

## FULL PAPER

## Double hybrid DFT calculations with Slater type orbitals

Arno Förster  | Lucas Visscher Theoretical Chemistry, Vrije Universiteit,  
Amsterdam, The Netherlands

## Correspondence

\*Arno Förster, Theoretical Chemistry, Vrije  
Universiteit, De Boelelaan 1083, NL-1081 HV,  
Amsterdam, The Netherlands.  
Email: a.t.l.foerster@vu.nl

## Funding information

Nederlandse Organisatie voor  
Wetenschappelijk Onderzoek

## Abstract

On a comprehensive database with 1,644 datapoints, covering several aspects of main-group as well as of transition metal chemistry, we assess the performance of 60 density functional approximations (DFA), among them 36 double hybrids (DH). All calculations are performed using a Slater type orbital (STO) basis set of triple- $\zeta$  (TZ) quality and the highly efficient pair atomic resolution of the identity approach for the exchange- and Coulomb-term of the KS matrix (PARI-K and PARI-J, respectively) and for the evaluation of the MP2 energy correction (PARI-MP2). Employing the quadratic scaling SOS-AO-PARI-MP2 algorithm, DHs based on the spin-opposite-scaled (SOS) MP2 approximation are benchmarked against a database of large molecules. We evaluate the accuracy of STO/PARI calculations for B3LYP as well as for the DH B2GP-PLYP and show that the combined basis set and PARI-error is comparable to the one obtained using the well-known def2-TZVPP Gaussian-type basis set in conjunction with global density fitting. While quadruple- $\zeta$  (QZ) calculations are currently not feasible for PARI-MP2 due to numerical issues, we show that, on the TZ level, Jacob's ladder for classifying DFAs is reproduced. However, while the best DHs are more accurate than the best hybrids, the improvements are less pronounced than the ones commonly found on the QZ level. For conformers of organic molecules and noncovalent interactions where very high accuracy is required for qualitatively correct results, DHs provide only small improvements over hybrids, while they still excel in thermochemistry, kinetics, transition metal chemistry and the description of strained organic systems.

## KEYWORDS

ADF, benchmark, double-hybrid, KS-DFT, STO

## 1 | INTRODUCTION

Due to its unrivaled price/performance-ratio,<sup>[1]</sup> Kohn-Sham (KS)<sup>[2]</sup> density functional theory (DFT),<sup>[3-5]</sup> is the workhorse of modern computational chemistry.<sup>[6-9]</sup> Despite recent tremendous progress in wave function (WF) theory, especially in the field of local Coupled Cluster (CC)<sup>[10-13]</sup> methods,<sup>[14-41]</sup> little doubt exist that this situation will not change in the foreseeable future.

No satisfactory way to systematically improve the exchange-correlation (XC) functional of KS-DFT is known<sup>[42]</sup> and in practice approximations are necessary. Density functional approximations (DFA) explicitly depending on both, the electron density and one-electron orbitals resolve many deficiencies of (meta-)generalized gradient approximations (GGA).<sup>[5,43-47]</sup> If only the exchange part of such DFAs depends on these orbitals, they are referred to as hybrid functionals. For an orbital-dependent treatment of correlation based on

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Journal of Computational Chemistry* published by Wiley Periodicals, Inc.

the KS reference Hamiltonian, a dependence of the DFA on virtual orbitals must be introduced. A particularly economic strategy to do so is via second order Møller-Plesset perturbation theory (MP2). As both, the exchange- and the correlation part of such DFAs are bridging standard DFAs and WF theory, they are commonly referred to as double-hybrid (DH)<sup>[48-55]</sup> functionals.

According to the famous Jacob's ladder classification scheme for DFAs,<sup>[56]</sup> the generally most accurate and robust functionals should be DH-DFAs. Indeed, rigorous validation on the comprehensive GMTKN55<sup>[57]</sup> database by Goerigk, Grimme, Martin and coworkers has substantiated this conjecture.<sup>[57-61]</sup> Unfortunately, blending WF theory into KS-DFT not only goes along with increased accuracy but also with increased computational cost. The calculation of the KS matrix for hybrid functionals is commonly associated with a  $N^4$  scaling operation count,<sup>[62]</sup> and the CPU time for the evaluation of the canonical MP2 correlation energy increases with  $N^5$  as a function of system size  $N$ .<sup>[63]</sup>

Given these scaling properties, the availability of cost-effective algorithms for both tasks is crucial to enable routine application of hybrid- and DH-DFAs to large molecules. For both, the Coulomb ( $J$ ) and the exchange ( $K$ )-term, exploiting sparsity in the rank-4 electron repulsion integral (ERI) tensor is key to reducing the computational cost of the KS matrix construction for hybrid functionals. Algorithms based on integral prescreening techniques<sup>[64-70]</sup> offer favorable scalability but suffer from a high prefactor. To circumvent this issue, modern codes additionally rely on density-fitting (DF)<sup>[71-82]</sup> approximations being an efficient approach to evaluate the  $J$ -term, especially in conjunction with the  $J$ -engine technique.<sup>[83-87]</sup> Global density fitting is less efficient for the  $K$ -term and promising approaches to achieve better performance are the pseudo-spectral method,<sup>[88-93]</sup> the auxiliary density matrix method<sup>[94-96]</sup> or different flavors of local DF (LDF)<sup>[97-99]</sup> approximations.<sup>[100]</sup> The most promising variant of the latter approach might be the pair-atomic resolution of the identity (PARI) approximation<sup>[101-103]</sup> (Known as PARI- $K$  when applied to the  $K$ -term<sup>[103]</sup>). In this most extreme variant of LDF methods, following the treatment of Baerends et al.<sup>[71]</sup> for the  $J$ -term, each pair product of AOs is expanded in a set of auxiliary basis functions centered on the same two atoms as the target pair of primitives.

To accelerate the evaluation of MP2 energies, DF is the only approach which has found widespread use.<sup>[104-115]</sup> Reducing the prefactor of canonical implementations by at least one order of magnitude, it retains their  $N^5$  scaling. However, relying on the short-rangedness of dynamical electron correlation,<sup>[116-119]</sup> many reduced-scaling MP2 algorithms, employing localized molecular orbitals (MO),<sup>[120-122]</sup> relying on fragmentation approaches,<sup>[113,114,123,124]</sup> or evaluating the MP2 energy in the AO basis and exploiting sparsity in the ERI tensor<sup>[125-143]</sup> have been developed.

Following the latter approach, we have shown that the PARI approach can not only be used to accelerate the KS matrix construction but also to efficiently compute MP2 correlation energies.<sup>[144]</sup> While we have demonstrated that an MO based PARI-MP2 algorithm can easily compete with DF-MP2 in terms of accuracy, we have also formulated the Spin opposite-scaled<sup>[145]</sup> (SOS)-AO-PARI-MP2

algorithm, enabling to obtain accurate SOS-MP2 energies with quadratic scaling operation count. Implemented in the STO-based Amsterdam density functional (ADF)<sup>[79,146,147]</sup> code, the SCF, although extensively accelerated via PARI- $K$ , represents the bottleneck (both CPU time and memory-wise) for single-point SOS-MP2 calculations on molecules of several hundreds of atoms using basis sets of TZ quality. We concluded that this enables quantum chemists to carry out DH calculations whenever a hybrid calculation is feasible too.

Employing any of the mentioned local approximations might lead to tremendous speed-up to the price of an additional source of error. For an approximation to be useful in practice, the introduced errors must be small and well controlled and the inequalities

$$\Delta_m > \Delta_b > \Delta_a, \quad (1)$$

where  $\Delta_m$ ,  $\Delta_b$ , and  $\Delta_a$  denote the inherent error of the method, basis set and additional approximations, respectively, should be fulfilled. The latter inequality is of high importance, as in most quantum-chemistry packages algorithmic details will differ and it is often non-obvious to users how to tweak them. In this sense, the second inequality is crucial to make comparisons of results between different codes meaningful.

Even for very small  $\Delta_m$  and  $\Delta_b$ , (1) is certainly met by global DF approximations. Is this also the case for PARI? Concerns regarding robustness<sup>[103,148,149]</sup> and accuracy<sup>[100]</sup> of PARI- $K$ , especially for the description of virtual orbital energies,<sup>[100,150]</sup> have recently been expressed. Additionally, the exact exchange energy from the PARI- $K$  algorithm is unbounded from below which might result in convergence to artificial low-lying states.<sup>[103,149]</sup> This should not be confused with errors due to nonconvergence of the SCF, an algorithm for which convergence is not guaranteed.<sup>[151-153]</sup> The appearance of fully converged, but unphysical results is much more problematic. These shortcomings have been found to get more pronounced with increasing number of AOs and molecular size.<sup>[100]</sup> Hybrid calculations with a QZ basis are unreliable for large molecules. For PARI-MP2, additional difficulties arise in the fitting of virtual orbitals,<sup>[144]</sup> leading to even larger numerical instabilities with increased number of primitives. Effectively, this limits the largest basis sets which can be safely be used to TZ quality with a moderate number of diffuse functions. In practice, this might be the most severe limitation for DH calculations. Although some exceptions are known,<sup>[154]</sup> TZ basis sets are usually sufficient to fulfill the first inequality in (1) for GGAs and hybrids.<sup>[155]</sup> Due to the exceptionally slow convergence to the CBS limit of MP2,<sup>[156-161]</sup> considerably larger basis sets are needed for DHs.<sup>[162]</sup>

As it only requires the evaluation of a subset of two-center Coulomb integrals, PARI is a crucial technique for efficient exact exchange algorithms in quantum chemistry packages based on Slater type orbitals (STO) or numerical atomic orbitals (NAOs)<sup>[163-166]</sup> relying on numerical integration techniques. Therefore, it is of utmost importance to investigate whether accurate, robust and efficient hybrid- and DH-KS calculations can be performed within the PARI-

**TABLE 1** List of DFAs benchmarked in this work. The functional name is given together with the amount of HF-exchange ( $a_x$  in (2)) and MP2 correlation energy ( $a_{c, os}/a_{c, ss}$  in (2), if only one number is given,  $a_{c, os} = a_{c, ss}$ ) as well as the variants of empirical dispersion correction with which it is combined. The corresponding parameters used herein are given in the ESI.

Functional	$a_x$	$a_c$	Dispersion correction
GGA			
PBE <sup>[171, 172]</sup>			D3(BJ) <sup>[173]</sup> /D4 <sup>[174]</sup>
BLYP <sup>[175–177]</sup>			D3(BJ) <sup>[173]</sup> /D4 <sup>[174]</sup>
REVPE <sup>[178]</sup>			D3(BJ) <sup>[173]</sup> /D4 <sup>[174]</sup>
Meta-GGA			
TPSS <sup>[179, 180]</sup>			D3(BJ) <sup>[173]</sup> /D4 <sup>[181]</sup>
revTPSS <sup>[182, 183]</sup>			D3(BJ) <sup>[57]</sup> /D4 <sup>[174]</sup>
SCAN <sup>[184]</sup>			D3(BJ) <sup>[185]</sup> /D4 <sup>[174]</sup>
Hybrid			
B97 <sup>[186]</sup>	0.19		D3(0) <sup>a</sup> /D4 <sup>[174]</sup>
B3LYP <sup>[187]</sup>	0.2		D3(BJ) <sup>[173]</sup> /D4 <sup>[181]</sup>
mPW1B95 <sup>[188]</sup>	0.31		D3(BJ) <sup>[189]</sup>
mPWB1K <sup>[188]</sup>	0.44		D3(BJ) <sup>[189]</sup> /D4 <sup>[174]</sup>
mPW1PW	0.25		D3(BJ) <sup>a</sup>
PBE0 <sup>[190,191]</sup>	0.25		D3(BJ) <sup>[173]</sup> /D4 <sup>[181]</sup>
SOGGA11-X <sup>[192]</sup>	0.40		D3(BJ) <sup>a</sup>
TPSSH <sup>[193]</sup>	0.10		D3(BJ) <sup>[189]</sup> /D4 <sup>[174]</sup>
PWB6K <sup>[194]</sup>	0.46		D3(0)
PW6B95 <sup>[194]</sup>	0.28		D3(0) <sup>[173]</sup> /D4 <sup>[174]</sup>
CAM-B3LYP	r.s.		D3(BJ) <sup>[189]</sup> /D4 <sup>[174]</sup>
$\omega$ B97-X <sup>[195]</sup>	r.s.		D3(0) <sup>a</sup>
M05-2X <sup>[196]</sup>	0.56		D3(0) <sup>a</sup>
M06 <sup>[197]</sup>	0.27		D3(0) <sup>a</sup> /D4 <sup>[174]</sup>
M06-2X <sup>[197]</sup>	0.54		D3(0) <sup>a</sup>
M08-HX <sup>[198]</sup>	0.52		D3(0) <sup>a</sup>
M08-SO <sup>[198]</sup>	0.57		D3(0) <sup>a</sup>
MN12-SX <sup>[199]</sup>	r.s.		D3(BJ) <sup>a</sup>
DH			
B2-PLYP <sup>[49]</sup>	0.53	0.27	D3(BJ) <sup>[189]</sup> /D4 <sup>[174]</sup>
B2GP-PLYP <sup>[200]</sup>	0.65	0.36	D3(BJ) <sup>[189]</sup> /D4 <sup>[174]</sup>
B2K-PLYP <sup>[55]</sup>	0.72	0.42	D3(BJ) <sup>b</sup>
B2T-PLYP <sup>[55]</sup>	0.60	0.31	D3(BJ) <sup>b</sup>
B2 $\pi$ -PLYP <sup>[201]</sup>	0.60	0.27	D3(BJ) <sup>b</sup>
B2NC-PLYP <sup>[202]</sup>	0.67	0.49	D3(BJ) <sup>[58]</sup>
mPW2-PLYP <sup>[50]</sup>	0.55	0.25	D3(BJ) <sup>[57]</sup>
mPW2K-PLYP <sup>[55]</sup>	0.72	0.42	D3(BJ) <sup>b</sup>
mPW2NC-PLYP <sup>[202]</sup>	0.67	0.49	D3(BJ) <sup>[58]</sup>
DH-BLYP <sup>[203]</sup>	0.65	0.42	D3(BJ) <sup>b</sup>
PBE0-DH <sup>[204]</sup>	0.5	0.125	D3(BJ) <sup>[205]</sup> /D4 <sup>[174]</sup>
PBE0-2 <sup>[206]</sup>	0.79	0.5	D3(BJ) <sup>[205]</sup> /D4 <sup>[174]</sup>
LS1-DH <sup>[207]</sup>	0.75	0.42	D3(BJ) <sup>[58]</sup>
LS1-TPSS <sup>[208]</sup>	0.85	0.61	D3(BJ) <sup>[58]</sup>
DS1-TPSS <sup>[208]</sup>	0.73	0.53	D3(BJ) <sup>2</sup>
PBE-QIDH <sup>[209]</sup>	0.69	0.33/0.33	D3(BJ) <sup>[210]</sup>
SOS1-PBE-QIDH <sup>[211]</sup>	0.69	0.44/0.00	D3(BJ) <sup>[210]</sup>

TABLE 1 (Continued)

Functional	$a_x$	$a_c$	Dispersion correction
SD-SCAN69 <sup>[60]</sup>	0.69	0.62/0.24	c
DOD-SCAN <sup>[60]</sup>	0.66	0.63/0.00	D3(BJ) <sup>[60]</sup>
revDSD-SCAN-D4 <sup>[60]</sup>	0.66	0.63/0.01	D4 <sup>[60]</sup>
revDOD-SCAN-D4 <sup>[60]</sup>	0.66	0.63/0.00	D4 <sup>[60]</sup>
DSD-BLYP <sup>[212]</sup>	0.69	0.46/0.37	D3(BJ) <sup>[189]</sup>
revDSD-BLYP <sup>[60]</sup>	0.71	0.55/0.20	D3(BJ) <sup>[60]</sup>
revDOD-BLYP <sup>[60]</sup>	0.71	0.62/0.00	D3(BJ) <sup>[60]</sup>
revDSD-BLYP-D4 <sup>[60]</sup>	0.71	0.56/0.20	D4 <sup>[60]</sup>
revDOD-BLYP-D4 <sup>[60]</sup>	0.71	0.63/0.00	D4 <sup>[60]</sup>
DSD-PBEP86 <sup>[213]</sup>	0.69	0.52/0.22	D3(BJ) <sup>[213]</sup>
revDSD-PBEP86 <sup>[60]</sup>	0.69	0.58/0.08	D3(BJ) <sup>[60]</sup>
revDOD-PBEP86 <sup>[60]</sup>	0.69	0.61/0.00	D3(BJ) <sup>[60]</sup>
revDSD-PBEP86-D4 <sup>[60]</sup>	0.69	0.59/0.06	D4 <sup>[60]</sup>
revDOD-PBEP86-D4 <sup>[60]</sup>	0.69	0.61/0.00	D4 <sup>[60]</sup>
DSD-PBE <sup>[213,214]</sup>	0.68	0.55/0.13	D3(BJ) <sup>[214]</sup>
revDSD-PBE <sup>[60]</sup>	0.68	0.58/0.07	D3(BJ) <sup>[60]</sup>
revDOD-PBE <sup>[60]</sup>	0.68	0.61/0.00	D3(BJ) <sup>[60]</sup>
revDSD-PBE-D4 <sup>[60]</sup>	0.68	0.60/0.04	D4 <sup>[60]</sup>
revDOD-PBE-D4 <sup>[60]</sup>	0.68	0.62/0.00	D4 <sup>[60]</sup>

<sup>a</sup>PBE-parameters have been used.

<sup>b</sup>Parametrization in this work.

