, does not describe the mathematical form of this functional. Many different functionals have been developed, either strictly from first-principles and suitable approximations, or with the inclusion of variable amounts of Hartree-Fock exchange and/or suitable parameters derived from experimental data or higher-level computations (reviewed by Sousa et al.[2]). The performance of DFT varies with the chemical system under study, as well with the functional itself. The choice of functional to tackle a chemical problem therefore requires the comparison of the results of DFT functionals with those obtained with costlier higher-level computations. The B3LYP[3-5] functional has been used traditionally as the “default functional” due to its good agreement with experiment in a large variety of chemical reactions, but it has become clear that it is not the best choice for several chemical systems and/or properties[e.g. 2 and references therein]. Moreover, the continuous development of increasingly sophisticated functionals, from the relatively simple local spin density approximation (LSDA) to the gradient-generalized (GGA), hybrid-GGA and meta-GGA functionals now offers a wide choice of functionals to the unwary user, who is confronted with a chemical variation of the “paradox of choice”. Indeed, as the number of possible choices increases, so does the cost of evaluating and choosing one of them, and users therefore tend to revert to traditional options even when superior alternatives are available. Many benchmarking studies of density functionals on different types of chemical systems are already available[2,6-9], but to our knowledge no systematic study of the performance of DFT functionals on the prediction of reaction energies of organic acid/base or redox pairs is available, despite their importance in the study of organic and enzymatic reactivity.

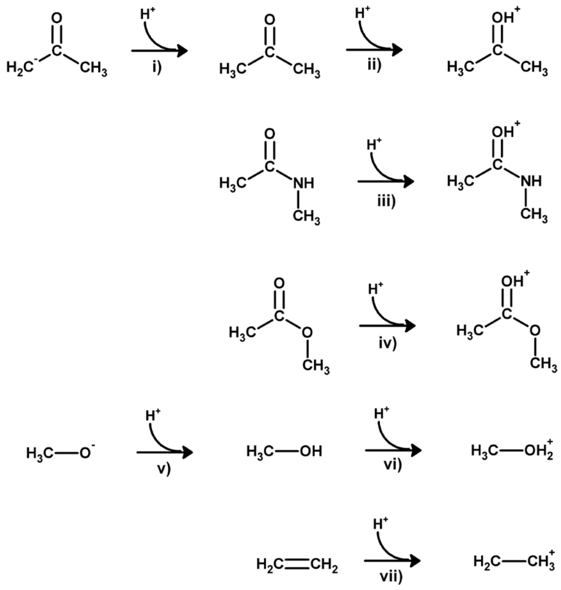
## Computational methods

The geometries of every molecule described were optimized at the MP2 level and with each of the tested density functionals. Autogenerated delocalized coordinates[10] were used in geometry optimizations with a medium-sized basis set, 6-31+G(d), since increasing the basis sets to triple-ζ quality gives very small additional corrections to the geometries while dramatically increasing the computational cost, and the use of such large basis sets for geometry optimization is often considered unnecessary from a computational point of view[11-13]. Accurate DFT energies of the optimized geometries obtained with each density functional were then calculated using the same functional with several triple-ζ quality basis sets: 6-311+G(d,p), 6-311+G(2d,p), 6-311+G(2d,2p), and 6-311+G(3d,2p). CCSD(T) and MP2 single-point energies were computed on the MP2 geometries using 6-31+G(d), aug-cc-pVDZ and aug-cc-pVTZ basis sets and (in the case of MP2), extrapolated to the complete basis set limit as described by Truhlar[14,15]. HF, MP3 and MP4 single-point energies were computed using the MP2-optimized geometries with the basis sets used for the single-point DFT energies. We used three GGA functionals (PBE96[16,17], PBEPW91[16,17] and PW91[16]), eight hybrid-GGA functionals (B3LYP[3,4,5], B3PW91[3,16], B97-1[18], B97-2[19], BHHLYP (50% HF exchange + plus 50% B88[20] exchange, with LYP correlation), PBE0[21], PBE1PW91[16,17], and X3LYP[22]), two meta-GGA functionals (TPSS[23,24] and TPSSm[25]) and five meta-hybrid GGA functionals (TPSSh[26], M06[27], M06-L[28], M06-2X[27] and M06-HF[29]). Computations involving CCSD(T), the B97-1, B97-2 and the M06 and TPSS families of functionals were performed with Gamess(US)[30]. All other computations were performed with the Firefly[31] quantum chemistry package.

