Supporting Information for: Møller-Plesset Adiabatic Connection Theory for Diverse Noncovalent Interactions

Etienne Palos,[*,†,@] Heng Zhao,[*,‡,@] Kimberly J. Daas,[¶] Eduardo Fabiano,[§]
Francesco Paesani,[†,⊥,#] and Stefan Vuckovic[*,‡]

†*Department of Chemistry and Biochemistry, University of California San Diego,*
La Jolla, California 92093, United States
‡*Department of Chemistry, Faculty of Science and Medicine, Université de Fribourg/Universitäat*
Freiburg, Chemin du Musée 9, CH-1700 Fribourg, Switzerland
¶*Department of Chemistry, University of California, Irvine, CA 92697, United States*
§*Institute for Microelectronics and Microsystems (CNR-IMM), 73100 Lecce, Italy*
‖*Center for Biomolecular Nanotechnologies, Istituto Italiano di Tecnologia, 73010 Arnesano,*
Italy
⊥*Materials Science and Engineering, University of California San Diego,*
La Jolla, California 92093, United States
#*San Diego Supercomputer Center, University of California San Diego,*
La Jolla, California 92093, United States
@*These authors contributed equally*

E-mail: epalos@ucsd.edu; heng.zhao@unifr.ch; stefan.vuckovic@unifr.ch

Supporting Information for B30 analysis

Table S1: Interaction energies (kcal/mol) for B30 using MP2, B2PLYP-D3(BJ), MPACF1, and MPAC25, compared to CCSD(T) reference values.[1]

Reaction index	CCSD(T)	MP2	B2PLYP-D3(BJ)	MPACF1	MPAC25
1	-43.85	-46.8356	-46.7026	-46.5858	-45.7886
2	-42.59	-46.3224	-46.0835	-45.6796	-44.9559
3	-46.38	-48.3412	-47.4271	-48.4906	-47.6765
4	-43.67	-46.1087	-45.1889	-45.8264	-45.1069
5	-12.07	-13.1323	-13.9420	-12.5971	-12.2910
6	-16.03	-17.1550	-17.3919	-16.8962	-16.4922
7	-31.49	-32.3426	-31.1734	-32.0170	-31.3841
8	-25.81	-26.8573	-26.0315	-26.2367	-25.7457
9	-40.63	-41.5188	-39.8007	-41.7130	-40.9304
10	-35.28	-36.4169	-34.9543	-36.1904	-35.5454
11	-8.00	-8.2414	-8.2585	-8.2076	-7.8845
12	-12.89	-13.4398	-13.0817	-13.5419	-13.0836
13	-9.52	-9.8734	-9.3704	-9.1048	-8.9372
14	-7.79	-8.2929	-7.7800	-7.5007	-7.3777
15	-13.52	-14.4191	-13.5009	-13.5414	-13.3174
16	-11.15	-12.2125	-11.2846	-11.2382	-11.0751
17	-1.68	-1.7198	-1.5002	-1.6736	-1.5821
18	-2.55	-2.6766	-2.4988	-2.6244	-2.5086
19	-8.41	-8.9070	-8.1822	-8.1918	-8.0574
20	-6.75	-7.3432	-6.6106	-6.6426	-6.5403
21	-15.26	-16.3620	-15.6634	-15.3584	-15.1436
22	-12.31	-13.5238	-12.7360	-12.4959	-12.3383
23	-1.42	-1.4981	-1.3547	-1.3877	-1.3092
24	-2.64	-2.8429	-2.7613	-2.7058	-2.5937
25	-21.33	-21.1093	-20.1727	-21.2128	-20.6434
26	-15.73	-15.7739	-15.1307	-15.6817	-15.2679
27	-34.06	-33.8694	-32.2801	-34.4649	-33.6673
28	-27.31	-27.4469	-26.2338	-27.6113	-26.9949
29	-4.86	-4.6529	-4.6127	-4.8820	-4.5928
30	-9.18	-8.9682	-8.7231	-9.3974	-8.9501