<sup>c</sup>Not evaluated with empirical dispersion correction.

approximation. The answer immediately implies the answer to a far more important question—is it possible to perform accurate, robust and efficient hybrid- and DH KS calculations using STOs? To the best of our knowledge, a final answer has not been given yet. Although we recently presented accurate PARI-MP2 energies using TZ basis sets,<sup>[144]</sup> a systematic study on large databases is yet to be conducted for hybrid- and DH-DFAs.

Different DFAs show different dependencies on the choice of basis set and numerical approximations.<sup>[154,167]</sup> The superior performance of a certain DFA under ideal conditions that is, using large basis sets and very small  $\Delta_o$  does not necessarily imply the same behavior when smaller basis sets and PARI are employed. Employing these ideal conditions, in the last years large numbers of DFAs have been ranked according to their performance on large databases.<sup>[57–59,154,168]</sup> It is certainly of great interest to investigate to what extent these rankings can be reproduced using STOs and PARI (STO/PARI in the following).

Against this background, we herein present benchmark calculations with 60 DFAs using a comprehensive database consisting of 1,644 data points, covering both, main-group and transition metal (TM) chemistry. This is the first comprehensive benchmark of DFAs using PARI and, to the best of our knowledge, the first one using STOs.

This paper is organized as follows: After giving an overview of our database and of the herein benchmarked DFAs we shortly describe

our computational approach and give representative timings before we discuss the results of our benchmarks. We will investigate the accuracy of the STO/PARI approach in comparison to GTO-type basis sets for two exemplary DFAs and subsequently present benchmarks for all 60 assessed functionals. In the end, we conclude and summarize this work.

## 2 | METHODS

### 2.1 | Test sets

Large compilations of diverse datasets covering a wide range of main-group chemistry have now a long-standing tradition in quantum chemistry. The importance of meta-databases like GMTKN55<sup>[57]</sup> (and its predecessors GMTKN24<sup>[169]</sup> and GMTKN30<sup>[162]</sup>) and MGCD84<sup>[168]</sup> compiled by Grimme and coworkers and Head-Gordon and coworkers, respectively, to assess the performance of DFAs and also wave-function methods<sup>[170]</sup> can barely be overemphasized.

In this work, we benchmark the DFAs in Table 1 against a comprehensive database comprised of 58 subsets, organized in five subcategories of different types of chemical problems with 1,644 data points in total. Additionally, we benchmark all SOS-MP2 based DHs as well as selected hybrids against a sixth subdatabase comprised of five test sets of large organic molecules with a total of 85 data points.

We do not present any new database or test set in this work. Instead, we selected nearly all subsets from GMTKN55<sup>[57]</sup> and MGCD84<sup>[168]</sup> and additionally include two databases featuring energies and barrier heights of reactions between TM containing species. An overview of all categories in which we organize our database is given in Table 2, while a short description of all subsets is given in Table 3. In the latter table we also reference for each test set the publication where it was introduced first. The excellent papers of Grimme and coworkers<sup>[57]</sup> and Head-Gordon and coworkers<sup>[168]</sup> relief us from the burden of a detailed description of all subsets and we refer to them for details. Nevertheless, we will briefly introduce all six subcategories employed herein.

While there is considerable overlap between MGCD84 and GMTKN55, the reference data in the latter one is generally newer and often more accurate.<sup>[226]</sup> In example, MGCD84 contains much data from GMTKN30,<sup>[162]</sup> which has been updated in GMTKN55.<sup>[57,226]</sup> Consequently, for test sets contained in both databases, we always compared our results to the reference values presented in the latter one and in total we selected 45 out of the 55 subsets in GMTKN55 (two of them, (MCONF, BUT14DIOL) with reduced size<sup>[265]</sup>) to benchmark all DFAs in Table 1 and included three more into our subcategory of large molecules.

### 2.1.1 | Transition metal containing systems

As they are not part of GMTKN55 or MGCD84, we describe the TM containing subsets in our database in some detail. We employ two subsets with 66 reaction energies combined, capturing different aspects of TM chemistry.

The MOR23 test set is a subset of the MOR41 database recently presented by Grimme and coworkers,<sup>[227]</sup> consisting of 41 reactions involving large molecules with up to 120 atoms and including 13 different TMs. All reference values have been obtained on the DLPNO-CCSD(T)<sup>[23,31]</sup>/CBS level of theory and all structures are chosen to have single-reference character. To keep the overall computational effort manageable, we excluded some of the largest systems from our benchmarks and herein only present all 23 remaining reaction energies with an average reference value of 35.57 kcal/mol.

The TMBH43 test set is a compilation of several smaller test sets<sup>[235–238]</sup> and contains 43 barrier heights of reactions catalyzed by

TMs, 6 are catalyzed by the 4d-elements Zr<sup>[236]</sup> or Mo,<sup>[238]</sup> and 37 are catalyzed by one of the 5d-elements Pt,<sup>[235]</sup> Au,<sup>[235]</sup> Ir,<sup>[235]</sup> Re<sup>[237]</sup> or W,<sup>[238]</sup> out of these 22 by Au. All reference values have been obtained at the CCSD(T)/CBS level of theory. The average barrier height in TMBH43 is 11.08 kcal/mol.

### 2.1.2 | Basic properties and reaction energies

This subcategory (TC for thermochemistry) contains 21 subsets with a total of 572 datapoints and covers many aspects of thermochemistry. With the exception of the MOR23 test set, all test sets in this category are taken from GMTKN55. While it mainly features basic properties like atomization energies (W4-11<sup>[224]</sup>), bond dissociation energies (BSR36,<sup>[57,221,222]</sup> SIE4x4<sup>[57]</sup>), electron affinities (G21EA<sup>[225]</sup>), ionization potentials (G21IP,<sup>[225]</sup> DIPCS10<sup>[57]</sup>) or proton affinities (PA26<sup>[57]</sup>) for small model systems, TC also contains subsets covering bonded interactions of various types like Diels-Alder reactions (DARC<sup>[169,218]</sup>) or H<sub>2</sub>-activation by frustrated Lewis-pairs (NBPRC<sup>[57,169,219]</sup>), both being of great practical interest,<sup>[219,266,267]</sup> as well as dissociation reactions (HEAVYSB9,<sup>[57]</sup> YBDE18,<sup>[217]</sup> ALKBDE10<sup>[57]</sup>). We especially emphasize the importance of the challenging SIE4x4 subset which is a compilation of self-interaction error (SIE)-related problems,<sup>[57]</sup> and DC13<sup>[57]</sup> (which also contains isomerization reactions), being a selection of difficult cases for many DFAs. The average reaction energy for this category is 132.34 kcal/mol.

### 2.1.3 | Reaction barrier heights

The BH (barrier heights) category consists of nine subsets, six being taken from GMTKN55, two from MGCD84 and one of them involving TM species, with a total of 264 datapoints and average barrier height of 21.92 kcal/mol. It contains the BH76<sup>[169,239,240]</sup> and BHDIV10<sup>[57]</sup> datasets featuring all kind of reactions involving small molecules as well as the CRBH20,<sup>[234]</sup> CR20,<sup>[233]</sup> and BHPER<sup>[57,169,229–231]</sup> test sets with reference values for cycloreversions and pericyclic reactions. It also contains rotational barriers around single bonds which are usually very low. To the best of our knowledge, the TM containing test set TMBH43 is herein included for the first time in a comprehensive benchmark.

Category	Description	#subsets	#	$\Delta E$
TC	Basic properties and reaction energies	21	572	132.34
BH	Reaction barrier heights	9	264	21.92
IE	Simple isomerization reactions	10	274	4.52
ID	Difficult isomerization reactions	6	159	25.96
INC	Intermolecular noncovalent interactions	12	375	11.84
	Total database	58	1,644	55.58
LM	Large molecules	5	85	23.60

**TABLE 2** List of all 6 subcategories comprising our database. The subsets of large molecules are analyzed separately

**TABLE 3** List of all subsets that comprise the database employed in this work together with a short description of the types of systems they contain. The third column displays the average reaction energy in a subset and the fourth column its number of datapoints

Name	Description	$ \Delta E $	#
Basic properties and reaction energies			
FH51 <sup>[57,215,216]</sup>	Diverse reaction energies in inorganic systems	31.01	51
YBDE18 <sup>[57,217]</sup>	Bond dissociation energies in Ylides	49.28	18
AL2X6 <sup>[57]</sup>	Dimerization energies for Al <sub>2</sub> X <sub>3</sub> -compounds	35.88	6
DARC <sup>[169,218]</sup>	Diels-Alder reactions	32.47	14
NBPRC <sup>[57,169,219]</sup>	H <sub>2</sub> activation reactions by frustrated Lewis pairs	27.71	12
HEAVYSB9 <sup>[220]</sup>	Dissociation energies in group 14–16 hydrides	58.02	9
BSR36 <sup>[57,221,222]</sup>	Bond-separation reactions of saturated hydrocarbons	16.20	36
RSE43 <sup>[57,223]</sup>	Radical-stabilization energies	7.60	43
W4-11 <sup>[57,224]</sup>	Atomization energies	306.91	140
G21EA <sup>[57,169,225]</sup>	Electron affinities	33.62	25
G21IP <sup>[57,169,225]</sup>	Ionization potentials	257.61	36
DIPCS10 <sup>[57]</sup>	Double ionization potentials of closed-shell systems	654.26	10
PA26 <sup>[57]</sup>	Adiabatic proton affinities	189.05	26
SIE4x4 <sup>[57]</sup>	Dissociation curves of very small molecules/ self-interaction error related problems	33.72	12
ALKBDE10 <sup>[57]</sup>	Dissociation energies of group 1–2 diatomic molecules	100.69	10
RC21 <sup>[57]</sup>	Fragmentation and rearrangement reactions in radical cations	35.70	21
ALK8 <sup>[57]</sup>	Reactions energies for alkaline compounds	62.60	8
DC13 <sup>[57]</sup>	Compilation of difficult reactions for DFT methods	54.98	13
G2RC <sup>[57]</sup>	Selected reaction energies of selected G2/97 systems	51.26	25
BH76RC <sup>[169]</sup>	Reaction energies of the BH76 test set	21.39	30
MOR23 <sup>[226,227]</sup>	Selected metalorganic reactions between single-reference species containing various Tis	35.57	23
Reaction barrier heights			
WCPT18 <sup>[169,228]</sup>	Proton-transfer barrier heights	34.99	18
BHROT27 <sup>[57]</sup>	Barrier heights for rotation around single bonds	6.37	27
BHPERI <sup>[57,169,229–231]</sup>	Barrier heights of pericyclic reactions	20.87	26
BHDIV10 <sup>[57]</sup>	Diverse reaction barrier heights	45.33	10
INV24 <sup>[57,232]</sup>	Inversion/racemisation barrier heights	32.85	24
CR20 <sup>[168,233]</sup>	Cyclo-reversion reaction energies	19.31	20
CRBH20 <sup>[168,234]</sup>	Barrier heights for cyclo-reversion of heterocyclic rings	46.13	20
TMBH43 <sup>[226,235–238]</sup>	Barrier heights for reactions catalyzed by 4d- and 5d-TMs	11.08	43
BH76 <sup>[57,169,239,240]</sup>	Diverse barrier heights for reactions involving small molecules	18.61	76
Simple isomerization reactions			
ISO34 <sup>[57,241]</sup>	Isomerisation energies of small and medium organic molecules	14.57	34
ICONF <sup>[57]</sup>	Isomers of inorganic systems	3.27	17
ACONF <sup>[57,242]</sup>	Relative energies of alkane conformers	1.83	15
TAUT15 <sup>[57]</sup>	Relative energies in tautomers	3.05	15
Amino20x4 <sup>[57,243]</sup>	Relative energies in amino acid conformers	2.44	80
PCONF <sup>[57,244,245]</sup>	Relative energies in tri- and tetrapeptide conformers	1.62	10
MCONF <sup>[57,246]</sup>	Relative energies in melatonin conformers	4.97	20
SCONF <sup>[57,169,247]</sup>	Relative energies of sugar conformers	4.60	10

(Continues)

TABLE 3 (Continued)

Name	Description	$ \Delta E $	#
PArel <sup>[57]</sup>	Relative energies in protonated isomers	4.63	20
BUT14DIOL <sup>[57,248]</sup>	Relative energies in butane-1,4-diol conformers	2.80	32
Difficult isomerization reactions			
EIE22 <sup>[168,249]</sup>	Isomerisation energies of enecarbonyls	5.44	22
Styrene45 <sup>[168,250]</sup>	Isomerisation energies of C <sub>8</sub> H <sub>8</sub>	62.64	45
ISOMERIZATION20 <sup>[168]</sup>	Compilation of difficult isomerisation energies	31.84	20
DIE60 <sup>[168,202]</sup>	Double-bond isomerisation energies in dienes	4.71	60
IDISP <sup>[51,57,162,169,241,251]</sup>	Intramolecular dispersion interactions for medium molecules	14.22	6
C20C24 <sup>[168,252]</sup>	Isomerisation energies of C <sub>20</sub> and C <sub>24</sub>	30.77	6
Intermolecular noncovalent interactions			
S66 <sup>[57,253]</sup>	NCI energies in organic molecules and biomolecules	5.47	66
S10x8 <sup>[253]</sup>	10 dimers from S66 in 8 nonequilibrium geometries	5.47	80
X40 <sup>[254]</sup>	Binding energies of NCIs involving halogenated molecules	3.76	40
HEAVY28 <sup>[57,255]</sup>	NCI energies for heavy element hydrides	1.24	28
CHB6 <sup>[57,256]</sup>	Interaction energies in cation–neutral dimers	26.79	6
AHB21 <sup>[57,256]</sup>	Interaction energies in anion–neutral dimers	22.49	21
IL16 <sup>[57,256]</sup>	Interaction energies in anion–cation dimers	109.04	16
PNICO23 <sup>[57,257]</sup>	Interaction energies in pnictogen-containing dimers	4.27	23
CT20 <sup>[168,258]</sup>	Binding energies of charge-transfer complexes	0.98	20
CARBHB12 <sup>[57]</sup>	Hydrogen-bonded complexes between carbene analogues and H <sub>2</sub> O, NH <sub>3</sub> , or HCl	6.04	12
ADIM6 <sup>[57,255]</sup>	Interaction energies of n-alkane dimers	3.36	6
3B-69-TRIM <sup>[168,259]</sup>	Binding energies of trimers, with three different orientations of 23 distinct molecular crystals	12.30	69
Isomerization reactions, NCIs and enzymatic reactions for large systems			
ISOL24 <sup>[57,260]</sup>	Isomerisation energies of large organic molecules	21.92	24
C60ISO <sup>[57,261]</sup>	Relative energies between C <sub>60</sub> isomers	98.25	9
L7 <sup>[262]</sup>	Dimerization energies in large organic systems	15.32	7
UPU23 <sup>[57,263]</sup>	Relative energies between RNA-backbone conformers	5.72	22
ENZYMES23 <sup>[264]</sup>	Reaction energies and barrier heights in enzymatic reactions	18.20	23

## 2.1.4 | Easy isomerization energies

This category is comprised of 10 subsets with a total of 274 relative energies of mainly organic molecules like alkane chains (ACONF<sup>[242]</sup>), sugars (SCONF<sup>[57,169,247]</sup>), peptides (PCONF<sup>[57,244,245]</sup>) and various kinds of other species (BUT14DIOL<sup>[248]</sup> MCONF<sup>[246]</sup> ISO34<sup>[241]</sup> TAUT15<sup>[57]</sup> PArel<sup>[57]</sup>). The biggest subset in this category is Amino20x4<sup>[57,243]</sup> containing the four relative energies between the five energetically lowest conformations of the 20 proteogenic amino acids. ICONF<sup>[57]</sup> is the only subset with isomers of inorganic systems. The average relative energy of this category is with 4.52 kcal/mol very low. The electronic structures of these species are rather simple

(isomerization easy [IE]) and are usually very well described already on the GGA + Dispersion level of theory. However, given the small energy differences in this subcategory, high accuracy is required to correctly reproduce the energetic ordering of conformers.

## 2.1.5 | Difficult isomerization reactions

Inspired by Head-Gordon and coworkers,<sup>[168]</sup> we decided to introduce a second category of isomerization reactions comprising of six test sets with 274 relative energies. The electronic structures of the systems in this category are notoriously difficult (isomerization difficult [ID]) to

describe for GGAs,<sup>[168]</sup> suffering from many-electron self-interaction error.<sup>[268]</sup> Among others, ID features strained ring systems (Styrene45,<sup>[250]</sup> C20C24<sup>[252]</sup>), Dienes (DIE60<sup>[202]</sup>) and isomerization reactions of enecarbonyls (EIE22<sup>[249]</sup>). With 25.96 kcal/mol, the average of the relative energies in this category is considerably higher than for IE.

### 2.1.6 | Intermolecular noncovalent interactions

The fifth category of our database contains the energies of 274 non-covalently bounded complexes, relative to the monomers of which they consist. This category contains 12 subsets and features the prominent S66<sup>[253]</sup> databases of dimerization reactions between organic molecules and biomolecules and the X40<sup>[254]</sup> test set of interaction energies between halogene-containing organic molecules, both compiled by Hobza and coworkers. While these systems are in their equilibrium geometry, Hobza and coworkers S66x8<sup>[253]</sup> test set additionally contains all dimers in the S66 test set at eight different non-equilibrium geometries and is also part of MGCDB84. While it is certainly insightful to benchmark nonequilibrium geometries also, we only selected 10 out the 66 potential energy curves in order to keep our dataset well balanced, resulting in the S10x8 subset.

