Minimally empirical double hybrid functionals trained against the GMTKN55 database: revDSD-PBEP86-D4, revDOD-PBE-D4, and DOD-SCAN-D4

Golokesh Santra, Nitai Sylvetsky, and Jan M.L. Martin

Department of Organic Chemistry, Weizmann Institute of Science, 7610001 Reḥovot, Israel.
Email: gershom@weizmann.ac.il

Abstract

We present a family of minimally empirical double-hybrid DFT functionals parametrized against the very large and diverse GMTKN55 benchmark. The very recently proposed ωB97M(2) empirical double hybrid (with 16 empirical parameters) has the lowest WTMAD2 (weighted mean absolute deviation over GMTKN55) ever reported at 2.19 kcal/mol. However, refits of the DSD-BLYP and DSD-PBEP86 spin-component-scaled, dispersion-corrected double hybrids can achieve WTMAD2 values as low as 2.37 with the very recent D4 dispersion correction, (2.55 kcal/mol with the D3BJ dispersion term) using just a handful of adjustable parameters. If we use full DFT correlation in the initial orbital evaluation, the xrevDSD-PBEP86-D4 functional reaches WTMAD2=2.22 kcal/mol, statistically indistinguishable from ωB97M(2) but using just four non-arbitrary adjustable parameters (and three semi-arbitrary ones). The changes from the original DSD parametrizations are primarily due to reaction and isomerization energies for large systems, which were undersampled in the original parametrization set. With the new parametrization, same-spin correlation can be eliminated at minimal cost in performance, which permits revDOD-PBEP86-D4 and revDOD-PBE-D4 functionals that scales as N^4 or even N^3 with the size of the system. Dependence of WTMAD2 for DSD functionals on the percentage of HF exchange is roughly quadratic; it is sufficiently weak that any reasonable value in the 64% to 72% range can be chosen semi-arbitrarily. Finally, DSD-SCAN and DOD-SCAN double hybrids involving the SCAN non-empirical meta-GGA as the semi-local component have also been considered, and offer a good alternative if one wishes to eliminate either the empirical dispersion correction or the same-spin correlation component. noDispSD-SCAN66 achieves WTMAD2=3.00 kcal/mol, while DOD-SCAN66-D4 reaches 2.61 kcal/mol. Finally, in the context of double-hybrid functionals, the very recent D4 dispersion correction is clearly superior over D3BJ.

Introduction

Large and chemically diverse standardized reference datasets play a crucial role in the validation of new approximate computational chemistry methods (not just density functional methods, but also semi-empirical molecular orbital methods, e.g.,[1] , composite wavefunction ab initio schemes,[2–6] and machine learning-assisted approaches[7]).

If these approaches are devoid of empirical parameters (such as the "non-empirical" DFT functionals PBE,[8] TPSS,[9] and SCAN,[10] as well as the ccCA,[11] W1,[5,12] and W1-F12[6] approaches), then the purpose of these datasets is only validation. If the methods include empirical parameters, however (such as in "empirical" DFT functionals, e.g., B97-1,[13] HCTH,[14] BMK,[15] M06,[16] MN15,[17] and many others), then such datasets take on the additional role of "training sets" or parametrization sets. In the earliest days, small sets of experimental data were used for this purpose, e.g. for B3LYP[18] and EDF-1;[19] as the practical limitations of this approach became apparent, Handy[14,20] pioneered the use of high-level

wavefunction ab initio data for the same purpose. This approach has perhaps been taken furthest by the Head-Gordon group in their combinatorially optimized ωB97X-V,[21] B97M-V,[22]ωB97M-V,[23] and ωB97M(2)[24] functionals.

In Perdew's "Jacob's Ladder" metaphor,[25] Hartree theory represents the "Earthly vale of tears" and the introduction of each new type of information one more rung on the "Jacob's Ladder" ascending to the Heaven of chemical accuracy. The first rung corresponds to the local density approximation, where the XC (exchange-correlation) functional only depends on the density ρ. The reduced density gradient is introduced on the second rung, leading to the various GGA (generalized gradient approximation) functionals. The introduction of the Laplacian (or the kinetic energy density, which contains similar information) creates the third rung, the meta-GGAs. The fourth rung introduces dependence on the occupied orbitals:[26] the most important special case are the different types of hybrid functionals. Finally, the fifth rung corresponds to dependence on virtual orbitals, such as double-hybrid functionals. The term "double hybrid" was first employed to denote the linear combination of GGA correlation and MP2 correlation from HF orbitals,[27] but since the landmark paper of Grimme,[28] the term has come to refer exclusively to the admixture of (meta)GGA DFT correlation with GLPT2 (2nd-order Görling-Levy[29] perturbation theory) correlation from hybrid (meta)GGA DFT orbitals.

In the first step of a double-hybrid calculation, a Kohn-Sham calculation is carried out for a given semilocal exchange-correlation (XC) functional with a fraction c'_x of Hartree-Fock exchange and $(1–c'_x)$ of XC exchange, plus XC correlation damped by a factor $c'_{C,DFA}$. With the converged Kohn-Sham orbitals at the end, the total energy is then evaluated in the second step as:

$$E = E_{N1e} + c_x E_{x,HF} + (1 - c_x) E_{x,XC} + c_{c,XC} E_{c,XC} + c_{2ab} E_{2ab} + c_{2ss} E_{2ss} + s_6 E_{disp}$$

where E_{N1e} stands for the nuclear repulsion and one-electron energy term; $E_{x,HF}$ is the Hartree-Fock exchange energy and c_x the fraction of Hartree-Fock-like exchange energy; $E_{x,XC}$ and $E_{c,XC}$ are the exchange and correlation energies, respectively, for the given semilocal XC functional with the converged density from the first step, and $c_{c,XC}$ the fraction of DFT correlation energy; E_{2ab} and E_{2ss} are the opposite-spin and same-spin MP2-like energies obtained in the basis of Kohn-Sham like orbitals from the first step, and c_{2ab} and c_{2ss} are the linear coefficients for the same; and, finally, E_{disp} is the dispersion energy obtained from a given empirical dispersion model (e.g., D2,[30] D3zero, [31] D3BJ, [31,32] or very recently,[33,34] D4) or nonlocal dispersion functional (such as VV10[35]), with an optional pre-factor s_6.

In the original Grimme approach,[28] $c'_x = c_x$, $c'_{c,XC} = c_{c,XC}$, and $c_{2ab} = c_{2ss}$; the latter constraint in practice restricts[36,37] the choice of correlation functionals to LYP. In the DSD functionals (dispersion-corrected, spin-component-scaled double hybrid[38]) of Kozuch and Martin, the constraint $c_{2ab} = c_{2ss}$ is relaxed: this was found[36,37] to enable a broader variety of exchange-correlation functionals, with DSD-PBEP86 the best performer at that point. In the GMTKN55 benchmark paper, the best two performers were DSD-PBEP86 and DSD-BLYP, followed by B2GP-PLYP.[39]

In the XYG3 approach of Goddard, Xu, and coworkers,[40] $c'_{c,XC} = 1 \neq c_{c,XC}$ and c'_x may differ from c_x : the implication of this choice have been discussed at length elsewhere.[41–43] With two exceptions (namely, ωB97M(2)[24] and xrevDSD-PBEP86-D4, see below), functionals of this form are not discussed in this paper. The special case $c'_x = c_x$ in a DSD context has been denoted xDSD.[42]

Double hybrid DFT has been reviewed by Goerigk and Grimme,[44] by Sancho-Garcia and Adamo,[45] and by Xu and coworkers.[40,46] An extensive comparative study between nonempirical[47] and semiempirical (e.g., Refs.[28,36,37,39,48–50]) double hybrids was recently made by Goerigk and coworkers[51], who found semiempirical functionals (at present) to be more accurate and more robust. For some recent perspectives on "the functional zoo" (Perdew's term), see, e.g., Refs.[52–54]

Perhaps the two most extensive and chemically diverse training/validation datasets around are GMTKN55 (General Main-group Thermochemistry, Kinetics, and Noncovalent interactions, 55 problem subsets) of Goerigk, Grimme, and coworkers[55] — which has about 1,500 nonredundant reaction energies and barrier heights — and the even larger MGCDB84 (Main Group Chemistry DataBase, 84 problem subsets) of Mardirossian and Head-Gordon,[53] which has close to 5,000 such nonredundant energy differences. These databases themselves incorporate and extend upon earlier work by these authors themselves (e.g.,[41,56]), by the Minnesota group (Refs.[54,57] and references therein), by the Hobza group (particularly for noncovalent interactions[58–61]) and by the present research team (e.g.,[39,62–73].)

Such large and unwieldy reference datasets have themselves inspired the statistical search for representative subsamples that would recover most of the variation in the underlying dataset yet be much easier to handle. To the authors' knowledge, the first such study was Ref.[74]; the two most recent ones are MG8 by Chan[75], a 60-reaction subset of MGCDB84 obtained through lasso regularization, and "Diet-GMTKN55" by Gould[76], the latter of which proposes 30-, 100-, and 150-reaction "Diet" versions of GMTKN55. Aside from "rapid prototyping", these could in principle serve as training sets for empirical functionals, with the full dataset then used for validation purposes.

Our explorations on the suitability of such reduced training sets for functional development will be discussed elsewhere. In the present paper, we focus instead on the full GMTKN55 benchmark as being sufficiently large and chemically varied that parametrization and validation against it is largely immune to sample bias.To the best of our knowledge, the present work is the first paper in which the full GMTKN55 dataset is used as a training set for DFT functionals, although we are also validating some new functionals not covered in the original GMTKN55 paper (for technical reasons).

We will show below that:

(a) the most accurate functional that does not entail the fifth rung of Perdew's "Jacob's Ladder"[25] is the combinatorially optimized, range-separated, hybrid meta-GGA wB97M-V, again by the Berkeley group;[23]

(b) if the search is widened to fifth-rung options, the combinatorially optimized, range-separated, double hybrid wB97M(2) by Mardirossian and Head-Gordon[24] is at present the most accurate functional available for general main-group chemistry;

(c) this having been said, reparametrized versions of DSD-BLYP-D3BJ[38] and DOD-PBEP86-D3BJ[36,37] fitted to GMTKN55 come quite close in performance with just one-third the number of empirical parameters;

(d) replacing the D3BJ dispersion correction by the more modern, partial-charge dependent D4 model significantly enhances performance;

(e) the xrevDSD-PBEP86-D4 model affords a statistically equivalent WTMAD2 to wB97M(2), as does its xrevDOD-PBEP86-D4 variant, which is amenable to reduced-scaling MP2 implementations;

(f) if one eschews empirical dispersion corrections, then the DSD-SCAN63-noDisp functional proposed in the present work offers the best performance;

(g) while performance over GMTKN55 is markedly improved from the original versus the refitted DSD functionals, performance for small-molecule atomization energies and barrier heights is barely affected — the improvements are seen in large-

molecule isomerization and reaction energies where there is an important dispersion component
(h) this presents a cautionary tale about "overfitting" to small and insufficiently diverse reference samples.

