SUPPORTING INFORMATION

General Optimization Procedure Towards the Design of a New Family of Minimal Parameter Spin-Component-Scaled Double-Hybrid Density Functional Theory

Loïc M. Roch and Kim K. Baldridge

Table of contents

I. General accuracy and basis set consideration.

Prior to this benchmarking study, an exhaustive basis set convergence study was carried out on a *reduced data set*, which is displayed in **Figure S1** and also referred to as the *training set* in the core manuscript (**Figure 1**).
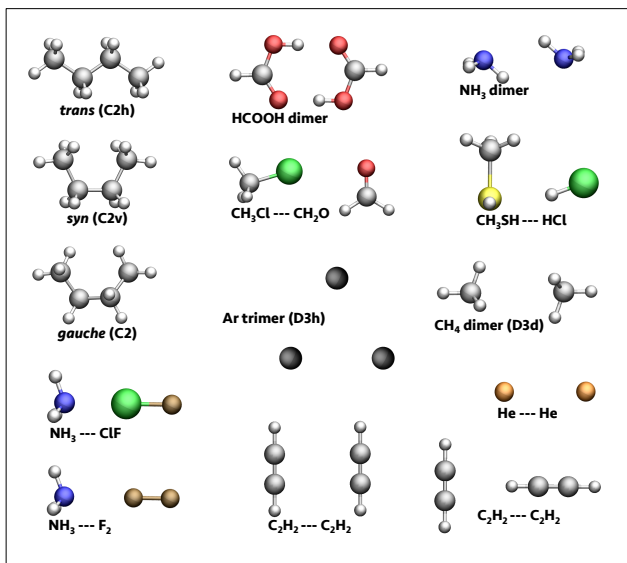

Figure S1. The reduced data set contains 13 complexes.

Three families of standard basis sets were investigated: (i) the augmented correlation-consistent aug-cc-pVnZ (n=D,T,Q,5,6) developed by T. H. Dunning et al., (ii) the augmented correlation-consistent with tight diffuse function aug-cc-pCVnZ (n=D,T,Q) developed by T. H. Dunning et al., and (iii) the Def2-style basis set developed by R. Ahlrichs et al.. The interaction energies of the *reduced data set* were computed at the PBE, DSD-PBEPBE and MP2 level of theory establishing convergence of the interaction energy with respect to increas- ing the basis set size. The choice of the computational methodologies shows the convergence of the density-based methods, the correlated models, and the mixed DSD-DFT. The N-electron methods combined with the T. H. Dunning n-tuple augmented correlation consistent aug-cc-pVnZ basis

set (n=D,T,Q,5,6) allow extrapolation to the CBS limit and gives information on the intrinsic errors of the different N-electron models, enabling the investigation of the accuracy of standard basis sets.

The aug-cc-pVnZ set is particularly well suited for an extrapolation to the CBS limit, since with each extension of the basis set, new functions are added that make similar contribution to the energy. Thus, given the hierarchical sequences of the Dunning aug-cc-pVnZ basis sets, a systematic improvement of the total energy is obtained when climbing the steps towards the completeness of the set.

Based on numerical evidences, and following earlier work by D. Feller, the exchange energy was extrapolated through a three-point exponential formula. The correlation energy, however, was extrapolated via a three-point inverse fitting functions, as suggested by C. Schwartz. At the CBS limit, the interaction energy E^{int} reads:

$$E^{int} = E^{A-B} - E^{A} - E^{B}$$

where E^{A-B}, E^{A}, and E^{B} are respectively the total energy of the complex A⋯B, the monomer A, and the monomer B at the CBS limit. The assessment of the basis set performance on the interaction energies of the strongest and weakest interacting complexes was the carried out by means of statistical analysis, through the normal distribution around the CBS limit. It has been observed that for high accuracy interaction energies, basis sets containing at least g functions for the 1-Row and 2-Row are required. Among the six quadruple-zeta basis sets included herein, the absolute energies of the HF and DFT components are best evaluated with the Def2-QZVPD basis, while the aug-cc-pCQZ provides the best results for the correlation component. In either