We note, that the very popular S22<sup>[269]</sup> subset (and its extension to nonequilibrium geometries in the S22x5 database), which is also a part of GMTKN55 is not part of our database. Despite its usefulness, it is rather imbalanced and does not represent important interactions like single hydrogen bonds or aliphatic-aliphatic dispersion interactions.<sup>[253]</sup> These shortcomings have been addressed with the compilation of S66 and we argue that with its introduction S22 is obsolete.

Other important subsets are CHB6,<sup>[256]</sup> AHB21,<sup>[256]</sup> and IL16<sup>[256]</sup> containing dimerization energies for charged systems and Hobza and coworkers 3B-69-TRIM,<sup>[259]</sup> being part of MGCDB84, containing 23 trimers of organic molecules in three different orientations each. The average energy of the reactions contained in this category is 11.84 kcal/mol.

### 2.1.7 | Large systems

The last category differs from all other categories in the sense that its subsets cannot be assigned to a certain type of chemical interactions. We refer to the systems in this category as *large* in the sense that it contains mainly molecules for which a canonical DH-DFA calculation requires a considerable amount of CPU-time. To give an example, calculation of the 10 isomers in the C60ISO<sup>[261]</sup> test set on the B2GP-PLYP/TZ2P level of theory already requires 58 hr on a single node with 24 cores in our implementation.

The category contains five subsets with 85 dimerization energies, relative conformational energies, reaction energies and barrier heights with an average energy of 23.60 kcal/mol and reference values are usually obtained on the DLPNO-CCSD(T)/CBS level of theory. Benchmarking large systems is crucial to assess the robustness and accuracy of local approximations which do not come into play for

small systems. Especially PARI-errors errors have already been reported to accumulate with increasing system size.<sup>[100]</sup> Furthermore, growing concern has been expressed that MP2 might diverge for weak interactions in large molecular complexes.<sup>[270,271]</sup>

ISOL24,<sup>[260]</sup> C60ISO, and UPU23<sup>[263]</sup> contain isomerization reactions and are already part of GMTKN55. Additionally, we selected Hobza and coworkers popular L7<sup>[262]</sup> test set of noncovalently bounded dimers with up to 120 atoms as well as another benchmark set which we designate herein ENZYMES23. It is based on recent work of Goerigk and coworkers<sup>[264]</sup> who published reference values for different model systems of five enzymatically catalyzed reactions, resulting in a dataset containing 28 datapoints. For this work, we considered the four enzymatic reactions with 23 data points in total for which the reference values have been obtained at the DLPNO-CCSD(T)/CBS level of theory.

## 2.2 | Selection of density functional approximations

In total, we selected 60 DFAs, among these 36 DHs, 18 hybrids and 6 (meta-)GGAs. We focus on DHs in this study as many of them have only been published recently and have not been comprehensively benchmarked so far. The performance of Hybrid-functionals and (meta-)GGAs, however, has been assessed excessively in the last decades and due to their rather fast convergence to the CBS limit, we do not expect significant deterioration of performance compared to results obtained with QZ basis sets. DHs, on the other hand, suffer significantly more from BSE and are more challenging for the PARI-approximation. Thus, we herein restrict ourselves to the best hybrids and (meta-)GGAs known to the literature and the results we present for them in this paper can be seen as a frame of reference to put the performance of the DHs into perspective.

The XC energy  $E_{xc}^{DFA}$  for all DFAs benchmarked in the current paper can be expressed as

$$E_{xc}^{DFA} = (1 - a_x)E_x^{GGA} + a_x E_x^{HF} + a_{c, gga} E_c^{GGA} + a_{c, os} E_{c, os}^{MP2} + a_{c, ss} E_{c, ss}^{MP2} + E_{empirical}^{DISP} \quad (2)$$

were  $E_x^{GGA}$  ( $E_c^{GGA}$ ) denote the exchange (correlation) part of some (meta-)GGA,  $E_x^{HF}$  the exact-exchange like energy contribution,  $E_{c, os}^{MP2}$  ( $E_{c, ss}^{MP2}$ ) denotes the opposite-spin (os) (same-spin [ss]) contribution to the MP2 correlation energy and finally,  $E_{empirical}^{DISP}$  is an empirical dispersion correction term. While many other forms of dispersion corrections have been suggested,<sup>[272-276]</sup> we restrict ourselves to the popular atom-atom dispersion potentials DFT-D3<sup>[173,255]</sup> and DFT-D4.<sup>[174,181]</sup>

For (meta-)GGAs ( $a_{c, os} = a_{c, ss} = a_x = 0$ ) and hybrids ( $a_{c, os} = a_{c, ss} = 0$ ) our selection is mainly based on their performance in the recent comprehensive benchmarks by Goerigk et al. on the GMTKN55 database<sup>[57,58]</sup> and Mardirossian et al. on the MGCDB84 database.<sup>[168]</sup> The only exceptions are the B3LYP and PBE0 hybrid functionals, which we included due to their continuing high popularity despite their now well established<sup>[9,57,162,168,264,277]</sup> average performance. As functionals



based on the B95<sup>[278]</sup> and B97 correlation functionals were not available to us (except for the hybrids  $\omega$ B97-X and B97 which were available to us through libxc<sup>[279]</sup>), we excluded the popular hybrids B97M-V<sup>[280]</sup> and  $\omega$ B97X-V<sup>[281]</sup> as well as the DHs  $\omega$ B97X-2,<sup>[282]</sup>  $\omega$ B97M(2),<sup>[283]</sup> PWPB95,<sup>[162]</sup> and DSD-PBEB95 from this study.<sup>[214]</sup>

The majority of the herein benchmarked DH-DFAs are of the DSD (Double hybrid, Spin-component scaled, Dispersion)-type.<sup>[212]</sup> These functionals are parametrized together with an empirical dispersion correction term as well as without any constraints on  $a_{c,ss}$ ,  $a_{c,os}$  and  $a_{c,sga}$ . Consequently, in this approach different forms of dispersion correction give rise to different overall parametrizations. Examples are the recently published<sup>[60]</sup> reparametrizations of DSD-type functionals by the Martin group, both for the D4<sup>[174,181]</sup> dispersion correction as well as for the older D3<sup>[255]</sup> version with Becke-Johnson damping (D3[BJ]).<sup>[173]</sup> While we consistently assess all other functionals with and without an empirical-dispersion correction term, we benchmark DSD-functionals in their dispersion corrected form only.<sup>[284]</sup>

We especially emphasize DOD-functionals in our study, a special variant of DSD-functionals with  $a_{c,ss} = 0$ .  $E_{c,os}^{MP2}$  can be evaluated efficiently using SOS-AO-PARI-MP2, and DOD-functionals can therefore be applied to molecules consisting of hundreds of atoms in a routine fashion.<sup>[144]</sup> As they can provide accuracies comparable to their DSD-counterparts, they are in our opinion the most-interesting DHs for practical applications.

Table 1 list the DFAs assessed herein together with the employed dispersion correction as well as the parameters  $a_x$  and  $a_c$  from (2). In total, 122 unique combinations of DFA and dispersion-correction are evaluated in this study.

## 2.3 | Computational details

The majority of our DFT calculations as well as all empirical dispersion corrections have been performed with a locally modified development version of the ADF code.<sup>[79,146,147]</sup> For a variety of subsets, we also performed B3LYP and B2GP-PLYP calculations with PSI4.<sup>[285]</sup>

For all test sets contained in GMTKN55, the structures and reference energies as available on the dedicated website<sup>[286]</sup> have been used. For all other datasets we employed the structures and reference energies from the ACCDB<sup>[226]</sup> database.

All energies for the S66, S66x10, X40, L7, AHB21, IL16, CHB6, and CT20 subsets have been computed using the counterpoise (CP) method of Boys and Bernardi<sup>[287]</sup> to correct for the basis set superposition error (BSSE).

### 2.3.1 | ADF calculations

PARI-K has been used for all hybrid and DH calculations and, if not stated otherwise, PARI-MP2 for the post-SCF energy correction required for the latter ones, except for all test sets in the category LM, for which SOS-AO-PARI-MP2 has been used instead. The accuracy of the latter algorithm depends on the quality of the numerical

approximation to an integral.<sup>[288]</sup> We use nine quadrature points, being the default in ADF and a number for which this approximation can be considered as sufficiently converged.<sup>[132,289]</sup> For the evaluation of the exact exchange as well as for all SOS-AO-PARI-MP2 calculations, the *Normal* tier of threshold qualities has been used. We refer to our recent work<sup>[144]</sup> for a detailed explanation and discussion of these algorithms as well as explicit threshold values.

The majority of the herein presented numbers have been calculated on the all-electron level using the TZ2P (TZ with two shells of polarization functions) STO-type basis set.<sup>[290]</sup> The *Normal* auxiliary fit set has been used for for PARI-K and all PARI-MP2 calculations with the only exception of the IDISP test set for which we employed the *Very Good* auxiliary fit set for both, PARI-K and PARI-MP2 in all DH calculations. For both, the numerical integration quality as well as the quality of the DF for the evaluation of the  $J$ -term,<sup>[291]</sup> we used *Good* quality.<sup>[291,292]</sup>

For all systems containing fourth-row elements or heavier, relativistic effects have been treated with the Zero Order Regular Approximation (ZORA)<sup>[293-296]</sup> in conjunction with ZORA-optimized basis sets and the Minimum of neutral Atomic potential Approximation (MAPA).

### 2.3.2 | PSI4 calculations

All PSI4 calculations have been performed using def2-TZVPP<sup>[297,298]</sup> and employing DF for the  $J$ - and  $K$ -terms as well as for the evaluation of MP2 energies; default auxiliary fits sets have been employed.<sup>[299]</sup> For all subsets, calculations have been performed on the all-electron level. Default settings have been used for the integration grids and all other numerical settings.

## 2.4 | Representative timings

To give an estimate on the CPU-time requirements for a prototypical GGA, hybrid and double-hybrid, respectively, we herein report timings for our whole database. Calculations have been performed sequentially on a  $2 \times 12$ -core 2.4 GHz Intel Xeon E5-2695 v2 (ivy Bridge) with 64 GB of memory. Sequential calculation is obviously not appropriate for small molecules as parallelization will be inefficient. In practice, the best performance will be achieved by running many jobs in parallel on one core each and only parallelize individual calculations for which the memory available on a single core is not sufficient. For the present context, this is irrelevant as we only aim to compare the relative timings. Results for all calculations required to obtain values for our whole database of 1,644 datapoints without large molecules are given in Table 4. For BLYP, B3LYP and revDOD-BLYP we also give timings for the subset of large systems.

For all systems in our database except for the large ones, all BLYP calculations can be performed in 33 hr; for B3LYP, less than twice CPU-time is required and less than thrice for B2GP-PLYP. The most-

time consuming part here are spin-unrestricted MP2 calculations as they are currently less-efficient than spin-restricted ones. For the S66 test set of organic molecules of between roughly 10–30 atoms, timings differ considerably less. With 3.5 hr, all 198 B3LYP-calculations can be performed in only 1 hr more than required for BLYP and the calculation of the post-SCF correlation energy only requires a negligible amount of time for both, PARI-MP2 and SOS-AO-PARI-MP2. A similar picture is obtained for large molecules. Moving from BLYP to B3LYP, CPU-time requirements increase by roughly 50% and double moving from BLYP to revDOD-BLYP when the AO-based algorithm is used. Here, the evaluation of  $E_{c,05}^{\text{MP2}}$  only consumes about 25% of the total CPU-time. Clearly, these timings show that CPU-time is hardly a concern when climbing Jacob's ladder to the top, provided a DOD-functional is used.

### 3 | RESULTS

In this section, the results of our calculations are presented. In the following discussion, we proceed in two steps. First, we investigate the basis set dependence of the STO/PARI-approach in detail. To do so, we present results for B3LYP as an exemplary hybrid, and B2GP-PLYP as an exemplary DH, respectively. Second, we analyze and discuss the results of our ADF calculations for all DFAs in Table 1 in order to identify the most accurate and robust DFAs. All numbers presented in the following figures can also be found in table form in the ESI.

#### 3.1 | Comparison of STO/PARI approach to Gaussian type basis sets

As a first step in the analysis of our results, we compare our TZ2P calculations to calculations using the GTO-type basis set def2-TZVPP for a selection of 30 subsets from our database. def2-TZVPP is comparable in size to the TZ2P basis set but has more shells of polarization functions for the first three rows of the periodic table. Table 5 shows the sizes of the herein employed basis sets for selected elements.

For our analysis we use two exemplary DFAs, the hybrid B3LYP and the DH B2GP-PLYP. The results of the respective functional as

**TABLE 4** CPU-times for our whole database obtained with BLYP, B3LYP, B2GP-PLYP and revDOD-BLYP

	BLYP	B3LYP	B2GP-PLYP	revDOD-BLYP
Small	33.10	57.00	89.08	
Large	18.42	26.56		36.03
S66	2.55	3.51	3.95	4.12

Note: All timings are given in hours and have been obtained on a single node with 24 cores. In the first row, the CPU-times for our whole database except for the subcategory of large molecules are given while the timings for this subcategory are displayed in the second row. B2GP-PLYP calculations have been performed using PARI-MP2. For revDOD-BLYP, SOS-AO-PARI-MP2 has been used.

calculated by Grimme and coworkers using an augmented def2-QZVP basis set (aug-def2-QZVP in the following, for details see Reference [57]) close to the CBS limit and available online<sup>[286]</sup> serve as reference values. This allows to investigate the basis set errors (BSE) without being flawed by cancellation between functional error and BSE.

While for hybrid-functionals basis sets of TZ quality are usually assumed to give results close to the CBS limit, a larger basis set dependence can be expected for B2-GPPLYP. Also due to possible difficulties in the description of virtual orbitals, the PARI approximation might introduce larger errors than for B3LYP. To this end, larger BSEs as well as larger differences between TZ2P and def2-TZVPP are to be expected.

Figure 1 shows the MADs obtained with B3LYP/TZ2P and B3LYP/def2-TZVPP with respect to B3LYP/aug-def2-QZVP for all 30 subsets considered herein. The subsets are sorted according to their average reaction energy  $|\overline{\Delta E}|$  in descending order and  $|\overline{\Delta E}|$  is given in parentheses after each subsets name. For comparison, the secondary x-axis displays the MAD of B3LYP/aug-def2-QZVP with respect to the usual reference values from high-level WF methods.

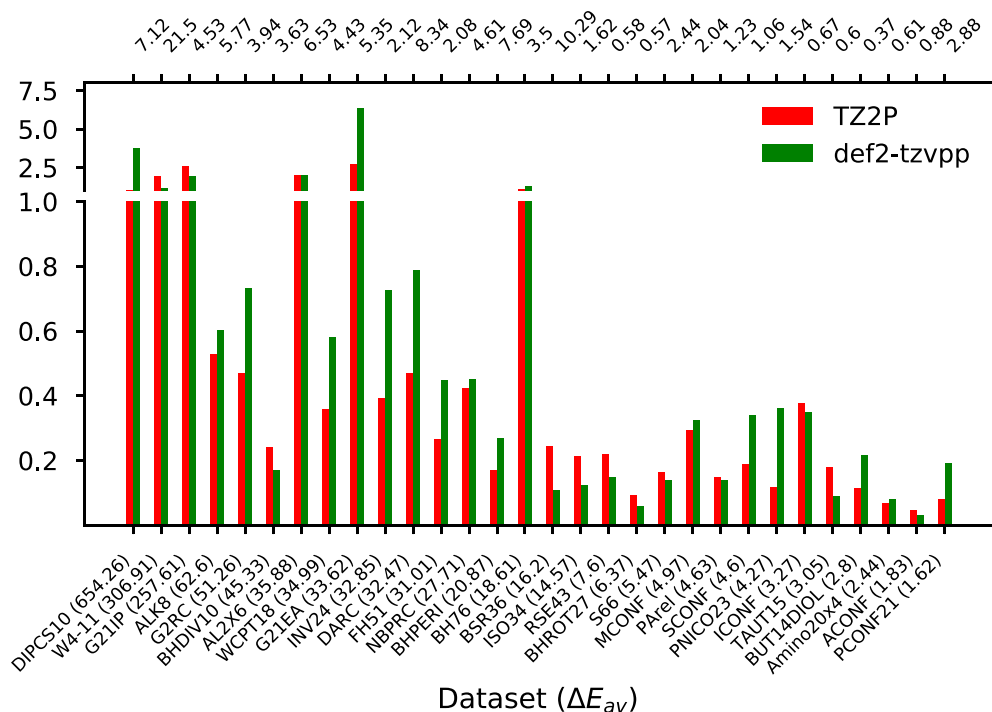
It can be verified immediately, that the average BSEs for both basis sets are of the same order of magnitude. One also identifies a rough correlation between  $|\overline{\Delta E}|$  and the error for both basis sets. On the left hand side of the plot, BSEs often exceed 1 kcal/mol, being most pronounced for ionization potentials (DIPCS10, G21IP) and electron affinities (G21EA) of atoms and small molecules. Possibly due to the more accurate description of the wave function close to the nuclei, the STO basis set outperforms the GTO one considerably for these subsets. The right hand side of the plot (BSR36 and subsets further to the right) features subsets with smaller  $|\overline{\Delta E}|$ , mostly containing isomerization energies and NCIs, but also smaller BSEs well below 0.4 kcal/mol.

Except for ionization potentials and electron affinities, TZ2P and def2-TZVP show comparable performance and no basis set can clearly be identified as superior to the other. There are subsets, where the def2-TZVPP BSE is roughly twice as large as the one from TZ2P (INV24, DARC, PNICO23) or the other way round (BSR36, ISO34, RSE43). This is also reflected in the average BSE for the entirety of all 30 subsets, which is 0.68 kcal/mol for TZ2P and 0.72 kcal/mol for the slightly larger def2-TZVPP.