Computational Methods

Reference data

As our primary parametrization and validation set, we used the comprehensive GMTKN55 benchmark[55] of Goerigk, Grimme, and coworkers. This set, itself a further expansion and update of earlier GMTKN[56] and GMTKN30[41] datasets, is a composite of fifty-five chemical problem types, ranging from small-molecule thermochemistry and barrier heights to large-molecule isomerization energies, noncovalent interactions, conformational equilibria, self-interaction errors, heavy p-block chemistry, ion chemistry,… intending to cover all aspects of main-group chemistry. Their reference data had been compiled from high-level ab initio benchmark studies in the literature, supplemented by some new benchmark calculations of their own. A detailed breakdown of the 55 subsets (and full source information for the original reference data) can be found in Table ESI-1 in the electronic supporting information (ESI): suffice to say a full evaluation entails 2,459 electronic structure calculations for 1,499 chemical energy differences.

The reference geometries, charge and multiplicity information, and reference data were extracted from the ACCDB database of Morgante and Peverati[77]. While initial runs were made with the help of the Snakemake[78] workflows defined as part of ACCDB, once we had a full set of input files we elected to use our own scripting. Data analysis was carried out using a Fortran program developed in-house and available on request from the authors.
The primary metric and "objective function" employed is the WTMAD2 (weighted mean absolute deviation, type 2) as defined by Goerigk et al.[55]It seeks to compensate both for the different energy scales various properties are on and for the different sizes of the various subsets:

$$\mathrm{WTMAD2} = \frac{1}{\sum_i^{55} N_i} \cdot \sum_i^{55} N_i \cdot \frac{56.84\ \mathrm{kcal/mol}}{\overline{|\Delta E|}_i} \cdot \mathrm{MAD}_i$$

In which $|\Delta E|_i$ is the mean absolute value of all the reference energies for subset i, N_i the number of systems in the subset, and MAD_i represents the mean absolute difference between calculated and reference reaction energies for subset *i*. We note that MAD is a more "robust statistic"[79] than the root mean square deviation, in the statistical sense that MAD is less prone to hypersensitivity to one or a few "outlier" points than the RMSD (root mean square deviation), even as the latter is more useful for spotting "troublemakers" for the exact same reason.

Electronic structure details

Reference geometries were used "as is" and not optimized further. The Weigend-Ahlrichs def2-QZVPP basis set[80] was used for most systems, except for the subsets WATER27, RG18, IL16, G21EA, and AHB21 where we used the diffuse-function augmented def2-QZVPPD instead,[81] and the large-molecule isomerization subsets C60ISO and UPU23, where we compromised on the def2-TZVPP basis set.[80]

All calculations were carried out using Q-CHEM 5.1.1[82] running on the ChemFarm HPC cluster of the Weizmann Institute Faculty of Chemistry. For GGAs and double hybrids derived from them, initial calculations employed the SG-2 integration grid,[83] which is a pruned (75,302) grid roughly comparable to the (Grid=Fine) in Gaussian; the notation stands for the direct product of a 75-point Euler-Maclaurin radial quadrature[84,85] and a 302-point Lebedev angular grid (see Ref.[86] and references therein). For meta-GGAs and double hybrids derived from them, we employed the larger SG-3 grid, which is a pruned (99,590) grid roughly comparable to (Grid=UltraFine) in Gaussian; ultimately, we also recalculated the GGAs and double-hybrid GGAs from them. As can be seen in the Electronic supporting information (Table ESI-12), the switch to an SG-3 grid makes a major difference for the RG18 rare gas complexes subset, and a minor but nontrivial one for the anionic subsets. For the SCAN (strongly constrained and appropriately normed[10]) meta-GGA — which exhibits a well-known[87] integration grid hypersensitivity — after some experimentation we decided on an unpruned (150,590) grid, which for a subset we checked for convergence against an even larger (200,974) grid.

The combinatorially optimized range-separated hybrid (RSH) GGA ωB97X-V,[21] its RSH meta-GGA successor ωB97M-V,[23] and finally its very recent double hybrid spinoffωB97M(2) were evaluated using their respective implementations in Q-CHEM.[24]
In the GMTKN55 paper, all electrons were correlated in the MP2-like steps of the double hybrids. While the def2-QZVPP basis set used there and in the present work is not really suitable for core-valence correlation, we have calculated statistics both with and without frozen inner-shell orbitals. In both cases, however, we have elected to correlate the subvalence electrons of the metal and metalloid atoms in subsets MB16-43, HEAVY28, HEAVYSB11, ALK8, and ALKBDE10 sets, as the core-valence gaps with default frozen orbital settings are smaller than 1 hartree. Indeed, for alkali and alkali earth oxides and halides, subvalence (n-1)p orbitals may otherwise intrude into the valence band, and thus result in nonsensical dissociation energies with standard frozen-core settings, as discussed at length in e.g., Refs.[88,89]

Optimization details

The BOBYQA (Bound Optimization BY Quadratic Approximation) derivative-free constrained optimizer[90] by Powell was used as the core of a computer program and collection of shell scripts developed in-house.
A DSD double hybrid, if fully optimized, has six empirical parameters:
(a) the fraction of HF exchange $c_{X,HF}$. (The fraction of semilocal DFT exchange is always $c_{X,DFT}=1-c_{X,HF}$.)
(b) the fraction of semilocal DFT correlation $c_{C,DFT}$
(c) the fraction of opposite-spin 2nd-order GLPT correlation energy c_{2ab}
(d) the fraction of same-spin 2nd-order GLPT correlation energy $c_{2ss}= c_{2aa+bb}$
(e) the prefactor s_6 for the D3BJ empirical dispersion correction[31,32,91]
(f) the length scale parameter a_2 for the D3BJ damping function. (As in Refs.[36,37], we are setting $a_1=0$; we are also setting $s_8=0$ as in Refs.[36,37], and in the SCAN-D3BJ[87] paper.)
For a given pair of values for (a,b), it is possible to obtain the optimal group of (c-f) parameters without re-evaluating any electronic structure calculations, simply by extracting individual energy components from the electronic structure calculations, evaluating total energies and hence WTMAD2 for a given combination of $\{c_{2ab},c_{2ss},s_6,a_2\}$, and minimizing WTMAD2 with respect to these four parameters using BOBYQA. This could then constitute an inner "microiteration" loop, while the outer "macroiteration" loop consists of varying$\{c_{X,DFT},c_{C,DFT}\}$ and rerunning all 2,459 calculations with the new parameters. We

considered, however, placing one or both variables in the microiteration loop, with the optimum values from the microiterations to be used in the macroiterations, and so forth until "self-consistency" has been reached. While the coupling between (a) and (c,d) proved to strong for this to be viable for $c_{X,DFT,}$we found that for a fixed value of $c_{X,DFT}$, convergence of $c_{C,DFT}$ to two decimal places or better typically does not require more than two macroiterations. Hence, we have adopted the practice of microiterating$\{c_{C,DFT},c_{2ab},c_{2ss},s_6,a_2\}$ at every macroiteration by means of BOBYQA.

The D3BJ corrections were computed for a thinly spaced grid in a_2 using the standalone DFTD3 program by Grimme and coworkers (https://www.chemie.uni-bonn.de/pctc/mulliken-center/software/dft-d3/). Values for intermediate a_2 were obtained by interpolation. We found, however (see below) that if s_6 is part of the microiterations, then the WTMAD2 surface is sufficiently flat in a2 that fixing a2 at semi-arbitrary values both stabilizes the optimization and has negligible effect on the final WTMAD2.

For a DOD double hybrid, c_{2ss}=0, leaving just four parameters $c_{C,DFT},c_{2ab}$, $s_6,a_2\}$ for the microiterations, while for a DSD-noDisp, s_6=0 and a_2 is irrelevant, leaving just three parameters $\{c_{C,DFT},c_{2ab},c_{2ss}\}$ in the inner loop.

Results and Discussion

WTMAD2 performance metrics over the GMTKN55 dataset are given in Table 1. Results for a large number of GGA, meta-GGAs, and hybrid functionals have been given in the GMTKN55 paper[55] and will not be repeated here. We have repeated the calculations for B3LYP, PBE0, and wB97X-V as a sanity check; in addition, we have evaluated the wB97M-V and wB97M(2) functionals, which were not included in the GMTKN55 study.

The switch from a combinatorially optimized range-separated hybrid GGA (coRSHGGA)in wB97X-V to a coRSH meta-GGA in wB97M-V represents a clear improvement over ωB97X-V, with WTMAD2 going down from 3.96 kcal/mol to 3.29 kcal/mol. This latter figure is the lowest WTMAD2 reported thus far for a hybrid functional: ωB97X-V was the previous best contender in the original GMTKN55 paper. Breakdown by components (Table ESI-11) reveals conspicuous accuracy gains for the pericyclic reaction barriers (BHPERI), for bond separation reactions of saturated hydrocarbons (BSR36), and large system reaction energies more generally. By way of data reduction, we may in fact consider the sums of WTMAD2 contributions for each of the five major subcategories in GMTKN55: Thermochemistry, Intermolecular interactions, Conformers, Barrier heights, and Reaction energies for large systems.

It then becomes apparent (Table 2) that the chief gain for ωB97M-V over ωB97X-V is in fact for thermochemistry and large system reaction energies. There is a small improvement for barrier heights, but no change for intermolecular interactions and in fact a slight deterioration for conformers; upon further inspection, the latter can be attributed primarily to the tri-and tetrapeptide conformers (PCONF21).The already excellent statistic for ωB97M-V can be brought down even further to 2.19 kcal/mol with the coRSH mGGA double hybrid ωB97M(2). This improvement is actually seen for all subcategories across the board. The WTMAD2 value for ωB97M(2) is, by some distance, the lowest reported for any functional thus far: the best performer from the original GMTKN55 paper,[55] DSD-BLYP-D3zero, clocked in at 3.0 kcal/mol.

Table1: WTMAD2 values (kcal/mol) for various functionals using the full GMTKN55 database.

