Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

The *p*-block challenge: assessing quantum chemistry methods for inorganic heterocycle dimerizations

Thomas Gasevic,[‡c] Markus Bursch,[*‡a] Qianli Ma,[b] Stefan Grimme,[c] Hans-Joachim Werner,[*‡b] and Andreas Hansen[*‡c]

April 19, 2024

[a] *Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany.*
E-mail: bursch@kofo.mpg.de

[b] *Institut für Theoretische Chemie, Universität Stuttgart, Pfaffenwaldring 55, D-70569 Stuttgart, Germany.*
E-mail: werner@theochem.uni-stuttgart.de

[c] *Mulliken Center for Theoretical Chemistry, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstr. 4, 53115 Bonn, Germany.*
E-mail: hansen@thch.uni-bonn.de

[‡] These authors contributed equally.

Contents

S1 Definition of Statistical Measures

Error: The error e for a given interaction i is given by the difference of the actual method and the reference.

$$e_i = \text{method}_i - \text{reference}_i \tag{1}$$

Mean Deviation (MD): Mean of the signed error, n corresponds to the number of data points of the respective set.

$$\text{MD} = \frac{1}{n}\sum_{i=1}^{n} e_i \equiv \bar{e} \tag{2}$$

Mean Absolute Deviation (MAD): The term mean absolute deviation (MAD) is used in this work interchangeably with the **mean absolute** (unsigned) **error** (**MAE**).

$$\text{MAD} \overset{\text{this work}}{\equiv} \text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |e_i| \tag{3}$$

Root Mean Square Deviation (RMSD): The term root mean square deviation (RMSD) in this work is used interchangeably with **root mean squared error** (**RMSE**).

$$\text{RMSD} \overset{\text{this work}}{\equiv} \text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(e_i\right)^2} \tag{4}$$

Standard Deviation (SD): The standard deviation is defined using Bessel's correction.

$$\text{SD} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} \left(e_i - \bar{e}\right)^2} \tag{5}$$

Largest And Smallest Errors (MaxDev, MinDev): Largest and smallest (signed) error of all n interactions i of a given set.

$$\text{MaxDev} = \max\left(\{e_i\}\right) \tag{6}$$

$$\text{MinDev} = \min\left(\{e_i\}\right) \tag{7}$$

$$\forall i \in \{1, 2, \ldots, n\}$$

S2 Computational Details for DFT Calculations

All DFT calculations discussed in the manuscript were performed with the ORCA 5.0.4 program package.[1,2] Ahlrichs quadruple-ζ basis set def2-QZVPP[3] with matching effective core potentials (ECPs)[4,5] for heavy elements with Z > 36 was generally employed. The *RIJCOSX*[6] approximation was used to accelerate the calculations. If not stated otherwise, the default *def2/J* auxiliary basis set option was applied as implemented in ORCA. The *DefGrid3* option was applied for the numerical integration grid as well as *TightSCF* convergence criteria as implemented in ORCA.

DFT-D4 dispersion corrections were calculated with the dftd4 3.4.0 standalone program,[7] DFT-D3 dispersion corrections with the s-dftd3 1.0.0 standalone program,[8] and for DFT-NL, the ORCA native, post-SCF implementation was used.

During the revision process of this work, we tested different basis sets and relativistic treatments. Covalent dimerization energies were computed with different density functionals in combination with the awcvtz basis set. The respective data can be found in the accompanying Excel sheet (data_dft_sqm.xlsx). Different program packages were applied for this study on relativistic effects:

- Calculations in ORCA 5.0.4 for the awcvtz, awcvqz, and avqz basis sets were conducted with the RIJK and the matching auxiliary basis set (*def2/JK*). The *DefGrid3* option was applied for the numerical integration grid. These settings were also applied in the computations of the spin contamination and the fractional occupation number weighted electron density (FOD).

- Calculations in Q-Chem 5.4[9] were conducted with the DIIS SCF algorithm and a SCF convergence threshold of 10^{-8}. A cutoff threshold of 10^{-14} was chosen for the neglect of two-electron integrals.

- Calculations in Molpro 2024.1[10,11] were conducted with the JKFIT density fitting and an energy threshold of 10^{-7}.

- Calculations in AMS 2023.104[12,13] were conducted without any symmetry and the *NumericalQuality* set to good.

- Calculations in Turbomole 7.6[14] were conducted with the RI-J approximation and the matching auxiliary basis set (*def2/J*). A grid size of *m4* was generally applied.

If not stated otherwise, default settings were applied.

S2.1 ORCA Input for B2NC-PLYP

! DEFGRID3 def2/J RIJCOSX def2-QZVPP def2-QZVPP/C

%method
Exchange X_B88
Correlation C_LYP
LDAOpt C_VWN3
ScalHFX=0.81
ScalDFX=0.19
ScalMP2C=0.55
ScalLDAC=0.45
ScalGGAC=0.45
end

* xyzfile 0 1 mol.xyz

S3 Further Details on the Reference Calculations

In the following test calculations, the CABS singles correction is included but it was computed without including core orbitals (`CORE_SINGLES=0`). As we found only after completing these calculations, this can slightly change the absolute values of the results, but the relevant differences for different F12 exponents, different RI basis sets, or different local approximations are not affected. For light elements with no pseudopotentials, cc-pVTZ-PP-F12 refers to the cc-pVTZ-F12 basis sets. The test systems are $\overline{\mathrm{Al-O-Al-S-Al-Se}}$ (**2**), $\overline{\mathrm{Ga-Te-In-Te-Ga-Se}}$ (**4**), $\overline{\mathrm{In-Te-Tl-Te-In-Se}}$ (**5**), $\overline{\mathrm{Tl-P-Tl-P-Tl-Bi}}$ (**6**), and $\overline{\mathrm{Pb-Sb-Si-Bi-Ge-N}}$ (**7**).

Table S1 Dependence of dimerization energies on the RI basis[a].

RI-basis	**2**		**4**		**5**		**6**		**7**	
	WDA	COV	WDA	COV	WDA	COV	WDA	COV	WDA	COV
cc-pVTZ-PP-F12:										
cc-VTZ-PP-F12/CABS	-13.40	-117.96	-21.24	-82.88	-27.60	-85.12	-23.88	-145.55	-18.22	-167.44
def2-QZVPP/JKFIT[b]	-13.40	-117.96	-21.32	-82.93	-27.58	-85.05	-23.88	-145.63	-18.24	-167.54
cc-VTZ-PP-F12/MP2FIT[c]	-13.40	-117.97	-21.25	-82.88	-27.64	-85.16	-23.93	-145.61	-18.24	-167.49
cc-VQZ-PP-F12/MP2FIT[d]	-13.39	-117.96	-21.22	-82.85	-27.58	-85.10	-23.85	-145.59	-18.20	-167.45
cc-pVQZ-PP-F12:										
cc-VQZ-PP-F12/CABS	-13.36	-117.94	-21.26	-83.07	-27.91	-86.50	-24.32	-145.65	-18.37	-168.01
def2-QZVPP/JKFIT[b]	-13.36	-117.95	-21.33	-83.15	-27.94	-86.55	-24.33	-146.69	-18.39	-168.06
cc-VQZ-PP-F12/MP2FIT[d]	-13.36	-117.95	-21.26	-83.05	-27.91	-86.48	-24.31	-145.65	-18.37	-168.01

a) Using three optimized geminal exponents from Table S3; `CORE_SINGLES=0`.
b) aug-cc-pVTZ/JKFIT for light atoms.
c) aug-cc-pVTZ/MP2FIT for light atoms.
d) aug-cc-pVQZ/MP2FIT for light atoms.

Table S2 PNO-LMP2 and PNO-LMP2-F12/cc-pVTZ-PP-F12 WDA interaction energies with and without counterpoise (CP) correction (in kcal mol^{-1}), γ(vv,cv,cc)=[0.7, 1.3, 1.3], `CORE_SINGLES=0`.