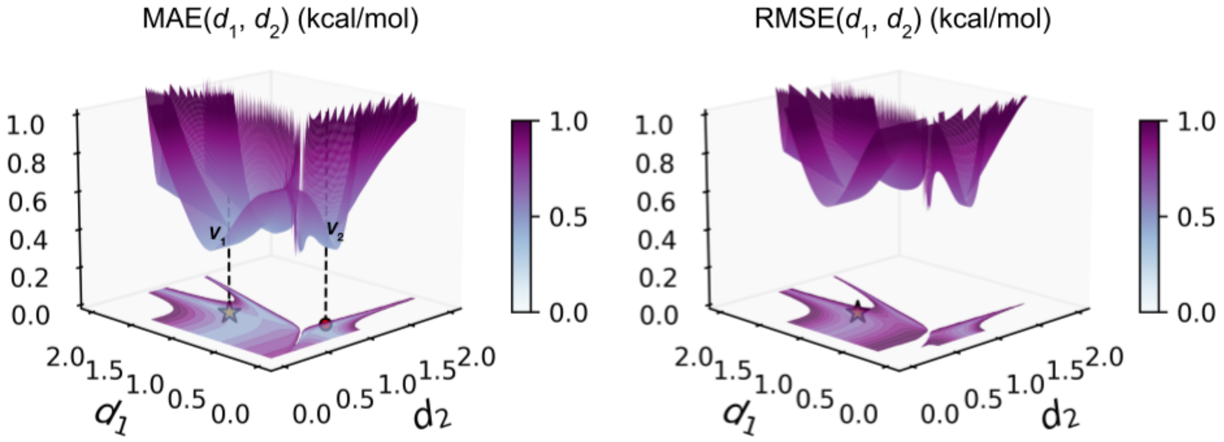
Table S2: Mean absolute errors (MAE) in kcal/mol for various methods plotted in Figure 4 in the main text.

Method	**MAE (kcal/mol)**	**Ref.**
MP2	0.87	this work
MPACF1	0.60	this work
SPL2	0.49	this work
MPAC25	**0.42**	This work
B2PLYP-D3	0.72	2
r^2SCAN@HF	1.73	3
C(HF)-dRPA@PBE	0.65	4
ωB97M-V	0.65	3
ωB97X-V	0.83	3
LC-ωPBE-D3	0.98	1
PBE0	1.69	1
PBE0-D3	1.98	1
PBE0-D3(BJ)	2.38	1
PBE0-XDM	2.46	1
B3LYP-D3	1.36	1

Table S3: Mean absolute error values for B30 for MP2, MPACF1, and MPAC25 grouped by system type.

System	**MP2**	**B2PLYP-D3(BJ)**	**MPACF1**	**MPAC25**
Halogen	2.2175	2.02	1.9143	1.2869
Chalcogen	0.6327	0.23	0.3005	0.1845
Pnictogen	0.1685	0.89	0.1852	0.3923

Supporting Information for MPAC25



B30

Figure S1: The MAE(d_1, d_2) (left) and RMSE(d_1, d_2) (right) landscapes of the B30 data set, associated with panel c of Figure 5 in the main text. We highlight that the curvature analysis for S22 holds, in that the curvature around the minimum of V_1 is smaller than the minimum in V_2.