case, the ΔE (define as the difference between the CBS energy and the energy obtained a given basis set level) challenges the aug-cc-pV5Z results. However, since the main focus is on interaction energies, the normal distribution around the E^{int} is heavier in the CBS assessment process. With an MAE between 0.011 (PBE level) and 0.165 kcal/mol (MP2 level), Def2-QZVPD outperforms aug-cc-pVQZ by a few cal/mol. A last element to consider in the assessment is the computational effort. **Figure S2** displays the computational cost in function of the interaction energies MAE, and this sheds light on the performance of the Def2-QZVPD basis set to provide the best ratio between the maximum amount of correlation energy recovered and low computational costs. As a consequence, the Def2-QZVPD basis set will be used to assess the performance of a wide range of N-electron models on weakly interacting complexes.
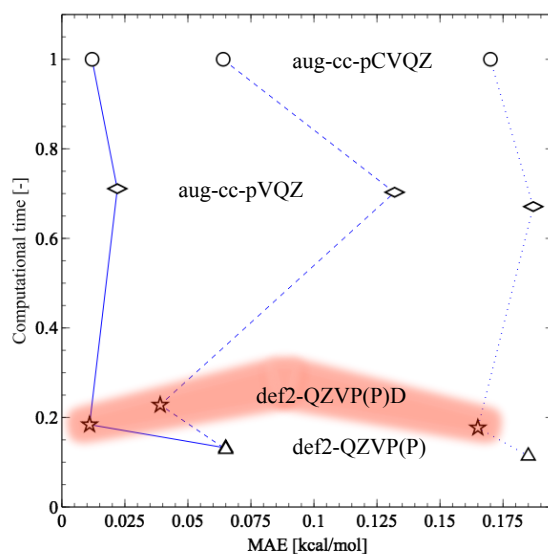

Figure S2. Computational time of the quadruple-zeta basis: aug-cc-pCVQZ (o), aug-cc-pVQZ (<>), Def2- QZVP and Def2-QZVPP (★), Def2-QZVPD and Def2-QZVPPD (Δ) in function of the MAE. To guide the eye, the solid line shows the PBE results, the dashed line shows the DSD-PBEPBE results, and the doted line shows the MP2 results.

This exhaustive basis set convergence study established the performance of Def2-QZVPD as basis set. Additionally, by moving to even larger basis set, such as aug-cc-pV6Z, no drastic improvement on the MAE is noticed, as shown in the table below, where the MAE (in kcal/mol) of PBE, DSD-PBEPBE, and MP2 with two basis sets is listed:

Method	Def2-QZVPD	aug-cc-pV6Z
PBE	0.011	0.007
DSD-PBEPBE	0.039	0.025
MP2	0.165	0.115

As evidenced by the rather small change in MAE from Def2-QZVPD to aug-cc-pV6Z, CP correction is not needed at the basis set level used in this benchmarking paper.

II. Interaction energies of the complexes in the training set.

Table S1. Experimental and/or theoretical interaction energies for the complexes displayed in **Figure 1** and in **Figure S1**. Values are in kcal/mol.

System	E_{int}	System	E_{int}
HCOOH --- HCOOH	-18.61[1]	Ar trimer	-0.850[2]
CH_3Cl --- CH_2O	-1.170[3]	He --- He	-0.022[4,5]
C_2H_2 --- C_2H_2	-1.537[6]	CH_4 --- CH_4	-0.530[7]
C_2H_2 --- C_2H_2	+1.074[6]	CH_3SH --- HCl	-4.160[8,9]
NH_3 --- ClF	-10.62[8,9]	Anti -> gauche	+0.67[10]
NH_3 --- F_2	-1.81[8,9]	Anti -> syn	+3.95[10]
NH_3 --- NH_3	-3.137[6]		