	LMP2	LMP2(CP)	LMP2-F12	LMP2-F12(CP)
$\overline{\mathrm{Al-O-Al-S-Al-Se}}_{\mathrm{WDA}}$ (**2**$_{\mathbf{WDA}}$)	-12.94	-12.52	-13.40	-13.31
$\overline{\mathrm{In-Te-Tl-Te-In-Se}}_{\mathrm{WDA}}$ (**5**$_{\mathbf{WDA}}$)	-29.23	-26.63	-27.62	-27.62
$\overline{\mathrm{Ga-Te-In-Te-Ga-Se}}_{\mathrm{WDA}}$ (**4**$_{\mathbf{WDA}}$)	-21.35	-19.76	-21.16	-21.13
$\overline{\mathrm{Tl-P-Tl-P-Tl-Bi}}_{\mathrm{WDA}}$ (**6**$_{\mathbf{WDA}}$)	-24.79	-23.00	-24.02	-24.12
$\overline{\mathrm{Pb-Sb-Si-Bi-Ge-N}}_{\mathrm{WDA}}$ (**7**$_{\mathbf{WDA}}$)	-17.76	-16.82	-18.19	-18.19
$\overline{\mathrm{Si-N-Si-N-Si-P}}_{\mathrm{WDA}}$ (**8**$_{\mathbf{WDA}}$)	-8.20	-7.88	-8.41	-8.31

Table S3 Optimized geminal exponents γ (in a_0^{-1}) for eight monomer structures using the cc-pVTZ-PP-F12 basis set. The (n-1)d shells are taken as core.

	vv	cv	cc
cc-pVTZ-PP-F12:			
B – Se – B – Se – B – Se (**1**)	0.8	1.7	2.0
Al – O – Al – S – Al – Se (**2**)	0.8	1.7	2.0
Ga – O – In – O – Tl – O (**3**)	1.0	1.3	1.4
Ga – Te – In – Te – Ga – Se (**4**)	0.7	1.3	1.7
In – Te – Tl – Te – In – Se (**5**)	0.6	1.2	1.4
Tl – P – Tl – P – Tl – Bi (**6**)	0.7	0.9	1.0
Pb – Sb – Si – Bi – Ge – N (**7**)	0.7	1.1	1.4
Si – N – Si – N – Si – P (**8**)	0.8	–	–
cc-pVQZ-PP-F12:			
B – Se – B – Se – B – Se (**1**)	1.0	1.8	2.2
Al – O – Al – S – Al – Se (**2**)	1.0	1.8	2.2
Ga – O – In – O – Tl – O (**3**)	1.3	1.4	1.5
Ga – Te – In – Te – Ga – Se (**4**)	0.8	1.4	1.6
In – Te – Tl – Te – In – Se (**5**)	0.8	1.3	1.4
Tl – P – Tl – P – Tl – Bi (**6**)	0.8	1.0	1.1
Pb – Sb – Si – Bi – Ge – N (**7**)	0.9	1.2	1.4
Si – N – Si – N – Si – P (**8**)	1.0	–	–

Table S4 PNO-LCCSD(T)-F12/vtz-f12 dimerization energies (in kcal mol^{-1}) for various choices of F12 exponents (in a_0^{-1}). γ=(opt3) are the optimized exponents from Table S3. If only 2 exponents are given, the first is γ_{vv} and the second $\gamma_{cc} = \gamma_{cv}$. The (n-1)$d$ shell is treated as core. `CORE_SINGLES=0`.

System	γ=(0.7,1.3)	γ=(0.8,1.3)	γ=(0.9,1.3)	γ=(0.9,1.4)	γ=(1.0,1.4)	γ=(opt3)
weak donor-acceptor dimers:						
B–Se–B–Se–B–Se$_{WDA}$ (**1$_{WDA}$**)	−5.5	−5.5	−5.5	−5.6	−5.5	−5.6
Al–O–Al–S–Al–Se$_{WDA}$ (**2$_{WDA}$**)	−12.0	−12.0	−12.0	−12.0	−12.0	−12.0
Ga–O–In–O–Tl–O$_{WDA}$ (**3$_{WDA}$**)	−11.2	−11.2	−11.2	−11.2	−11.2	−11.2
Ga–Te–In–Te–Ga–Se$_{WDA}$ (**4$_{WDA}$**)	−16.1	−16.1	−16.1	−16.1	−16.1	−16.2
In–Te–Tl–Te–In–Se$_{WDA}$ (**5$_{WDA}$**)	−21.2	−21.2	−21.2	−21.3	−21.3	−21.1
Tl–P–Tl–P–Tl–Bi$_{WDA}$ (**6$_{WDA}$**)	−17.0	−17.0	−17.0	−17.1	−17.1	−16.8
Pb–Sb–Si–Bi–Ge–N$_{WDA}$ (**7$_{WDA}$**)	−11.8	−11.8	−11.8	−11.8	−11.8	−11.8
covalent dimers:						
B–Se–B–Se–B–Se$_{COV}$ (**1$_{COV}$**)	−2.3	−2.4	−2.4	−2.4	−2.4	−2.4
Al–O–Al–S–Al–Se$_{COV}$ (**2$_{COV}$**)	−112.6	−113.0	−113.2	−113.2	−113.2	−113.0
Ga–O–In–O–Tl–O$_{COV}$ (**3$_{COV}$**)	−115.4	−115.5	−115.6	−115.7	−115.5	−115.7
Ga–Te–In–Te–Ga–Se$_{COV}$ (**4$_{COV}$**)	−68.7	−68.8	−68.9	−69.1	−68.9	−69.2
In–Te–Tl–Te–In–Se$_{COV}$ (**5$_{COV}$**)	−70.6	−70.7	−70.7	−71.0	−70.7	−70.4
Tl–P–Tl–P–Tl–Bi$_{COV}$ (**6$_{COV}$**)	−132.6	−132.8	−132.9	−133.1	−132.8	−132.1
Pb–Sb–Si–Bi–Ge–N$_{COV}$ (**7$_{COV}$**)	−155.3	−155.4	−155.5	−155.6	−155.3	−155.3

Table S5 demonstrates the dependence of the dimerization energies on the local approximations. Parameter `REXT` determines the size of the PAO domains; `THRPNO_CC` and `THREN_CC` determine the PNO domain sizes in PNO-LCCSD(T)-F12. Note: the corresponding LMP2 parameters `THRPNO_LMP2` and `THREN_LMP2` only affect the domain corrections. `PAIROPT=TIGHT` tightens the pair approximations (close, weak, distant pairs) and also extends the triples list. For more details see Refs. [15,16].

Table S5 Dependence of PNO-LCCSD(T)-F12/vtz-f12 dimerization energies on options for local approximations;[a] $\gamma = (1.0, 1.4)$.

RI-basis	**2**		**4**		**5**		**6**		**7**	
	WDA	COV	WDA	COV	WDA	COV	WDA	COV	WDA	COV
`DOMOPT=DEFAULT`	-11.89	-113.13	-15.92	-69.00	-21.14	-70.84	-16.92	-133.32	-11.56	-155.49
`DOMOPT=TIGHT`	-11.97	-113.30	-16.13	-69.13	-21.34	-71.07	-17.09	-133.18	-11.82	-155.71
`DOMOPT=TIGHT,REXT=10`	-12.05	-113.30	-16.12	-69.13	-21.25	-71.07	-16.96	-133.18	-11.70	-155.71
`DOMOPT=TIGHT,THREN_CC=0.997`	-12.07	-113.36	-16.19	-69.19	-21.40	-71.12	-17.16	-133.24	-11.89	-155.76
`DOMOPT=TIGHT,PAIROPT=TIGHT`	-11.96	-113.34	-16.18	-69.23	-21.41	-71.17	-17.17	-133.02	-11.96	-156.04

a) `PAIROPT=DEFAULT` unless noted otherwise;
`DOMOPT=DEFAULT` implies: `REXT=5`, `THRPNO_LMP2=1d-8`, `THREN_LMP2=0.997` `THRPNO_CC=1d-7`, `THREN_CC=0.990`;
`DOMOPT=TIGHT` implies: `REXT=7`, `THRPNO_LMP2=1d-8`, `THREN_LMP2=0.997` `THRPNO_CC=1d-8`, `THREN_CC=0.990`.