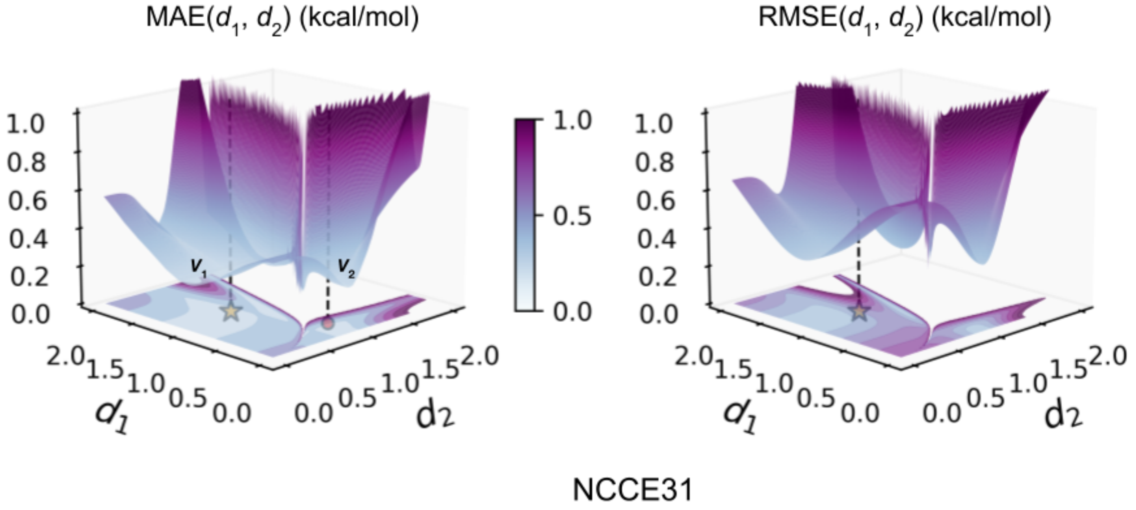


Figure S2: The MAE(d_1,d_2) (left) and RMSE(d_1,d_2) (right) landscapes of the NCCE31 data set, associated with panel c of Figure 5 in the main text. We highlight that the curvature around the minimum of V_1 is smaller than the minimum in V_2 holds, in that the curvature analysis for S22 holds, in that the curvature around the minimum of V_1 is smaller than the minimum in V_2.

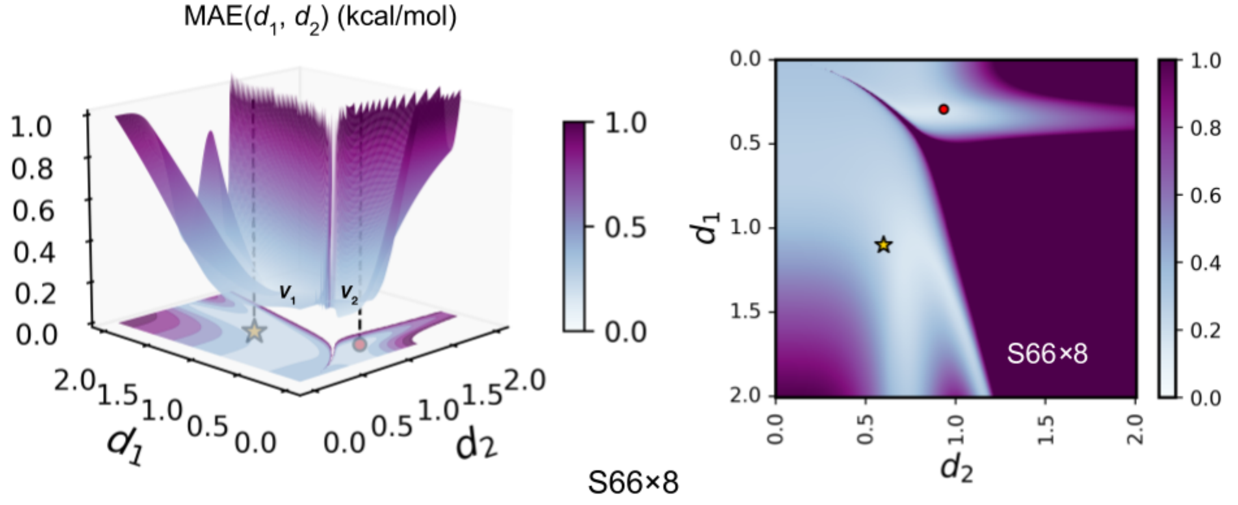


Figure S3: The MAE(d_1,d_2) (left) and RMSE(d_1,d_2) (right) landscapes of the S66×8 data set, associated with panel c of Figure 5 in the main text. We highlight that the curvature analysis for S22 holds, in that the curvature around the minimum of V_1 is smaller than the minimum in V_2.

Supporting Information for DES15K subset

Computational Details for DES15 subset calculations.