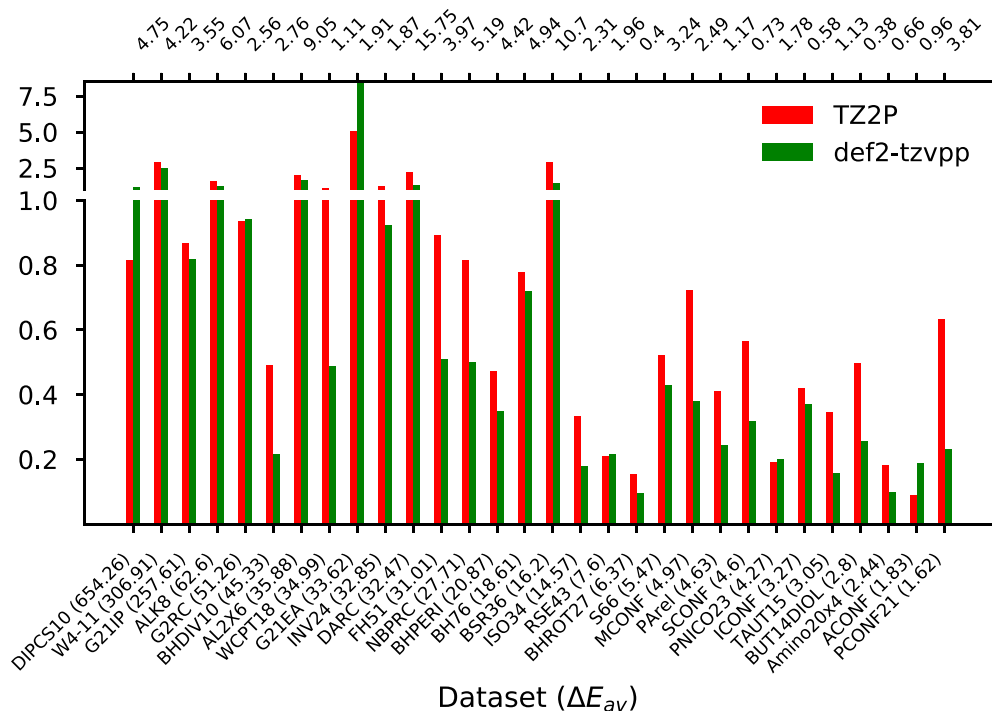
For all subsets, the BSE is at least one order of magnitude smaller than  $|\overline{\Delta E}|$  which is of course a necessary prerequisite for qualitatively correct reaction energies. On the other hand, the BSE is sometimes even larger than the error of the functional at the basis set limit. In

**TABLE 5** Composition and basis set sizes (in parentheses) for selected elements of the STO-type TZ2P and the GTO-type def2-TZVPP basis sets<sup>[298]</sup>

Element	TZ2P	def2-TZVPP
H	3s,1p,1d (11)	3s,2p,1d (14)
C	5s,3p,1d,1f (26)	5s,3p,2d,1f (31)
Ge	9s,7p,4d,1f (57)	6s,5p,4d,1f (48)



**FIGURE 1** Comparison of the BSE with respect to aug-def2-QZVP for B3LYP. For each of the selected subsets the sign-corrected average BSE is shown. The subsets are sorted according to their average reaction energy ( $|\Delta E|$ , in parentheses after each subsets name) in descending order. For each subset, the MAD obtained with aug-def2-QZVP with respect to the higher level reference is given on the secondary x-axis on top. All values are in kcal/mol [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 2** Comparison of the BSE with respect to aug-def2-QZVP for B2-GPPLYP. For each of the selected subsets the sign-corrected average BSE is shown. The subsets are sorted according to their average reaction energy ( $|\Delta E|$ , in parentheses after each subsets name) in descending order. For each subset, the MAD obtained with aug-def2-QZVP with respect to the higher level reference is given on the secondary x-axis on top. All values are in kcal/mol [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

example, def2-TZVPP gives a BSE of 6.34 kcal/mol for G21EA while B3LYP/aug-def2-QZVP gives a MAD of 5.35 kcal/mol. Clearly, this is a rather special case of arguably little practical relevance. However, especially for isomerization energies, where method errors are sometimes of the same order of magnitude than energetic differences between isomers, already small BSEs can lead to skewed results. For

ICONF, both basis sets give an error larger than 0.4 kcal/mol while B3LYP/aug-def2-QZVP only gives a slightly larger MAD of 0.67 kcal/mol. These examples seriously question the common belief that basis set of TZ quality are sufficient for hybrid calculations.

In the same way as Figure 1, Figure 2 shows the BSEs of TZ2P and def2-TZVPP for B2GP-PLYP for all subsets. While essentially the

same trends can be observed than for B3LYP, for the subsets featuring smaller  $|\Delta E|$ , def2-TZVPP often outperforms TZ2P. While this could hint on the importance of a large set of polarization functions for the calculation of orbital-dependent correlation energies which is known to be considerably larger for DHs than for hybrids, it could also be due to a larger PARI-error which is more pronounced when virtual orbitals are involved. Both hypotheses cannot be verified. Increasing the basis set leads to unreliable and therefore flawed PARI-MP2 energies for the subsets containing medium-sized or larger molecules while it is technically impossible for us to perform STO-calculations without PARI.

While it would of course be useful to be able to identify the exact source of error, it is of little relevance in practice. As already outlined, the use of STO-type basis sets and PARI go hand in hand and of practical relevance is only the combined error of the approach. With 0.96 kcal/mol for def2-TZVPP and 1.12 kcal/mol for TZ2P, respectively, the errors for both basis sets are of the same order of magnitude, although TZ2P produces larger errors than def2-TZVPP. These numbers suggest that the STO/PARI approach can clearly compete with standard GTO-type calculations also for DHs and the slightly decreased accuracy might be seen as a small price which is to be paid for the increased efficiency attainable with PARI-K and PARI-MP2.

Given the large number of reactions considered herein, we can state with high confidence that the data presented in this section clearly demonstrate that the STO/PARI approach can easily compete with GTO-type basis sets in terms of accuracy for a variety of chemical problems, for both, hybrid and DH calculations.

Of course, for both types of DFAs, the possibility that the good agreement of the STO/PARI calculations with the ones using GTOs are the result of fortuitous error cancellation cannot be excluded. Whether this is of practical relevance might be more of a philosophical question. While being unpleasant from a purists point of view, the success of approximate KS-DFT is to some extent rooted in error cancellation. In a field which is mostly based on heuristics, a method can be regarded as reliable if it only works well in virtually all cases in which it is applied. In this sense, the STO/PARI-approach clearly works. In our introduction we asked whether it possible to perform accurate, robust and efficient hybrid- and DH-KS calculations using STOs. We think that an affirmative answer to this question can already be given.

On the other hand, our data also show that the results obtained with the STO/PARI approach for individual subsets can differ considerably from GTO calculations. Depending on the details of their construction, different DFAs of the same rung will give quite different results when paired with basis sets not sufficiently close to the CBS limit. Furthermore, it is not clear if method error and BSE add up or cancel each other. Consequently, the trends observed here cannot be expected to be the same when changing the DFA for the same rung. In this sense, the results of this subsection strongly suggest that a comprehensive benchmark of DFAs using STO-type basis sets is of high interest. This benchmark is the subject of the next subsection.

## 3.2 | Discussion of the whole database and its categories

As we have performed well over 150.000 single point calculations in total, we can herein only highlight the most important trends and conclusions while we refer to Data S1 for details. We will discuss the performance of the assessed DFAs for each of the five subcategories. Results for large molecules will also be discussed whenever appropriate. Arguably, the information given in the following is most relevant for users of the ADF code. However, having established that STO/PARI calculations offer accuracies comparable to ones attainable using GTO-type TZ basis sets, we hope the herein presented results can also serve as an important guideline for users of GTO-based quantum chemistry codes who are not always willing or able to afford QZ sized basis sets. Final recommendations will be given based on WTMAD-2 values as well as on a best-worse analysis, that is, we recommend functionals based on their accuracy but also on their robustness. Before we dive into the discussion of the benchmarked DFAs performance, both metrics will shortly be introduced.

## 3.3 | Statistical criteria

### 3.3.1 | Weighted mean absolute deviations

We follow the approach of Grimme and coworkers<sup>[57]</sup> and base the discussion of our data on weighted mean absolute deviations (WTMAD-2),

$$\text{WTMAD-2} = \frac{1}{N} \sum_i N_i \times \text{MAD}_i \times \frac{|\Delta E|}{|\Delta E|_i}, \quad (3)$$

where the sum runs over all subsets in a subcategory,  $\text{MAD}_i$  denotes the mean absolute deviation of subset  $i$ ,  $|\Delta E|_i$  its average reaction energy and  $N_i$  the number of datapoints in it. The benefits of this scheme have been addressed in detail by Grimme and coworkers.<sup>[57]</sup> We prefer this metric over simple MADs as it does not focus on absolute accuracies but on relative ones, being clearly of higher relevance to predict and reproduce chemical trends.

### 3.3.2 | Best-Worse-Analysis

For this we also adopt the approach of Grimme and coworkers<sup>[57]</sup> and present best-worse analyses for the total database and subcategories. Instead of simply counting how often a functional yields the lowest (best) and the highest (worse) MAD, we rather count how often a certain functional yields a MAD which is not more than 15% higher (lower) than the MAD of the best (worse)-performing functional.

We consider an illustrative example to point out the advantage of this approach. The functional with the lowest MAD for the Amino20x4 subset is revDSD-PBEP86-D4 with 0.13 kcal/mol. For six other functionals we obtain MADs not more than 15% worse. All of

these functionals perform well for this subset which means that there are seven reliable (within our arbitrary criterion) DHs to describe it. As another example, we consider the DARC test set. The best DH here is B2GP-PLYP-D4 with a MAD of 0.23 kcal/mol, followed by mPW2K-PLYP with a MAD of 0.57 kcal/mol. Thus, B2GP-PLYP-D4 is the only logical choice for this test set.

### 3.4 | Discussion of all subcategories

#### 3.4.1 | Basic properties and reaction energies

To start with, we discuss the category of basic properties of small systems and bonded interactions. Figure 3 shows the WTAMD-2s obtained with 60 pairings of a DFA with a variant of empirical dispersion correction; the best 25 DHs and hybrids as well as the best 10 (meta-)GGAs. Due to its large average reaction energy of 132.34 kcal/mol, all WTAMD-2s are relatively small, ranging from about 2.5 kcal/mol for the best DH B2T-PLYP-D3(BJ) to nearly 7.0 kcal/mol for PBE-D4 being the worst GGA displayed here. Overall, Figure 3 clearly reveals that double-hybrids are the method of choice for this category. The Minnesota functional M06-2X-D3(0) clearly outperforms all other hybrids by nearly 0.5 kcal/mol but the obtained WTAMD-2 is comparable to dispersion uncorrected mPW2NC-PLYP being ranked 25th and having been parametrized for NCIs. To our surprise, B2 $\pi$ -PLYP-D3(BJ), parametrized for  $\pi$ - $\pi$  interactions, is the best DH in this category. As it has been optimized for thermochemistry,<sup>[55]</sup> it is not surprising that B2T-PLYP-D3(BJ) performs well too.

Taking a closer look at the performance of the benchmarked hybrid-functionals reveals that the Minnesota functionals M06-2X, M08-SO and M08-HX with different flavors of dispersion correction are the methods of choice on the hybrid level. The amount of exact exchange contained in these functionals is with more than 50% rather large. This is in line with results for GMTKN55 where the two best performing functionals are M06-2X-D3(0) and M08-HX-D3(0) and in contrast to the common belief that, due to the ability of the SIE to mimic static electron correlation,<sup>[300,301]</sup> a smaller fraction of exact exchange should be ideal for thermochemistry.<sup>[302,303]</sup>

After this general analysis we proceed by taking a closer look at selected individual test sets. SIE4x4 and DC13 are interesting benchmark sets as they are composed of systems which are challenging for DFT methods. Possibly due to their large amount of exact exchange, DHs outperform all hybrids and (meta-)GGAs by far. The only hybrid with a MAD below 6 kcal/mol is M05-2X, while LS1-TPSS is the best DH with a MAD below 1 kcal/mol. The best DH at the aug-def-QZVP level is PBE0-2 with 1.81 kcal/mol for which we obtain a smaller MAD of 1.49 kcal/mol. Also for DC13, DHs clearly outperform all other functionals; revDOD-PBE-D4 performs best with 2.50 kcal/mol.

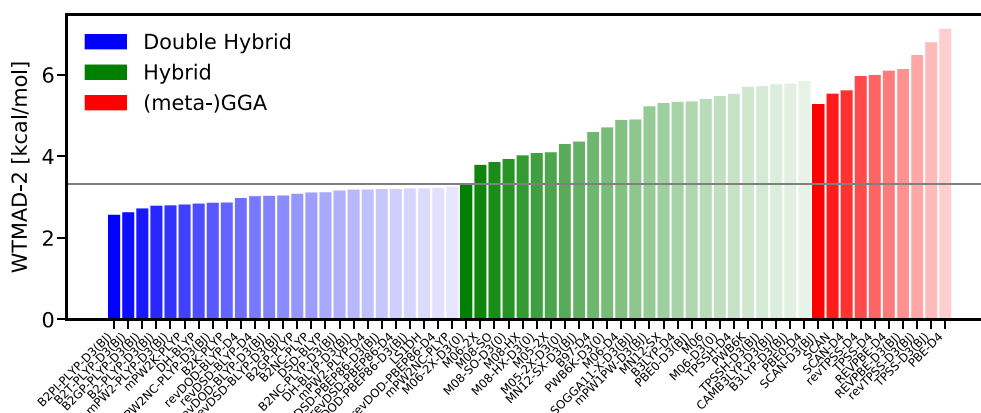
Another interesting subset is MOR23 containing TMs. SOS1-PBE-QIDH performs best with a MAD of 1.8 kcal/mol (followed by PBE-DH and the hybrid mPW1PW-D3(BJ)) and a maximum error of 6.16 kcal/mol. Given the average reaction energy of 35.57 kcal/mol for this test set, this performance is quite satisfactory.

Figure 4 presents the results of a best-worse analysis. The good performance of B2T-PLYP-D3(BJ) is substantiated further, being six times among the best DH-DFAs and never among the worst. Also revDSD-BLYP-D3(BJ) reaches MADs not worse more than 15% than the best DH for five times and can be recommended as a reliable functional for thermochemistry. On the hybrid level, B97-D4 and M08-HX are five times among the best functionals. Again, this does not come as a surprise for the latter one, given its fourth position in the WTAMD-2 ranking, while the former only ranks tenth there. The hybrid yielding by far the lowest WTAMD-2, M08-SO ranks among the best hybrids four times, but once it can also be found among the worst ones.

In conclusion, we recommend B2T-PLYP-D3(BJ) as the method of choice for the description of bonded interaction and basic properties and we generally advise to use DH-DFAs whenever feasible. If CPU-time is a concern, we recommend the best DOD-DH, revDOD-BLYP-D3(BJ)/D4 as well as the Minnesota functional M06-2X-D3(0).

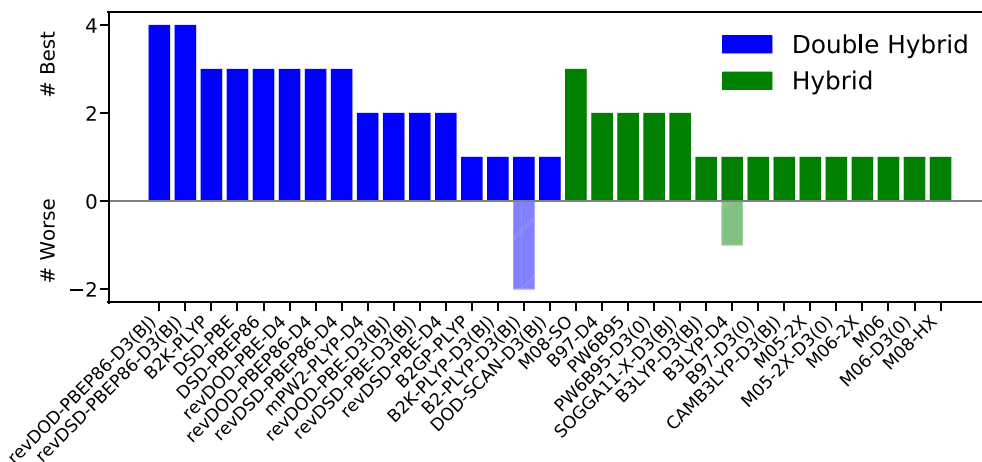
#### 3.4.2 | Reaction barrier heights

The next category consists of nine subsets with 264 data points, containing barrier heights of 21.92 kcal/mol on average. The WTAMD-2s for the best functionals of each rung are displayed in Figure 5. As

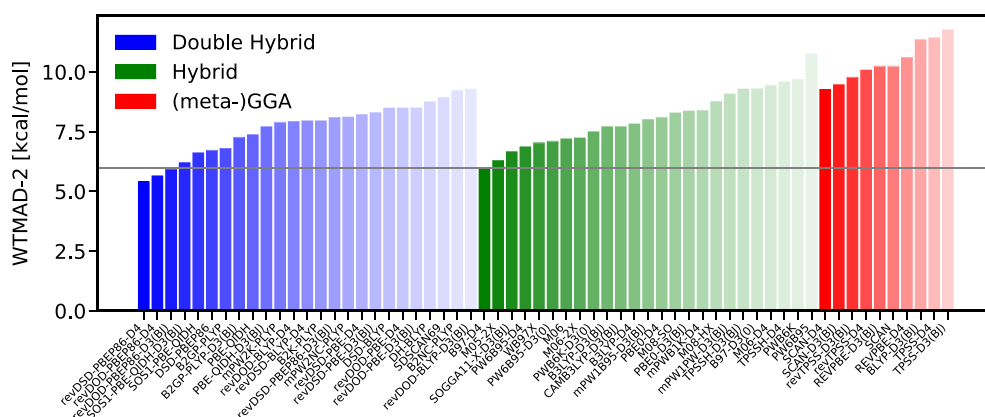


**FIGURE 3** The best 25 DHs (blue) and hybrids (green) as well as the best 10 (meta-)GGAs (red) according to their WTAMD-2s for the category of basic properties and reaction energies [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]





**FIGURE 6** Best (full colored bars, positive x-range)-worse (shaded bars, negative x-range) analysis for the best 15 DHs (blue) and hybrids (green) for reaction barrier heights [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 7** The best 25 DHs (blue) and hybrids (green) as well as the best 10 (meta-)GGAs (red) according to their WTMAD-2 s for the category of easy isomerization reactions [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

As it contains many barrier heights, we briefly discuss our results for ENZYMES23 here. We generally observe good performance of functionals with a large fraction of exact exchange. The best hybrid functional is M08-HX-D3(0). It is only slightly outperformed by revDOD-PBEP86-D3(BJ) with a MAD of roughly 1.2 kcal/mol. Note that Goerigk and coworkers obtained a only slightly better MAD of 1.07 kcal/mol with the herein not benchmarked SOS0-PBE-2-D3(BJ) functional in their benchmark study on this test set.<sup>[264]</sup>

In conclusion, we cannot unreservedly recommend to use DH-DFAs for the calculation of BHs unless TMs are involved. In this case, the method of choice is dispersion uncorrected B2K-PLYP, having been optimized for kinetic properties, and, when CPU time is a concern, revDOD-PBE-D4. Otherwise, hybrid functionals seem to provide the best price/performance-ratio for the calculation of barrier heights and we recommend M08-HX as the most robust and accurate alternative.