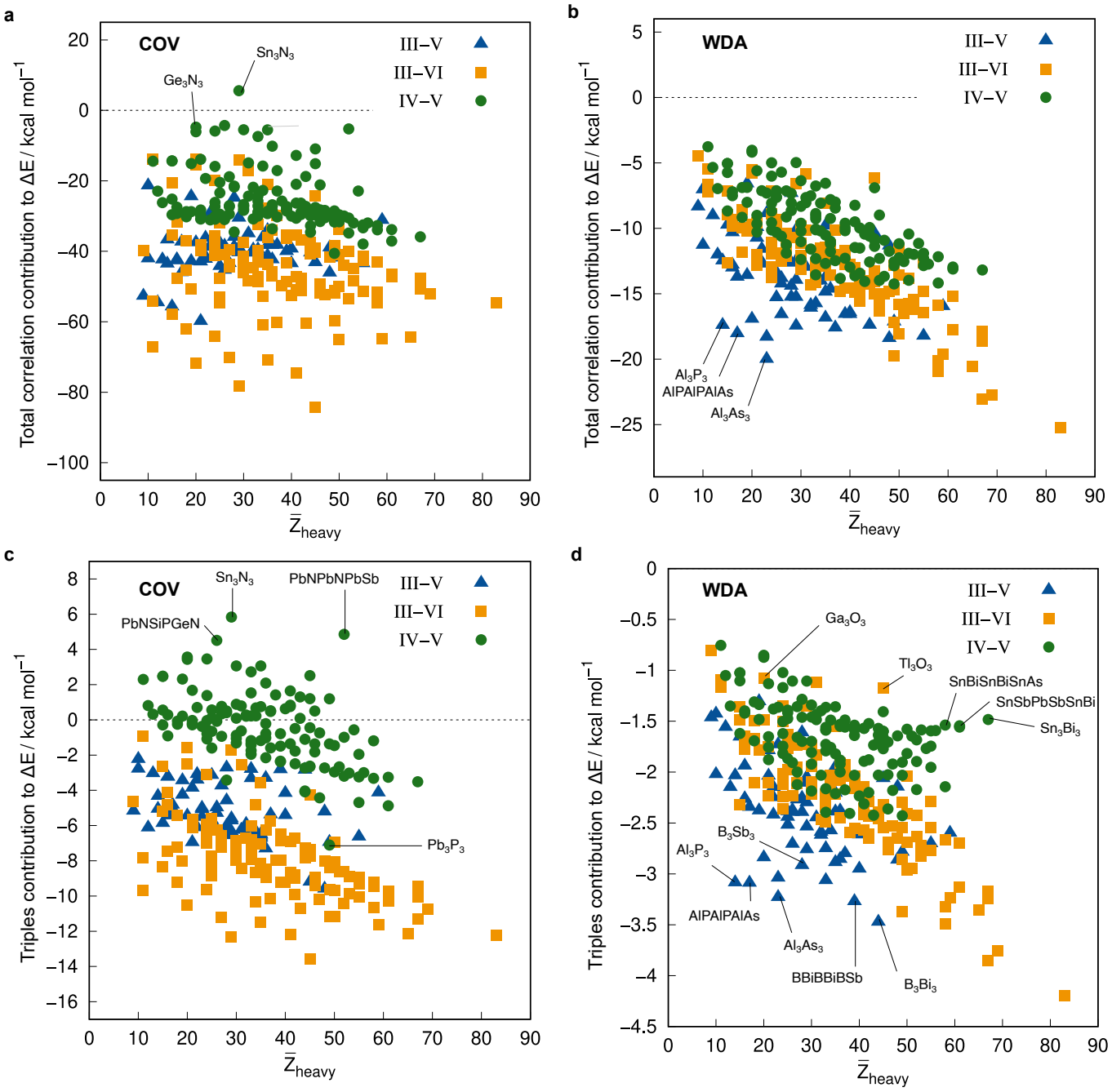


Figure S1 Correlation between the mean atomic number of all atoms except H in the molecule ($\overline{Z}_{heavy}$) and the total correlation contribution (a,b) (PNO-LCCSD(T)-F12/vtz-f12(corr.) level of theory), as well as the correlation between $\overline{Z}_{heavy}$ and the total triples contribution (c,d) (PNO-LCCSD(T)-F12(PNO)/vtz-f12 [THRPNO_MP2=1d-8] level of theory) to the dimerization energies for the covalent (COV) and the weak donor-acceptor (WDA) subsets.

In addition to the analysis presented in section 3.6 of the paper, we want to analyze the total correlation and triples contributions to the respective dimerization energies. For ΔE_{corr} and $\Delta E_{triples}$, the III-V systems give surprisingly large contributions for these relatively small molecules (up to -84.3 kcal·mol^{-1} and -13.6 kcal·mol^{-1} for $\overline{\mathrm{B-Po-B-Po-B-Po}}_{\mathrm{COV}}$, respectively), while the IV-V systems even more often give positive $\Delta E_{triples}$ contributions (up to $+5.8$ kcal·mol^{-1} for $\overline{\mathrm{Sn-N-Sn-N-Sn-N}}_{\mathrm{COV}}$). This system is also the only one with a positive ΔE_{corr} contribution ($+5.6$ kcal·mol^{-1}). The ΔE_{corr} and $\Delta E_{triples}$ contributions for the WDA systems behave similarly (with -25.3 kcal·mol^{-1} and -4.2 kcal·mol^{-1} for $\overline{\mathrm{Tl-Po-Tl-Po-Tl-Po}}_{\mathrm{WDA}}$ as the largest bonding contributions). The III-V systems show on average twice as large ΔE_{corr} and $\Delta E_{triples}$ contributions compared to the IV-V systems.