Since the computation of gas phase Gibbs free energies through any computational method entails the determination of the electronic energies and zero-point/vibrational energies (ZPVE) of each reactant and each product, the performance of a functional depends not only on the accuracy of the computed electronic energies but also on the quality of the vibrational frequencies provided by the method. In this work, we decided to focus on the electronic reaction energies *only*, to prevent spurious results arising from mutual cancellation of errors in energy/ZPVE and because no ZPVE scaling factors are available for many of the method/basis set combinations tested. Electronic-only acidities were computed as $E_{acid}$-$E_{base}$. Electronic-only redox energies were computed as $E_{reduced\ species}$ - $E_{oxidized\ species}$

Protonation of the acetone carbonyl and deprotonation of acetone α-carbon atoms was selected as model for ketone acidities; N-methylacetamide and methyl acetate were used as generic amide and ester models, respectively. For the redox reaction models, isoalloxazine was used as a model for flavin nucleotide, the oxidation of 1,4-butanedithiol to 1,2-dithiane was selected as model for disulfide bridge formation, and N-hydroxymethylnicotinamide as a model for $NAD^+$. Other models used are self-explanatory (Schemes 1 and 2).

## Results and Discussion



Scheme 1

The organic acid/base model reactions selected for our study are depicted in Scheme 1. They include the protonation of acetonate to acetone, the protonation of the ketone group in acetone, the protonation of amides and esters, the deprotonation of an alcohol, the protonation of the hydroxyl group and the protonation of ethane. The energies of this set of reactions range from -166.7 kcal.mol$^{-1}$ to -388.6 kcal.mol$^{-1}$ at the MP2/CBS // MP2/6-31+G(d) level, and from -168.0 kcal.mol$^{-1}$ to -391.1 kcal.mol$^{-1}$ at the CCSD(T) /aug-cc-pVTZ // MP2/6-31+G(d) level (Table 1). Energies predicted by CCSD(T) /aug-cc-pVTZ lie a few kcal.mol$^{-1}$ below MP2 in all instances. MP3 results deviated appreciably from MP2, and were always predicted to be lower than CCSD(T)/aug-cc-pVTZ. In keeping with the usually observed convergence behavior of the Møller-Plesset series, MP4 results lie between those from MP2 and MP3, and in this set are quite close to the coupled-cluster results.

The geometries computed with density functional theory were generally in good agreement with MP2 geometries (root-mean-squared errors were most often below 0.02 Å). Larger differences were observed in the staggered/eclipsed conformation of a methyl group in: O-protonated ester (B3LYP and B3PW91);

O-protonated amide (all functionals except M06-2X and M06); amide (M06-L, PW91, PBE0, PBE96, PBEPW91, PBE1PW91 and X3LYP); and acetonate (B3PW91). In spite of these differences, all tested functionals correctly predict the relative ordering of the reaction energies in this test set as dictated by MP2/CBS and CCSD(T) (with the exception of M06, which predicts the protonation of an ester to be less favorable than the protonation of acetonate). The reaction energies obtained with the DFT functionals are generally under-estimated relative to the MP2/CBS energies; the most striking exceptions are the protonation energies of methoxide and acetonate computed with the GGA-only functionals (overestimated by >2.5 kcal.mol$^{-1}$ and >1.5 kcal.mol$^{-1}$, respectively) and the protonation energy of acetonate computed with the meta-hybrid M06-HF functional (underestimated by 3.1 kcal.mol$^{-1}$) (see Supporting Information).

Table 1: Computed gas-phase energies of the tested acid/base reactions. All values in kcal.mol$^{-1}$. ZPVE not included.

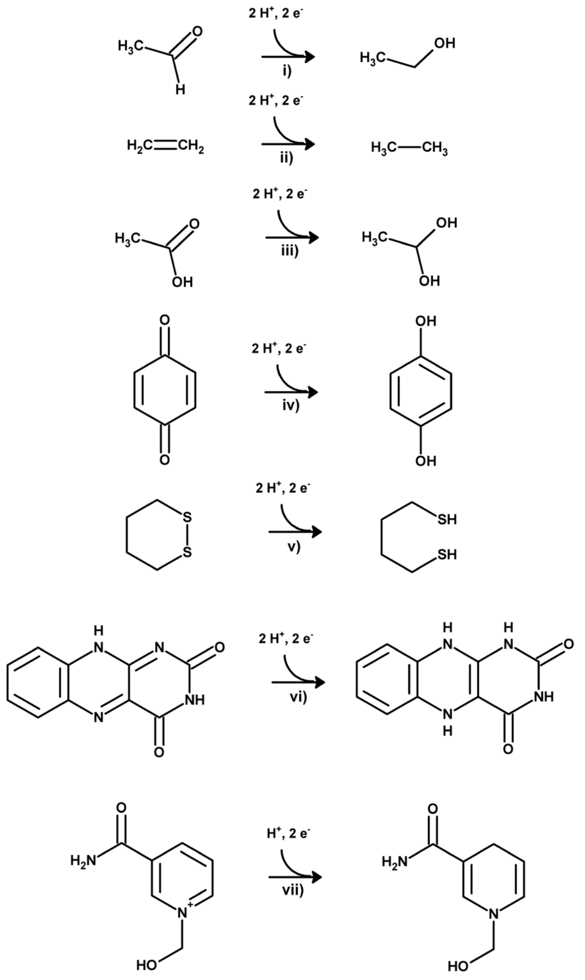
| | Reaction | | MP2/CBS | CCSD(T)/aug-cc-pVTZ |
|---|---|---|---|---|
| i) | $H_3C\text{-}C\text{=}O\ CH_2^- + H^+$ | $\rightarrow (CH_3)_2C\text{=}O$ | -373.8 | -377.2 |
| ii) | $(CH_3)_2C\text{=}O + H^+$ | $\rightarrow (CH_3)_2C\text{=}OH^+$ | -198.7 | -201.3 |
| iii) | amide + $H^+$ | → Protonated amide | -217.4 | -220.2 |
| iv) | ester + $H^+$ | → Protonated ester | -200.4 | -203.2 |
| v) | $CH_3O^- + H^+$ | $\rightarrow CH_3OH$ | -388.6 | -391.1 |
| vi) | $CH_3OH + H^+$ | $\rightarrow CH_3OH_2^+$ | -185.5 | -187.2 |
| vii) | $H_2C\text{=}CH_2 + H^+$ | $\rightarrow H_2C\text{=}CH_3^+$ | -166.7 | -168.0 |