### 3.4.3 | Easy isomerization reactions

The availability of a plethora of benchmark sets containing isomers of organic molecules illustrates the great relevance of this type of

systems. In total, our database contains 433 relative conformational energies of which only 15 are not organic. In this work, we consider two categories of isomers, and in this paragraph we are concerned with the category containing 274 relative energies of isomers with an easy electronic structures and an average isomerization energy of only 4.52 kcal/mol.

Given these subtle energy differences, obtaining qualitatively correct results requires high precision and Figure 7 reveals large differences in the accuracies of the benchmarked DFAs. While the best DH has a WTMAD-2 of 5.4 kcal/mol, the WTMAD-2 of the 25th is with more than 9 kcal/mol already 71% higher, for hybrids, the corresponding ratio is 79%. This might be contrasted with the corresponding number for BH, where we obtain a ratio of 25% for DHs and of 29% for hybrids, respectively. We already mention at this point that the WTMAD-2s for IE are higher than for all other subcategories, that is, the largest relative errors are obtained here.

Going from the best hybrid, B97-D4, to the best DH, revDSD-PBEP86-D4, the WTMAD-2 only improves by roughly 10%, only three DHs yield improvements at all and only 16 perform better than the best meta-GGA, SCAN-D4. Thus, some care must be taken when selecting a DFA for these systems and with an uneducated choice of DH, one easily ends up with a worse performance than with a GGA.

Hybrids, although the best of them performing slightly worse than the best DH, seem to be the most robust class of DFAs for IE. Also the best-worse analysis in Figure 8 suggests B97-D4 to be the most reliable DFA being never among the worst-performing DFAs, while each of the three best DHs, revDSD-PBEP86-D4, revDOD-PBEP86-D4 and revDSD-PBEP86-D3(BJ), can be found among the worst performing DHs in this subcategory twice.

The source of the difficulties of DHs for this category is most likely the BSE for MP2 as exemplified by the PCONF test set of relative energies of oligopeptide conformers. Recently, Goerigk et al. investigated the basis set dependence of MP2 for parts of this subset and found a tremendous BSE for MP2 calculations on the TZ level.<sup>[245]</sup> Although to a lesser extent, the same is true for SCONF.<sup>[247]</sup> Indeed, eliminating the 35 data points of these two test set from our analysis, we observe that the WTMAD-2 of the best DH is now more than 17% smaller than the one of the best hybrid. Furthermore, instead of three DHs when PCONF and SCONF are included, 11 DHs outperform the best hybrid when they are excluded (See the ESI).

On the other hand, this implies that DHs must outperform DFAs of lower rung for some test sets. For the large Amino20x4 subset, revDSD-PBEP86-D4 is the clear winner with 0.13 kcal/mol while B97-D4 yields the lowest MAD of all hybrids with 0.21 kcal/mol. For the diverse ISO34 test set, the best hybrid  $\omega$ B97X yields with 0.87 kcal/mol a MAD 91% worse than the 0.45 kcal/mol obtained for revDSD-PBEP86-D3(BJ). For ISO34, the unrevised version of the latter functional is also the winner in the GMTKN55 study, where Grimme and coworkers calculated a MAD of 0.41 kcal/mol.

From the subcategory of large molecules, the UPU23 and ISOL24 subsets also fall into the category of easy isomerization reactions, although with  $|\Delta E| = 21.92$  kcal/mol the latter one features considerably larger relative energies than other subsets in IE. Our results for both subsets are in line with the numbers obtained for other subsets in this category, DHs do not perform particularly well here, although revDOD-PBEP86-D3(BJ)/D4 yield MADs below 1.0 kcal/mol for ISOL24, whereas the best hybrid is SOGGA11-X-D3(BJ) with nearly 1.5 kcal/mol. With SCAN-D4, the best functional for UPU23 is actually a meta-GGA.

Finally, we recommend the B97-D4 functionals as a reliable, robust and cheap DFA for easy isomerization reactions. Also revDSD-PBEP86-D3(BJ)/D4 and the cheaper alternative revDOD-PBEP86-D3(BJ) can be recommended, although some care must be taken for test sets for which a slow convergence to the CBS limit is documented.

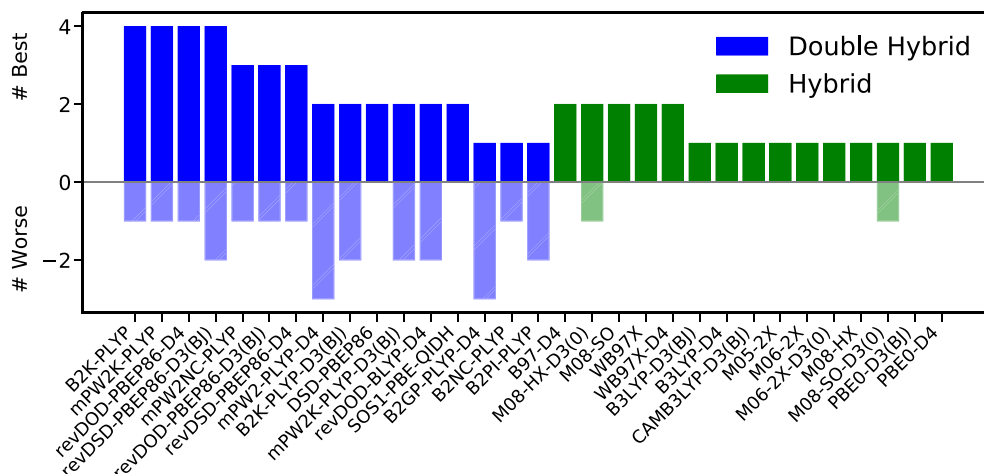
### 3.4.4 | Difficult isomerization reactions

We proceed with our discussion of relative conformational energies by analyzing our results for the 159 isomerization reactions in the category of difficult isomerization reactions. As pointed out by Head-Gordon and coworkers, the systems in this category, mainly consisting of strained and conjugated organic molecules with a difficult electronic structure and an average energy of 25.96, pose a big challenge for DFT methods.<sup>[168]</sup> For the corresponding subcategory in MGCD84, essentially being identical to ours, no GGA or hybrid yields a RMSD lower than 2 kcal/mol, while for their dataset of easy isomerization reactions 23 functionals yield RMSDs lower than 0.5 kcal/mol.<sup>[168]</sup>

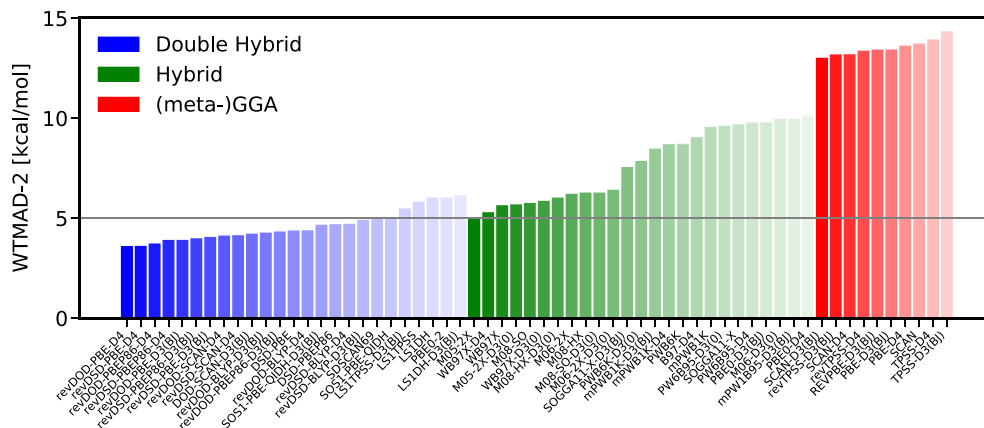
The failure of common DFAs originates from the pronounced many-electron SIE in the systems contained in ID<sup>[168]</sup> and DHs, usually featuring large fractions of exact exchange, should provide a better description of their electronic structures. This expectation is immediately confirmed by the WTMAD-2s presented in Figure 9. As for barrier heights, with WTMAD-2s of up to more than 15 kcal/mol, the (meta-)GGAs benchmarked by us show a devastating performance and are clearly inadequate for ID. The best hybrids all feature a large amount of exact exchange—M05-2X with 56%, the range-separated  $\omega$ B97-X with 100% in the long-range-regime, and M08-SO with 57%. However, the WTMAD-2 of 2.35 kcal/mol obtained with M05-2X is still 41% higher than the 1.67 kcal/mol obtained with revDOD-PBE-D4. The latter functional has 68% exact exchange, a feature common to all other well-performing DHs in this category.

Notwithstanding the great performance of DHs for this subcategory, they are generally outperformed by hybrids in the

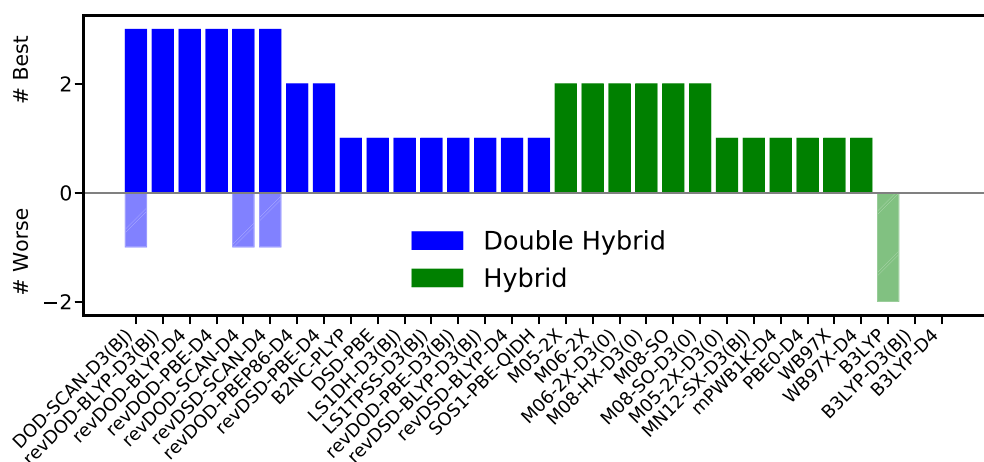
**FIGURE 8** Best (full colored bars, positive x-range)–worse (shaded bars, negative x-range) analysis for the best 15 DHs (blue) and hybrids (green) for easy isomerization reactions [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]







**FIGURE 9** The best 25 DHs (blue) and hybrids (green) as well as the best 10 (meta-)GGAs (red) according to their WTMAD-2 s for the category of difficult isomerization reactions [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 10** Best (full colored bars, positive x-range)–worse (shaded bars, negative x-range) analysis for the best 15 DHs for difficult isomerization reactions [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

description of strained carbon clusters. It was already noted by Martin et al. more than 20 years ago<sup>[307]</sup> that MP2 tends to overestimate the resonance effect in these systems and indeed,  $\omega$ B97-X outperforms all DHs for C20C24. Furthermore, for reasons remaining elusive to us, compared to DFAs of lower rung, DHs show a devastating performance for the C60ISO subset, confirming previous results.<sup>[57,252]</sup> As all C60 isomers are nonaromatic,<sup>[308]</sup> the overestimation of the resonance effect is ruled out as possible cause. The best-performing functional here is revDOD-BLYP-D3(BJ) with a MAD of nearly 3 kcal/mol. Given the average reaction energy of nearly 100 kcal/mol of this subset, this number is clearly satisfying. However, a MAD of less than 2 kcal/mol can be obtained using MO6.

Also DHs based on the SCAN meta-GGA, usually not performing well in our benchmarks, are among the best DFAs for ID. On the other hand, the best-worse analysis shown in Figure 10 reveals that DOD-SCAN-D3(BJ), revDSD-SCAN-D4 and revDOD-SCAN-D4 are both among the worst DHs once (for the very small IDISP subset), while they are among the best performing functionals three times. These functionals are especially suitable for the description of strained ring systems, that is, Styrene45, C20C24 and also DIE60. revDOD-PBE-D4, the functional with the smallest WTMAD-2, is three times among the best functional but never among the worst.

We also emphasize the excellent performance of revDOD-BLYP-D3 (BJ) being the best functional for EIE22, DIE60 and ISOMERIZATION20 but only showing a mediocre performance for the remaining subsets.

Our final recommendation for this category is therefore clear—DOD-functionals, especially revDOD-PBE-D4 and revDODPBP86-D4, but also revDOD-BLYP-D3(BJ) and DOD-SCAN-D3(BJ) in case of strained ring systems, should always be used.

### 3.4.5 | Intermolecular noncovalent interactions

The interactions stabilizing the structures of biomacromolecules like DNA and proteins<sup>[269]</sup> or governing the self-assembly process of nanomaterials, but also interactions between a drug and a protein or intramolecular interactions in donor-acceptor complexes are mostly of noncovalent nature.<sup>[309]</sup> Consequently, with 11 subsets and a total of 375 datapoints, this interaction type comprises a large fraction of our database. The low average reaction energy of 11.84 kcal/mol nicely illustrates why NCIs are often denoted as weak interactions.

The fact that semi-local and hybrid-functionals do not reproduce the correct  $1/r^6$ -behaviour of the London-dispersion

interaction<sup>[310–312]</sup> has often been used as an argument for the importance of explicit electron correlation to accurately describe NCIs.<sup>[48,132,134,135,144]</sup> Indeed, Engel et al. could show<sup>[313]</sup> that the exchange-correlation functional from second order Levy–Görling perturbation theory (LG-PT2) contains the leading contribution to the van der Waals interaction which might have served as the main motivation for the construction of an early DH which includes an MP2 term in the long-range regime only.<sup>[314]</sup>

Modern DHs can be seen as semiempirical “corrections” of LG-PT2 and it seems to be reasonable to expect them to yield substantial improvements over hybrids and (meta-)GGAs for NCIs. In practice however, NCIs can be described very well by partnering a GGA or a hybrid with some form of empirical dispersion correction.<sup>[57]</sup> It has been pointed out by Goerigk et al.<sup>[311]</sup> that a plethora of DFAs performs well on the S66 and the S66x8 test sets if only some form of empirical dispersion correction is considered.

The WTMAD-2s we present in Figure 11 partly confirm these findings. Dispersion corrections are crucial for the accurate description of NCIs. The only exception are Minnesota functionals already containing terms taking into account the interaction of overlapping electron densities of individual monomers. As already pointed out by others,<sup>[57,154,315]</sup> they generally do not perform well for the present category. Many GGAs perform well here; the best GGA, BLYP-D4, is only outperformed by four hybrid functionals, the best of them being B3LYP-D4. In general, we find that D4 is to be preferred over D3(BJ), although the best DH is B2T-PLYP-D3(BJ) not being assessed in conjunction with D4. DHs do not give an improved description of NCIs, although they are more robust—the 25th ranked DH is outperformed by 11 hybrids only. With the exception of mPW2-PLYP-D4, the best eight DHs contain B88 exchange and LYP correlation.

While the analysis of the WTMAD-2s favors D4 over D3(BJ) for B3LYP, as shown in Figure 12, the latter combination is three times among the best hybrid functionals but the former only twice. Also B97-D4, CAM-B3LYP-D3(BJ) and PW6B95-D3(O) are among the best DFAs thrice. B2K-PLYP-D3(BJ) is the only DH being three times among the best ones of this rung on Jacob's ladder.