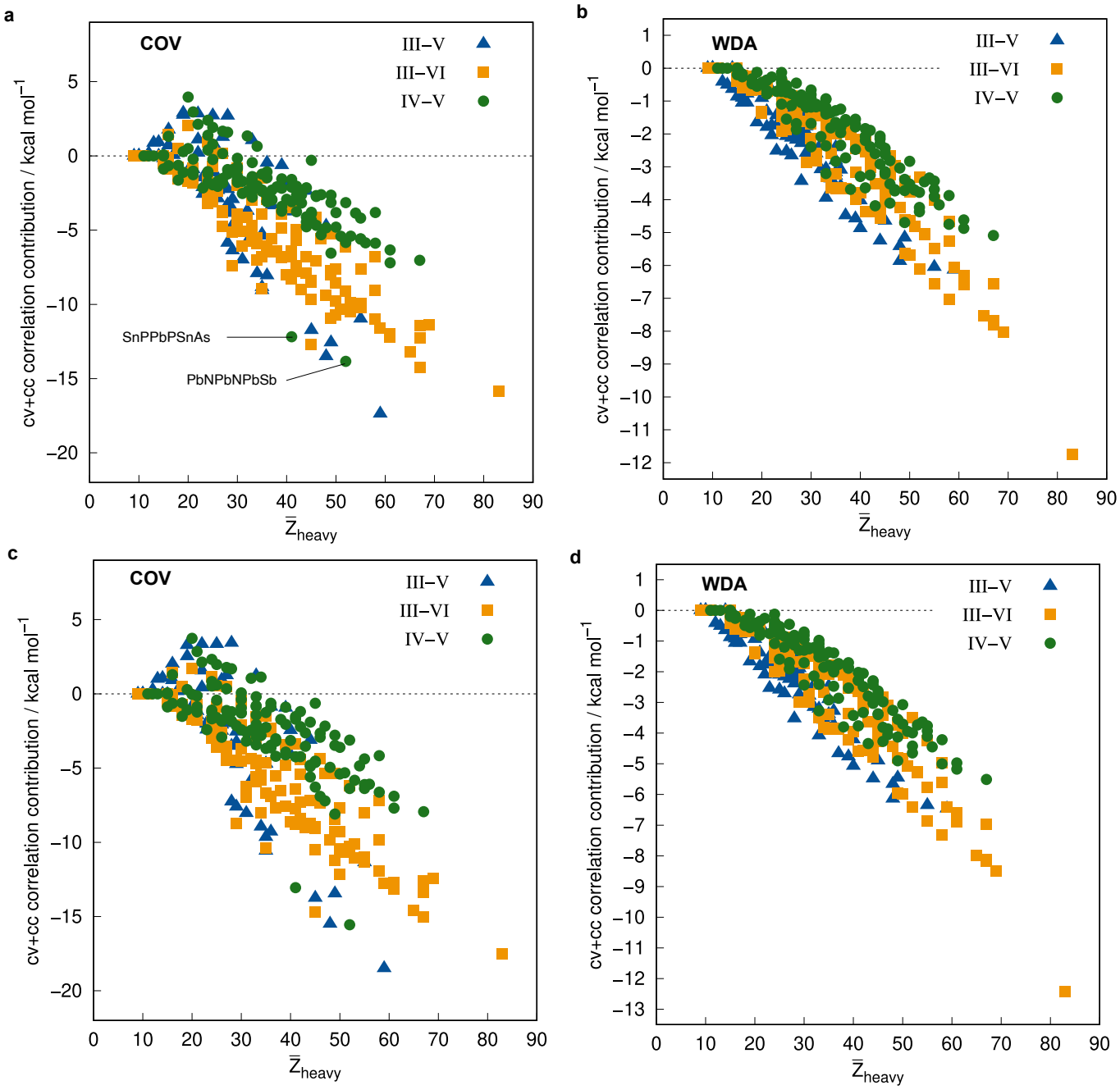


Figure S2 Correlation between the mean atomic number of all atoms except H in the molecule ($\overline{Z}_{heavy}$) and the cv+cc correlation contribution for PNO-LMP2-F12(PAO)/vtz-f12 (`THRPNO_MP2=1d-8`) (a,b) and PNO-LMP2-F12(PAO)/awcvtz (`THRPNO_MP2=1d-10`) (c,d), respectively, for the covalent (COV) and the weakly donor-acceptor (WDA) subsets.

For the COV structures, there is, however, no clear trend of the ΔE_{corr} and $\Delta E_{triples}$ contributions to the dimerization energies, respectively, with increasing $\overline{Z}_{heavy}$. For the WDA structures, only the III-VI systems show a rough trend towards a linear increase of the ΔE_{corr} and $\Delta E_{triples}$ contributions to the dimerization energies, respectively, with $\overline{Z}_{heavy}$, which is not recognizable for the III-V and IV-V systems.

Furthermore, we want to look at the trends in the cv+cc correlation and F12 contributions with PNO-LMP2-F12(PAO)/vtz-f12 vs. awcvtz as well as the basis set correction (cf. Eq. 2 in the paper), whereby it should be noted that the cv+cc contributions were calculated from the corresponding PNO-LMP2-F12 pair energies, which are considerably too large compared to PNO-LCCSD-F12

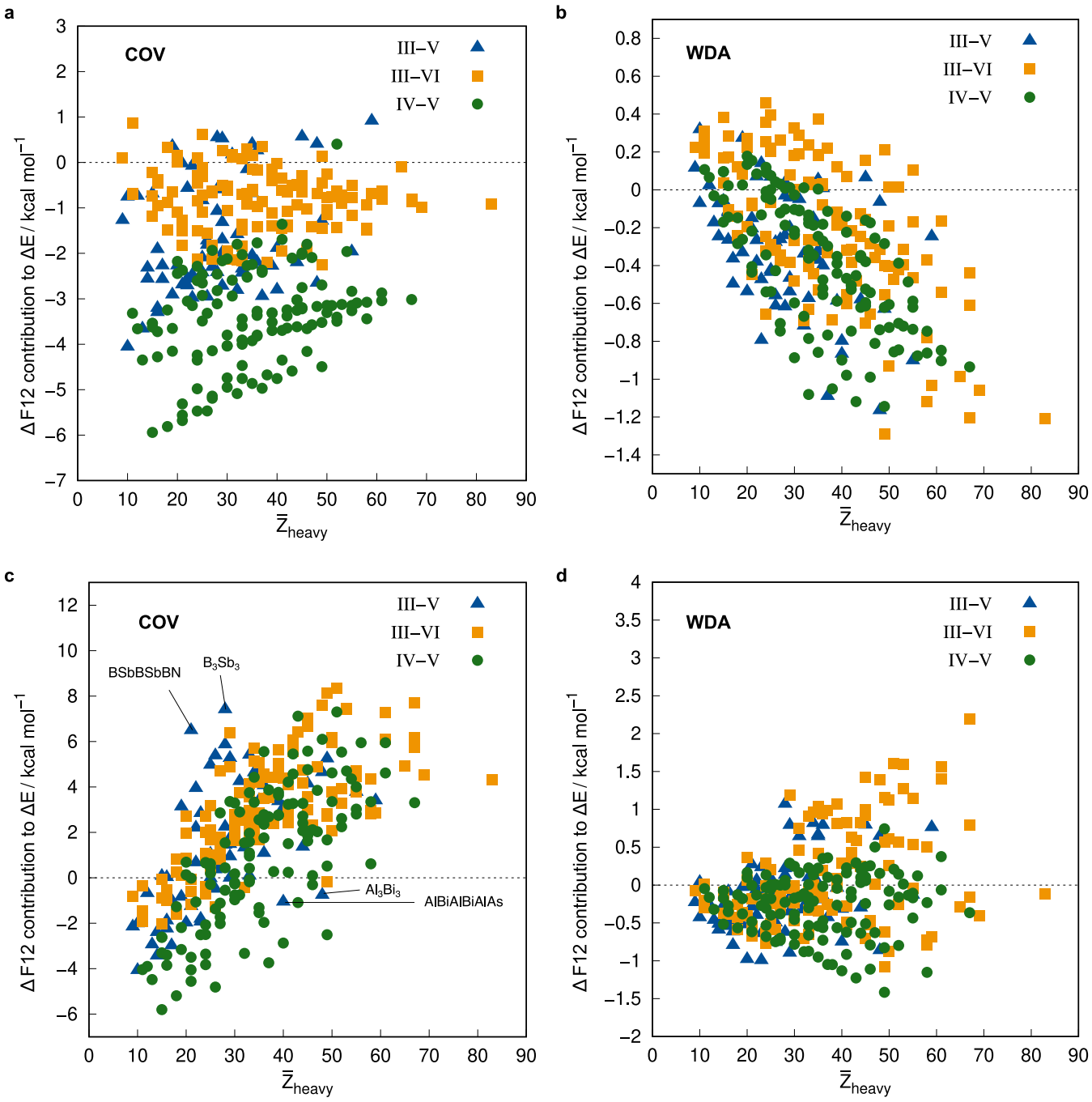


Figure S3 Correlation between the mean atomic number of all atoms except H in the molecule ($\overline{Z}_{heavy}$) and the F12 contribution to the dimerization energies computed with PNO-LMP2-F12(PAO)/vtz-f12 (`THRPNO_MP2=1d-8`) (a,b) and PNO-LMP2-F12(PAO)/awcvtz (`THRPNO_MP2=1d-10`) (c,d), respectively, for the covalent (COV) and the weakly donor-acceptor (WDA) subsets.

and thus only allow qualitative statements to be made. The difference between the vtz-f12 and awcvtz cc+cv contributions is small and the chemical trends are identical (cf. Figure S2). For the WDA systems, the corresponding contributions are negative for all systems and increase linearly with $\overline{Z}_{heavy}$. III-V systems usually have a slightly larger cv+cc correlation than III-VI systems and IV-V systems tend to show the smallest cv+cc contributions. The covalent systems show a comparable trend with greater cv+cc correlation, although not quite as systematic, and also partly positive cv+cc contributions.

Figure S3 shows the total F12 contributions to the respective dimerization energies. Opposed to the results with the vtz-f12 basis, the ΔF12 contributions with awcvtz are mostly negative and small in magnitude for the WDA structures (up to -1.3 kcal·mol^{-1} for