With the largest basis sets we used, the best DFT functionals provide a mean unsigned error≈80 % below the error afforded by non-correlated Hartree-Fock computations, and less than 1 kcal.mol$^{-1}$ above that of the much more expensive MP4 level of theory. The best agreement between the DFT values and the MP2/CBS reaction energies was found for the hybrid X3LYP functional and the meta-hybrid M06 and M06-2X functionals (with all tested basis sets) and the popular B3LYP functional (provided that less than 3 polarization *d* functions and only one *p* polarization function are included). Surprisingly, the simpler GGA-only functionals afforded very satisfactory results with the smaller 6-311+G(d,p) basis set (mean unsigned error=1.7 kcal.mol$^{-1}$ for PBE96, PBEPW91 and PW91). It is clear from these results that for these systems no systematic improvement is achieved simply by increasing the DFT functional complexity or the basis set, as each generation of functionals (GGA-only, hybrid GGA, and meta-hybrid GGA) contains both satisfactory and inadequate functional/basis set combinations. The picture that emerges upon comparison of the DFT computations with the CCSD(T)/aug-cc-pVTZ reaction energies is slightly different: since the CCSD(T)/aug-cc-pVTZ reaction energies are lower than the MP2/CBS energies for all reactions in this set (Table 1), the under-estimation of DFT energies *vs.* MP2/CBS means that DFT errors *vs.* CCSD(T)/aug-cc-pVTZ are generally smaller than errors *vs.* MP2/CBS. B3LYP now approaches the performance of X3LYP and M06-2X with all tested basis sets. The mean unsigned errors for every functional (and corresponding standard deviations) are shown in Table 2.

Table 2: Mean Unsigned Errors in the computed gas-phase energies of the tested protonation reactions. All values in kcal.mol$^{-1}$. ZPVE not included. Error values below 2.0 kcal.mol$^{-1}$ *vs.* MP2/CBS or below 1.0 kcal.mol$^{-1}$ *vs.* CCSD(T)/aug-cc-pVTZ are highlighted in bold. For functionals M06-L and the TPSS family, the energies of the reactions involving acetone could not be computed with the largest basis set due to non-convergence of the SCF cycles (even with very precise integral evaluation and the largest Lebedev integration grids available). For these combinations of functional and basis set, the errors shown do not include these reactions.