It is instructive to take a closer look at the important S66, S10x8 and 3B-69-TRIM subsets. With 215 datapoints in total, they comprise nearly two third of the NCI category and serve as a representative benchmark for NCIs between neutral organic molecules, including dimers in nonequilibrium as well as in their equilibrium geometry and trimers. B2K-PLYP-D3(BJ) describes noncovalently bounded complexes in their equilibrium geometry by far most accurately and its fifth position for S10x8 demonstrates that it is reliable for non-equilibrium geometries also. B3LYP-D3(BJ) is the best hybrid functional for 3B-69-TRIM and S66 but cannot be found among the best 15 DFAs for S10x8. On the GGA level, we also emphasize the great performance of BLYP-D3(BJ) for all three subsets.

For the large molecular complexes in the L7 test sets, BLYP-D3(BJ) is the second best of all benchmarked DFAs, only being outperformed by B3LYP-D3(BJ). As for S66 and 3B-69-Trim and S66x10, a DOD-functional cannot be found among the best 15 DFAs, indicating the importance of a balanced inclusion of same-spin and opposite-spin correlation for the description of NCIs.

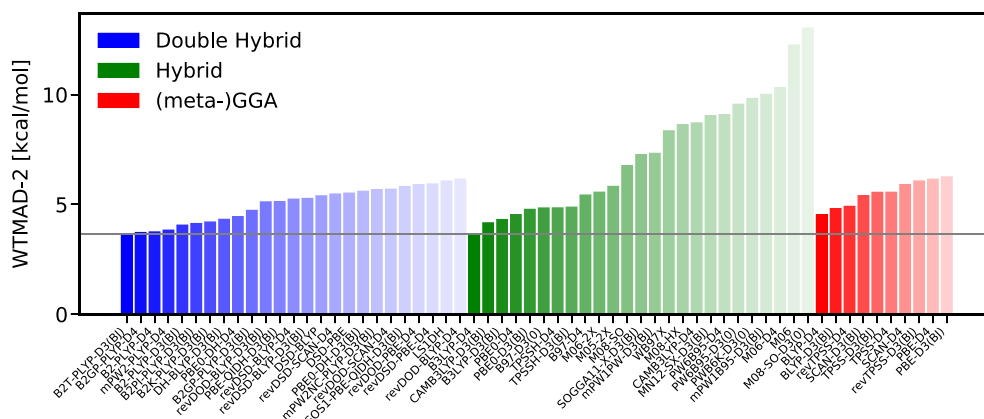
In conclusion, B2K-PLYP-D3(BJ) is the method of choice for the description of neutral, noncovalently bounded organic complexes, although B3LYP-D3(BJ) and BLYP-D3(BJ) perform quite good as well and especially the latter provides an excellent price/performance-ratio. With the exception of these (quite important) systems, D4 dispersion correction is generally to be preferred over D3(BJ).

### 3.4.6 | The entire database

To conclude the discussion in this subsection, we shortly comment on the performance of the benchmarked DFAs for all 1,644 datapoints in our database. As for all subcategories, Figure 13 shows the WTMAD-2 s of the 25 best performing DHs (blue) and hybrids (green) as well as the best 10 meta-GGAs (red) together with their empirical dispersion correction.

We find that, also on the TZ level, using STOs and PARI, Jacob's ladder is still reproduced—the best hybrid functionals clearly

**FIGURE 11** The best 25 DHs (blue) and hybrids (green) as well as the best 10 (meta-)GGAs (red) according to their WTMAD-2 s for the category of intermolecular noncovalent interactions [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]





(2) outperforms all other functionals, it is closely followed by revDSD-PBEP86-D4 and the same functional in its DOD-variant and with different flavors of dispersion correction. Our benchmark favors DHs based on B88 exchange and LYP correlation, revDSD-BLYP-D4 being the best functional and revDSD-PBEP86-D4 only ranked sixth. While we can confirm the DSD-approach (with the exception of range-separated DHs based on  $\omega$ B97) to be the method of choice towards the construction of robust DHs, B2GP-PLYP-D3(BJ), one of the first DHs ever constructed, also performs surprisingly well in our benchmark, being only average in studies on GMTKN55.<sup>[61]</sup>

Although our database is not identical to GMTKN55, it has considerable overlap. The highlighted differences show that care must be taken in the interpretation of benchmarks of DFAs and result obtained at the CBS limit are not necessarily transferable to smaller basis sets.

It is also interesting to observe, that the best three hybrid functionals, M05-2X, M06-2X and M08-S0 do not feature an empirical dispersion-correction term unlike, with the exception of LS1DH, the best 25 DHs. revDSD-BLYP-D4 outperforms all other functionals in this study, followed by B2GP-PLYP and revDOD-BLYP-D3(BJ). Using PARI-AO-MP2, the latter one is faster than the best hybrid-functionals M05-2X and M06-2X for medium systems and only negligibly slower for large molecules.

However, we hesitate to recommend it as a general purpose functional. The best-worse analysis in Figure 14 reveals that it can also be found once among the worst DHs and the same holds for all other DHs. On the hybrid level, B97-D4 is 12 times among the best and never among the worst functionals. On the other hand, it shows a considerably worse WTMA-D2 than all of the best 25 DHs. Given the large number of subsets investigated herein, 12 subsets out of 58 does actually not indicate great general performance: It rather means that B97-D4 is only a mediocre choice for nearly 80% of all subsets.

Given these observations, we do not recommend a general-purpose functional here. For different chemical problems, different functionals should be used. This insight highlights the utmost importance of the availability of high-quality benchmarks on large and diverse datasets which can serve as a guideline for the selection of the right functional for the particular problem at hand.

## 4 | CONCLUSION

In this work, we demonstrated that accurate PARI-MP2 and PARI-K calculations at the TZ level can be performed in a numerically robust way. As shown for the hybrid-DFA B3LYP and the DH-functional B2GP-PLYP, the BSE of the STO-type TZ2P basis set in conjunction with PARI is comparable to the one from the slightly larger GTO-type def2-TZVPP basis set used with global DF. While we obtained BSEs of 0.68 kcal/mol (0.72 for def2-TZVPP) on average for B3LYP, the same quantity is with 1.12 kcal/mol (0.96 for def2-TZVPP) considerably larger for B2GP-PLYP.

These BSEs strongly suggest that the use of QZ basis sets is clearly indispensable for DHs to exploit their full potential. This can be seen most clearly for ground state energies of conformers of organic

molecules with easy electronic structure and NCI where much higher accuracy than the often cited 1 kcal/mol criterion of chemical accuracy is crucial for predictive quantum chemistry. It is the most pressing problem of PARI-MP2, that QZ calculations are not doable at the moment. Research towards a strategy to overcome the limiting numerical issues is currently pursued in our group.

Nevertheless, we could demonstrate on a large and diverse database of 1,644 datapoints, that Jacob's ladder is still reproduced—DHs are the most accurate and robust DFAs available and many of them outperform the best hybrid functionals. Possibly due to the large BSEs for DHs and maybe also due to larger PARI-errors in the computation of four-center integrals, the increase in accuracy of DHs over hybrids is less pronounced than the one found in recent studies on GMTKN55.

Among the 14 DHs outperforming all hybrid functionals, also five DOD-functionals can be found which enable DH calculations to essentially the price of a hybrid calculation. For five test sets of large molecules we could show that their accuracy is also retained for molecules of more than 100 atoms, including biologically relevant enzymes and isomers of large organic molecules. Consequently, DOD-functionals are the method of choice when CPU-time and memory requirements are an issue and highly accurate energies are required.

In summary, the use of DHs is recommended for various chemical problems. On the TZ level, they especially excel in thermochemistry, the calculation of barrier heights and the calculation of relative conformational energies of organic molecules with challenging electronic structure. This is not only true for main group chemistry but especially when TMs are involved. In this case, highly accurate energies can be obtained from DHs in conjunction with the ZORA formalism.

As an important results of this work, we shortly recapitulate our recommendations for the individual subcategories of our database.

- 1 Basic properties and reaction energies:** DHs should always be used here. B2T-PLYP-D3(BJ) provides the best balance between accuracy and robustness. If CPU-time is a concern, revDOD-BLYP-D3(BJ) should be used.
- 2 Reaction barrier heights:** revDOD-PBEP86-D3(BJ) can always be recommended, but in the absence of TMs, M08-HX is a robust and accurate alternative. When TMs are involved, B2K-PLYP provides excellent performance.
- 3 Simple isomerization reactions:** DHs cannot be recommended here. We recommend B97-D4 due to its excellent price/performance ratio.
- 4 Difficult isomerization reactions:** DOD-functionals are the method of choice here, especially revDOD-PBE-D4 and revDOD-PBEP86-D4. Note, however, that DHs fail badly in the description of the electronic structure of fullerenes.
- 5 Intermolecular noncovalent interactions:** For neutral, noncovalently bounded complexes, the DH B2K-PLYP-D3(BJ) can be recommended. BLYP-D3(BJ) on the other hand offers an excellent price/performance ratio and NCIs are the only category where we can safely recommend a GGA. Empirical dispersion correction should always be used, and D4 is generally to be preferred over D3(BJ).

Apart from these findings, we could draw two other important conclusions from the results presented herein. First, a general purpose functional remains to be developed. None of the herein benchmarked functionals can safely be applied to all chemical problems assessed here. Even for a single subcategory, there is almost never a functional which can be recommended for all subsets. Thus, the recommendations given herein might serve as important guidelines but they should never be understood as black-box solutions for a problem at hand.

Second, in our benchmarks we could not reproduce the ranking of functionals according to their accuracy on GMTKN55 when large QZ basis sets are used. While we can confirm some important results presented previously, for example, the outstanding success of the Minnesota-functional M05-2X, M06-2X and M08-HX on the hybrid level, the great performance of revDSD-functionals<sup>[60,61]</sup> or the considerably worse performance of one-parameter functionals,<sup>[58]</sup> revDSD-PBEP86-D4 is not the frontrunner in our study but rather revDSD-BLYP-D4. On the other hand, the second best DFA in our study, B2GP-PLYP, only shows an average performance on GMTKN55 with QZ basis sets. This indicates that benchmark results obtained with QZ basis sets do not necessarily apply to the TZ level also, especially for DHs, where BSEs can become quite large.

As the certainly most important outcome of this study, we conclude this work by giving a definite answer to the question asked in the very beginning—is it possible to perform accurate, robust and efficient hybrid- and DH-KS calculations using STOs? We are convinced to have sufficiently demonstrated that this is indeed possible. Clearly, the numerical instability of PARI-MP2 for QZ basis sets and larger is still an important issue which must be tackled to unleash the full potential of the PARI-approach and research in this direction is currently pursued in our group.

PARI-K and PARI-MP2 as implemented in ADF essentially retain the accuracy of their parent algorithms and enable accurate and reliable quantum chemistry with STOs on the TZ level. At the same time, they tremendously accelerate hybrid- as well as DH calculations for DOD-functionals and therefore enable routine application of both, hybrids and DHs to molecules of hundreds of atoms.

## ACKNOWLEDGMENTS

This research received funding from the Netherlands Organization for Scientific Research (NWO) in the framework of the Innovation Fund for Chemistry and from the Ministry of Economic Affairs in the framework of the “TKI/PPS-Toeslagregeling”. The authors thank Erik van Lenthe for fruitful discussions.

## ORCID

Arno Förster  <https://orcid.org/0000-0002-0957-4081>

Lucas Visscher  <https://orcid.org/0000-0002-7748-6243>

## REFERENCES

- [1] E. H. Rubensson, E. Rudenberg, P. Salek, *J. Chem. Theory Comput.* **2001**, *7*, 340–350.
- [2] W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133.
- [3] P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, *136*, 864.
- [4] M. Levy, *Physics (College Park, Md.)*. **1979**, *76*, 6062.
- [5] E. Engel, R. M. Dreizler, *Density functional theory: An advanced course*, Springer, Berlin, **2013**.
- [6] A. D. Becke, *J. Chem. Phys.* **2014**, *140*, 18A301.
- [7] A. Pribram-Jones, D. A. Gross, and K. Burke, *Annu. Rev. Phys. Chem.* **66**, 283 (2015), 1408.4826.
- [8] S. Grimme, P. R. Schreiner, *Angew. Chemie - Int. Ed.* **2018**, *57*, 4170.
- [9] L. Goerigk, N. Mehta, *Aust. J. Chem.* **2019**, *72*, 563.
- [10] J. Čížek, J. Paldus, *Int. J. Quantum Chem.* **1971**, *5*, 359.
- [11] G. D. Purvis, R. J. Bartlett, *J. Chem. Phys.* **1982**, *76*, 1910.
- [12] R. J. Bartlett, *J. Phys. Chem.* **1989**, *93*, 1697.
- [13] K. Raghavachari, G. Trucks, J. A. Pople, M. Head-Gordon, *Chem. Phys. Lett.* **1989**, *157*, 479.
- [14] C. Hampel, H. J. Werner, *J. Chem. Phys.* **1996**, *104*, 6286.
- [15] M. Schütz, H. J. Werner, *Chem. Phys. Lett.* **2000**, *318*, 370.
- [16] M. Schütz, *J. Chem. Phys.* **2002**, *116*, 8772.
- [17] M. Schütz, F. R. Manby, *Phys. Chem. Chem. Phys.* **2003**, *5*, 3349.
- [18] F. Neese, F. Wennmohs, A. Hansen, *J. Chem. Phys.* **2009**, *130*, 114108.
- [19] F. Neese, A. Hansen, D. G. Liakos, *J. Chem. Phys.* **2009**, *131*, 064103.
- [20] S. Kossmann, F. Neese, *J. Phys. Chem. A* **2010**, *114*, 11768.
- [21] J. Yang, G. K. L. Chan, F. R. Manby, M. Schütz, H. J. Werner, *J. Chem. Phys.* **2012**, *136*, 144105.
- [22] C. Riplinger, F. Neese, *J. Chem. Phys.* **2013**, *138*, 034106.
- [23] C. Riplinger, B. Sandhoefer, A. Hansen, F. Neese, *J. Chem. Phys.* **2013**, *139*, 134101.
- [24] G. Schmitz, B. Helmich, C. Hättig, *Mol. Phys.* **2013**, *111*, 2463.
- [25] M. Schütz, J. Yang, G. K. L. Chan, F. R. Manby, H. J. Werner, *J. Chem. Phys.* **2013**, *138*, 054109.
- [26] G. Schmitz, C. Hättig, D. P. Tew, *Phys. Chem. Chem. Phys.* **2014**, *16*, 22167.
- [27] M. Sparta, F. Neese, *Chem. Soc. Rev.* **2014**, *43*, 5032.
- [28] M. Schwilk, D. Usvyat, H. J. Werner, *J. Chem. Phys.* **2015**, *142*, 121102.
- [29] D. G. Liakos, M. Sparta, M. K. Kesharwani, J. M. Martin, F. Neese, *J. Chem. Theory Comput.* **2015**, *11*, 1525.
- [30] D. G. Liakos, F. Neese, *J. Chem. Theory Comput.* **2015**, *11*, 4054.
- [31] C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, F. Neese, *J. Chem. Phys.* **2016**, *144*, 024109.
- [32] G. Schmitz, C. Hättig, *J. Chem. Phys.* **2016**, *145*, 234107.
- [33] M. Schwilk, Q. Ma, C. Köppl, H. J. Werner, *J. Chem. Theory Comput.* **2017**, *13*, 3650.
- [34] F. Pavošević, C. Peng, P. Pinski, C. Riplinger, F. Neese, E. F. Valeev, *J. Chem. Phys.* **2017**, *146*, 174108.
- [35] Q. Ma, M. Schwilk, C. Köppl, H. J. Werner, *J. Chem. Theory Comput.* **2017**, *13*, 4871.
- [36] P. R. Nagy, M. Kállay, *J. Chem. Phys.* **2017**, *146*, 214106.
- [37] Y. Guo, C. Riplinger, U. Becker, D. G. Liakos, Y. Minenkov, L. Cavallo, F. Neese, *J. Chem. Phys.* **2018**, *148*, 011101.
- [38] P. R. Nagy, G. Samu, M. Kállay, *J. Chem. Theory Comput.* **2018**, *14*, 4193.
- [39] Q. Ma, H. J. Werner, *J. Chem. Theory Comput.* **2018**, *14*, 198.
- [40] Q. Ma, H. J. Werner, *J. Chem. Theory Comput.* **2019**, *15*, 1044.
- [41] P. R. Nagy, M. Kállay, *J. Chem. Theory Comput.* **2019**, *15*, 5275.
- [42] R. J. Bartlett, *J. Chem. Phys.* **2019**, *151*, 160901.
- [43] A. D. Becke, *J. Chem. Phys.* **2005**, *122*, 064101.
- [44] A. Marini, P. García-González, A. Rubio, *Phys. Rev. Lett.* **2006**, *96*, 2.
- [45] H. Jiang, E. Engel, *J. Chem. Phys.* **2007**, *127*, 184108.
- [46] H. Jiang, E. Engel, *Zeitschrift für Phys. Chemie* **2010**, *224*, 455.
- [47] A. J. Cohen, P. Mori-Sánchez, W. Yang, *Chem. Rev.* **2012**, *112*, 289.
- [48] S. Grimme, *J. Chem. Phys.* **2006**, *124*, 034108.
- [49] S. Grimme, C. Mück-Lichtenfeld, E. U. Würthwein, A. W. Ehlers, T. P. Goumans, K. Lammertsma, *J. Phys. Chem. A* **2006**, *110*, 2583.
- [50] T. Schwabe, S. Grimme, *Phys. Chem. Chem. Phys.* **2006**, *8*, 4398.
- [51] T. Schwabe, S. Grimme, *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397.