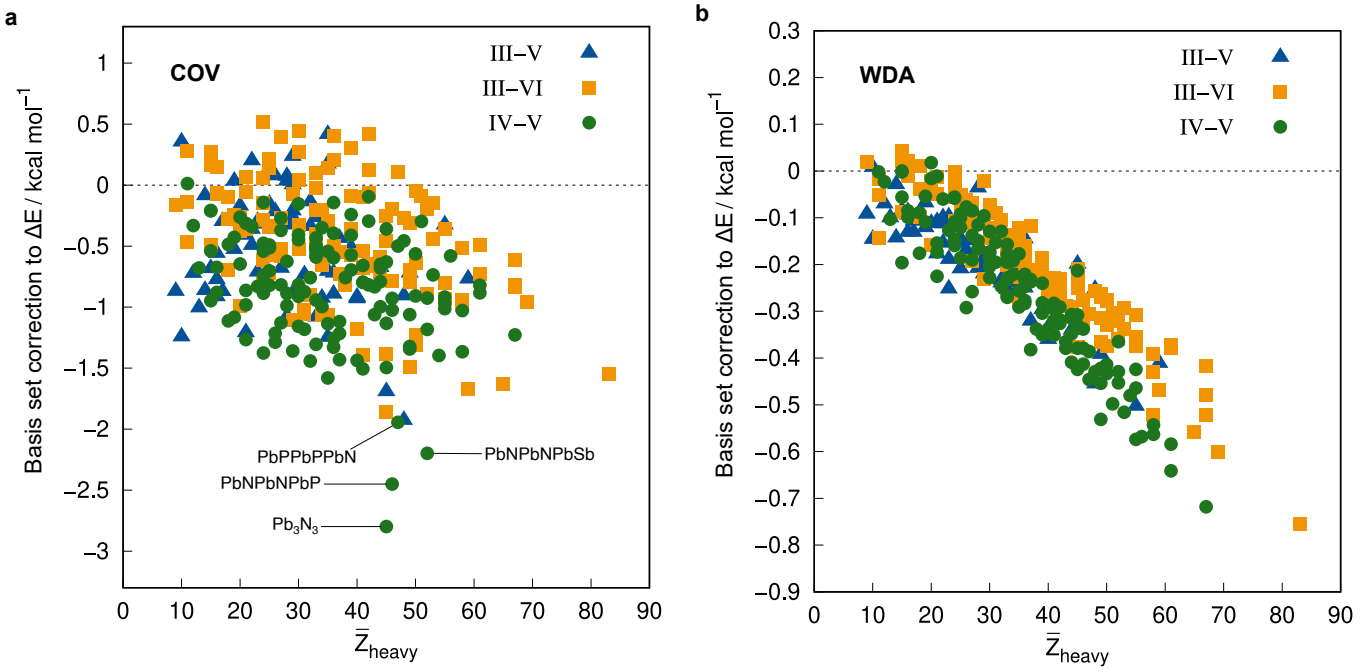


Figure S4 Correlation between the mean atomic number of all atoms except H in the molecule ($\overline{Z}_{heavy}$) and the MP2-F12 basis set correction (a,b) (PNO-LMP2-F12(PAO)/awcvtz (`THRPNO_MP2=1d-10`) - PNO-LMP2-F12(PAO)/vtz-f12 (`THRPNO_MP2=1d-8`), cf. Eq. 2 in the paper) for the covalent (COV) and the weak donor-acceptor (WDA) subsets.

$\overline{Al-Po-Al-Po-Al-Po}_{WDA}$). The positive ΔF12 contributions artificially increased by BSSE for vtz-f12 are also significantly smaller with awcvtz (e.g. for $\overline{Tl-Te-Tl-Te-Tl-Te}_{WDA}$ from 2.2 kcal·mol^{-1} to −0.4 kcal·mol^{-1}). The increase in ΔF12 with increasing $\overline{Z}_{heavy}$ for the WDA systems is probably due to an increase in cc+cv contributions with Z, although the trend is not as consistent. In contrast, for the covalent systems with awcvtz there is almost no dependence on $\overline{Z}_{heavy}$, but the IV-V systems show the largest F12 contributions (up to −5.9 kcal·mol^{-1}). For the covalent systems, the reduction in BSSE with awcvtz compared to vtz-f12 is even more pronounced. Almost all ΔF12 contributions are negative with awcvtz, even for the systems with the very large ΔF12 contribution with vtz-f12 (8.4 kcal·mol^{-1} for $\overline{In-Te-In-Te-In-Te}_{COV}$ becomes −0.8 kcal·mol^{-1} with awcvtz).

Although the differences in the F12 contributions with vtz-f12 and awcvtz are quite large, the PNO-LMP2-F12(PAO)/awcvtz(1d-10) - PNO-LMP2-F12(PAO)/vtz-f12(1d-8) basis set corrections are comparatively small (up to −2.8 kcal·mol^{-1} for $\overline{Pb-N-Pb-N-Pb-N}_{COV}$ and −0.8 kcal·mol^{-1} for $\overline{Tl-Po-Tl-Po-Tl-Po}_{WDA}$). The basis set correction to ΔE is illustrated in Figure S4. It can be concluded that F12 strongly reduces the basis set dependency and the BSSE. The residual errors for awcvtz relative to awcvqz are mainly due to the basis set errors of the triples contributions, which are not corrected by F12. The basis set corrections are largely negative and behave similarly to the ΔF12 contributions in terms of the $\overline{Z}_{heavy}$ trend (no trend visible for covalent systems, (linear) increase of the basis set correction with $\overline{Z}_{heavy}$ for the WDA structures), whereby this trend is even more pronounced here. This systematic behavior confirms our conclusions from the tests discussed in Section 3 of the paper.

S4 Additional DFT Investigations

In addition to the evaluations of the error in ΔE as function of $\overline{Z}_{heavy}$ presented for ωB97M-V and PWPB95-D4 in the paper, data for r^{2}SCAN0-D4 and r^{2}SCAN-3c can be found in Figure S5. Further, the absolute relative deviation as a function of $\overline{Z}_{heavy}$ for r^{2}SCAN0-D4 is depicted in Figure S6. These plots show that there are no clear trends of the error with increasing atomic number. Nevertheless, for the covalent subset, larger errors are observed for elements with small atomic numbers, which can be attributed to their comparably larger Lewis basicity/acidity of the light elements causing stronger covalent bonds upon dimerization. Large absolute relative deviations observed for boron primarily originate from its relatively small dimerization energies. For r^{2}SCAN0-D4 some remarkably large errors are observed for systems containing 4*p* elements like germanium or gallium. These can be partly attributed to the lack of ECPs for these elements in the applied def2-QZVPP basis set.