| | | MUE *vs.* MP2/CBS | | | | MUE *vs.* CCSD(T)/aug-cc-pVTZ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 6-311+G (d,p) | 6-311+G (2d,p) | 6-311+G (2d,2p) | 6-311+G (3d,2p) | 6-311+G (d,p) | 6-311+G (2d,p) | 6-311+G (2d,2p) | 6-311+G (3d,2p) |
| HF | *ab initio* | 8.7±2.0 | 9.0±2.2 | 9.5±2.2 | 9.8±2.1 | 6.2±1.7 | 6.5±1.8 | 7.1±1.7 | 7.3±1.8 |
| MP2 | *ab initio* | 1.8±1.7 | 0.8±0.5 | 0.8±0.7 | 0.4±0.4 | 1.5±1.0 | 2.7±0.9 | 1.8±0.8 | 2.2±0.8 |
| MP3 | *ab initio* | 6.1±2.4 | 4.0±1.9 | 4.9±2.1 | 4.6±1.6 | 3.7±2.0 | 1.6±1.4 | 2.5±1.6 | 2.2±1.1 |
| MP4 | *ab initio* | 3.1±2.0 | 1.0±1.3 | 1.9±1.2 | 1.4±0.8 | 1.8±0.8 | 1.5±1.0 | 0.9±0.5 | 1.1±0.8 |
| PBE96 | GGA | **1.7±1.3** | 2.1±1.1 | 2.2±1.2 | 2.2±1.2 | 2.5±2.5 | 2.4±2.4 | 2.1±2.3 | 2.2±2.2 |
| PBEPW91 | GGA | **1.7±1.3** | 2.1±1.1 | 2.1±1.2 | 2.2±1.2 | 2.4±2.5 | 2.3±2.3 | 2.1±2.2 | 2.1±2.2 |
| PW91 | GGA | **1.7±1.2** | 2.1±1.1 | 2.1±1.2 | 2.1±1.2 | 2.5±2.5 | 2.3±2.4 | 2.0±2.2 | 2.1±2.2 |
| B3LYP | hybrid-GGA | **1.6±0.8** | **1.9±1.3** | 2.3±1.4 | 2.3±1.3 | **1.0±1.0** | **1.0±0.8** | **1.0±0.7** | **1.0±0.6** |
| B3PW91 | hybrid-GGA | 3.7±1.2 | 4.1±1.5 | 4.5±1.5 | 4.5±1.4 | 1.4±0.8 | 1.8±1.1 | 2.1±1.2 | 2.1±1.2 |
| B97-1 | hybrid-GGA | 3.3±0.9 | 3.7±1.3 | 4.1±1.3 | 4.1±1.3 | 1.1±0.8 | 1.4±1.0 | 1.7±1.2 | 1.7±1.3 |
| B97-2 | hybrid-GGA | 4.9±0.9 | 5.3±1.3 | 5.8±1.3 | 5.8±1.3 | 2.5±1.1 | 2.9±1.3 | 3.3±1.2 | 3.3±1.3 |
| BHHLYP | hybrid-GGA | 4.2±1.0 | 4.5±1.3 | 5.0±1.3 | 5.1±1.2 | 1.8±0.7 | 2.1±0.8 | 2.6±0.8 | 2.6±0.8 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PBE0 | hybrid-GGA | 2.8±1.0 | 3.2±1.4 | 3.7±1.4 | 3.7±1.4 | **0.9±0.8** | 1.3±0.9 | 1.4±1.2 | 1.4±1.3 |
| PBE1PW91 | hybrid-GGA | 2.9±0.9 | 3.3±1.3 | 3.8±1.4 | 3.8±1.3 | **0.9±0.8** | 1.2±0.9 | 1.3±1.3 | 1.3±1.3 |
| X3LYP | hybrid-GGA | **1.3±0.7** | **1.6±1.2** | **2.0±1.4** | **2.0±1.3** | 1.3±1.1 | 1.1±1.1 | **1.0±0.8** | **1.0±0.8** |
| TPSS | meta-GGA | 2.8±1.4 | 3.2±1.6 | 3.6±1.6 | 3.3±1.7 | 1.4±1.1 | 1.7±1.0 | 1.8±1.0 | 1.8±1.0 |
| TPSSm | meta-GGA | 3.0±1.5 | 3.4±1.7 | 3.7±1.7 | 3.5±1.8 | 1.5±1.2 | 1.8±1.1 | 1.9±1.2 | 2.0±1.1 |
| M06 | meta-hybrid-GGA | **1.1±0.9** | **1.6±1.0** | **1.8±1.5** | **1.8±1.5** | 2.1±1.3 | 2.0±1.0 | 1.6±0.6 | 1.5±0.7 |
| M06-2X | meta-hybrid-GGA | **0.9±0.8** | **1.4±0.9** | **1.6±1.0** | **1.5±0.8** | 1.6±1.0 | 1.1±0.7 | **0.9±0.6** | **0.9±0.8** |
| M06-HF | meta-hybrid-GGA | **1.9±1.4** | 2.4±1.2 | 2.3±1.3 | **2.0±1.2** | 2.0±2.4 | 1.6±1.9 | 1.6±1.8 | 1.7±2.2 |
| M06-L | meta-hybrid-GGA | 2.6±1.8 | 3.0±2.1 | 3.7±2.0 | 2.8±1.4 | 1.5±0.6 | 1.7±0.8 | 1.7±1.3 | 1.1±1.0 |
| TPSSh | meta-hybrid-GGA | 3.6±1.1 | 4.0±1.3 | 4.4±1.2 | 4.1±1.3 | 1.4±1.2 | 1.8±1.2 | 2.0±1.3 | 1.9±1.4 |



Scheme 2

Table 3: Computed gas-phase energies of the tested reduction reactions. All values in kcal.mol$^{-1}$. ZPVE not included. n.c. : not computed due to excessive computational cost.

| | Reaction | MP2 / CBS | CCSD(T) / aug-cc-pVTZ |
|---|---|---|---|
| i) | aldehyde→ alcohol | -759.0 | -760.1 |
| ii) | ethene → ethane | -775.6 | -775.7 |
| iii) | carboxylic acid → diol | -739.9 | -742.1 |
| iv) | quinone → quinol | -782.3 | n.c. |
| v) | disulfide → dithiol | -741.2 | n.c. |
| vi) | flavin reduction | -762.7 | n.c. |
| vii) | nicotinamide reduction | -512.2 | n.c. |