Functionals	WTMAD2(kcal/mol)				
		With Refitted a2			
	Standard/Original	DSD	DOD(c2ss=0)	noDispSD (s6=0)	noDispOD(c2ss=s6=0)
SCAN-D3BJ	7.943				
M06-D3zero	7.749				
SCAN0	7.685				
PBE0-D3BJ	6.551				
B3LYP-D3BJ	6.503				
SCAN0-D3BJ	6.229				
MN15-D3BJ	5.772				
PW6B95-D3BJ	5.488				
revPBE0-D3BJ	5.429				
M062X-D3zero	4.843				
M062X	4.784				
ωB97X-V	3.959				
B2GP-PLYP	3.329				
ωB97M-V	3.286				
Double hybrids without inner-shell correlation					
ωB97M(2)	2.186				
DSD-BLYP-D3BJ	3.522	2.586	2.736	4.134	4.818
DSD-PBEP86-D3BJ	3.282	2.575	2.590	3.577	4.043
DSD66-PBEP86-D3BJ		2.616	2.614	3.823	
DSD-PBE-D3BJ	3.341	2.840	2.848	4.296	
DSD-TPSS-D3BJ	3.296	3.068	3.152	4.240	
DSD-PBEB95-D3BJ	3.444	2.853	2.928	3.456	3.814
DSD-SCAN0-2-D3BJ	4.300[a]	2.891	3.016	3.007	3.306
DSD74-SCAN-D3BJ		2.765	2.834	2.942	3.242
DSD69-SCAN-D3BJ	4.538[b]	2.726	2.752	2.965	3.277
DSD66-SCAN-D3BJ		2.722	2.731	3.038	3.344
DSD63-SCAN-D3BJ		2.744	2.744	3.141	3.437
DSD58-SCAN-D3BJ		2.832	2.832	3.343	3.621
DSD55-SCAN-D3BJ		2.933	2.931	3.536	3.790
DSD50-SCAN-D3BJ	6.087[c]	3.125	3.125	3.854	4.068
Double hybrids, parameters optimized with inner-shell correlation					
SCAN0-2 [standard]	4.686				
ωB97M(2) [standard]	3.506				
DSD-revPBEP86		2.663			
DSD-BLYP-D3BJ		2.470	2.590	3.877	
DSD-PBEP86-D3BJ		2.463	2.470	3.345	
DSD-PBE-D3BJ		2.718	2.724	4.013	
DSD-PBEB95-D3BJ		2.866	2.878	3.381	
DSD-SCAN0-2-D3BJ		2.880	2.926	2.926	
DSD74-SCAN-D3BJ		2.727	2.771	2.831	
DSD69-SCAN-D3BJ		2.660	2.683	2.827	
DSD66-SCAN-D3BJ		2.649	2.658	2.881	
DSD63-SCAN-D3BJ		2.666	2.666	2.969	

DSD58-SCAN-D3BJ		2.756	2.756	3.163	
DSD55-SCAN-D3BJ		2.860	2.864	3.354	
DSD50-SCAN-D3BJ		3.071	3.073	3.682	

* SCAN0-2 can also be written DSD79-SCAN; DSD69-SCAN can also be written as DSD-SCAN-QIDH; DSD55-SCAN as DSD-SCAN-CIDH; (a) SCAN0-2;[92] (b) SCAN0-QIDH;[92] (c) SCAN0-DH.[92]

Table 2: WTMAD2 contribution (kcal/mol) for each of five major subcategories in cases of B3LYP-D3BJ, ωB97X-V, ωB97M-V & ωB97M(2) functionals.

Subcategories	ΔWTMAD2(kcal/mol)			
	B3LYP-D3BJ	ωB97X-V	ωB97M-V	ωB97M(2)
Intermolecular interactions (Intermol)	1.238	0.578	0.565	0.492
Conformers/Intramolecular (Conformer)	1.147	0.729	0.897	0.578
Barrier heights (Barrier)	1.141	0.561	0.454	0.258
Thermochemistry (Thermo)	1.314	1.020	0.730	0.442
Large-species reaction energies (REAClarge)	1.662	1.070	0.640	0.418
Total WTMAD2	**6.503**	**3.959**	**3.286**	**2.187**

ωB97M(2) was parametrized for frozen subvalence orbitals: if correlation from such orbitals were to be included, c2ab=c2ss for this functional would have to be slightly reduced to compensate.[38]Indeed, if we evaluated WTMAD2 with all orbitals correlated and original parametrization, we saw an *increase* to 2.36 kcal/mol.

Sensitivity to the percentage of Hartree-Fock exchange: DSD-SCAN as a case study

The costliest parameter to vary in the refit of a DSD functional would be the percentage of Hartree-Fock exchange. As already shown in Figure 1 of Ref.[37], only minimal changes in performance statistics result from varying the fraction of HF-like exchange $c_{X,HF}$ of a DSD functional within a fairly broad range—but a relatively small training set was sampled there. Presently, we will consider the case of DSD-SCAN$_x$-D3BJ (where x stands for the percentage of HF exchange) in more detail: for the evaluation points, we have chosen $c_{X,HF}=n^{-1/3}$, where n=2,2 ½, 3, 3 ½, 4, 5, 6, and 8. [In the "nonempirical" double hybrids of Adamo and coworkers, these choices would correspond to $c_{2ab}=c_{2ss}=1/n$, and $c_{C,DFT}=1-1/n$ owing to the putative cubic dependence of the integrand.[93,94]] For the cases of n=3, 3½, and 4, this yields values close to percentage points 0.69, 0.66, and 0.63, respectively, and we elected to round off to these latter values.

The WTMAD2 for these trial functionals, with orbitals where $c_{C,DFT}$ was fixed at 0.50 during the iterations (but not during the linear optimization) is depicted in Figure 1.

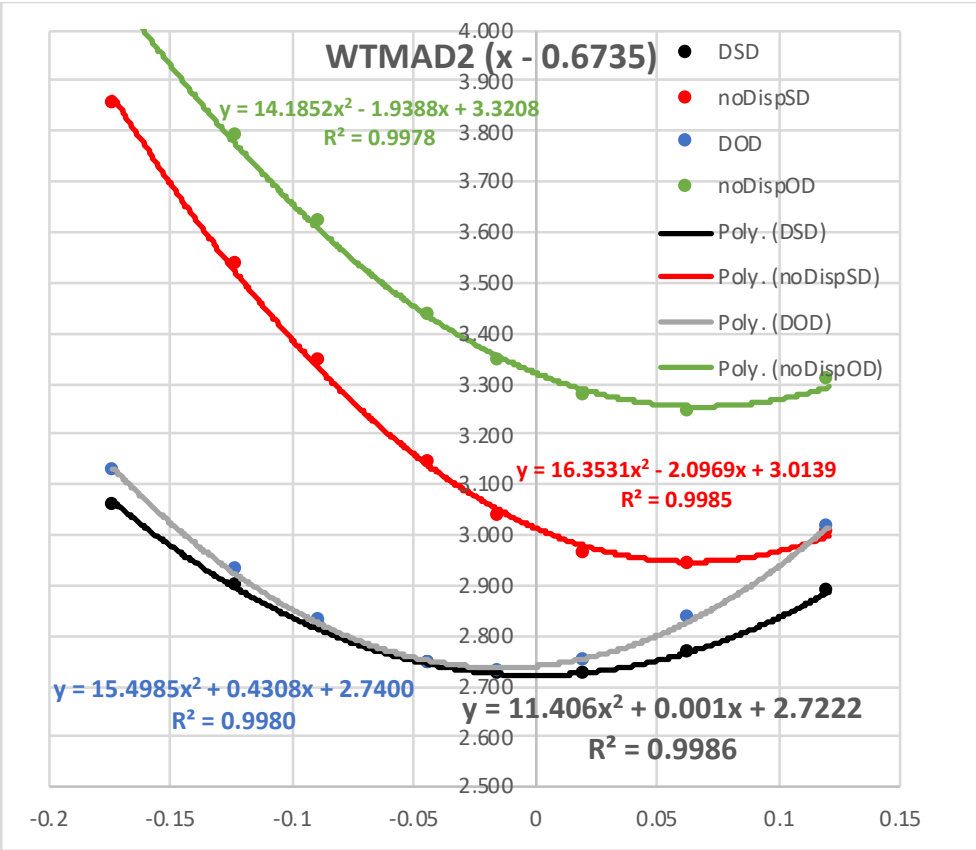

Figure 1: dependence on the fraction of HF exchange x/100 % of WTMAD2 over the GMTKN55 dataset for the dispersion-corrected, spin-component-scaled, double hybrid DSD-SCAN$_x$-D3BJ, as well as the constrained versions DOD (i.e, c_{2ss}=0), noDispSD (i.e., s_6=0), and noDispOD (i.e., s_6=c_{2ss}=0)

As can be seen there, these can be fit excellently (R^2=0.998 or better) by a simple parabola: for the DSD-SCAN-D3BJ curve (black), the minimum is at $c_{X,HF}$=0.6735 and the quadratic coefficient is 11.406. The latter implies a remarkably weak sensitivity of WTMAD2 to the percentage of HF exchange: it will vary by just 0.01 kcal/mol from 0.644 to 0.704, by 0.02 kcal/mol over a range from 0.632 to 0.716, and by just 0.05 kcal/mol between 0.607 and 0.740. What this implies is that nonlinear optimization for a specific minimum $c_{X,HF}$ becomes a somewhat academic exercise: one can choose any sensible fixed value in those ranges, such as 0.69 or 0.66 (or, if one prefers, $3^{-1/3}$ or $3.5^{-1/3}$).

The DOD-SCAN-D3BJ curve (blue) — where c_{2ss}=0 throughout — is only somewhat less flat, with a minimum at x=0.66. We note that where x<0.621 (the crossing point between DSD and DOD WTMAD2 curves), an unconstrained DSD fit actually has a *negative* c_{2ss} — if c_{2ss} were constrained to be non-negative, then DSD would follow the blue line left of the crossing point. If the empirical dispersion correction is eliminated (which we denote here by noDispSD-SCAN), the curve does become a little steeper and the minimum shifts up to x=0.738. If we in addition constrain c_{2ss}=0 (the yellow noDispOD-SCAN curve), we pay a relatively modest accuracy premium — but this is specific to the underlying SCAN semilocal functional. We will address this point further below.

Of course, a single global performance metric such as WTMAD2 does not tell the whole story. A breakdown by components is given in Tables 2–10 of the ESI. A number of these subsets are essentially indifferent to the fraction of HF exchange (particularly the noncovalent interaction sets), while others prefer small HF exchange (e.g. DC13=thirteen

"difficult cases"), yet others (e.g., SIE=self-interaction error) will prefer large HF exchange by design, and a number of the thermochemistry subsets have clearly defined minima.

Following the original GMTKN55 paper and the previous section, we can partition WTMAD2 into the same five primary components. A plot of these as a function of $c_{X,HF}$ is given in Figure 2.