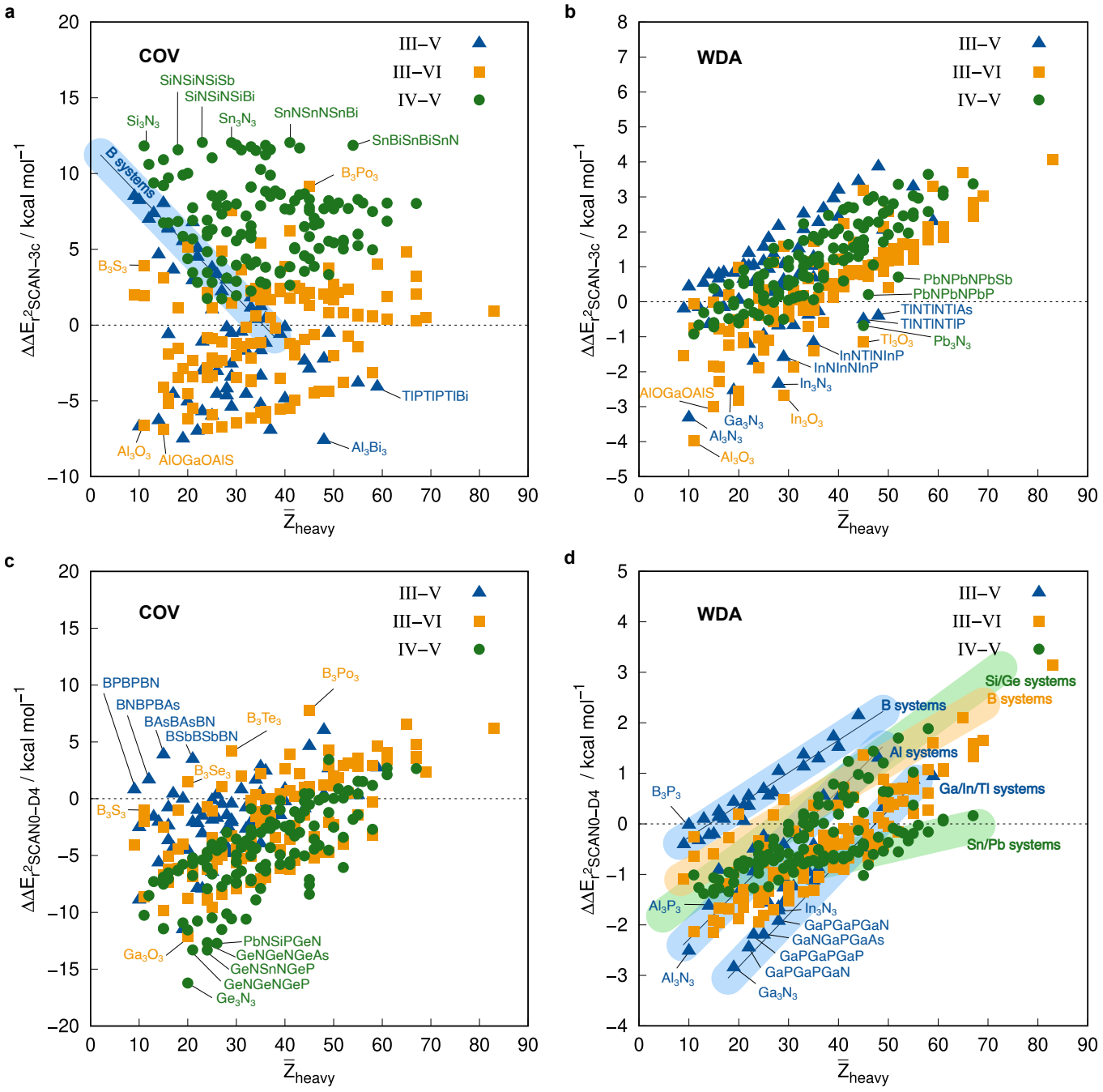


Figure S5 Correlation between the mean atomic number of all atoms except H in the molecule ($\overline{Z}_{heavy}$) and the deviations computed with r^{2}SCAN-3c (a,b), and with r^{2}SCAN0-D4 (c,d) in combination with the def2-QZVPP basis set for the covalent (COV) and the weakly donor-acceptor (WDA) subsets with respect to the reference dimerization energies.

DFT results for selected well-performing functionals obtained with the awcvtz basis sets are depicted in Figure S7. Further, a comparison of awcvtz and def2-QZVPP is given in Table S6. While no improvement by using the awcvtz basis set is observed for the (meta-)GGA functionals, slightly better results are obtained for (double-)hybrid functionals. The expected general improvement due to the now considered scalar-relativistic effects (via ECPs in the awcvtz basis set) for 4th-row elements is partly outweighed by the generally (slightly) worse performance of the awcvtz basis compared to def2-QZVPP and partly also by the elimination of random error compensation for 4th-row elements.

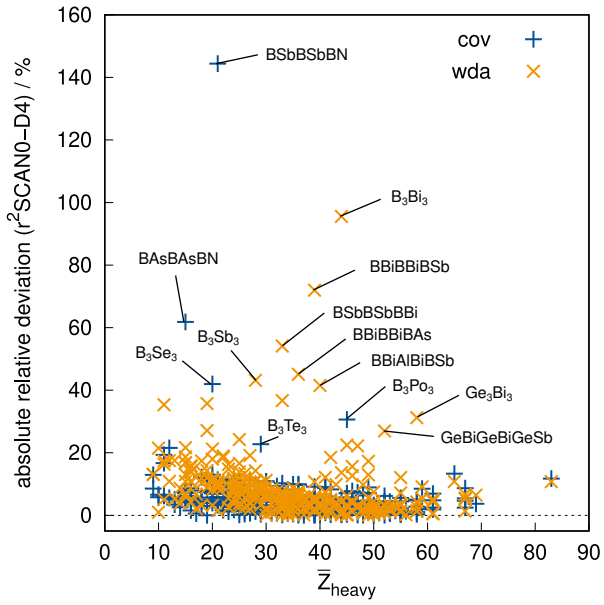


Figure S6 Correlation between the mean atomic number of all atoms except H in the molecule ($\overline{Z}_{heavy}$) and the absolute relative deviation of the dimerization energy computed with r^{2}SCAN0-D4 with respect to the final reference method.

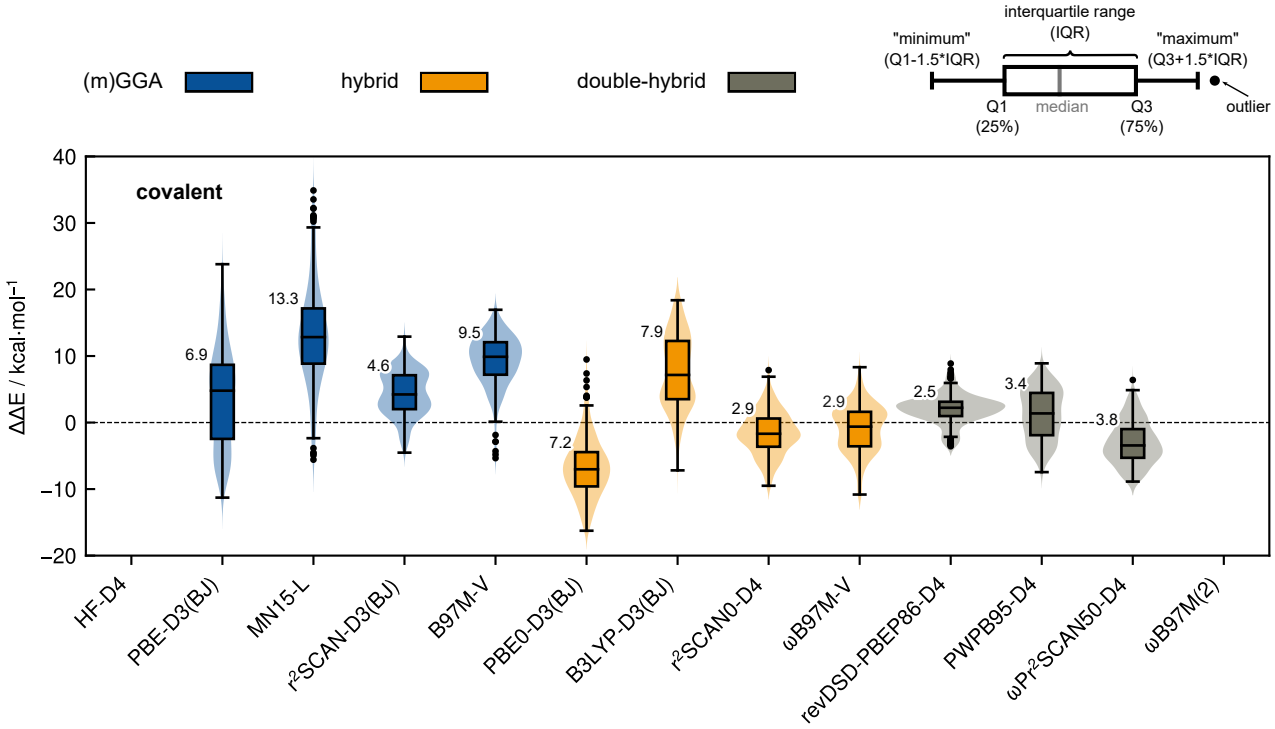


Figure S7 Violinplots with boxplots of the deviations from the PNO-LCCSD(T)-F12/vtz-f12(corr.) reference for selected DFT methods for covalent dimerizations in kcal·mol^{-1} calculated with the awcvtz-pp basis set. The numbers at the boxes are the MADs of the respective method.

Table S6 Root-mean-square deviation (RMSD) for the covalent IHD302 benchmark set computed with different density functional approximations.

	RMSD / kcal mol^{-1}								
	PBE-D3BJ	MN15-L	r^{2}SCAN-D3BJ	B97M-V	PBE0-D3BJ	r^{2}SCAN0-D4	ωB97M-V	revDSD-PBEP86-D4	ωPr^{2}SCAN50-D4
awcvtz	8.2	15.1	5.4	10.1	8.1	3.6	3.7	3.0	4.4
def2-QZVPP	7.3	14.4	4.3	8.8	9.1	5.1	4.8	2.8	7.1

S5 MAD vs. Computation Time

Figure S8 visualizes the cost-accuracy ratio of selected methods. The good performance of the ωB97X-3c and r^{2}SCAN-3c composite methods compared to the much more computationally demanding revDSD-PBEP86-D4 double-hybrid functional is specifically remarkable.