We also selected several organic oxidation reactions often encountered in the study of enzyme reaction mechanisms: the reduction of intercellular organic electron carriers (flavin, nicotinamide, quinol), the conversions of carboxylic acids in diols and aldehydes in alcohols, as well as the reduction of the carbon-carbon double bond and the formation of a disulfide bridge (Scheme 2). The geometries predicted from the density functionals were in good (RMSD<0.05Å) or very good (RMSD<0.025Å) agreement with those obtained by MP2 in most cases (ethane, ethylene, acetaldehyde, 1,1-ethanedithiol, acetic acid, FMN, 1,4-benzoquinone, 1,4-benzoquinol, 1,4-butanedithiol). Larger deviations are observed for $NAD^+$ (RMSD<0.15 Å), $FMNH_2$ (RMSD between 0.10 and 0.14 Å; predicted by most functionals to be slightly less bent than shown by MP2), and 1,2-dithiane (due to its inherent flexibility, B3LYP, PW91, PBE96 and PBEPW91 converge to a different local minimum, RMSD between 0.40 and 0.65 Å vs. MP2). The heterocyclic ring in NADH is predicted by all density functionals to be planar, whereas MP2 predicts its N1 atom and C4 atoms to lie above the plane of the remaining atoms.

The energies from these model reactions were computed at the MP2/CBS level and (when computationally feasible) at the CCSD(T)/ aug-cc-pVTZ level (Table 3). The performance of the DFT functionals in this reaction test set proved to be poorer than expected from the acid/base reaction test set, as several functionals (B97-1, M06-L and the TPSS family of functionals) were unable to order correctly their reaction energies and erroneously predicted ethene to be easier to reduce than quinone. Mean unsigned errors (Table 4) were also larger than observed in the acid/base reaction test set. The most common sources of error were the computations of reduction energies of 1,4-benzoquinone to 1,4-benzoquinol (underpredicted by 8.8 to 3 $kcal.mol^{-1}$ by 12 of the 18 tested functionals) and 1,2-dithiane to 1,4-butanedithiol (overpredicted by 4 to 12.2 $kcal.mol^{-1}$ by 12 of the 18 tested functionals). Most functionals lead to similar mean unsigned errors, below those computed by MP2 at the same basis set level, but still above the desired accuracy (mean unsigned errors between 2 – 4 $kcal.mol^{-1}$). PBE0 emerges as the most suitable choice, with errors below 1.5 $kcal.mol^{-1}$ (i.e. 0.07 eV). The results reported on Tables 2 and 4 do not provide a full appraisal of the merits of the tested DFT functionals: the results from a hypothetical functional with a systematic overestimation of 2 kcal.mol$^{-1}$ for every reaction energy are much easier to correct than those of another hypothetical functional that overestimates the energies of half of the reactions by 2 kcal.mol$^{-1}$ and underestimates those of the other reactions by 2 kcal.mol$^{-1}$, despite the fact that the mean unsigned errors are equal in both cases. We have therefore computed the mean unsigned errors for the *differences* in reaction energies (which we call *relative* reaction energies) for all 7×6/2 reactions obtained by coupling any two acid/base pairs tested and all 7×6/2 reactions obtained by coupling any two redox pairs for each method, with the 6-311+G(3d,2p) basis sets (Table 5).

Table 4: Mean Unsigned Errors in the computed gas-phase energies of the tested redox reactions. All values in kcal.mol$^{-1}$. ZPVE not included. Error values below 2.0 kcal.mol$^{-1}$+ vs. MP2/CBS are highlighted in bold. The energy of FMN reduction proved to be too expensive for MP4 and MP3. For these theory levels the errors shown do not include this reaction.

| | MUE *vs.* MP2/CBS | | | |
|---|---|---|---|---|
| | 6-311+G (d,p) | 6-311+G (2d,p) | 6-311+G (2d,2p) | 6-311+G (3d,2p) |
| HF | 24.3 ±6.3 | 25.5 ±6.3 | 24.8 ±6.1 | 24.8 ±6.4 |
| MP2 | 7.7 ±3.4 | 8.0 ±3.0 | 5.7 ±2.6 | 6.0 ±2.6 |
| MP3 | 3.4 ±3.1 | 3.2 ±2.9 | 2.7 ±2.8 | 2.8 ±3.2 |
| MP4 | 6.4 ±4.6 | 6.0 ±4.6 | 4.5 ±3.8 | 4.8 ±4.1 |
| PBE96 | 4.7 ±2.9 | 4.7 ±3.0 | 4.1 ±2.6 | 4.3 ±2.8 |
| PBEPW91 | 4.1 ±2.6 | 3.9 ±2.7 | 3.4 ±2.2 | 3.6 ±2.3 |
| PW91 | 3.0 ±2.3 | 2.6 ±2.2 | 2.2 ±1.8 | 2.4 ±1.9 |
| B3LYP | 4.2 ±4.0 | 3.6 ±3.5 | 3.9 ±3.4 | 4.0 ±3.7 |
| B3PW91 | 4.4 ±3.0 | 3.8 ±2.4 | 4.5 ±2.2 | 4.3 ±2.5 |
| B97-1 | 2.3 ±2.8 | 2.3 ±2.2 | 1.9 ±2.1 | 2.1 ±2.3 |
| B97-2 | 2.9 ±3.0 | 2.4 ±2.4 | 2.6 ±2.4 | 2.7 ±2.6 |