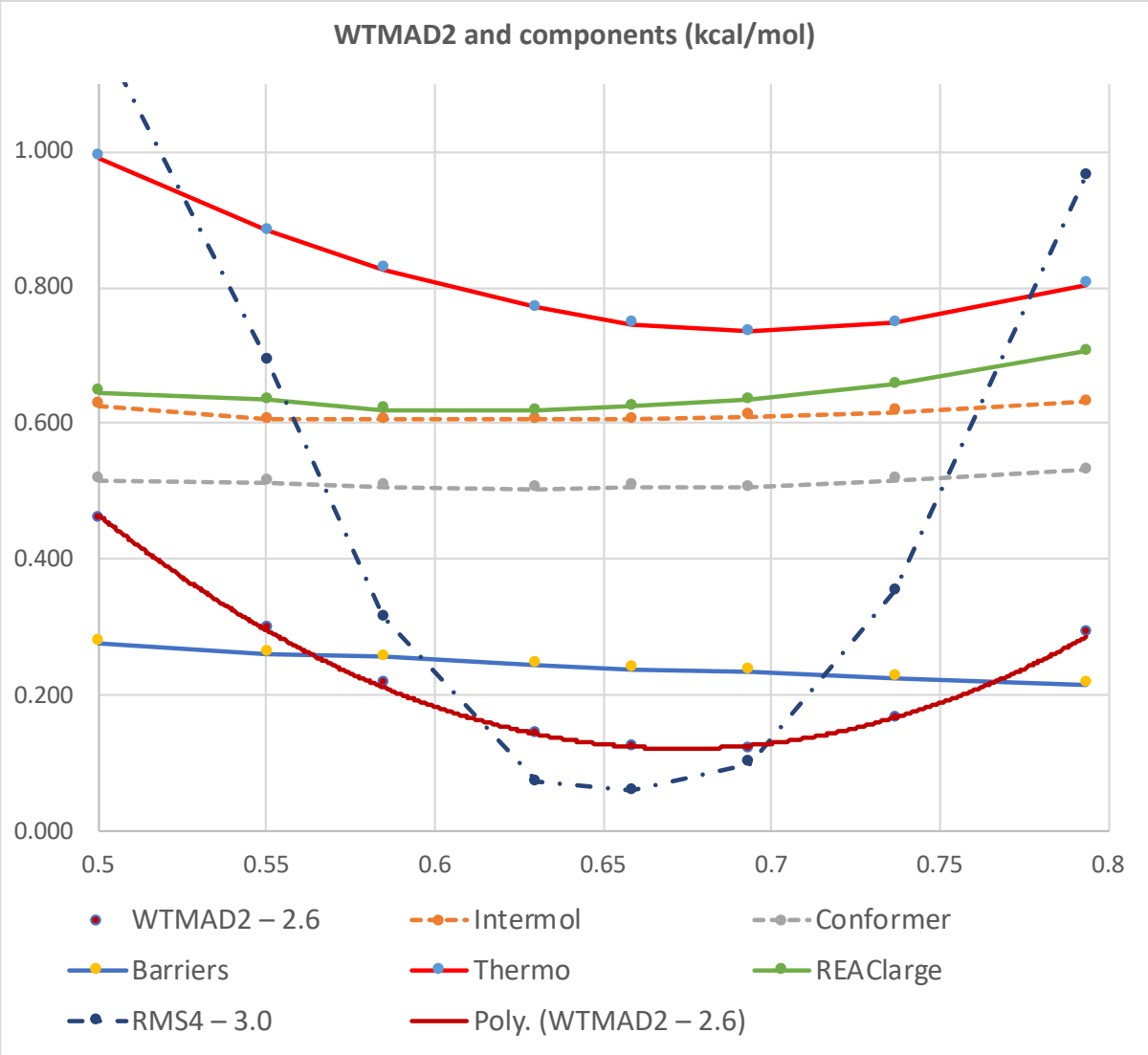

Figure 2: dependence on the fraction of HF exchange x/100 % of WTMAD2 over the GMTKN55 dataset for the dispersion-corrected, spin-component-scaled, double hybrid DSD-SCAN$_x$-D3BJ, as well as of the five major subdivisions thereof and of an objective function similar to Ref.[37] ("Opus 251")

We see there that of these five primary subsets, two are remarkably insensitive to the percentage of HF exchange, namely intermolecular interactions and conformer energies (which are to a large extent driven by intramolecular noncovalent interactions). The greatest variation is seen for thermochemistry, where a clear "valley" exists, which is however nearly flat between 0.66 and 0.74. For barrier heights, error goes down slowly but monotonically as $c_{X,HF}$ goes up. Finally, for the large reaction energies, the profile stays fairly flat but error starts increasing for the largest values. The combination of these factors creates something of a "Goldilocks zone" between 0.65 and 0.70, or even between 0.62 and 0.74, as we noted

earlier: for instance, while 0.69 yields slightly better thermochemistry and barrier heights than 0.66, this is offset by increased errors in the large system isomerizations.

The GMTKN55 minimum is even more shallow than what we previously found[37] in the DSD paper. We do note that instead of weighted MADs, that paper used unweighted RMSDs, which tends to amplify differences; moreover, it focused on a training set with just six subsets. Two of these were transition metal reaction prototypes; the remaining four are identical or similar to the W4-11, BH76, S22, and MB18 subsets from GMTKN55. RMSD4, an average of the latter four's RMSDs, which should behave quite similarly to the training set used in Ref.[37], is displayed as the dot-dashed black line in Figure 2. One indeed sees a much more pronounced variation here, as differences are neither smoothed out by the more robust (in the statistical sense) MAD averaging nor diluted by many noncovalent interaction-driven subsets.

We hence conclude that we may avoid the costly nonlinear optimization of $c_{X,HF}$ for the other DSD-XC functionals, and that we may instead semi-arbitrarily fix $c_{X,HF}$ at either the same value as the original,[37] or choose $0.69 \approx 3^{-1/3}$ as a sensible compromise.

DSD double hybrids and refits

Performance statistics for DSD functionals with original parametrization were already given in the GMTKN55 paper: differences with the values reported there are principally due to slight differences in the basis set and choices of frozen-cores. In the present paper, we will use the notation, e.g., DSD-PBEP86-D3BJ for the original functional and revDSD-PBEP86-D3BJ for the present refit. WTMAD2 statistics can be found in Table 1.

With valence electrons correlated, revDSD-PBE86-D3BJ and revDSD-BLYP-D3BJ are essentially tied at WTMAD2=2.58 and 2.59 kcal/mol, respectively. While this is still higher than ωB97M(2), it should be kept in mind that DSD-BLYP-D3refit and DSD-PBEP86-D3refit only entail six adjustable parameters rather than sixteen, rendering them less "empirical". Of these six, $c_{X,HF}$ can be fixed at a semi-arbitrary value; furthermore, we found here that the same is true of the damping function turnover point a_2 of the D3BJ dispersion term. In fact, the surface is so flat in a2 that its inclusion as an optimization parameter leads to poor convergence of WTMAD2: we hence fixed a_2 at semi-arbitrary value of 5.5 for most functionals, 5.2 for short-ranged correlation functionals like LYP and P86, and 5.75 for the longer-ranged SCAN (Table ESI-16).

This leaves us arguably with only four true empirical parameters.

It was previously noted, in the original DSD papers, that "upgrading" the underlying semilocal functional from a GGA to a meta-GGA is not necessarily beneficial, with DSD-TPSS-D3refit for instance being among the poorer performers. At first sight, this observation holds true here as well: it is also notable that DSD-PBE-D3refit does noticeably more poorly than DSD-PBEP86-D3refit and DSD-BLYP-D3refit. What the two best performers have in common are fairly "short-ranged" semilocal correlation functionals, which at long range at least "do no harm" and leave the treatment of dispersion up to the MP2-like correlation and the D3BJ correction. In contrast, as noted in Ref.[95], the PBE_c correlation functional exhibits a spurious attraction at long range, and so does TPSS.

The very recent SCAN (strongly constrained and appropriately normed[10]) meta-GGA, in contrast, exhibits much better performance in a DSD context than TPSS — second only to DSD-BLYP and DSD-PBEP86..

The great improvement from the original DSD-PBEP86-D3 (WTMAD2=3.28 kcal/mol) to revDSD-PBEP86-D3serves as a cautionary tale against small and idiosyncratic training sets. All four of the main-group training sets for the DSD functional are part of GMTKN55: if we were to instead consider the sum of the four WTMAD2 contributions for

W4-11, RG18, S22, and BH76, we would actually find essentially no improvement from DSD-PBEP86-D3 to revDSD-PBEP86-D3: the large-system subsets are what makes the difference.

Is there any benefit to be gained from correlating the inner-shell orbitals? (We have noted previously[38] that DSD-BLYP parameters fitted with and without inner-shell correlation are slightly different.) We have only considered this for a subset of double hybrids. As seen in Table 1, about 0.1 kcal/mol can be gained in WTMAD2: in-depth analysis (Table ESI-14) reveals that this improvement is primarily confined to the subsets BSR36 (bond separation reactions), HAL59 (halogen bonding), and HEAVY28 (heavy p-block compounds). The importance of subvalence correlation for halogen-bonded species has been shown previously[63] for the X40x10 dataset.[60]

The popular simple hybrids B3LYP-D3BJ and PBE0-D3BJ are nearly tied at WTMAD2=6.5 kcal/mol. Replacing PBE exchange by Weitao Yang's revision (revPBE[96]) leads to a significant improvement to WTMAD2=5.43 kcal/mol for revPBE0-D3BJ. (A detailed comparison can be found in Table ESI-15. This functional was not considered in the GMTKN55 paper.) Intriguingly, making a similar substitution in the double hybrid to create DSD-revPBEP86-D3 does *not* yield improved performance (WTMAD2=2.66 kcal/mol, compared to 2.46 for DSD-PBEP86-D3 — both without frozen cores).

At the end of this section, we can report that *revDSD-BLYP-D3 and revDSD-PBEP86-D3 are essentially tied for best performer*. We shall see below that the tie is broken when the D4 correction is introduced.

Eliminating the semiempirical dispersion correction

The presence of the D3BJ dispersion correction[31,32] in DSD functionals exposes them to the criticism of "mixing DFT with molecular mechanics". We have earlier considered the option of eliminating D3BJ entirely: in practice, this entails an increased percentage of same-spin correlation as compensation.

As can be seen in Table 1, all the DSD-noD3 functionals exhibit significantly degraded performance compared to their DSD-D3 siblings. In addition, however, the ordering is upended: DSD-SCAN now exhibits significantly better performance than the other options. With core-valence correlation included, DSD-SCAN-QIDH reaches WTMAD2=2.84 kcal/mol, still significantly better than the original DSD-PBEP86-D3.

Why, in a DSD-XC-noDisp functional, does XC=SCAN outperform all other options? At long range, three scenarios are possible: (a) the functional tapers away quickly (like in BLYP and PBEP86), which at least does no harm but in the absence of a D3BJ correction leaves PT2 to handle all the long-range dispersion effects (for which it is inadequate); (b) the functional does not decay quickly but has the wrong behavior, leading to poor performance for noncovalent interactions; (c) the functional can at least partly recover the correct behavior, in which case PT2 may be sufficient to handle the remainder. It appears that (c) is the case for SCAN.

As a proxy for behavior at intermediate distance, we may consider the s_8 coefficient for the r^{-8} term in D3BJ: for MP2, this was found[69] to be large and negative, while for DSD double hybrids and functionals like M06, s8 was found to be zero or statistically insignificantly different from zero. We note in particular that SCAN-D3BJ, unlike other GGAs and meta-GGA, has a fitted $s_8=0$ in the D3BJ correction[87]; in the present work, we also fitted a D3BJ correction to the SCAN0 hybrid, and obtained $a_1=0$, $a_2=7.9042$, $s_8=0$.

DOD double hybrids and refit

If only opposite-spin MP2 correlation is included, then the cost scaling of the post-KS step can be reduced to $\propto N^4$ formally by means of a Laplace transform algorithm.[97–99] In fact, Song and Martinez[100] achieved further reduction to $\propto N^3$ using tensor hypercontraction techniques. In our previous work,[37] we have denoted such functionals DOD, short for Dispersion-corrected, Opposite-spin, Double-hybrids.

(As was expected and can be seen in Table 1, the results of trying to eliminate both same-spin correlation and the dispersion correction were dismal for most functionals, with the exception of noDispOD-SCAN$_{74}$ where WTMAD=3.3 kcal/mol. Still, this is a performance comparable with wB97M-V, which does not require evaluation of E_{2ab}.)