- [52] S. Grimme, F. Neese, *J. Chem. Phys.* **2007**, *127*, 154116.
- [53] S. Kossmann, B. Kirchner, F. Neese, *Mol. Phys.* **2007**, *105*, 2049.
- [54] F. Neese, T. Schwabe, S. Grimme, *J. Chem. Phys.* **2007**, *126*, 124115.
- [55] A. Tarnopolsky, A. Karton, R. Sertchook, D. Vuzman, J. M. L. Martin, *J. Phys. Chem. A* **2008**, *112*, 3.
- [56] J. P. Perdew, K. Schmidt, *AIP Conf. Proc.* **2001**, *577*, 1.
- [57] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, S. Grimme, *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184.
- [58] N. Mehta, M. Casanova-Páez, L. Goerigk, *Phys. Chem. Chem. Phys.* **2018**, *20*, 23175.
- [59] A. Najibi, L. Goerigk, *J. Chem. Theory Comput.* **2018**, *14*, 5725.
- [60] G. Santra, N. Sylvetsky, J. M. Martin, *J. Phys. Chem. A* **2019**, *123*, 5129.
- [61] J. M. L. Martin, G. Santra, *Israel J. Chem.* **2019**, *60*, 1.
- [62] The notion of  $N^\times$  always implies  $\mathcal{O}(N^\times)$ : That is, the function  $f$ , mapping the system size to the runtime of an algorithm, does asymptotically not grow faster than  $N^\times$ .
- [63] We follow the usual convention and denote with  $N$  a generalized measure of the system size, proportional to the number of atoms.
- [64] M. Häser, R. Ahlrichs, *J. Comput. Chem.* **1989**, *10*, 104.
- [65] C. A. White, B. G. Johnson, P. M. Gill, M. Head-Gordon, *Chem. Phys. Lett.* **1994**, *230*, 8.
- [66] D. L. Strout, G. E. Scuseria, *J. Chem. Phys.* **1995**, *102*, 8448.
- [67] E. Schwegler, M. Challacombe, M. Head-Gordon, *J. Chem. Phys.* **1997**, *106*, 9708.
- [68] P. Safek, S. Høst, L. Thøgersen, P. Jørgensen, P. Manninen, J. Olsen, B. Jansik, S. Reine, F. Pawłowski, E. Tellgren, T. Helgaker, S. Coriani, *J. Chem. Phys.* **2007**, *126*, 114110.
- [69] F. Aquilante, R. Lindh, T. Bondo Pedersen, *J. Chem. Phys.* **2007**, *127*, 114107.
- [70] S. Reine, T. Helgaker, R. Lindh, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 290.
- [71] E. J. Baerends, D. E. Ellis, P. Ros, *Chem. Phys.* **1973**, *2*, 41.
- [72] J. L. Whitten, *J. Chem. Phys.* **1973**, *58*, 4496.
- [73] H. Sambe, R. H. Felton, *J. Chem. Phys.* **1975**, *62*, 1122.
- [74] B. I. Dunlap, J. W. Connolly, J. R. Sabin, *J. Chem. Phys.* **1979**, *71*, 3396.
- [75] B. I. Dunlap, J. W. Connolly, J. R. Sabin, *J. Chem. Phys.* **1979**, *71*, 4993.
- [76] C. Van Alsenoy, *J. Comput. Chem.* **1988**, *9*, 620.
- [77] G. Te Velde, E. J. Baerends, *Phys. Rev. B* **1991**, *44*, 7888.
- [78] O. Vahtras, J. Almlöf, M. W. Feyereisen, *Chem. Phys. Lett.* **1993**, *213*, 514.
- [79] C. F. Guerra, J. G. Snijders, G. te Velde, E. J. Baerends, *Theor. Chem. Acc.* **1998**, *99*, 391.
- [80] B. I. Dunlap, *J. Mol. Struct. THEOCHEM* **2000**, *529*, 37.
- [81] B. I. Dunlap, *Phys. Chem. Chem. Phys.* **2000**, *2*, 2113.
- [82] B. I. Dunlap, N. Rösch, S. B. Trickey, *Mol. Phys.* **2010**, *108*, 3167.
- [83] G. Reza Ahmadi, J. Almlöf, *Chem. Phys. Lett.* **1995**, *246*, 364.
- [84] C. A. White, M. Head-Gordon, *J. Chem. Phys.* **1996**, *104*, 2620.
- [85] Y. Shao, M. Head-Gordon, *Chem. Phys. Lett.* **2000**, *323*, 425.
- [86] Y. Shao, C. A. White, M. Head-Gordon, *J. Chem. Phys.* **2001**, *114*, 6572.
- [87] F. Neese, *J. Comput. Chem.* **2003**, *24*, 1740.
- [88] R. A. Friesner, *Chem. Phys. Lett.* **1985**, *116*, 39.
- [89] R. A. Friesner, *J. Chem. Phys.* **1986**, *85*, 1462.
- [90] R. A. Friesner, *J. Chem. Phys.* **1987**, *86*, 3522.
- [91] M. N. Ringnalda, M. Belhadj, R. A. Friesner, *J. Chem. Phys.* **1998**, *3397*, 3397.
- [92] F. Neese, F. Wennmohs, A. Hansen, U. Becker, *Chem. Phys.* **2009**, *356*, 98.
- [93] R. Izsák, F. Neese, *J. Chem. Phys.* **2011**, *135*, 144105.
- [94] M. Guidon, J. Hutter, J. Vandevondele, *J. Chem. Theory Comput.* **2010**, *6*, 2348.
- [95] P. Merlot, R. Izsák, A. Borgoo, T. Kjærgaard, T. Helgaker, S. Reine, *J. Chem. Phys.* **2014**, *141*, 094104.
- [96] C. Kumar, H. Fliegl, F. Jensen, A. M. Teale, S. Reine, T. Kjærgaard, *Int. J. Quantum Chem.* **2018**, *118*, 25639.
- [97] R. Polly, H. J. Werner, F. R. Manby, P. J. Knowles, *Mol. Phys.* **2004**, *102*, 2311.
- [98] A. Sodt, J. E. Subotnik, M. Head-Gordon, *J. Chem. Phys.* **2006**, *125*, 194109.
- [99] A. Sodt, M. Head-Gordon, *J. Chem. Phys.* **2008**, *128*, 104106.
- [100] E. Rebolini, R. Izsák, S. S. Reine, T. Helgaker, T. B. Pedersen, *J. Chem. Theory Comput.* **2016**, *12*, 3514.
- [101] M. A. Watson, N. C. Handy, A. J. Cohen, *J. Chem. Phys.* **2003**, *119*, 6475.
- [102] M. Krykunov, T. Ziegler, E. Van Lenthe, *Int. J. Quantum Chem.* **2009**, *109*, 1676.
- [103] P. Merlot, T. Kjærgaard, T. Helgaker, R. Lindh, F. Aquilante, S. Reine, T. B. Pedersen, *J. Comput. Chem.* **2013**, *34*, 1486.
- [104] M. Feyereisen, G. Fitzgerald, A. Komornicki, *Chem. Phys. Lett.* **1993**, *208*, 359.
- [105] F. Weigend, M. Häser, H. Patzelt, R. Ahlrichs, *Chem. Phys. Lett.* **1998**, *294*, 143.
- [106] H. J. Werner, F. R. Manby, *J. Chem. Phys.* **2006**, *124*, 054114.
- [107] T. Nakajima, K. Hirao, *Chem. Phys. Lett.* **2006**, *427*, 225.
- [108] L. Maschio, D. Usvyat, F. R. Manby, S. Casassa, C. Pisani, M. Schütz, *Phys. Rev. B* **2007**, *76*, 1.
- [109] L. Vogt, R. Olivares-Amaya, S. Kermes, Y. Shao, C. Amador-Bedolla, A. Aspuru-Guzik, *J. Phys. Chem. A* **2008**, *112*, 2049.
- [110] X. Ren, P. Rinke, V. Blum, J. Wierferink, A. Tkatchenko, A. Sanfilippo, K. Reuter, M. Scheffler, *New J. Phys.* **2012**, *14*, 053020.
- [111] T. Ishikawa, K. Kuwata, *J. Phys. Chem. Lett.* **2012**, *3*, 375.
- [112] M. Katouda, T. Nakajima, *J. Chem. Theory Comput.* **2013**, *9*, 5373.
- [113] P. Baudin, P. Ettenhuber, S. Reine, K. Kristensen, T. Kjærgaard, *J. Chem. Phys.* **2016**, *144*, 054102.
- [114] T. Kjærgaard, *J. Chem. Phys.* **2017**, *146*, 044103.
- [115] B. Cooper, S. Girdlestone, P. Burovskiy, G. Gaydadjiev, V. Averbukh, P. J. Knowles, W. Luk, *J. Chem. Theory Comput.* **2017**, *13*, 5265.
- [116] P. Pulay, *Chem. Phys. Lett.* **1983**, *100*, 151.
- [117] S. Sæbø, P. Pulay, *Chem. Phys. Lett.* **1985**, *113*, 13.
- [118] W. Kohn, *Phys. Rev. Lett.* **1996**, *76*, 3168.
- [119] E. Prodans, W. Kohn, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 11635.
- [120] P. Pinski, C. Riplinger, E. F. Valeev, F. Neese, *J. Chem. Phys.* **2015**, *143*, 034108.
- [121] H. J. Werner, G. Knizia, C. Krause, M. Schwilk, M. Dornbach, *J. Chem. Theory Comput.* **2015**, *11*, 484.
- [122] P. R. Nagy, G. Samu, M. Kállay, *J. Chem. Theory Comput.* **2016**, *12*, 4897.
- [123] M. Kobayashi, Y. Imamura, H. Nakai, *J. Chem. Phys.* **2007**, *127*, 074103.
- [124] K. Kristensen, T. Kjærgaard, I. M. Høyvik, P. Ettenhuber, P. Jørgensen, B. Jansik, S. Reine, J. Jakowski, *Mol. Phys.* **2013**, *111*, 1196.
- [125] J. Almlöf, *Chem. Phys. Lett.* **1991**, *181*, 319.
- [126] M. Häser, J. Almlöf, *J. Chem. Phys.* **1992**, *96*, 489.
- [127] M. Haser, *Theor. Chim. Acta* **1993**, *87*, 147.
- [128] P. Y. Ayala, G. E. Scuseria, *J. Chem. Phys.* **1999**, *110*, 3660.
- [129] G. E. Scuseria, P. Y. Ayala, *J. Chem. Phys.* **1999**, *111*, 8330.
- [130] S. Schweizer, B. Doser, C. Ochsenfeld, *J. Chem. Phys.* **2008**, *128*, 154101.
- [131] J. Zienau, L. Clin, B. Doser, C. Ochsenfeld, *J. Chem. Phys.* **2009**, *130*, 204112.
- [132] B. Doser, D. S. Lambrecht, J. Kussmann, C. Ochsenfeld, *J. Chem. Phys.* **2009**, *130*, 064107.
- [133] B. Doser, J. Zienau, L. Clin, D. S. Lambrecht, C. Ochsenfeld, *Zeitschrift für Phys. Chemie* **2010**, *224*, 397.
- [134] S. A. Maurer, D. S. Lambrecht, J. Kussmann, C. Ochsenfeld, *J. Chem. Phys.* **2013**, *138*, 014101.
- [135] S. A. Maurer, M. Beer, D. S. Lambrecht, C. Ochsenfeld, *J. Chem. Phys.* **2013**, *139*, 184104.

- [136] S. A. Maurer, J. Kussmann, C. Ochsenfeld, *J. Chem. Phys.* **2014**, *141*, 051106.
- [137] S. A. Maurer, L. Clin, C. Ochsenfeld, *J. Chem. Phys.* **2014**, *140*, 224112.
- [138] S. Vogler, M. Ludwig, M. Maurer, C. Ochsenfeld, *J. Chem. Phys.* **2017**, *147*, 024101.
- [139] T. J. Martínez, E. A. Carter, T. J. Martínez, E. A. Carter, *J. Chem. Phys.* **1993**, *100*, 3631.
- [140] E. G. Hohenstein, R. M. Parrish, T. J. Martínez, *J. Chem. Phys.* **2012**, *137*, 1085.
- [141] R. M. Parrish, E. G. Hohenstein, N. F. Schunck, C. D. Sherrill, T. J. Martínez, *Phys. Rev. Lett.* **2013**, *111*, 1.
- [142] C. Song, T. J. Martínez, *J. Chem. Phys.* **2016**, *144*, 174111.
- [143] C. Song, T. J. Martínez, *J. Chem. Phys.* **2017**, *146*, 034104.
- [144] A. Förster, M. Franchini, E. van Lenthe, L. Visscher, *J. Chem. Theory Comput.* **2020**, *16*, 875.
- [145] Y. Jung, R. C. Lochan, A. D. Dutoi, M. Head-Gordon, *J. Chem. Phys.* **2004**, *121*, 9793.
- [146] G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. van Gisbergen, J. G. Snijders, T. Ziegler, *J. Comput. Chem.* **2001**, *22*, 931.
- [147] E. Baerends, T. Ziegler, A. Atkins, J. Autschbach, O. Basergio, D. Bashford, A. Bérces, F. Bickelhaupt, C. Bo, P. Boerrigter, et al., *ADF2019.3, SCM, theoretical chemistry*, Vrije Universiteit, Amsterdam, The Netherlands **2019**.
- [148] D. S. Hollman, H. F. Schaefer, E. F. Valeev, *J. Chem. Phys.* **2014**, *140*, 064109.
- [149] L. N. Wirz, S. S. Reine, T. B. Pedersen, *J. Chem. Theory Comput.* **2017**, *13*, 4897.
- [150] S. F. Manzer, E. Epifanovsky, M. Head-Gordon, *J. Chem. Theory Comput.* **2015**, *11*, 518.
- [151] E. Cancès, C. Le Bris, *ESAIM Math. Model. Numer. Anal.* **2000**, *34*, 749.
- [152] K. N. Kudin, G. E. Scuseria, E. Cancès, *J. Chem. Phys.* **2002**, *116*, 8255.
- [153] X. Liu, X. Wang, Z. Wen, Y. Yuan, *SIAM J. Matrix Anal. Appl.* **2014**, *35*, 546.
- [154] N. Mardirossian, M. Head-Gordon, *J. Chem. Theory Comput.* **2016**, *12*, 4303.
- [155] A. D. Boese, J. M. Martin, N. C. Handy, *J. Chem. Phys.* **2003**, *119*, 3005.
- [156] M. R. Nyden, G. A. Petersson, *J. Chem. Phys.* **1981**, *75*, 1843.
- [157] G. A. Petersson, M. R. Nyden, *J. Chem. Phys.* **1981**, *75*, 3423.
- [158] G. A. Petersson, A. Bennett, T. G. Tensfeldt, M. A. Al-Laham, W. A. Shirley, J. Mantzaris, *J. Chem. Phys.* **1988**, *89*, 2193.
- [159] W. Kutzelnigg, W. Klopper, *J. Chem. Phys.* **1991**, *94*, 1985.
- [160] K. D. Vogiatzis, E. C. Barnes, W. Klopper, *Chem. Phys. Lett.* **2011**, *503*, 157.
- [161] A. Irmeler, A. Grüneis, *J. Chem. Phys.* **2019**, *151*, 104107.
- [162] L. Goerigk, S. Grimme, *J. Chem. Theory Comput.* **2011**, *7*, 291.
- [163] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, The Fritz Haber Institute ab initio molecular simulations package (FHI-aims), <http://www.fhi-berlin.mpg.de/aims>, **2009**.
- [164] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, M. Scheffler, *Comput. Phys. Commun.* **2009**, *180*, 2175.
- [165] I. Y. Zhang, X. Ren, P. Rinke, V. Blum, M. Scheffler, *New J. Phys.* **2013**, *15*, 123033.
- [166] A. C. Ihrig, J. Wierferink, I. Y. Zhang, M. Ropo, X. Ren, P. Rinke, M. Scheffler, V. Blum, *New J. Phys.* **2015**, *17*, 093020.
- [167] N. Mardirossian, M. Head-Gordon, *J. Chem. Theory Comput.* **2013**, *9*, 4453.
- [168] N. Mardirossian, M. Head-Gordon, *Mol. Phys.* **2017**, *115*, 2315.
- [169] L. Goerigk, S. Grimme, T. O. Chemie, *J. Chem. Theory Comput.* **2009**, *6*, 107.
- [170] D. Liakos, Y. Guo, F. Neese, *J. Phys. Chem. A* **2019**, *124*, 90.
- [171] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865.
- [172] J. P. Perdew, M. Ernzerhof, K. Burke, *Phys. Rev. Lett.* **1996**, *77*, 3865.
- [173] S. Grimme, S. Ehrlich, L. Goerigk, *J. Comput. Chem.* **2011**, *32*, 1456.
- [174] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, S. Grimme, *J. Chem. Phys.* **2019**, *150*, 154122.
- [175] A. D. Becke, R. M. Dickson, *J. Chem. Phys.* **1988**, *89*, 2993.
- [176] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785.
- [177] B. Miehlich, A. Savi, H. Stoll, H. Preuss, *J. Pediatr. Urol.* **1989**, *157*, 200.
- [178] Y. Zhang, *165*, 9007 (**1998**).
- [179] J. Tao, J. P. Perdew, V. N. Staroverov, G. E. Scuseria, *Phys. Rev. Lett.* **2003**, *91*, 3.
- [180] J. P. Perdew, J. Tao, V. N. Staroverov, G. E. Scuseria, *J. Chem. Phys.* **2004**, *120*, 6898.
- [181] E. Caldeweyher, C. Bannwarth, S. Grimme, *J. Chem. Phys.* **2017**, *147*, 034112.
- [182] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, L. A. Constantin, J. Sun, *Phys. Rev. Lett.* **2009**, *103*, 026403.
- [183] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, L. A. Constantin, J. Sun, *Phys. Rev. Lett.* **2011**, *106*, 179902.
- [184] J. Sun, A. Ruzsinszky, J. Perdew, *Phys. Rev. Lett.* **2015**, *115*, 1.
- [185] J. G. Brandenburg, J. E. Bates, J. Sun, J. P. Perdew, *Phys. Rev. B* **2016**, *94*, 17.
- [186] A. D. Becke, *J. Chem. Phys.* **1997**, *107*, 8554.
- [187] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* **1994**, *98*, 11623.
- [188] Y. Zhao, D. G. Truhlar, *J. Phys. Chem. A* **2004**, *108*, 6908.
- [189] L. Goerigk, S. Grimme, *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670.
- [190] M. Ernzerhof and G. E. Scuseria, *J. Chem. Phys.* **1999**, *110*, 5029.
- [191] C. Adamo, V. Barone, *J. Chem. Phys.* **1999**, *110*, 6158.
- [192] R. Peverati, D. G. Truhlar, *J. Chem. Phys.* **2011**, *135*, 191102.
- [193] G. I. Csonka, J. P. Perdew, A. Ruzsinszky, *J. Chem. Theory Comput.* **2010**, *6*, 3688.
- [194] Y. Zhao, D. G. Truhlar, *J. Phys. Chem. A* **2005**, *109*, 5656.
- [195] J. D. Chai, M. Head-Gordon, *J. Chem. Phys.* **2008**, *128*, 084106.
- [196] Y. Zhao, N. E. Schultz, D. G. Truhlar, *J. Chem. Theory Comput.* **2006**, *2*, 364.
- [197] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215.
- [198] Y. Zhao, D. G. Truhlar, *J. Chem. Theory Comput.* **2008**, *4*, 1849.
- [199] R. Peverati, D. G. Truhlar, *Phys. Chem. Chem. Phys.* **2012**, *14*, 16187.
- [200] A. Karton, A. Tarnopolsky, J. F. Lamère, G. C. Schatz, J. M. Martin, *J. Phys. Chem. A* **2008**, *112*, 12868.
- [201] J. C. Sancho-García, A. J. Prez-Jimnez, *J. Chem. Phys.* **2009**, *131*, 084108.
- [202] F. Yu, *J. Phys. Chem. A* **2014**, *118*, 3175.
- [203] K. Sharkas, J. Toulouse, A. Savin, *J. Chem. Phys.* **2011**, *134*, 064113.
- [204] E. Brémond and C. Adamo, *J. Chem. Phys.* **2011**, *135*, 0 (2011), 024106.
- [205] L. Goerigk, S. Grimme, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 576.
- [206] J. D. Chai, S. P. Mao, *Chem. Phys. Lett.* **2012**, *538*, 121.
- [207] J. Toulouse, K. Sharkas, E. Brémond, C. Adamo, *J. Chem. Phys.* **2011**, *135*, 1.
- [208] S. M. Souvi, K. Sharkas, J. Toulouse, *J. Chem. Phys.* **2014**, *140*, 084107.
- [209] É. Brémond, J. C. Sancho-García, Á. J. Pérez-Jiménez, C. Adamo, *J. Chem. Phys.* **2014**, *141*, 031101.
- [210] J. C. Sancho-García, M. S. Brémond, A. J. Pérez-Jiménez, C. Adamo, *Phys. Chem. Chem. Phys.* **2017**, *19*, 13481.
- [211] É. Brémond, M. Savarese, J. C. Sancho-García, Á. J. Pérez-Jiménez, C. Adamo, *J. Chem. Phys.* **2016**, *144*, 124104.
- [212] S. Kozuch, D. Gruzman, J. M. Martin, *J. Phys. Chem. C* **2010**, *114*, 20801.
- [213] S. Kozuch, J. M. Martin, *Phys. Chem. Chem. Phys.* **2011**, *13*, 20104.