| | | | | |
|---|---|---|---|---|
| BHHLYP | 2.3 ±3.3 | 2.1 ±2.9 | 2.3 ±2.5 | 2.3 ±2.8 |
| PBE0 | **1.8 ±1.5** | **1.3 ±1.1** | **1.3 ±0.9** | **1.4 ±1.1** |
| PBE1PW91 | 2.5 ±1.8 | 1.8 ±1.2 | 2.1 ±1.2 | 2.2 ±1.3 |
| X3LYP | 2.6 ±3.4 | 2.4 ±2.9 | 1.9 ±2.9 | 2.2 ±3.1 |
| TPSS | 3.3 ±3.3 | 3.2 ±2.7 | 2.9 ±2.7 | 3.1 ±2.9 |
| TPSSm | 3.7 ±3.2 | 3.7 ±2.8 | 3.2 ±2.8 | 3.5 ±2.9 |
| M06 | 3.2 ±2.0 | 2.7 ±1.6 | 2.2 ±1.5 | 2.6 ±1.7 |
| M06-2X | 2.2 ±1.5 | **1.9 ±1.0** | **1.5 ±1.1** | **1.8 ±1.4** |
| M06-HF | 2.7 ±1.9 | 2.5 ±2.2 | 2.9 ±1.7 | 2.3 ±1.8 |
| M06-L | 3.7 ±3.2 | 3.8 ±2.5 | 3.2 ±2.6 | 3.4 ±2.8 |
| TPSSh | 3.0 ±3.2 | 2.8 ±2.5 | 2.8 ±2.5 | 2.9 ±2.6 |

Several functionals which behave poorly in the prediction of *absolute* gas-phase acidities (such as B97-2 and BHHLYP) yielded very good results in the prediction of *relative* acidities, due to systematically overestimating individual reaction energies by similar amounts. On the other hand, the significant under-estimation of methanol and acetone acidities by the GGA-functionals and M06-HF (whereas the other acidities are over-estimated) relegates these functionals to a worse ranking in the prediction of *relative* acidities. The popular B3LYP functional and the (slightly older) B3PW91 perform remarkably well, as does the meta-hybrid M06-2X functional. The prediction of *relative* reduction energies is more problematic, as even the best functional (PBE0) still returns a relatively large mean unsigned error of 2.0 kcal.mol$^{-1}$. Still, this error is ≈40% smaller than that obtained from MP2 computations with the same basis set. Elimination of the problematic reductions of 1,2-dithiane and 1,4-benzoquinone from the tested set brings the errors from most DFT hybrid functionals to values <1.5 kcal.mol$^{-1}$ (0.8 ± 0.7 for BHHLYP, B3LYP and X3LYP). The errors from meta-GGA and meta-hybrid-GGA functionals remain relatively large due to their poor performance on the prediction of the reduction energy of the flavin model (TPSS family), ethylene (M06-HF and M06-L) and nicotinamide (M06).

Table 5: Mean Unsigned Errors in the computation of energy differences of paired tested reactions using the 6-311+G(3d,2p) basis set. All values in kcal.mol$^{-1}$. ZPVE not included. Error values below 1.5 kcal.mol$^{-1}$ are highlighted in bold. The energy of FMN reduction proved to be too expensive for MP4 and MP3. For these theory levels the errors shown do not include this reaction.

| | Acid/base reactions | | Redox reactions | Redox reactions (reduced test set) |
|---|---|---|---|---|
| | MUE vs. MP2/CBS | MUE vs. CCSD(T) | MUE vs. MP2/CBS | MUE vs. MP2/CBS |
| HF | 2.6±1.6 | 2.1±1.4 | 7.2±5.4 | 2.6±2.2 |
| MP2 | 0.6±0.4 | 0.9±0.6 | 3.2±2.9 | 0.9±0.6 |
| MP3 | 1.9±1.1 | 1.3±0.9 | 5.1±3.6 | 1.5±0.7 |
| MP4 | 0.9±0.7 | 0.9±0.6 | 6.7±5.1 | 1.3±0.6 |
| PBE96 | 3.1±1.9 | 3.2±2.1 | 3.5±2.2 | 1.9±1.2 |
| PBEPW91 | 3.0±1.8 | 3.1±2.1 | 3.6±2.4 | 1.8±1.2 |
| PW91 | 2.9±1.8 | 3.1±2.1 | 3.4±2.2 | 1.7±1.1 |
| B3LYP | 1.6±0.9 | **1.5±0.9** | 4.7±4.3 | **0.8±0.6** |
| B3PW91 | 1.7±1.1 | **1.4±0.9** | 2.9±2.1 | **1.1±0.9** |
| B97-1 | **1.5±0.9** | **1.5±0.9** | 3.7±2.7 | **1.2±0.9** |
| B97-2 | **1.5±1.0** | **1.5±0.9** | 3.8±2.6 | 1.8±1.2 |
| BHHLYP | **1.4±1.0** | **0.8±0.7** | 4.0±3.4 | **0.8±0.7** |
| PBE0 | 1.7±1.0 | 1.7±1.1 | 2.0±1.3 | **1.0±0.6** |
| PBE1PW91 | 1.6±0.9 | 1.6±1.0 | 2.2±1.5 | **1.0±0.6** |
| X3LYP | 1.6±1.0 | **1.5±1.0** | 4.1±3.6 | **0.8±0.5** |
| TPSS | 2.0±1.3 | 2.3±1.5 | 5.2±3.3 | 2.7±1.5 |
| TPSSm | 2.0±1.3 | 2.5±1.6 | 5.6±3.5 | 2.9±1.7 |
| M06 | 1.9±1.2 | 1.9±1.3 | 3.6±2.2 | 2.3±1.2 |
| M06-2X | **1.0±0.6** | **1.0±0.6** | 2.5±1.7 | 1.7±1.5 |
| M06-HF | 2.3±1.9 | 2.7±2.3 | 4.0±2.7 | 3.8±2.7 |
| M06-L | 2.4±1.4 | 1.7±1.2 | 5.1±3.4 | 3.0±2.4 |