The tie between revDSD-BLYP and revDSD-PBEP86 is broken in favor of the latter: Inspection reveals that revDSD-PBEP86-D3 functionals has a c_{2ss} coefficient close to zero, which is not the case for revDSD-BLYP. (The latter is plausible when one considers that BLYP does not treat opposite-spin and same-spin correlation on the same footing: in fact, it is easily seen from eq. (2) in Ref.[101] that the BLYP correlation energy for a fully polarized uniform electron gas is zero, which is clearly an unphysical answer.) Hence, we see that revDOD-PBEP86-D3 "pulls ahead of" revDOD-BLYP-D3in the "WTMAD2 race", at 2.59 kcal/mol with frozen cores and 2.47 kcal/mol with core correlation (Table 1).

Final recommended D3BJ functionals

In light of the above, we only are retaining three exchange-correlation combinations for the semilocal part: BLYP, PBEP86, and SCAN. For PBEP86, both DSD and DOD combinations are given: while DSD represents only a very small improvement over DOD, it comes at zero additional computational expense when using a code that cannot exploit reduced-scaling algorithms for opposite-spin-only MP2. (The most commonly used such codes are Gaussian 09 and Gaussian 16.) Hence we have elected to recommend both DSD-PBEP86-D3rev2 and DOD-PBEP86-D3rev2.

In view of the weak dependence of performance on the percentage of HF exchange, we have elected to retain the original percentages of 71% for DSD-BLYP and 69% for DSD-PBEP86, in order to simplify nonstandard inputs for codes such as Gaussian, ORCA, and Q-CHEM. For DSD-SCAN and DOD-SCAN, we have no such incentive: following inspection of the minima of the parabolic fits to WTMAD2, we have chosen 66% for DOD-SCAN and 69% for DSD-SCAN-noDisp.

Finally, we have noted above that, as long as the dispersion prefactor s6 is self-consistently optimized with the other parameters, WTMAD2 is only weakly dependent on the chosen turnover parameter a_2. With a given fixed a_2, DSD-BLYP has the largest s6, followed by DSD-PBEP86: for DSD-SCAN and DOD-SCAN, s_6 is much smaller—this reflects that SCAN is better able to handle long-range effects than the two others. We have hence semi-arbitrarily fixed a_2 at a "short" value of 5.2 for DSD-BLYP, and at a "longer" values of 5.75 for DSD-SCAN, while for DSD-PBEP86 we chose an intermediate a_2=5.5. Comparison with full optimizations including a_2 revealed that WTMAD2 differences are on the other of a few "small" calories per mole: hence fixing these parameters is considered justifiable in light of a more smoothly converging optimization for the remaining ones.

Sample input files for most major codes are given in the ESI. We wish to point out that, while DSD-SCAN-noD is inferior to the DSDrev2 offerings, its WTMAD2 is still superior to the original B2GP-PLYP and DSD double hybrids, and this without any empirical dispersion correction and with just three non-arbitrary parameters.

Table 3: Final parameters for revised DSD-D3BJ functionals. Original parameters, if any, given for comparison

Functionals	$c_{X,HF}$	$c_{C,DFT}$	c_{2ab}	c_{2ss}	s_6	a_1	a_2	WTMAD2 (kcal/mol)
noDispSD-SCAN$_{69}$	0.69	0.4409	0.6228	0.2424	[0]	—	—	2.965
DOD-SCAN$_{66}$-D3BJ	0.66	0.5014	0.6302	[0]	0.2935	[0]	5.75	2.733
revDSD-BLYP-D3BJ	0.71	0.5402	0.5411	0.1944	0.5233	[0]	5.2	2.578
revDSD-PBEP86-D3	0.69	0.4362	0.5693	0.0801	0.4283	[0]	5.5	2.572
revDOD-PBEP86-D3	0.69	0.4475	0.5984	[0]	0.4727	[0]	5.5	2.613
Original[37] DSD-BLYP-D3	0.71	0.54	0.47	0.40	0.57	[0]	5.4	3.522
Original[37] DSD-PBEP86-D3	0.69	0.44	0.52	0.22	0.48	[0]	5.6	3.282
B2GP-PLYP-D3BJ[39,55]	0.65	0.64	0.36	0.36	0.56	0.2597	6.333	3.329

Considering the D4 dispersion correction: Final recommended D4 functionals

As the present manuscript was being prepared for publication, a preprint[34] by Grimme and coworkers was posted on ChemRXiv, in which they propose a next-generation D4 dispersion correction (see also Ref.[33]). The reader is referred to these references for details; for the purpose of our discussion, the most significant difference between D3BJ and D4 is that the latter introduces dependence on atomic partial charges, which (by default) are evaluated using the electronegativity equalization principle.[102] (For the general theory, see Refs.[103,104] and references therein.)

As a first step, we substituted the D4 correction for D3BJ in the original DSD functionals from Ref.[37] as a "drop-in replacement" using parameters optimized for these functionals and published by Grimme et al.[34] The results can be found in the third numerical column of Table 4 and, for individual GMTKN55 subsets, in Table ESI-13. Across the board, the WTMAD2 values are significantly better than those with the original, in the case of DSD-PBE even superior to the refitted revDSD-PBE-D3BJ!

We then proceeded to reoptimize the DSD functionals in the presence of D4 and adjusting the latter's parameters. It quickly became clear that setting s_8 to zero had negligible impact on the WTMAD2: furthermore, that the other parameters settled around a_1=0.4 and a_2=3.6, and that one could actually choose these 'semi-arbitrary values' across the board, leaving the same four adjustable parameters $c_{C,DFT}$, c_{2ab}, c_{2ss}, and s_6 as in the revDSD-XC-D3BJ cases.

revDSD-PBEP86-D4 in particular shines, with WTMAD2=2.375 kcal/mol, quite close to the wB97M(2) functional with its 16 adjustable parameters. But revDSD-SCAN$_{66}$-D4, revDSD-PBE-D4, and revDSD-PBEB95-D4 all likewise outperform their revDSD-XC-D3BJ counterparts, and revDSD-BLYP-D4 marginally bests revDSD-BLYP-D3BJ.

Three of the refitted functionals have c_{2ss} values close to zero, hence we also performed revDOD-PBE-D4, DOD-SCAN$_{66}$-D4, and revDOD-PBEB95-D4 fits in case one wants to exploit the reduced-scaling algorithms for opposite-spin-only MP2. For revDOD-BLYP-D4, there is substantial loss in performance, but revDOD-PBEP86-D4, at WTMAD2=2.395 kcal/mol, only sacrifices 0.02 kcal/mol compared to its DSD counterpart. Next in performance is revDOD-PBE-D4 at WTMAD2=2.49 kcal/mol, then followed by DOD-SCAN-D4 at 2.61 kcal/mol.

Inspection of the contributions to WTMAD2 reveals that the improvement from revDSD-PBEP86-D3BJ to revDSD-PBEP86-D4 is mostly due to the intermolecular interaction components, and somewhat due to conformers. In DOD-SCAN66-D4, on the other hand, intermolecular interactions and conformers are improved about equally.

Iron and Janes[105] have very recently examined the performance of hybrid and double-hybrid functionals for their newly developed MOBH35 transition metal barrier heights benchmark as well as for the older MOR41 organometallic reaction energy benchmark.[106] There, the SCAN-based functionals were found to be superior to the others for these applications, even though overall the ωB97M-V functional outperformed all double hybrids except PWPB95.[107] Detailed inspection of the double-hybrid results revealed a number of outliers for system that exhibit some degree of static correlation: apparently, the PT2 correlation is insufficiently resilient to that. The use of dRPA (direct random phase approximation) as an alternative, as proposed by Mezei et al.,[108,109] will be explored in future work. (The use of perturbation theory higher than second order was considered by Chan and Radom[110] and found to yield essentially no performance benefit.) Results for the ωB97M(2) functional were not given in that paper. For the sake of completeness, we carried out these calculations ourselves using the def2-TZVPP basis set used in Ref. [105] for the other double hybrids: for MOR41, [106] we obtain MAD=2.8 and RMSD=3.5 kcal/mol, while for MOBH35,[105] we obtain MAD=2.4 and RMSD=4.0 kcal/mol.

Table 4: Final parameters for revDSD-D4 functionals and comparison of WTMAD2 (kcal/mol) with original double-hybrids (D3BJ), ditto with drop-in replacement of D3BJ by D4, and revD3BJ

Functionals	WTMAD2(kcal/mol)				Parameters							
	D3BJ	revD3BJ	D4	revD4	$c_{X,HF}$	$c_{C,DFT}$	c_{2ab}	c_{2ss}	s_6	s_8	a_1	a_2
DSD-PBEP86	3.282	2.575	2.678	2.376	0.69	0.4251	0.5857	0.0580	0.5027	0	0.44	3.60
With core corr. DSD-PBEP86				2.307	0.69	0.4038	0.5979	0.0571	0.4612	0	0.44	3.60
DSD-PBE	3.341	2.840	2.747	2.488	0.68	0.4484	0.6009	0.0293	0.6608	0	0.4	3.6
DSD-BLYP	3.522	2.586	2.900	2.575	0.71	0.5250	0.5547	0.1890	0.5995	0	0.38	3.52
DSD-SCAN	—	2.722	—	2.610	0.66	0.4844	0.6309	0.0125	0.3110	0	0.4	3.6
DSD-PBEB95	3.444	2.853	3.043	2.670	0.66	0.4614	0.5290	0.0469	0.4587	0	0.42	2.93
xDSD-PBEP86				**2.225**	0.69	0.3989[a]	0.6073	0.0513	0.4238	0	0.44	3.60
xDOD-PBEP86				**2.247**	0.69	0.4071[a]	0.6261	0	0.4561	0	0.44	3.60
DOD-PBEP86				2.395	0.69	0.4335	0.6079	0	0.5407	0	0.44	3.60
DOD-PBE				2.490	0.68	0.4540	0.6097	0	0.6771	0	0.4	3.6
DOD-SCAN				2.612	0.66	0.4895	0.6343	0	0.3190	0	0.4	3.6
DOD-PBEB95				2.685	0.66	0.4677	0.5521	0	0.4786	0	0.42	2.93
DOD-BLYP				2.768	0.71	0.5598	0.6312	0	0.6958	0	0.38	3.52

a. During iterations, $c_{C,DFT}$=1.00 as for all xDSD functionals

Coming back to GMTKN55: can we improve over revDSD-PBEP86-D4, at WTMAD2=2.376 kcal/mol? As above, we found a minor improvement over revDSD-PBEP86-D3BJ when the frozen-core approximation was not made, we attempted the same here, and found that revDSD-PBEP86-D4(noFC) has a slightly lower WTMAD2=2.307 kcal/mol. Inspection of the components (Table ESI-17) reveals that, as above, most of the improvement derives from just three subsets: BSR36 bond separation reactions (the lion's share of the improvement), HAL59 (halogen bonding), and HEAVY28 (heavy p-block compounds).