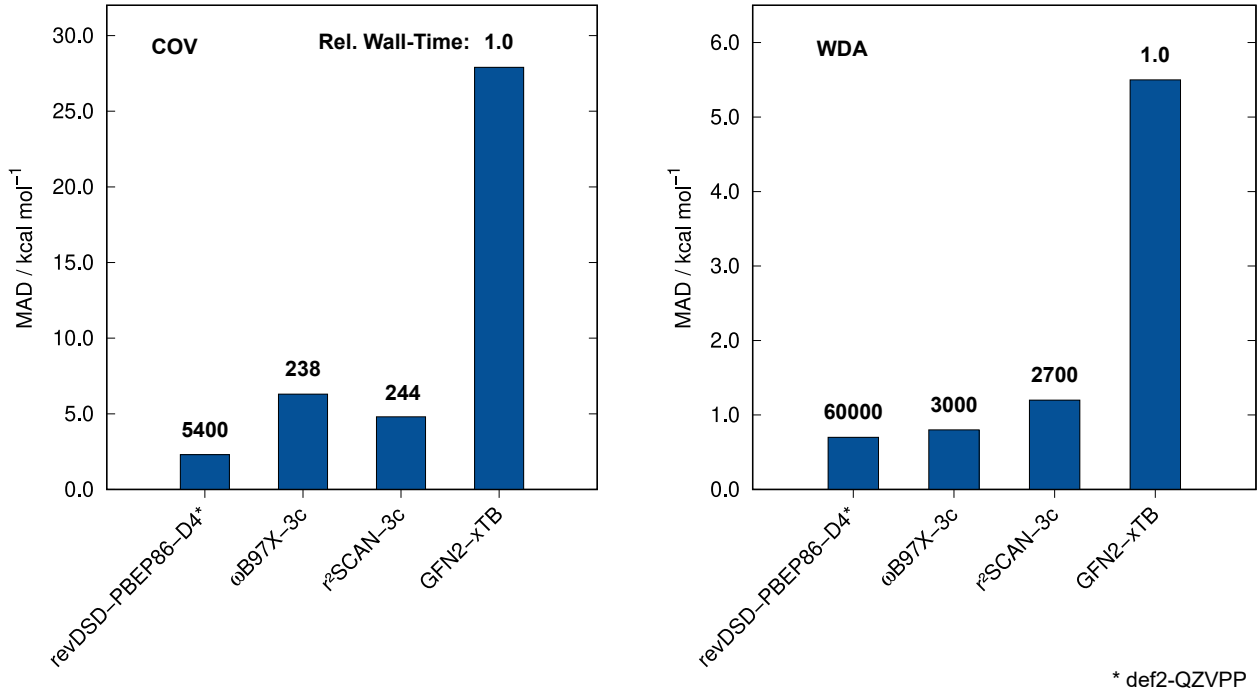


Figure S8 Mean absolute deviation (MAD) of the whole IHD302 test set and relative wall-times for one single-point calculation of the respective covalent or WDA system $\overline{\text{Tl}-\text{Po}-\text{Tl}-\text{Po}-\text{Tl}-\text{Po}}$. Computations were conducted on the Intel® Xeon® E3-1270 v5 @ 3.60 GHz with 4 cores.

S6 Relativistic Effects

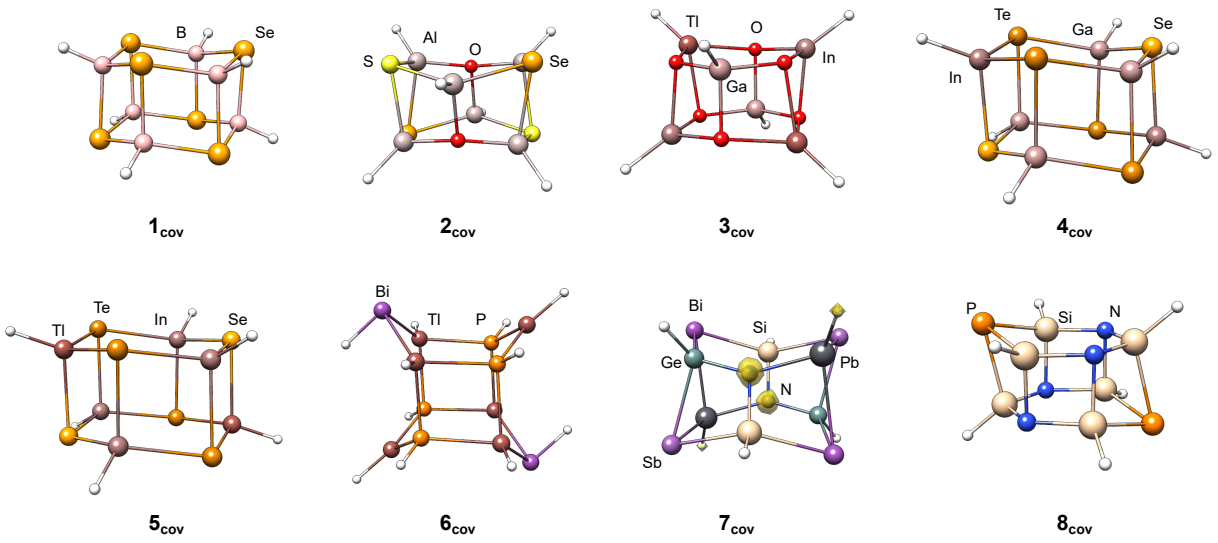
Relativistic effects beyond scalar contributions, specifically spin-orbit coupling (SOC) contributions, were estimated by computing the dimerization energy of the heaviest structure of the IHD302 test set ($\overline{\text{Tl}-\text{Po}-\text{Tl}-\text{Po}-\text{Tl}-\text{Po}}$) with the exact two-component relativistic treatment (X2C) and the zeroth-order regular approximation (ZORA) (Table S7). Since the X2C dimerization energies

without SOC contribution and the respective values calculated with effective core potentials (ECPs) agree well, the EPCs cannot effectively cover SOC contribution via their fit to fully relativistic (or experimental) data in this case. The in this way estimated SOC contributions are about 18.4 and 5.9 kcal·mol^{-1} for the covalent and WDA dimers, respectively computed at the PBE0-D4/x2c-QZVPPall-2c level of theory and have a destabilizing effect for the investigated dimers. An unusually large deviation upon increasing the basis set size from TZ2P to QZ4P is observed for the Slater-type orbital basis sets.

Table S7 Dimerization energies for $\overline{\mathrm{Tl-Po-Tl-Po-Tl-Po}}_{\mathrm{WDA}}$ computed with different relativistic treatments in combination with PBE0-D4. The comparison includes scalar as well as spin-orbit coupling (SOC) contributions.

			Dimerization Energy / kcal·mol^{-1}		
System	Relativistic	Basis Set	Scalar Relativistic	SOC Relativistic	SOC Contribution
Covalent	ZORA	ZORA/TZ2P	-55.0	-41.2	13.8
	X2C	TZ2P	-55.2	-41.2	14.0
	X2C	QZ4P	-53.7	-35.4	18.3
	X2C	x2c-QZVPPall-2c	-53.5	-35.1	18.4
	PP	aug-cc-pwCVQZ-PP	-53.1		
	PP	def2-QZVPP	-53.4		
WDA	ZORA	ZORA/TZ2P	-31.2	-27.5	3.7
	X2C	TZ2P	-31.5	-27.5	4.0
	X2C	QZ4P	-30.0	-24.0	6.0
	X2C	x2c-QZVPPall-2c	-30.3	-24.4	5.9
	PP	aug-cc-pwCVQZ-PP	-30.0		
	PP	def2-QZVPP	-30.1		

S7 Multi-reference Character



$\mathbf{1}_{\mathrm{cov}}$ $\mathbf{2}_{\mathrm{cov}}$ $\mathbf{3}_{\mathrm{cov}}$ $\mathbf{4}_{\mathrm{cov}}$

$\mathbf{5}_{\mathrm{cov}}$ $\mathbf{6}_{\mathrm{cov}}$ $\mathbf{7}_{\mathrm{cov}}$ $\mathbf{8}_{\mathrm{cov}}$

Figure S9 FOD plots of example structures at σ = 0.005 e Bohr^{-3} and 5000K computed with TPSS/def2-TZVP. The FOD is shown in yellow.