- [214] S. Kozuch, J. M. Martin, *J. Comput. Chem.* **2013**, *34*, 2327.
- [215] J. Friedrich, J. Hänchen, *J. Chem. Theory Comput.* **2013**, *9*, 5381.
- [216] J. Friedrich, *J. Chem. Theory Comput.* **2015**, *11*, 3596.
- [217] Y. Zhao, H. T. Ng, R. Peverati, D. G. Truhlar, *J. Chem. Theory Comput.* **2012**, *8*, 2824.
- [218] E. R. Johnson, P. Mori-Sánchez, A. J. Cohen, W. Yang, *J. Chem. Phys.* **2008**, *129*, 204112.
- [219] S. Grimme, H. Kruse, L. Goerigk, G. Erker, *Angew. Chemie - Int. Ed.* **2010**, *49*, 1402.
- [220] This test set is a modified version of HEAVYSB11<sup>[57]</sup>. In Addition to HEAVYSB9, HEAVYSB11 also contains the bonding energies of dibromide and dichloride which we removed here.
- [221] S. N. Steinmann, G. Csonka, C. Corminboeuf, *Society* **2009**, *5*, 2950.
- [222] H. Krieg, S. Grimme, *Mol. Phys.* **2010**, *108*, 2655.
- [223] F. Neese, T. Schwabe, S. Kossmann, B. Schirmer, S. Grimme, *J. Chem. Theory Comput.* **2009**, *5*, 3060.
- [224] A. Karton, S. Daon, J. M. Martin, *Chem. Phys. Lett.* **2011**, *510*, 165.
- [225] L. A. Curtiss, K. Raghavachari, G. W. Trucks, J. A. Pople, *J. Chem. Phys.* **1991**, *94*, 7221.
- [226] P. Morgante, R. Peverati, *J. Comput. Chem.* **2019**, *40*, 839.
- [227] S. Dohm, A. Hansen, M. Steinmetz, S. Grimme, M. P. Checinski, *J. Chem. Theory Comput.* **2018**, *14*, 2596.
- [228] A. Karton, R. J. O'Reilly, L. Radom, *J. Phys. Chem. A* **2012**, *116*, 4211.
- [229] T. C. Dinadayalane, R. Vijaya, A. Smitha, G. N. Sastry, *J. Phys. Chem. A* **2002**, *106*, 1627.
- [230] V. Guner, K. S. Khuong, A. G. Leach, P. S. Lee, M. D. Bartberger, K. N. Houk, *J. Phys. Chem. A* **2003**, *107*, 11445.
- [231] D. H. Ess, K. N. Houk, *J. Phys. Chem. A* **2005**, *109*, 9542.
- [232] L. Goerigk, R. Sharma, *Can. J. Chem.* **2016**, *94*, 1133.
- [233] L. J. Yu, F. Sarrami, R. J. O'Reilly, A. Karton, *Mol. Phys.* **2016**, *114*, 21.
- [234] L. J. Yu, F. Sarrami, R. J. O'Reilly, A. Karton, *Chem. Phys.* **2015**, *458*, 1.
- [235] R. Kang, W. Lai, J. Yao, S. Shaik, H. Chen, *J. Chem. Theory Comput.* **2012**, *8*, 3119.
- [236] Y. Sun, H. Chen, *J. Chem. Theory Comput.* **2013**, *9*, 4735.
- [237] Y. Sun, H. Chen, *J. Chem. Theory Comput.* **2014**, *10*, 579.
- [238] L. Hu, H. Chen, *J. Chem. Theory Comput.* **2015**, *11*, 4601.
- [239] Y. Zhao, B. J. Lynch, D. G. Truhlar, *Phys. Chem. Chem. Phys.* **2005**, *7*, 43.
- [240] Y. Zhao, N. González-Garda, D. G. Truhlar, *J. Phys. Chem. A* **2005**, *109*, 2012.
- [241] S. Grimme, M. Steinmetz, M. Korth, *J. Org. Chem.* **2007**, *72*, 2118.
- [242] D. Gruzman, A. Karton, J. M. Martin, *J. Phys. Chem. A* **2009**, *113*, 11974.
- [243] M. K. Kesharwani, A. Karton, J. M. Martin, *J. Chem. Theory Comput.* **2016**, *12*, 444.
- [244] D. Řeha, H. Valdés, J. Vondrášek, P. Hobza, A. Abu-Riziq, B. Crews, M. S. De Vries, *Chem. A Eur. J.* **2005**, *11*, 6803.
- [245] L. Goerigk, A. Karton, J. M. Martin, L. Radom, *Phys. Chem. Chem. Phys.* **2013**, *15*, 7028.
- [246] U. R. Fogueri, S. Kozuch, A. Karton, J. M. Martin, *J. Phys. Chem. A* **2013**, *117*, 2269.
- [247] G. I. Csonka, A. D. French, G. P. Johnson, C. A. Stortz, *J. Chem. Theory Comput.* **2009**, *5*, 679.
- [248] S. Kozuch, S. M. Bachrach, J. M. Martin, *J. Phys. Chem. A* **2014**, *118*, 293.
- [249] L. J. Yu, F. Sarrami, A. Karton, R. J. Oreilly, *Mol. Phys.* **2015**, *113*, 1284.
- [250] A. Karton, J. M. Martin, *Mol. Phys.* **2012**, *110*, 2477.
- [251] S. Grimme, *Angew. Chemie - Int. Ed.* **2006**, *45*, 4460.
- [252] D. Manna, J. M. Martin, *J. Phys. Chem. A* **2016**, *120*, 153.
- [253] J. Řezáč, K. E. Riley, P. Hobza, *J. Chem. Theory Comput.* **2011**, *7*, 2427.
- [254] J. Řezáč, K. E. Riley, P. Hobza, *J. Chem. Theory Comput.* **2012**, *8*, 4285.
- [255] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
- [256] K. U. Lao, R. Schäffer, G. Jansen, J. M. Herbert, *J. Chem. Theory Comput.* **2015**, *11*, 2473.
- [257] D. Setiawan, E. Kraka, D. Cremer, *J. Phys. Chem. A* **2015**, *119*, 1642.
- [258] S. N. Steinmann, C. Piemontesi, A. Delachat, C. Corminboeuf, *J. Chem. Theory Comput.* **2012**, *8*, 1629.
- [259] J. Řezáč, Y. Huang, P. Hobza, G. J. Beran, *J. Chem. Theory Comput.* **2015**, *11*, 3065.
- [260] R. Huenerbein, B. Schirmer, J. Moellmann, S. Grimme, *Phys. Chem. Chem. Phys.* **2010**, *12*, 6940.
- [261] R. Sure, A. Hansen, P. Schwerdtfeger, S. Grimme, *Phys. Chem. Chem. Phys.* **2017**, *19*, 14296.
- [262] R. Sedlak, T. Janowski, M. Pitoňák, J. Řezáč, P. Pulay, P. Hobza, *J. Chem. Theory Comput.* **2013**, *9*, 3364.
- [263] H. Kruse, A. Mladek, K. Gkionis, A. Hansen, S. Grimme, J. Sponer, *J. Chem. Theory Comput.* **2015**, *11*, 4972.
- [264] D. A. Wappett, L. Goerigk, *J. Phys. Chem. A* **2019**, *123*, 7057.
- [265] To avoid the evaluation of the performance of DFAs to depend too much on relative conformational energies of organic molecules, we artificially reduced the size of these subsets by a factor of 2. Instead of randomly selecting a smaller number of molecules out of these sets, we preferred to calculate the full sets and only include them with a weight of 0.5 in our overall analysis. MCONF: 26 (52), BUT14DIOL: 32 (64).
- [266] A. K. Narsaria, T. A. Hamlin, K. Lammertsma, F. M. Bickelhaupt, *Chem. - A Eur. J.* **2019**, *25*, 9902.
- [267] T. A. Hamlin, B. J. Levandowski, A. K. Narsaria, K. N. Houk, F. M. Bickelhaupt, *Chem. A Eur. J.* **2019**, *25*, 6342.
- [268] P. Mori-Sánchez, A. J. Cohen, W. Yang, *J. Chem. Phys.* **2006**, *125*, 201102.
- [269] P. Jurečka, J. Šponer, J. Černý, P. Hobza, *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.
- [270] S. Grimme, *Chem. - A Eur. J.* **2012**, *18*, 9955.
- [271] B. Nguyen, G. P. Chen, M. M. Agee, A. M. Burrow, M. Tang, and F. Furche (2019).
- [272] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, B. I. Lundqvist, *Phys. Rev. Lett.* **2004**, *92*, 22.
- [273] A. D. Becke, E. R. Johnson, *J. Chem. Phys.* **2005**, *123*, 154101.
- [274] A. D. Becke, E. R. Johnson, *J. Chem. Phys.* **2006**, *124*, 600.
- [275] A. Tkatchenko, M. Scheffler, *Phys. Rev. Lett.* **2009**, *102*, 6.
- [276] O. A. Vydrov, T. Van Voorhis, *J. Chem. Phys.* **2010**, *133*, 244103.
- [277] H. Kruse, S. Grimme, *J. Chem. Phys.* **2012**, *136*, 154101.
- [278] A. D. Becke, *J. Chem. Phys.* **1996**, *104*, 1040.
- [279] M. A. Marques, M. J. Oliveira, T. Burnus, *Comput. Phys. Commun.* **2012**, *183*, 2272.
- [280] N. Mardirossian, M. Head-Gordon, *J. Chem. Phys.* **2015**, *142*, 074111.
- [281] N. Mardirossian, M. Head-Gordon, *Phys. Chem. Chem. Phys.* **2014**, *16*, 9904.
- [282] J. D. Chai, M. Head-Gordon, *J. Chem. Phys.* **2009**, *131*, 174105.
- [283] N. Mardirossian, J. D. McClain, G. K. L. Chan, *J. Chem. Phys.* **2018**, *148*, 044106.
- [284] SD-SCAN69 is the only exception as it has been designed explicitly without dispersion correction.
- [285] R. M. Parrish, L. A. Burns, D. G. Smith, A. C. Simmonett, A. E. DePrince, E. G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. Di Remigio, R. M. Richard, et al., *J. Chem. Theory Comput.* **2017**, *13*, 3185.
- [286] S. Grimme, <https://www.chemie.uni-bonn.de/pctc/mulliken-center/software/GMTKN/gmtkn55>.
- [287] S. F. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553.
- [288] B. Helmich-Paris, L. Visscher, *J. Comput. Phys.* **2016**, *321*, 927.
- [289] E. Van Lenthe, J. E. Baerends, *J. Comput. Chem.* **2003**, *24*, 1142.
- [290] M. Franchini, P. H. T. Philipsen, E. Van Lenthe, L. Visscher, *J. Chem. Theory Comput.* **2014**, *10*, 1994.



- [291] M. Franchini, P. H. T. Philipsen, L. Visscher, *J. Comput. Chem.* **2013**, *34*, 1819.
- [292] E. Van Lenthe, E. J. Baerends, J. G. Snijders, *J. Chem. Phys.* **1993**, *99*, 4597.
- [293] E. Van Lenthe, E. J. Baerends, J. G. Snijders, *J. Chem. Phys.* **1994**, *101*, 9783.
- [294] E. Van Lenthe, J. G. Snijders, E. J. Baerends, *J. Chem. Phys.* **1996**, *105*, 6505.
- [295] E. Van Lenthe, A. Ehlers, J. E. Baerends, *J. Chem. Phys.* **1999**, *110*, 8943.
- [296] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
- [297] B. P. Pritchard, D. Altarawy, B. Didier, T. D. Gibson, T. L. Windus, *J. Chem. Inf. Model.* **2019**, *59*, 4814.
- [298] F. Weigend, *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057.
- [299] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.
- [300] D. Cremer, *Mol. Phys.* **2001**, *99*, 1899.
- [301] J. P. Perdew, M. Ernzerhof, K. Burke, *J. Chem. Phys.* **1996**, *105*, 9982.
- [302] B. G. Janesko, G. E. Scuseria, *J. Chem. Phys.* **2008**, *128*, 194105.
- [303] B. J. Lynch, P. L. Fast, M. Harris, D. G. Truhlar, *J. Phys. Chem. A* **2000**, *104*, 4813.
- [304] S. Patchkovskii, T. Ziegler, *J. Chem. Phys.* **2002**, *116*, 7806.
- [305] D. Hait, M. Head-Gordon, *J. Phys. Chem. Lett.* **2018**, *9*, 6280.
- [306] J. M. Martin, P. R. Taylor, *J. Phys. Chem.* **1996**, *100*, 6047.
- [307] A. Hirsch, Z. Chen, H. Jiao, *Angew. Chem., Int. Ed.* **2000**, *39*, 3915.
- [308] E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen, W. Yang, *J. Am. Chem. Soc.* **2010**, *132*, 6498.
- [309] S. Kristyán, P. Pulay, *Chem. Phys. Lett.* **1994**, *229*, 175.
- [310] L. Goerigk, H. Kruse, S. Grimme, *ChemPhysChem* **2011**, *12*, 3421.
- [311] C. J. Mackie, J. F. Gonthier, M. Head-Gordon, *J. Chem. Phys.* **2020**, *152*, 024112.
- [312] E. Engel, A. Facco Bonetti, S. Keller, I. Andrejkovics, R. M. Dreizler, *Phys. Rev. A* **1998**, *58*, 964.
- [313] J. G. Angyan, I. C. Gerber, A. Savin, J. Toulouse, *Phys. Rev. A* **2005**, *72*, 1.
- [314] E. G. Hohenstein, S. T. Chill, C. D. Sherrill, *J. Chem. Theory Comput* **2008**, *4*, 1996.
- [315] A. Najibi, L. Goerigk, *J. Phys. Chem. A* **2018**, *122*, 5610.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Förster A, Visscher L. Double hybrid DFT calculations with Slater type orbitals. *J Comput Chem.* 2020;41:1660–1684. <https://doi.org/10.1002/jcc.26209>