Then we attempted one more thing that also is present in wB97M(2): we evaluated the KS orbitals with full DFT correlation, akin to the XYG3 family of functionals[40]. Previously, we have found[42] for much smaller training sets that (a) typically error metrics go through a minimum when the percentage of HF exchange in the final energy evaluation is at or near that used for determining the orbitals (leading to what we have termed[42] xDSD functionals); (b) the improvement seen from DSD to xDSD was small and its statistical significance uncertain. Hait and Head-Gordon have discussed some downsides of the xDSD and XYG3 type functionals for nonequilibrium geometries.[43]

In the present work, we have obtained an xrevDSD-PBEP86-D4(noFC) functional fitted to the GMTKN55 dataset. Parameters are given in Table 4: the WTMAD2 obtained, 2.22 kcal/mol, is the lowest of any functional optimized here, and *is statistically equivalent to the 2.18 kcal/mol of the highly empirical ωB97M(2) functional despite the much smaller number of parameters*. Detailed inspection of Table ESI-17 reveals that the improvement over the revDSD-PBEP86-D4 functional mostly comes from just one subset, namely, the radical stabilization energies in RSE43. Turning to the five major categories, ωB97M(2) outperforms xrevDSD-PBEP86-D4(noFC) for thermochemistry and is in turn outperformed for conformer energies, while there is little to choose between them for intermolecular interaction energies, barrier heights, and large system reactions.

Can we eliminate the same-spin correlation and thus enable the reduced-scaling MP2 algorithms (as well as one more empirical parameter)? As seen in Table 4, the WTMAD2 for xrevDOD-PBEP86-D4(noFC) is just 0.02 kcal/mol higher at 2.244 kcal/mol, We have, hence, a functional comparable to ωB97M(2) in quality that is amenable to reduced $O(N^4)$ or $O(N^3)$ MP2 scaling, unlike ωB97M(2) which has $c_{2ab}=c_{2ss}$. This may be relevant in application to larger systems than considered presently. (The largest species in GMTKN55 has 'only' 83 atoms.)

Conclusions

Having made an extensive survey of DSD double hybrid (and some other) functionals with the aid of the GMTKN55 dataset, we are in a position to state the following observations:

• the combinatorially optimized wB97M-V is "best in class" for fourth-rung exchange-correlation functionals, and approaches performance of double hybrids like B2GP-PLYP-D3;

• the combinatorially optimized wB97M(2) yields the lowest WTMAD2 metric of any functional in existence, making it best in class for double hybrids

• in a DSD double hybrid context, the very recent D4 dispersion correction is clearly superior over D3BJ, presumably owing to the newly introduced partial charge dependence;

• while wB97M(2) has sixteen empirical parameters, refitted revDSD-PBEP86-D3BJ comes close in performance, while xrevDSD-PBEP86-D4 and xrevDOD-PBEP86-D4 are essentially equivalent in quality to xB97M(2). Out of their reduced number of parameters, the percentage of HF exchange $c_{X,HF}$and the damping function parameters a_1,a_2 can be fixed at semi-arbitrary values (as the WTMAD2 surface is fairly flat in them), leaving just four true optimization parameters for revDSD and three for revDOD. The revDOD option permits the use of reduced-scaling MP2 algorithms, which might prove useful for large systems;'

• for the underlying semilocal functional in double hybrids, any good exchange functional appears to work well, while simple correlation functionals that rapidly "get out of the way" at long distances appear to work best (e.g., P86[111] and LYP[112]).

• if one wishes to avoid the D3 or D4 corrections, however, DSD-SCAN appears to work by far the best. Here, the number of empirical parameters is down to four, one of which semi-arbitrary.

• refitting of the DSD functionals to the GMTKN55 database very substantially improves their accuracy particularly for noncovalent interactions and large-system reactions. This

serves as a cautionary tale about the use of small, idiosyncratic training sets for empirical functionals.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpca.xxyyyyy.

Tables ESI-1 through ESI-17 (PDF).

Acknowledgments

This research was supported by the Israel Science Foundation (grant 1358/15) and by the Minerva Foundation, Munich, Germany, as well as by two internal Weizmann Institute funding sources: the Helen and Martin Kimmel Center for Molecular Design and a research grant from the estate of Emile Mimran. N.S. acknowledgesa doctoral fellowship from the Feinberg Graduate School (WIS).

Author Information

Corresponding Author
J. M. L. Martin. E-mail: gershom@weizmann.ac.il. Fax: +972 8 934 3029.
ORCID
Jan M. L. Martin: 0000-0002-0005-5074
Notes
The authors declare no competing financial interest.

References

(1) Dral, P. O.; Wu, X.; Spörkel, L.; Koslowski, A.; Thiel, W. Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Benchmarks for Ground-State Properties. *J. Chem. Theory Comput.* **2016**, *12* (3), 1097–1120. https://doi.org/10.1021/acs.jctc.5b01047.

(2) DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. The Correlation Consistent Composite Approach (CcCA): Efficient and Pan-Periodic Kinetics and Thermodynamics. In *Advances in the Theory of Atomic and Molecular Systems (Progress in Theoretical Chemistry and Physics, Vol. 19)*; Piecuch, P., Maruani, J., Delgado-Barrio, G., Wilson, S., Eds.; Progress in Theoretical Chemistry and Physics; Springer Netherlands: Dordrecht, 2009; Vol. 19, pp 197–224. https://doi.org/10.1007/978-90-481-2596-8_9.

(3) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 Theory. *J. Chem. Phys.* **2007**, *126* (8), 084108. https://doi.org/10.1063/1.2436888.

(4) Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A Complete Basis Set Model Chemistry. VI. Use of Density Functional Geometries and Frequencies. *J. Chem. Phys.* **1999**, *110* (6), 2822. https://doi.org/10.1063/1.477924.

(5) Martin, J. M. L.; de Oliveira, G. Towards Standard Methods for Benchmark Quality Ab Initio Thermochemistry—W1 and W2 Theory. *J. Chem. Phys.* **1999**, *111* (5), 1843–1856. https://doi.org/10.1063/1.479454.

(6) Karton, A.; Martin, J. M. L. Explicitly Correlated W n Theory: W1-F12 and W2-F12. *J.*

Chem. Phys. **2012**, *136* (12), 124114. https://doi.org/10.1063/1.3697678.
(7) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11* (5), 2087–2096. https://doi.org/10.1021/acs.jctc.5b00099.
(8) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77* (18), 3865–3868. https://doi.org/10.1103/PhysRevLett.77.3865.
(9) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta–Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91* (14), 146401. https://doi.org/10.1103/PhysRevLett.91.146401.
(10) Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.* **2015**, *115* (3), 036402. https://doi.org/10.1103/PhysRevLett.115.036402.
(11) DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. The Correlation Consistent Composite Approach (CcCA): An Alternative to the Gaussian-n Methods. *J. Chem. Phys.* **2006**, *124* (11), 114104. https://doi.org/10.1063/1.2173988.
(12) Martin, J. M. L.; Parthiban, S. W1 and W2 Theories, and Their Variants: Thermochemistry in the KJ/Mol Accuracy Range. In *Quantum-Mechanical Prediction of Thermochemical Data*; Cioslowski, J., Ed.; Understanding Chemical Reactivity; Kluwer Academic Publishers: Dordrecht, 2002; Vol. 22, pp 31–65. https://doi.org/10.1007/0-306-47632-0_2.
(13) Becke, A. D. Density-Functional Thermochemistry. V. Systematic Optimization of Exchange-Correlation Functionals. *J. Chem. Phys.* **1997**, *107* (20), 8554–8560. https://doi.org/10.1063/1.475007.
(14) Hamprecht, F. A.; Cohen, A. J.; Tozer, D. J.; Handy, N. C. Development and Assessment of New Exchange-Correlation Functionals. *J. Chem. Phys.* **1998**, *109* (15), 6264–6271. https://doi.org/10.1063/1.477267.
(15) Boese, A. D.; Martin, J. M. L. Development of Density Functionals for Thermochemical Kinetics. *J. Chem. Phys.* **2004**, *121* (8), 3405–3416. https://doi.org/10.1063/1.1774975.
(16) Zhao, Y.; Truhlar, D. G. Density Functionals with Broad Applicability in Chemistry. *Acc. Chem. Res.* **2008**, *41* (2), 157–167. https://doi.org/10.1021/ar700111a.
(17) Yu, H. S.; He, X.; Li, S. L.; Truhlar, D. G. MN15: A Kohn–Sham Global-Hybrid Exchange–Correlation Density Functional with Broad Accuracy for Multi-Reference and Single-Reference Systems and Noncovalent Interactions. *Chem. Sci.* **2016**, *7* (8), 5032–5051. https://doi.org/10.1039/C6SC00705H.
(18) Becke, A. D. Density-functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652. https://doi.org/10.1063/1.464913.
(19) Adamson, R. D.; Gill, P. M. W.; Pople, J. A. Empirical Density Functionals. *Chem. Phys. Lett.* **1998**, *284* (1–2), 6–11. https://doi.org/10.1016/S0009-2614(97)01282-7.
(20) Boese, A. D.; Handy, N. C. A New Parametrization of Exchange-Correlation Generalized Gradient Approximation Functionals. *J. Chem. Phys.* **2001**, *114* (13), 5497–5503. https://doi.org/10.1063/1.1347371.
(21) Mardirossian, N.; Head-Gordon, M. ΩB97X-V: A 10-Parameter, Range-Separated Hybrid, Generalized Gradient Approximation Density Functional with Nonlocal Correlation, Designed by a Survival-of-the-Fittest Strategy. *Phys. Chem. Chem. Phys.*

2014, *16* (21), 9904–9924. https://doi.org/10.1039/c3cp54374a.
(22) Mardirossian, N.; Head-Gordon, M. Mapping the Genome of Meta-Generalized Gradient Approximation Density Functionals: The Search for B97M-V. *J. Chem. Phys.* **2015**, *142* (7). https://doi.org/10.1063/1.4907719.
(23) Mardirossian, N.; Head-Gordon, M. ΩB97M-V: A Combinatorially Optimized, Range-Separated Hybrid, Meta-GGA Density Functional with VV10 Nonlocal Correlation. *J. Chem. Phys.* **2016**, *144* (21), 214110. https://doi.org/10.1063/1.4952647.
(24) Mardirossian, N.; Head-Gordon, M. Survival of the Most Transferable at the Top of Jacob's Ladder: Defining and Testing the ω B97M(2) Double Hybrid Density Functional. *J. Chem. Phys.* **2018**, *148* (24), 241736. https://doi.org/10.1063/1.5025226.
(25) Perdew, J. P.; Schmidt, K. Jacob's Ladder of Density Functional Approximations for the Exchange-Correlation Energy. In *AIP Conference Proceedings*; Van Doren, V., Van Alsenoy, C., Geerlings, P., Eds.; AIP Conference Proceedings; AIP: Antwerp (Belgium), 2001; Vol. 577, pp 1–20. https://doi.org/10.1063/1.1390175.
(26) Kümmel, S.; Kronik, L. Orbital-Dependent Density Functionals: Theory and Applications. *Rev. Mod. Phys.* **2008**, *80* (1), 3–60. https://doi.org/10.1103/RevModPhys.80.3.
(27) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. Doubly Hybrid Meta DFT: New Multi-Coefficient Correlation and Density Functional Methods for Thermochemistry and Thermochemical Kinetics. *J. Phys. Chem. A* **2004**, *108* (21), 4786–4791. https://doi.org/10.1021/jp049253v.
(28) Grimme, S. Semiempirical Hybrid Density Functional with Perturbative Second-Order Correlation. *J. Chem. Phys.* **2006**, *124* (3), 034108. https://doi.org/10.1063/1.2148954.
(29) Görling, A.; Levy, M. Exact Kohn-Sham Scheme Based on Perturbation Theory. *Phys. Rev. A* **1994**, *50* (1), 196–204. https://doi.org/10.1103/PhysRevA.50.196.
(30) Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27* (15), 1787–1799. https://doi.org/10.1002/jcc.20495.
(31) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15), 154104. https://doi.org/10.1063/1.3382344.
(32) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32* (7), 1456–1465. https://doi.org/10.1002/jcc.21759.
(33) Grimme, S.; Bannwarth, C.; Caldeweyher, E.; Pisarek, J.; Hansen, A. A General Intermolecular Force Field Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Phys.* **2017**, *147* (16), 161708. https://doi.org/10.1063/1.4991798.
(34) Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Grimme, S. A Generally Applicable Atomic-Charge Dependent London Dispersion Correction Scheme. *ChemRXivt* **2019**.
(35) Vydrov, O. A.; Van Voorhis, T. Nonlocal van Der Waals Density Functional: The Simpler the Better. *J. Chem. Phys.* **2010**, *133* (24), 244103. https://doi.org/10.1063/1.3521275.
(36) Kozuch, S.; Martin, J. M. L. DSD-PBEP86: In Search of the Best Double-Hybrid DFT with