To assess the multi-reference character of the investigated systems, we conducted unrestricted PBE0-D4/def2-QZVPP single-point calculations for the IHD302 set. No spin-symmetry was found ($\Delta\langle S^2\rangle_{\mathrm{PBE0}} = 0.000$) for all systems. Further, we performed a fractional occupation number weighted electron density (FOD) analysis employing TPSS/def2-TZVP level with default electronic temperature (5000 K) for visualization of the FOD as recommended in the original publication [17] and the size-intensive non-dynamic correlation

measure r_{ND} as proposed by Martin *et al.*[18] and Matito *et al.*[19] was computed at the r^{2}SCAN0-D4/def2-QZVPP level of theory for 10000 K (see data_dft_sqm.xlsx for individual values). Overall, the maximum r_{ND} value is 0.449, and hence, we can conclude that static correlation does not play a role for the systems comprised in the *IHD302* benchmark. The FOD plots for the test molecules used in Section 3 of the paper are shown in Figure S9 (for the compound abbreviations, cf. the paper). Since the FOD is very small here, it is barely visible in the plots (only slightly for $\mathbf{7_{cov}}$).

S8 AVQZ-PP-KS Basis Set

The aug-cc-pVQZ-PP-KS (avqz-pp-ks) basis sets derived from aug-cc-pVQZ-PP using atomic RKS,PBE0 coefficients can be found in a separate file in the Supporting Information (avqz-pp-ks.basis).

Notes and references

1 F. Neese, F. Wennmohs, U. Becker and C. Riplinger, *J. Chem. Phys.*, 2020, **152**, 224108.

2 F. Neese, *WIREs Comp. Mol. Science*, 2022, **12**, e1606.

3 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297.

4 D. Andrae, U. Häußermann, M. Dolg, H. Stoll and H. Preuß, *Theor. Chim. Acta*, 1990, **77**, 123–141.

5 K. A. Peterson, D. Figgen, E. Goll, H. Stoll and M. Dolg, *J. Chem. Phys.*, 2003, **119**, 11113–11123.

6 F. Neese, F. Wennmohs, A. Hansen and U. Becker, *Chem. Phys.*, 2009, **356**, 98–109.

7 *DFT-D4 at GitHub, https://github.com/dftd4/dftd4 (accessed 20.03.2024).*

8 *s-DFT-D3 at GitHub, https://github.com/dftd3/simple-dftd3 (accessed 20.03.2024).*

9 E. Epifanovsky, A. T. B. Gilbert, X. Feng, J. Lee, Y. Mao, N. Mardirossian, P. Pokhilko, A. F. White, M. P. Coons, A. L. Dempwolff, Z. Gan, D. Hait, P. R. Horn, L. D. Jacobson, I. Kaliman, J. Kussmann, A. W. Lange, K. U. Lao, D. S. Levine, J. Liu, S. C. McKenzie, A. F. Morrison, K. D. Nanda, F. Plasser, D. R. Rehn, M. L. Vidal, Z.-Q. You, Y. Zhu, B. Alam, B. J. Albrecht, A. Aldossary, E. Alguire, J. H. Andersen, V. Athavale, D. Barton, K. Begam, A. Behn, N. Bellonzi, Y. A. Bernard, E. J. Berquist, H. G. A. Burton, A. Carreras, K. Carter-Fenk, R. Chakraborty, A. D. Chien, K. D. Closser, V. Cofer-Shabica, S. Dasgupta, M. de Wergifosse, J. Deng, M. Diedenhofen, H. Do, S. Ehlert, P.-T. Fang, S. Fatehi, Q. Feng, T. Friedhoff, J. Gayvert, Q. Ge, G. Gidofalvi, M. Goldey, J. Gomes, C. E. González-Espinoza, S. Gulania, A. O. Gunina, M. W. D. Hanson-Heine, P. H. P. Harbach, A. Hauser, M. F. Herbst, M. Hernández Vera, M. Hodecker, Z. C. Holden, S. Houck, X. Huang, K. Hui, B. C. Huynh, M. Ivanov, Á. Jász, H. Ji, H. Jiang, B. Kaduk, S. Kähler, K. Khistyaev, J. Kim, G. Kis, P. Klunzinger, Z. Koczor-Benda, J. H. Koh, D. Kosenkov, L. Koulias, T. Kowalczyk, C. M. Krauter, K. Kue, A. Kunitsa, T. Kus, I. Ladjánszki, A. Landau, K. V. Lawler, D. Lefrancois, S. Lehtola, R. R. Li, Y.-P. Li, J. Liang, M. Liebenthal, H.-H. Lin, Y.-S. Lin, F. Liu, K.-Y. Liu, M. Loipersberger, A. Luenser, A. Manjanath, P. Manohar, E. Mansoor, S. F. Manzer, S.-P. Mao, A. V. Marenich, T. Markovich, S. Mason, S. A. Maurer, P. F. McLaughlin, M. F. S. J. Menger, J.-M. Mewes, S. A. Mewes, P. Morgante, J. W. Mullinax, K. J. Oosterbaan, G. Paran, A. C. Paul, S. K. Paul, F. Pavošević, Z. Pei, S. Prager, E. I. Proynov, Á. Rák, E. Ramos-Cordoba, B. Rana, A. E. Rask, A. Rettig, R. M. Richard, F. Rob, E. Rossomme, T. Scheele, M. Scheurer, M. Schneider, N. Sergueev, S. M. Sharada, W. Skomorowski, D. W. Small, C. J. Stein, Y.-C. Su, E. J. Sundstrom, Z. Tao, J. Thirman, G. J. Tornai, T. Tsuchimochi, N. M. Tubman, S. P. Veccham, O. Vydrov, J. Wenzel, J. Witte, A. Yamada, K. Yao, S. Yeganeh, S. R. Yost, A. Zech, I. Y. Zhang, X. Zhang, Y. Zhang, D. Zuev, A. Aspuru-Guzik, A. T. Bell, N. A. Besley, K. B. Bravaya, B. R. Brooks, D. Casanova, J.-D. Chai, S. Coriani, C. J. Cramer, G. Cserey, A. E. DePrince, R. A. DiStasio, A. Dreuw, B. D. Dunietz, T. R. Furlani, W. A. Goddard, S. Hammes-Schiffer, T. Head-Gordon, W. J. Hehre, C.-P. Hsu, T.-C. Jagau, Y. Jung, A. Klamt, J. Kong, D. S. Lambrecht, W. Liang, N. J. Mayhall, C. W. McCurdy, J. B. Neaton, C. Ochsenfeld, J. A. Parkhill, R. Peverati, V. A. Rassolov, Y. Shao, L. V. Slipchenko, T. Stauch, R. P. Steele, J. E. Subotnik, A. J. W. Thom, A. Tkatchenko, D. G. Truhlar, T. Van Voorhis, T. A. Wesolowski, K. B. Whaley, H. L. Woodcock, P. M. Zimmerman, S. Faraji, P. M. W. Gill, M. Head-Gordon, J. M. Herbert and A. I. Krylov, *J. Chem. Phys.*, 2021, **155**, 084801.

10 H.-J. Werner, P. J. Knowles, F. R. Manby, J. A. Black, K. Doll, A. Heßelmann, D. Kats, A. Köhn, T. Korona, D. A. Kreplin, Q. Ma, I. Thomas F. Miller, A. Mitrushchenkov, K. A. Peterson, I. Polyak, G. Rauhut and M. Sibaev, *J. Chem. Phys.*, 2020, **152**, 144107.

11 H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györffy, D. Kats, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, S. J. Bennie, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, S. J. R. Lee, Y. Liu, A. W. Lloyd, Q. Ma, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, T. F. Miller III, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang and M. Welborn, *MOLPRO, version 2023.1, a package of ab initio programs*.

12 AMS 2023.104*, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands,* ***2020****, http://www.scm.com/.*

13 G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders and T. Ziegler, *J. Comput. Chem.*, 2001, **22**, 931–967.

14 *TURBOMOLE V7.6 2023, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from* `http://www.turbomole.com`.

15 Q. Ma and H.-J. Werner, *WIREs Comput. Mol. Sci.*, 2018, **8**, e1371.

16 Q. Ma and H.-J. Werner, *J. Chem. Theory Comput.*, 2021, **17**, 902–926.

17 S. Grimme and A. Hansen, *Angew. Chem. Int. Ed.*, 2015, **54**, 12308–12313.

18 M. K. Kesharwani, N. Sylvetsky, A. Köhn, D. P. Tew and J. M. L. Martin, *J. Chem. Phys.*, 2018, **149**, 154109.

19 E. Ramos-Cordoba, P. Salvador and E. Matito, *Phys. Chem. Chem. Phys.*, 2016, **18**, 24015–24023.