III. Grid refinement and scaled RMSD at different basis set level.

First optimization round. For general feeling of the potential energy surface at the SCS-PBEPBE level, $C_{HF} = [0.1, 0.2, ... , 0.9]$ and $C_{GGA} = [0.1, 0.2, ... , 0.9]$ regardless of the basis set used. **Figure S3** displays the corresponding $RMSD_{SC}$.
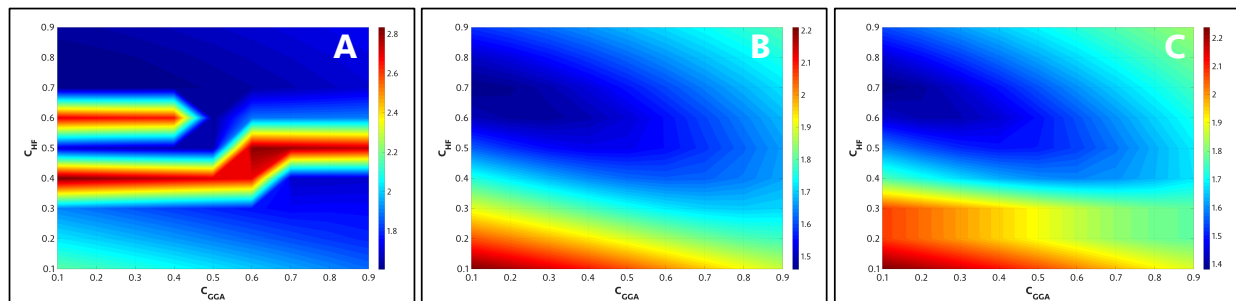

Figure S3. Results of the $RMSD_{SC}$ for the first round of the SCS-PBEPBE functional optimization. The $RMSD_{SC}$ are displayed at three different basis set level: (A) cc-pVDZ, (B) cc-pVTZ, and (C) cc-pVQZ.

Second optimization round. Based on the results obtained in the first round, the coefficients are refined with respect to the basis set used. **Figure S4** displays the corresponding $RMSD_{SC}$.

- the coefficients with the cc-pVDZ basis set are refined to:

$C_{HF} = [0.62, 0.64, 0.66, 0.68, 0.72, 0.74, 0.76, 0.78]$

$C_{GGA} = [0.12, 0.14, 0.16, 0.18, 0.22, 0.24, 0.26, 0.28]$

- the coefficients with the cc-pVTZ basis set are refined to:

$C_{HF} = [0.62, 0.64, 0.66, 0.68, 0.72, 0.74, 0.76, 0.78]$

$C_{GGA} = [0.02, 0.04, 0.06, 0.08, 0.12, 0.14, 0.16, 0.18]$

- the coefficients with the cc-pVQZ basis set are refined to

$C_{HF} = [0.62, 0.64, 0.66, 0.68, 0.72, 0.74, 0.76, 0.78]$

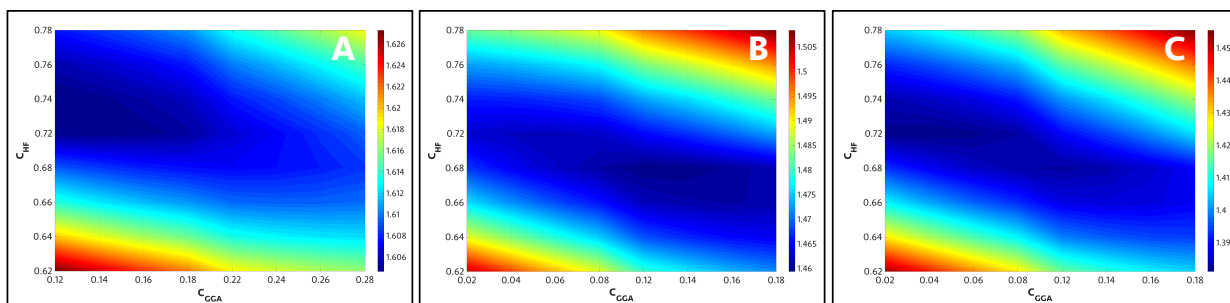$C_{GGA} = [0.02, 0.04, 0.06, 0.08, 0.12, 0.14, 0.16, 0.18]$

Figure S4. Results of the $RMSD_{sc.}$ for the second round of the SCS-PBEPBE functional optimization. The $RMSD_{SC}$ are displayed at three different basis set level: (A) cc-pVDZ, (B) cc-pVTZ, and (C) cc-pVQZ.