All calculations have been performed with the program package TURBOMOLE,[? ?] using the aug-cc-pVDZ and aug-cc-pVTZ basis sets.[5–9] Basis set superposition error effects have been corrected using the counterpoise procedure of Boys-Bernardi.[10] In MP2 calculations we employed the resolution of identity (RI) approximation.[11,12]

All calculations have been performed with the program package TURBOMOLE,[13] using the aug-cc-pVDZ and aug-cc-pVTZ basis sets.[5–7,14] Basis set superposition error effects have been corrected using the counterpoise procedure of Boys-Bernardi.[10] The MP2 calculations employed the resolution of identity (RI) approximation,[11,12] and all computed energies were extrapolated to the complete basis set (CBS) limit, starting from DZ and TZ results. In order to do so, we have prescreened different possible extrapolation formulas.[15,16] Finally, for Hartree-Fock (E^{HF}, HF, $E_{\mathrm{x}}^{\mathrm{HF}}$, W_∞ and W'_∞) and MP2 components respectively, we have considered the following expressions:[9,15]

$$E_{\mathrm{CBS}}^{\mathrm{HF}} = \frac{\exp(-2\alpha)E_{\mathrm{TZ}}^{\mathrm{HF}} - \exp(-3\alpha)E_{\mathrm{DZ}}^{\mathrm{HF}}}{\exp(-2\alpha) - \exp(-3\alpha)}, \tag{1}$$

$$\tag{2}$$

$$\tag{3}$$

$$E_{\mathrm{CBS}}^{\mathrm{MP2}} = \frac{(3+\beta)^3 E_{\mathrm{TZ}}^{\mathrm{MP2}} - (2+\beta)^3 E_{\mathrm{DZ}}^{\mathrm{MP2}}}{(3+\beta)^3 - (2+\beta)^3}, \tag{4}$$

where α and β are parameters to optimize. The parameters have been fixed by considering a subset of the DES15K database[17] consisting of 574 randomly chosen systems and fitting to the HF and MP2 CBS energies reported in the database. As a result, we have obtained the optimized parameters $\alpha = 2.085$ and $\beta = 0.3939$. These yield mean absolute errors (MAEs) of 0.02 and 0.06 kcal/mol for HF and MP2 energies, respectively. The data used for the optimization are reported in the `csv` file attached in the supporting material.

Statistical results for DES15K subset.

Table S4: DES15k statistics. Mean average error (MAE), Root Mean Square Error (RMSE), Mean Signed Error (MSE) and maximum error (MAX) are all reported in kcal/mol. The DFT+DISP data is from Ref. 18.

DES15k	**Full**				**Charged**				**Neutral**			
Method	**MAE**	**RMSE**	**MSE**	**MAX**	**MAE**	**RMSE**	**MSE**	**MAX**	**MAE**	**RMSE**	**MSE**	**MAX**
MP2	0.76	1.244	-0.697	7.091	0.834	1.273	-0.733	6.789	0.736	1.235	-0.686	7.091
B86bPBE25X-XDM	0.62	1.01	-0.29	7.49	1.13	1.57	-1.03	7.49	0.4	0.62	0.03	3.78
PBE0-D3(BJ)	0.83	1.3	-0.64	9.34	1.58	2.04	-1.54	9.34	0.5	-0.27	0.8	1.28
PBE0-D4	0.79	1.26	-0.59	7.17	1.48	1.96	-1.43	7.17	0.49	0.79	-0.23	5.86
PBE0-XDM	0.74	1.18	-0.36	7.85	1.31	1.8	-1.22	7.85	0.49	0.77	0	5.87
MPACF1	0.36	0.531	0.037	2.461	0.51	0.748	-2.272	2.461	0.312	0.439	0.137	2.258
MPAC25	0.343	0.455	0.232	1.756	0.372	0.499	0.095	1.756	0.333	0.439	0.276	1.48

Spin-opposite scaled MPAC25

In Scaled Opposite-Spin MP2 (SOS-MP2), only the opposite-spin term of $E_{\mathrm{c}}^{\mathrm{MP2}}$ is computed and scaled by a factor c_{os}, while the same-spin ($E_{\mathrm{ss}}^{\mathrm{MP2}}$) term is neglected. This is, $E^{\mathrm{SOS-MP2}} = c_{\mathrm{os}} E_{\mathrm{os}}^{\mathrm{MP2}}$. Neglecting the $E_{\mathrm{ss}}^{\mathrm{MP2}}$ drops all fifth-order computations, reducing the computational cost of MP2 $\sim N^5$ to $E^{\mathrm{SOS-MP2}} \sim N^4$.[2] In itself, c_{os} may reduce the scaling and improve the accuracy of MP2 for targeted applications. The same principles may be applied to reduce the computational cost of MPAC functionals, whose scaling is determined by their MP2 ingredient.