Spin-Component Scaled MP2 and Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2011**, *13* (45), 20104. https://doi.org/10.1039/c1cp22592h.
(37) Kozuch, S.; Martin, J. M. L. Spin-Component-Scaled Double Hybrids: An Extensive Search for the Best Fifth-Rung Functionals Blending DFT and Perturbation Theory. *J. Comput. Chem.* **2013**, *34* (27), n/a-n/a. https://doi.org/10.1002/jcc.23391.
(38) Kozuch, S.; Gruzman, D.; Martin, J. M. L. DSD-BLYP: A General Purpose Double Hybrid Density Functional Including Spin Component Scaling and Dispersion Correction. *J. Phys. Chem. C* **2010**, *114* (48), 20801–20808. https://doi.org/10.1021/jp1070852.
(39) Karton, A.; Tarnopolsky, A.; Lamère, J.-F.; Schatz, G. C.; Martin, J. M. L. Highly Accurate First-Principles Benchmark Data Sets for the Parametrization and Validation of Density Functional and Other Approximate Methods. Derivation of a Robust, Generally Applicable, Double-Hybrid Functional for Thermochemistry and Thermochemical . *J. Phys. Chem. A* **2008**, *112* (50), 12868–12886. https://doi.org/10.1021/jp801805p.
(40) Su, N. Q.; Xu, X. The XYG3 Type of Doubly Hybrid Density Functionals. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2016**, *6* (6), 721–747. https://doi.org/10.1002/wcms.1274.
(41) Goerigk, L.; Grimme, S. A Thorough Benchmark of Density Functional Methods for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670–6688. https://doi.org/10.1039/c0cp02984j.
(42) Kesharwani, M. K.; Kozuch, S.; Martin, J. M. L. Comment on “Doubly Hybrid Density Functional XDH-PBE0 from a Parameter-Free Global Hybrid Model PBE0” [J. Chem. Phys. 136, 174103 (2012)]. *J. Chem. Phys.* **2015**, *143* (18), 187101. https://doi.org/10.1063/1.4934819.
(43) Hait, D.; Head-Gordon, M. Communication: XDH Double Hybrid Functionals Can Be Qualitatively Incorrect for Non-Equilibrium Geometries: Dipole Moment Inversion and Barriers to Radical-Radical Association Using XYG3 and XYGJ-OS. *J. Chem. Phys.* **2018**, *148* (17), 171102. https://doi.org/10.1063/1.5031027.
(44) Goerigk, L.; Grimme, S. Double-Hybrid Density Functionals. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4* (6), 576–600. https://doi.org/10.1002/wcms.1193.
(45) Sancho-García, J. C.; Adamo, C. Double-Hybrid Density Functionals: Merging Wavefunction and Density Approaches to Get the Best of Both Worlds. *Phys. Chem. Chem. Phys.* **2013**, *15* (35), 14581. https://doi.org/10.1039/c3cp50907a.
(46) Su, N. Q.; Zhu, Z.; Xu, X. Doubly Hybrid Density Functionals That Correctly Describe Both Density and Energy for Atoms. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (10), 2287–2292. https://doi.org/10.1073/pnas.1713047115.
(47) Brémond, E.; Ciofini, I.; Sancho-García, J. C.; Adamo, C. Nonempirical Double-Hybrid Functionals: An Effective Tool for Chemists. *Acc. Chem. Res.* **2016**, *49* (8), 1503–1513. https://doi.org/10.1021/acs.accounts.6b00232.
(48) Benighaus, T.; Distasio, R.; Lochan, R.; Chai, J.-D.; Head-Gordon, M. Semiempirical Double-Hybrid Density Functional with Improved Description of Long-Range Correlation. *J. Phys. Chem. A* **2008**, *112* (12), 2702–2712.
(49) Graham, D.; Menon, A.; Goerigk, L.; Grimme, S.; Radom, L. Optimization and Basis-Set Dependence of a Restricted-Open-Shell Form of B2-PLYP Double-Hybrid Density Functional Theory. *J. Phys. Chem. A* **2009**, *113* (36), 9861–9873.
(50) Chan, B.; Radom, L. Accurate Quadruple-ζ Basis-Set Approximation for Double-Hybrid Density Functional Theory with an Order of Magnitude Reduction in Computational

Cost. *Theor. Chem. Acc.* **2013**, *133* (2), 1426. https://doi.org/10.1007/s00214-013-1426-9.

(51) Mehta, N.; Casanova-Páez, M.; Goerigk, L. Semi-Empirical or Non-Empirical Double-Hybrid Density Functionals: Which Are More Robust? *Phys. Chem. Chem. Phys.* **2018**, *20* (36), 23175–23194. https://doi.org/10.1039/C8CP03852J.

(52) Goerigk, L.; Mehta, N. A Trip to the Density Functional Theory Zoo: Warnings and Recommendations for the User*. *Aust. J. Chem.* **2019**. https://doi.org/10.1071/CH19023.

(53) Mardirossian, N.; Head-Gordon, M. Thirty Years of Density Functional Theory in Computational Chemistry: An Overview and Extensive Assessment of 200 Density Functionals. *Mol. Phys.* **2017**, *115* (19), 2315–2372. https://doi.org/10.1080/00268976.2017.1333644.

(54) Peverati, R.; Truhlar, D. G. Quest for a Universal Density Functional: The Accuracy of Density Functionals across a Broad Spectrum of Databases in Chemistry and Physics. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2014**, *372* (2011), 20120476–20120476. https://doi.org/10.1098/rsta.2012.0476.

(55) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A Look at the Density Functional Theory Zoo with the Advanced GMTKN55 Database for General Main Group Thermochemistry, Kinetics and Noncovalent Interactions. *Phys. Chem. Chem. Phys.* **2017**, *19* (48), 32184–32215. https://doi.org/10.1039/C7CP04913G.

(56) Goerigk, L.; Grimme, S. A General Database for Main Group Thermochemistry, Kinetics, and Noncovalent Interactions − Assessment of Common and Reparameterized (Meta -)GGA Density Functionals. *J. Chem. Theory Comput.* **2010**, *6* (1), 107–126. https://doi.org/10.1021/ct900489g.

(57) Yu, H. S.; Zhang, W.; Verma, P.; He, X.; Truhlar, D. G. Nonseparable Exchange–Correlation Functional for Molecules, Including Homogeneous Catalysis Involving Transition Metals. *Phys. Chem. Chem. Phys.* **2015**, *17* (18), 12146–12160. https://doi.org/10.1039/C5CP01425E.

(58) Jurecka, P.; Sponer, J.; Cerný, J.; Hobza, P. Benchmark Database of Accurate (MP2 and CCSD(T) Complete Basis Set Limit) Interaction Energies of Small Model Complexes, DNA Base Pairs, and Amino Acid Pairs. *Phys. Chem. Chem. Phys.* **2006**, *8* (17), 1985–1993. https://doi.org/10.1039/b600027d.

(59) Rezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-Balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7* (8), 2427–2438. https://doi.org/10.1021/ct2002946.

(60) Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules. *J. Chem. Theory Comput.* **2012**, *8* (11), 4285–4292. https://doi.org/10.1021/ct300647k.

(61) Hobza, P. Calculations on Noncovalent Interactions and Databases of Benchmark Interaction Energies. *Acc. Chem. Res.* **2012**, *45* (4), 663–672. https://doi.org/10.1021/ar200255p.

(62) Karton, A.; Daon, S.; Martin, J. M. L. W4-11: A High-Confidence Benchmark Dataset for Computational Thermochemistry Derived from First-Principles W4 Data. *Chem. Phys. Lett.* **2011**, *510*, 165–178. https://doi.org/10.1016/j.cplett.2011.05.007.

(63) Kesharwani, M. K.; Manna, D.; Sylvetsky, N.; Martin, J. M. L. The X40×10 Halogen Bonding Benchmark Revisited: Surprising Importance of (n –1)d Subvalence Correlation. *J. Phys. Chem. A* **2018**, *122* (8), 2184–2197.