Third optimization round. Based on the results obtained in the first round, the coefficients are refined with respect to the basis set used. **Figure S5** displays the corresponding $RMSD_{SC}$.

- the coefficients with the cc-pVDZ basis set are refined to:

$C_{HF} = [0.72, 0.73, 0.74, 0.75, 0.76]$

$C_{GGA} = [0.07, 0.08, 0.09, 0.10, 0.11, 0.12, 0.13, 0.14]$

- the coefficients with the cc-pVTZ basis set are refined to:

$C_{HF} = [0.66, 0.67, 0.68, 0.69, 0.70, 0.71, 0.72]$

$C_{GGA} = [0.08, 0.09, 0.10, 0.11, 0.12, 0.13, 0.14]$

- the coefficients with the cc-pVQZ basis set are refined to:

$C_{HF} = [0.68, 0.69, 0.70, 0.71, 0.72, 0.73, 0.74]$

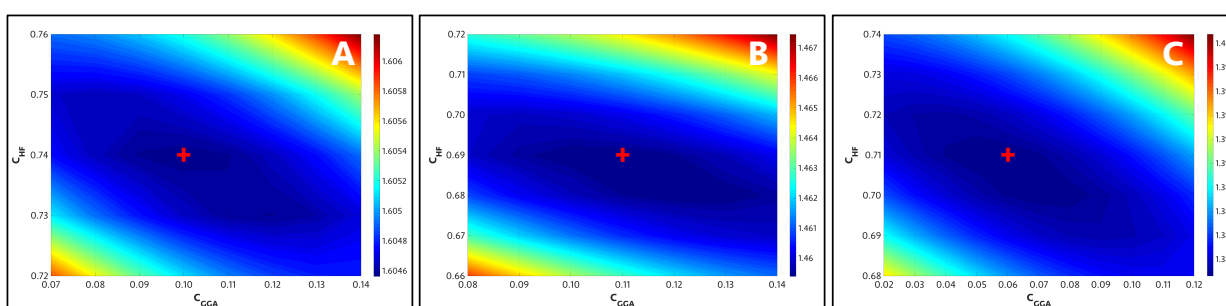$C_{GGA} = [0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10, 0.11, 0.12]$

Figure S5. Results of the $RMSD_{SC}$ for the third/last round of the SCS-PBEPBE functional optimization. The $RMSD_{SC}$ are displayed at three different basis set level: (A) cc-pVDZ, (B) cc-pVTZ, and (C) cc-pVQZ. The best C_{GGA} and C_{HF} combination is highlighted by the red cross.

REFERENCES

1. P. Jurečka, J. Šponer, J. Černý and P. Hobza, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1985-1993.
2. R. Peverati, M. Macrina and K. K. Balridge, *J. Chem. Theory Comput.*, 2010, **6**, 1951-1965.
3. J. Řezáč, K. E. Riley and P. Hobza, *J. Chem. Theory Comput.*, 2012, **8**, 4285-4292.
4. J. F. Ogilvie and F. Y. H. Wang, *J. Mol. Struct.*, 1992, **273**, 277-277.
5. J. F. Ogilvie and F. Y. H. Wang, *J. Mol. Struct.*, 1993, **291**, 313-313.
6. J. Řezáč and P. Hobza, *J. Chem. Theory Comput.*, 2013, **9**, 2151-2155.
7. S. Tsuzuki, K. Honda, T. Uchimaru and M. Mikami, *J. Chem. Phys.*, 2006, **124**, 114304-114304.
8. Y. Zhao and D. G. Truhlar, *J. Chem. Theory Comput.*, 2005, **1**, 415-432.
9. Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A*, 2005, **109**, 5656-5667.
10. N. L. Allinger, J. T. Fermann, W. D. Allen and I. I. I. H. F. Schaefer, *J. Chem. Phys.*, 1997, **12**, 5143-5150.