To calibrate SOS-MPAC, a constrained optimization is performed to find c_{os}-MPAC25 beginning from MPAC25. The objective is to minimize the deviation from reference energies under physical constraints, such as ensuring the scaling remains within a plausible range, e.g. $1 \leq c_{\mathrm{os}} \leq 2$. This avoids overfitting and ensures the parameter retains general transferability across chemical systems. Using only S22+B30(ab) as the training set, we test a working model SOS-MPAC25 defined by d_1= 0.8324, d_2 = 0.8257 with an associated $c_{os} \simeq 1.95$ optimal value.

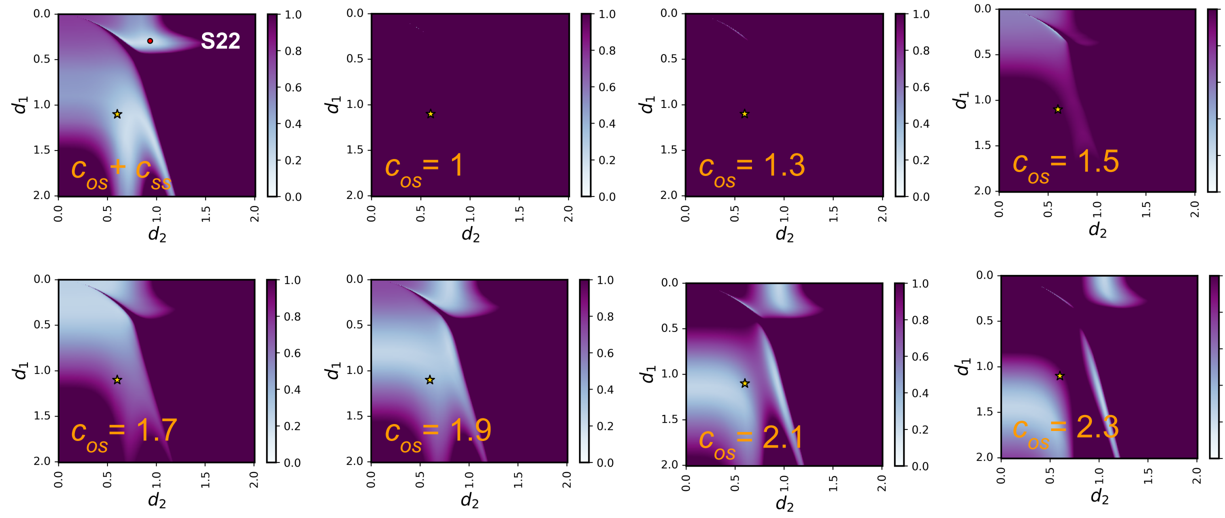


Figure S4: The MAE($d_1, d_2; c_{\mathrm{os}}$) landscape projections for the S22 dataset. Beginning from MPAC25, this is, with $c_{\mathrm{os}} = c_{\mathrm{ss}} = 1$, we sweep values of c_{os} in range [1, 2.3] ($c_{\mathrm{ss}} = 0$) exploring N^4-scaling models.

In this case, we find the optimal $c_{os} \simeq 1.95$. This number can be rationalized by the fact that, as shown in Ref. 19 , SOS-MP2 is found to "mimick" MP2/CBS at $c_{os} = 1.95$. Therefore, it is plausible that for an SOS-MPAC25 model parameterized following the protocol introduced for MPAC25, $c_{os} \to 2$ (i.e. the MPAC funcitonal *itself* does not properly know about OS/SS – it simply takes in $E_{\mathrm{c,i}}^{\mathrm{MP2}}$ where $i = [\mathrm{os}, \mathrm{ss}]$.