https://doi.org/10.1021/acs.jpca.7b10958.
(64) Fogueri, U. R.; Kozuch, S.; Karton, A.; Martin, J. M. L. The Melatonin Conformer Space: Benchmark and Assessment of Wave Function and DFT Methods for a Paradigmatic Biological and Pharmacological Molecule. *J. Phys. Chem. A* **2013**, *117* (10), 2269–2277. https://doi.org/10.1021/jp312644t.
(65) Kozuch, S.; Martin, J. M. L. Halogen Bonds: Benchmarks and Theoretical Analysis. *J. Chem. Theory Comput.* **2013**, *9*, 1918–1931. https://doi.org/10.1021/ct301064t.
(66) Martin, J. M. L. What Can We Learn About Dispersion from the Conformer Surface of N-Pentane? *J. Phys. Chem. A* **2013**, *117* (14), 3118–3132. https://doi.org/10.1021/jp401429u.
(67) Kozuch, S.; Bachrach, S. M.; Martin, J. M. L. Conformational Equilibria in Butane-1,4-Diol: A Benchmark of a Prototypical System with Strong Intramolecular H-Bonds. *J. Phys. Chem. A* **2014**, *118* (1), 293–303. https://doi.org/10.1021/jp410723v.
(68) Karton, A.; Schreiner, P. R.; Martin, J. M. L. Heats of Formation of Platonic Hydrocarbon Cages by Means of High-Level Thermochemical Procedures. *J. Comput. Chem.* **2016**, *37* (1), 49–58. https://doi.org/10.1002/jcc.23963.
(69) Kesharwani, M. K.; Karton, A.; Martin, J. M. L. Benchmark Ab Initio Conformational Energies for the Proteinogenic Amino Acids through Explicitly Correlated Methods. Assessment of Density Functional Methods. *J. Chem. Theory Comput.* **2016**, *12* (1), 444–454. https://doi.org/10.1021/acs.jctc.5b01066.
(70) Manna, D.; Martin, J. M. L. What Are the Ground State Structures of C 20 and C 24 ? An Explicitly Correlated Ab Initio Approach. *J. Phys. Chem. A* **2016**, *120* (1), 153–160. https://doi.org/10.1021/acs.jpca.5b10266.
(71) Brauer, B.; Kesharwani, M. K.; Kozuch, S.; Martin, J. M. L. The S66x8 Benchmark for Noncovalent Interactions Revisited: Explicitly Correlated Ab Initio Methods and Density Functional Theory. *Phys. Chem. Chem. Phys.* **2016**, *18* (31), 20905–20925. https://doi.org/10.1039/C6CP00688D.
(72) Manna, D.; Kesharwani, M. K.; Sylvetsky, N.; Martin, J. M. L. Conventional and Explicitly Correlated Ab Initio Benchmark Study on Water Clusters: Revision of the BEGDB and WATER27 Data Sets. *J. Chem. Theory Comput.* **2017**, *13* (7), 3136–3152. https://doi.org/10.1021/acs.jctc.6b01046.
(73) Karton, A.; Sylvetsky, N.; Martin, J. M. L. W4-17: A Diverse and High-Confidence Dataset of Atomization Energies for Benchmarking High-Level Electronic Structure Methods. *J. Comput. Chem.* **2017**, *38* (24), 2063–2075. https://doi.org/10.1002/jcc.24854.
(74) Lynch, B. J.; Truhlar, D. G. Small Representative Benchmarks for Thermochemical Calculations. *J. Phys. Chem. A* **2003**, *107* (42), 8996–8999. https://doi.org/10.1021/jp035287b.
(75) Chan, B. Formulation of Small Test Sets Using Large Test Sets for Efficient Assessment of Quantum Chemistry Methods. *J. Chem. Theory Comput.* **2018**, *14* (8), 4254–4262. https://doi.org/10.1021/acs.jctc.8b00514.
(76) Gould, T. 'Diet GMTKN55' Offers Accelerated Benchmarking through a Representative Subset Approach. *Phys. Chem. Chem. Phys.* **2018**, *EarlyView*, 1–9. https://doi.org/10.1039/C8CP05554H.
(77) Morgante, P.; Peverati, R. ACCDB: A Collection of Chemistry Databases for Broad Computational Purposes. *J. Comput. Chem.* **2019**, *40* (6), 839–848. https://doi.org/10.1002/jcc.25761.

(78) Koster, J.; Rahmann, S. Snakemake--a Scalable Bioinformatics Workflow Engine. *Bioinformatics* **2012**, *28* (19), 2520–2522. https://doi.org/10.1093/bioinformatics/bts480.
(79) Huber, P. J.; Ronchetti, E. M. *Robust Statistics*; Wiley Series in Probability and Statistics; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2009. https://doi.org/10.1002/9780470434697.
(80) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7* (18), 3297–3305. https://doi.org/10.1039/b508541a.
(81) Rappoport, D.; Furche, F. Property-Optimized Gaussian Basis Sets for Molecular Response Calculations. *J. Chem. Phys.* **2010**, *133* (13), 134105. https://doi.org/10.1063/1.3484283.
(82) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; et al. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.* **2015**, *113* (2), 184–215. https://doi.org/10.1080/00268976.2014.952696.
(83) Dasgupta, S.; Herbert, J. M. Standard Grids for High-Precision Integration of Modern Density Functionals: SG-2 and SG-3. *J. Comput. Chem.* **2017**, *38* (12), 869–882. https://doi.org/10.1002/jcc.24761.
(84) Gill, P. M. .; Johnson, B. G.; Pople, J. A. A Standard Grid for Density Functional Calculations. *Chem. Phys. Lett.* **1993**, *209* (5–6), 506–512. https://doi.org/10.1016/0009-2614(93)80125-9.
(85) Murray, C.; Handy, N.; Laming, G. Quadrature Schemes for Integrals of Density Functional Theory. *Mol. Phys.* **1993**, *78* (4), 997–1014. https://doi.org/10.1080/00268979300100651.
(86) Lebedev, V. I.; Laikov, D. N. A Quadrature Formula for the Sphere of the 131st Algebraic Order of Accuracy. *Dokl. Math.* **1999**, *59* (3), 477–481.
(87) Brandenburg, J. G.; Bates, J. E.; Sun, J.; Perdew, J. P. Benchmark Tests of a Strongly Constrained Semilocal Functional with a Long-Range Dispersion Correction. *Phys. Rev. B - Condens. Matter Mater. Phys.* **2016**, *94* (11), 17–19. https://doi.org/10.1103/PhysRevB.94.115144.
(88) Sullivan, M. B.; Iron, M. a.; Redfern, P. C.; Martin, J. M. L.; Curtiss, L. a.; Radom, L. Heats of Formation of Alkali Metal and Alkaline Earth Metal Oxides and Hydroxides: Surprisingly Demanding Targets for High-Level Ab Initio Procedures. *J. Phys. Chem. A* **2003**, *107* (29), 5617–5630. https://doi.org/10.1021/jp034851f.
(89) Haworth, N. L.; Sullivan, M. B.; Wilson, A. K.; Martin, J. M. L.; Radom, L. Structures and Thermochemistry of Calcium-Containing Molecules. *J. Phys. Chem. A* **2005**, *109* (40), 9156–9168. https://doi.org/10.1021/jp052889h.
(90) Powell, M. J. D. *The BOBYQA Algorithm for Bound Constrained Optimization without Derivatives (DAMPT Report 2009/NA06)*; Department of Applied Mathematics and Theoretical Physics, University of Cambridge, UK, 2009. https://doi.org/10.1.1.443.7693.
(91) Goerigk, L. A Comprehensive Overview of the DFT-D3 London-Dispersion Correction. In *Non-Covalent Interactions in Quantum Chemistry and Physics*; Elsevier, 2017; pp 195–219. https://doi.org/10.1016/B978-0-12-809835-6.00007-4.
(92) Hui, K.; Chai, J.-D. SCAN-Based Hybrid and Double-Hybrid Density Functionals from

Models without Fitted Parameters. *J. Chem. Phys.* **2016**, *144* (4), 044114. https://doi.org/10.1063/1.4940734.
(93) Toulouse, J.; Sharkas, K.; Brémond, E.; Adamo, C. Communication : Rationale for a New Class of Double-Hybrid Approximations in Density-Functional Theory. *J. Chem. Phys.* **2011**, *135*, 101102. https://doi.org/10.1063/1.3640019.
(94) Sharkas, K.; Toulouse, J.; Savin, A. Double-Hybrid Density-Functional Theory Made Rigorous. *J. Chem. Phys.* **2011**, *134* (6), 064113. https://doi.org/10.1063/1.3544215.
(95) Austin, A.; Petersson, G. A.; Frisch, M. J.; Dobek, F. J.; Scalmani, G.; Throssell, K. A Density Functional with Spherical Atom Dispersion Terms. *J. Chem. Theory Comput.* **2012**, *8* (12), 4989–5007.
(96) Zhang, Y.; Yang, W. Comment on "Generalized Gradient Approximation Made Simple." *Phys. Rev. Lett.* **1998**, *80* (4), 890–890. https://doi.org/10.1103/PhysRevLett.80.890.
(97) Almlöf, J. Elimination of Energy Denominators in Møller-Plesset Perturbation Theory by a Laplace Transform Approach. *Chem. Phys. Lett.* **1991**, *181* (4), 319–320. https://doi.org/10.1016/0009-2614(91)80078-C.
(98) Häser, M. Møller-Plesset (MP2) Perturbation Theory for Large Molecules. *Theor. Chim. Acta* **1993**, *87* (1–2), 147–173. https://doi.org/10.1007/BF01113535.
(99) Häser, M.; Almlöf, J. Laplace Transform Techniques in Møller-Plesset Perturbation Theory. *J. Chem. Phys.* **1992**, *96* (1), 489–494. https://doi.org/10.1063/1.462485.
(100) Song, C.; Martínez, T. J. Atomic Orbital-Based SOS-MP2 with Tensor Hypercontraction. I. GPU-Based Tensor Construction and Exploiting Sparsity. *J. Chem. Phys.* **2016**, *144* (17), 174111. https://doi.org/10.1063/1.4948438.
(101) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. Results Obtained with the Correlation Energy Density Functionals of Becke and Lee, Yang and Parr. *Chem. Phys. Lett.* **1989**, *157* (3), 200–206.
(102) Ghasemi, S. A.; Hofstetter, A.; Saha, S.; Goedecker, S. Interatomic Potentials for Ionic Systems with Density Functional Accuracy Based on Charge Densities Obtained by a Neural Network. *Phys. Rev. B - Condens. Matter Mater. Phys.* **2015**, *92* (4), 1–6. https://doi.org/10.1103/PhysRevB.92.045131.
(103) Rappé, A. K.; Goddard, W. A. Charge Equilibration for Molecular Dynamics Simulations. *J. Phys. Chem.* **1991**, *95* (8), 3358–3363. https://doi.org/10.1021/j100161a070.
(104) Verstraelen, T.; Van Speybroeck, V.; Waroquier, M. The Electronegativity Equalization Method and the Split Charge Equilibration Applied to Organic Systems: Parametrization, Validation, and Comparison. *J. Chem. Phys.* **2009**, *131* (4). https://doi.org/10.1063/1.3187034.
(105) Iron, M. A.; Janes, T. Evaluating Transition Metal Barrier Heights with the Latest DFT Exchange – Correlation Functionals – the MOBH35 Benchmark Dataset. *J. Phys. Chem. A* **2019**, *submitted*, 1–71.
(106) Dohm, S.; Hansen, A.; Steinmetz, M.; Grimme, S.; Checinski, M. P. Comprehensive Thermochemical Benchmark Set of Realistic Closed-Shell Metal Organic Reactions. *J. Chem. Theory Comput.* **2018**, *14* (5), 2596–2608. https://doi.org/10.1021/acs.jctc.7b01183.
(107) Goerigk, L.; Grimme, S. Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals—Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J. Chem. Theory*

Comput. **2011**, *7* (2), 291–309. https://doi.org/10.1021/ct100466k.
(108) Mezei, P. D.; Csonka, G. I.; Ruzsinszky, A.; Kállay, M. Construction and Application of a New Dual-Hybrid Random Phase Approximation. *J. Chem. Theory Comput.* **2015**, *11* (10), 4615–4626. https://doi.org/10.1021/acs.jctc.5b00420.
(109) Mezei, P. D.; Csonka, G. I.; Ruzsinszky, A.; Kállay, M. Construction of a Spin-Component Scaled Dual-Hybrid Random Phase Approximation. *J. Chem. Theory Comput.* **2017**, *13* (2), 796–803. https://doi.org/10.1021/acs.jctc.6b01140.
(110) Chan, B.; Goerigk, L.; Radom, L. On the Inclusion of Post-MP2 Contributions to Double-Hybrid Density Functionals. *J. Comput. Chem.* **2016**, *37* (2), 183–193. https://doi.org/10.1002/jcc.23972.
(111) Perdew, J. P. Density-Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas. *Phys. Rev. B* **1986**, *33* (12), 8822–8824. https://doi.org/10.1103/PhysRevB.33.8822.
(112) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37* (2), 785–789. https://doi.org/10.1103/PhysRevB.37.785.

Table of Contents Graphic