| | | | | |
|---|---|---|---|---|
| TPSSh | 1.5±0.9 | 1.9±1.2 | 4.6±2.9 | 2.4±1.4 |

## Conclusions

The choice of a functional depends critically on the type of information the computation is expected to provide. The best functional in the computation of *single* reaction energies (“absolute” reaction energies) is often not the best for the computation of energies involving *pairs* of reactions (“relative” reaction energies), and this must be kept in mind when deciding the functional to use. For example, M06 and M06-2X perform best in the prediction of absolute basicities, but the non-systematic nature of the deviation of the M06 energies from MP2/CBS benchmark values makes this functional less accurate in the computation of *relative* basicities. The errors provided by BHHLYP are much more homogeneous, and therefore this functional performs very well in the computation of *relative* basicities, in spite of its very poor performance in the computation of *absolute* basicities.

Most hybrid-GGA perform adequately in the computation of *relative* reduction energies of flavins, aldehydes, carboxylic acids, $NAD^+$ and alkenes. Functional performance does not depend on the size of the system: a functional may provide small errors in the reduction energies of $NAD^+$, benzoquinone or 1,2-dithiane, yet fail dramatically in the reduction energy of a small molecule like ethylene (as occurred with M06-HF and M06-L). To help the user in the choice of functional, we present the most suitable functionals for the tested reactions in Tables 6 and 7. For most of the proton transfers between different acid/base pairs (or electron transfer between redox pairs) tested we could find at least one functional with very high accuracy (error <0.2 kcal.mol$^{-1}$). The reactions involving electron transfer between quinone (or disulfide) and other redox groups stand out as the clearest example of the shortcomings of DFT methods, as the best functionals are most often wrong by 1-3 kcal.mol$^{-1}$. The electron transfer between quinone and a disulfide is very well described by only one functional (M06-HF), which underpredicts both electron transfers by roughly the same small amount (0.7 kcal.mol$^{-1}$) but performs inadequately for most other electron transfer reactions studied.

Table 6: Suggested DFT functionals for the study of reactions involving proton transfer from an organic acid to an organic base. Absolute errors (kcal.mol$^{-1}$) (DFT/6-311+G(3d,2p) vs. MP2/CBS) shown in parentheses.

| | α-deprotonation of a ketone | Deprotonation of a protonated ketone | deprotonation of a protonated amide | deprotonation of a protonated ester | deprotonation of a protonated Alkene | Deprotonation of an alcohol |
|---|---|---|---|---|---|---|
| Ketone protonation | M06-L (0.2) | | | | | |
| Amide protonation | M06-2X (<0.1) | BHHLYP (0.2) | | | | |
| Ester protonation | M06-2X (0.2) | M06-HF (0.2) | B97-1 (<0.1) | | | |
| Alkene protonation | B3LYP (<0.1) | M06-HF (<0.1) | M06-2X (0.3) | M06 (0.2) | | |
| Alkoxide protonation | TPSS (<0.1) | M06-HF (<0.1) | TPSS (0.8) | M06-HF (0.2) | M06-HF (0.1) | |
| Alcohol protonation | BHHLYP (0.9) | BHHLYP (0.2) | BHHLYP (<0.1) | BHHLYP (0.8) | X3LYP (1.2) | B97-1 (<0.1) |

Table 7: Suggested DFT functionals for the study of reactions involving electron transfer between organic redox pairs. Absolute errors (kcal.mol$^{-1}$) (DFT/6-311+G(3d,2p) vs. MP2/CBS) shown in parentheses.