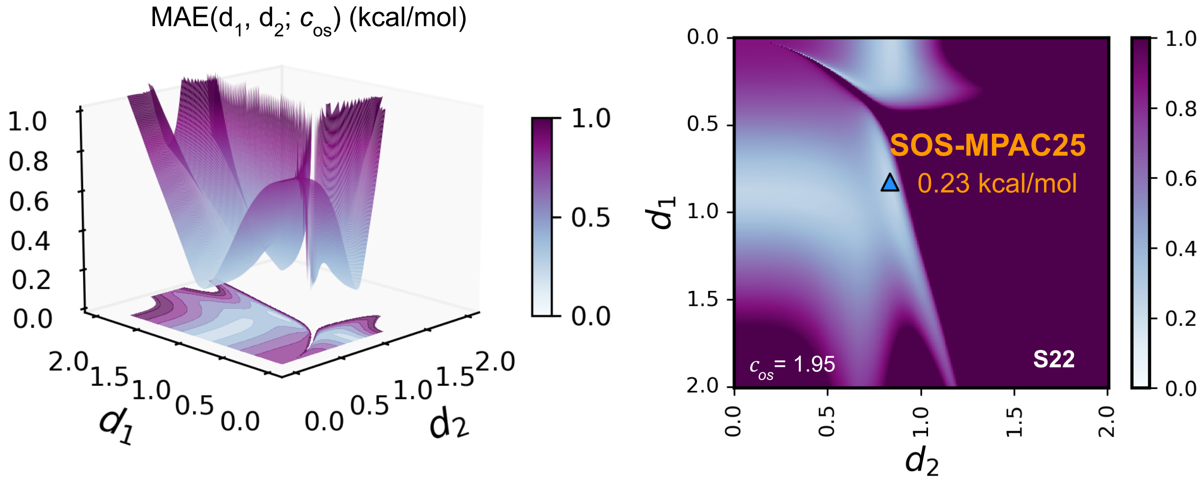


Figure S5: SOS-MPAC25: The MAE($d_1, d_2; c_{\mathrm{os}}$) landscape projections for the S22 test set, with $c_{\mathrm{os}} = 1.95$. The MAE of SOS-MPAC25 on the S22 test set is of 0.23 kcal/mol.

SOS-MPAC25 achieves minimal change in accuracy on S22.

Table S5: Comparison of errors for different SOS/c_{os} methods across datasets. Errors are reported in kcal/mol.

Dataset	c_{os}-MP2	c_{os}-MPACF1	c_{os}-MPAC25
S22	1.06	0.22	0.23
B30	1.42	2.05	1.08
NCCE31	0.47	0.76	0.36
S22+B30(ab)	1.07	1.66	0.80
B30(ab)+CT7	0.71	3.08	1.8
DES15K	2.04	0.57	0.98

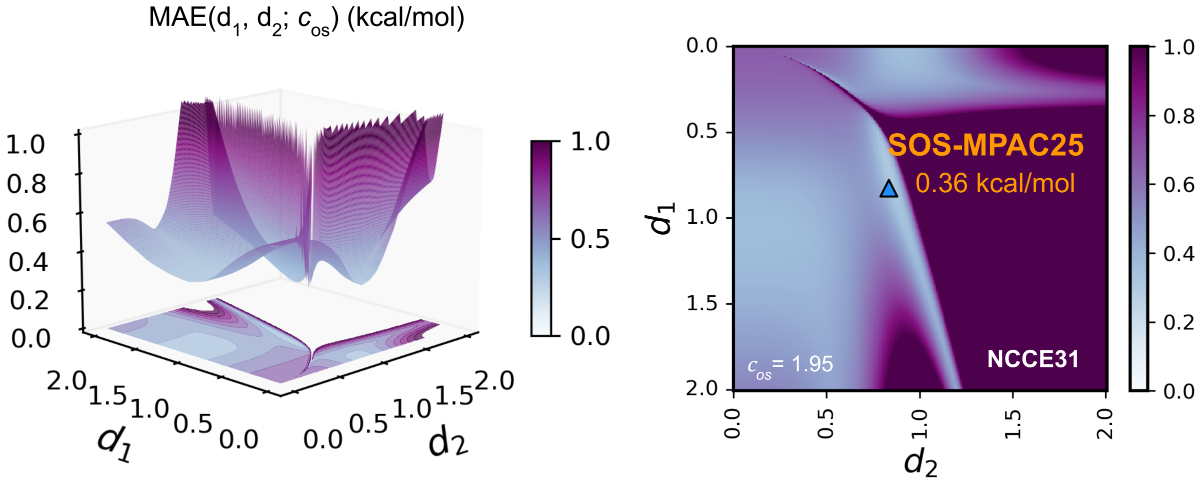


Figure S6: SOS-MPAC25: The MAE($d_1, d_2; c_{\mathrm{os}}$) landscape projections for the NCCE31 test set, with $c_{\mathrm{os}} = 1.95$. The MAE of SOS-MPAC25 on the NCCE31 test set is of 0.36 kcal/mol.

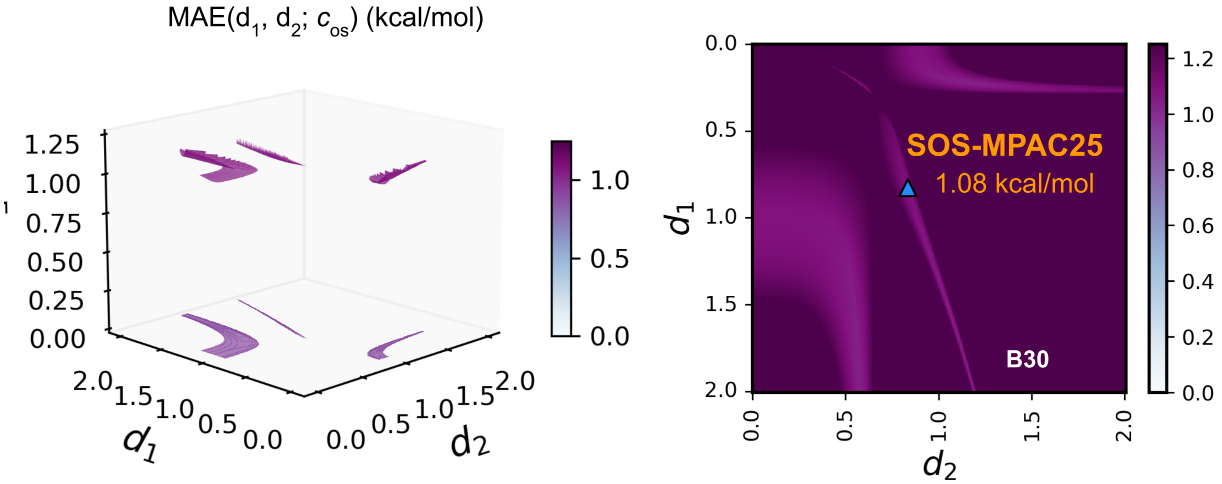


Figure S7: SOS-MPAC25: The MAE($d_1, d_2; c_{\mathrm{os}}$) landscape projections for the B30 test set, with $c_{\mathrm{os}} = 1.95$. The MAE of SOS-MPAC25 on the B30 test set is of 1.08 kcal/mol.

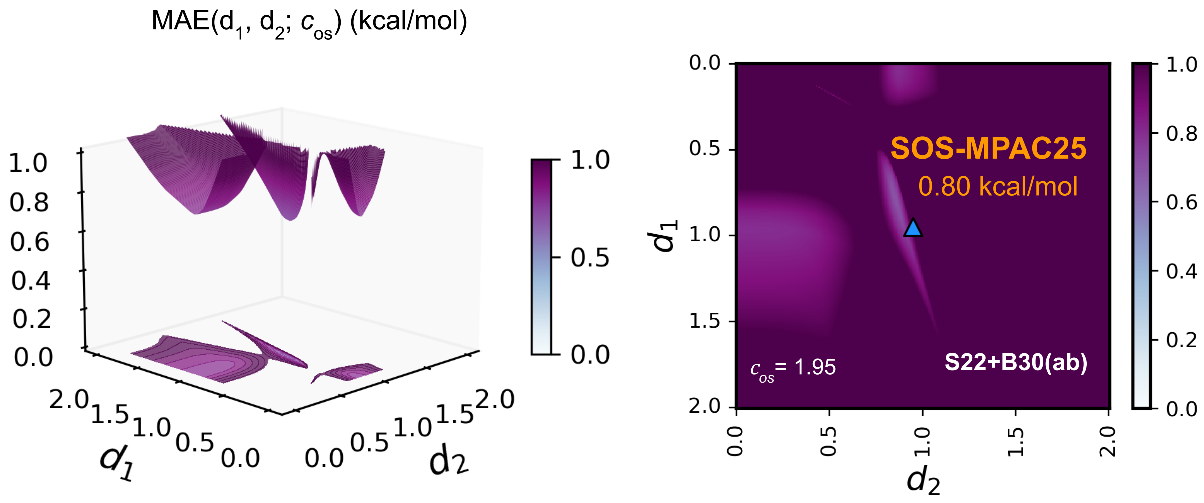


Figure S8: SOS-MPAC25: The MAE($d_1, d_2; c_{\text{os}}$) landscape projections for the combined S22+B30(ab) training set, with $c_{\text{os}} = 1.95$. The MAE of SOS-MPAC25 on the S22+B30(ab) training set is of 0.80 kcal/mol.