| | Oxidation of an alcohol to an Aldehyde | Oxidation of a geminal diol to a carboxylic acid | Oxidation of an alkane to an alkene | Oxidation of reduced flavin | Oxidation of NADH | Oxidation of a quinol |
|---|---|---|---|---|---|---|
| Carboxylic acid reduction | B3LYP (<0.1) | | | | | |
| Alkene reduction | BHHLYP (0.3) | BHHLYP (0.3) | | | | |
| Flavin reduction | B3PW91 (<0.1) | M06-2X (<0.1) | BHHLYP (0.5) | | | |
| NAD$^+$ reduction | B3LYP (<0.1) | B3PW91 (<0.1) | X3LYP (0.2) | B97-1 (0.2) | | |
| Quinone reduction | M06-HF (1.1) | PBE0 (2.0) | M06-2X (1.9) | M06-HF (2.5) | B97-2 (1.9) | |
| Disulfide reduction | M06-HF (1.2) | M06-2X (3.0) | PBE0 (1.1) | PBE0 (2.1) | M06-2X (0.2) | M06-HF (<0.1) |

SUPPORTING INFORMATION Geometries of every described molecule optimized with each density functional and with MP2, as well as their electronic energies with different basis sets.

## REFERENCES

1 P. Hohenberg, W. Kohn, Phys. Rev. B, 136 (1964) 864

2 S. F. Sousa, P.A. Fernandes, M.J. Ramos, J. Phys. Chem. A, 111 (2007) 10439

3 A. D. Becke, J. Chem. Phys. I, 98 (1993) 5648

4 C. Lee, W. Yang, R. Parr, J. Phys. Rev. B, 37 (1998) 785

5 R. W. Hertwig, W. Koch, J. Comp. Chem. , 16 (1995) 576.

6 C. Flener-Lovitt, D. E. Woon, T.M. Dunning, G. S. Girolami, J. Phys. Chem. A, 114 (2010) 1843

7 G. Csonka, A.D. French, G.P. Johnson, C.A. Stortz, J. Chem. Theory Comput., 5 (2009) 679

8 S.E. Wheeler, A. Moran, S.N. Pieniazek, K.N. Houk, J. Phys. Chem. A, 113 (2009) 10376

9 M. Korth and S. Grimme, J. Chem. Theory Comput., 5 (2009) 993

10 J. Baker, A. Kessi, B. Delley, J. Chem. Phys. 105 (1996) 192

11 P.E.M. Siegbahn, L. Eriksson, F. Himo, M. Pavlov, J. Phys. Chem. B. , 102 (1998) 10622.

12 P.A. Fernandes, M.J. Ramos, J. Am. Chem. Soc. , 125 (2003) 6311

13 K.E. Riley, B. T. Op't Holt, K. M. Merz Jr. , J. Chem. Theory Comput. , 3 (2007) 407

14 D. G. Truhlar, Chem. Phys. Lett. , 294 (1998) 45

15 Y. Zhao, D. G. Truhlar, J. Phys. Chem. A, 109 (2005) 6624

16 J.P. Perdew, "Unified Theory of Exchange and Correlation Beyond the Local Density Approximation. " In Electronic Structure of Solids '91 pp 11-20; Ziesche, P. , Eschig, H. , Eds. ; Akademie Verlag: Berlin, Germany, 1991

17 J.P. Perdew, K. Burke, M. Ernzerhof, Phys. Rev. Lett. , 77 (1996) 3865

18 F. A. Hamprecht, A. J. Cohen, D. J. Tozer, N.C. Handy, J. Chem. Phys. , 109 (1998) 6264

19 P.J. Wilson, T. J. Bradley, D. J. Tozer, J. Chem. Phys.,115 (2001) 9233

20 A. D. Becke, Phys. Rev. A, 38 (1988) 3098

21 C. Adamo, V. Barone, J. Chem. Phys. , 110 (1999) 6158

22 X. Xu, Q. Zhang, R. P. Muller, W. A. Goddard, J. Chem. Phys. , 122 (2005) 014105

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

24 J. P. Perdew, J. Tao, V. N. Staroverov, G. E. Scuseria, J. Chem. Phys. , 120 (2004) 6898

25 J. P. Perdew, A. Ruzsinszky, J. Tao, G. I. Csonka, G. E. Scuseria, Phys. Rev. A, 76 (2007) 042506

26 V. N. Staroverov, G. E. Scuseria, J. Tao, J. P. Perdew, J. Chem. Phys. , 119(2003) 12129; erratum in J. Chem. Phys. , 121, 11507

27 Y. Zhao, D. G. Truhlar, Theoret. Chem. Acc. , 120 (2008) 215

28 Y. Zhao, D. G. Truhlar, J. Chem. Phys. , 125 (2006) 194101

29 Y. Zhao, D. G. Truhlar, J. Phys. Chem. A , 110 (2006) 13126

30 M.W. Schmidt, K.K. Baldridge, J. A. Boatz, S.T. Elbert, M.S. Gordon, J.J. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S. Su, T.L. Windus, M. Dupuis, J.A. Montgomery, J. Comput. Chem. , 14 (1993) 1347

31 A. A. Granovsky, PC GAMESS/Firefly version 7.0, http://classic.chem.msu.su/gran/gamess/index.html