References

(1) Kim, Y.; Song, S.; Sim, E.; Burke, K. Halogen and chalcogen binding dominated by density-driven errors. *J. Phys. Condens. Matter* **2018**, *10*, 295–301.

(2) Daas, K. J.; Fabiano, E.; Della Sala, F.; Gori-Giorgi, P.; Vuckovic, S. Noncovalent interactions from models for the Møller–Plesset adiabatic connection. *J. Phys. Condens. Matter* **2021**, *12*, 4867–4875.

(3) Hernandez, D. J.; Rettig, A.; Head-Gordon, M. A New View on Density Corrected DFT: Can One Get a Better Answer for a Good Reason? *arXiv preprint arXiv:2306.15016* **2023**,

(4) Graf, D.; Thom, A. J. Corrected density functional theory and the random phase approximation: Improved accuracy at little extra cost. *J. Chem. Phys.* **2023**, *159*, 174106.

(5) Kendall, R. A.; Dunning Jr, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

(6) Woon, D. E.; Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358–1371.

(7) Woon, D. E.; Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties. *J. Chem. Phys.* **1994**, *100*, 2975–2988.

(8) Prascher, B. P.; Woon, D. E.; Peterson, K. A.; Dunning Jr, T. H.; Wilson, A. K. Gaussian basis sets for use in correlated molecular calculations. VII. Valence, core-valence, and scalar relativistic basis sets for Li, Be, Na, and Mg. *Theoretical Chemistry Accounts* **2011**, *128*, 69–82.

(9) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Olsen, J. Basis-set convergence of the energy in molecular Hartree–Fock calculations. *Chem. Phys. Lett.* **1999**, *302*, 437–446.

(10) Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553–566.

(11) Hättig, C.; Weigend, F. CC2 excitation energy calculations on large molecules using the resolution of the identity approximation. *J. Chem. Phys.* **2000**, *113*, 5154–5161.

(12) Hättig, C.; Hellweg, A.; Köhn, A. Distributed memory parallel implementation of energies and gradients for second-order Møller–Plesset perturbation theory with the resolution-of-the-identity approximation. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1159–1169.

(13) Franzke, Y. J.; Holzer, C.; Andersen, J. H.; Begušić, T.; Bruder, F.; Coriani, S.; Della Sala, F.; Fabiano, E.; Fedotov, D. A.; Fürst, S.; others TURBOMOLE: Today and tomorrow. *J. Chem. Theory Comput.* **2023**, *19*, 6859–6890.

(14) Dunning Jr, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(15) Okoshi, M.; Atsumi, T.; Nakai, H. Revisiting the extrapolation of correlation energies to complete basis set limit. *J. Comput. Chem.* **2015**, *36*, 1075–1082.

(16) Huh, S. B.; Lee, J. S. Basis set and correlation dependent extrapolation of correlation energy. *J. Chem. Phys.* **2003**, *118*, 3035–3042.

(17) Donchev, A. G.; Taube, A. G.; Decolvenaere, E.; Hargus, C.; McGibbon, R. T.; Law, K.-H.; Gregersen, B. A.; Li, J.-L.; Palmo, K.; Siva, K.; others Quantum chemical benchmark databases of gold-standard dimer interaction energies. *Scientific data* **2021**, *8*, 55.

(18) Nickerson, C. J.; Bryenton, K. R.; Price, A. J.; Johnson, E. R. Comparison of density-functional theory dispersion corrections for the DES15K database. *J. Phys. Chem. A* **2023**, *127*, 8712–8722.

(19) Jung, Y.; Head-Gordon, M. A fast correlated electronic structure method for computing interaction energies of large van der Waals complexes applied to the fullerene–porphyrin dimer. *Physical Chemistry Chemical Physics* **2006**, *8*, 2831–2840.