Supporting Information for

"Quantifying and Understanding Errors in Molecular Geometries"

Stefan Vuckovic and Kieron Burke

Departments of Chemistry and of Physics, University of California, Irvine, CA 92697, USA

October 23, 2020

Contents

List of Tables

List of Figures

S1. Additional GEO mathematical details

For an approximate quantum-mechanical (QM) solver $\tilde{E}(\mathbf{G})$, the total error is given by:

$$\Delta E = \tilde{E}(\tilde{\mathbf{G}}) - E(\mathbf{G}_0), \tag{S1}$$

where $E(\mathbf{G})$ is the exact ground-state energy at geometry $\mathbf{G}$, $\mathbf{G}_0$ is the exact and $\tilde{\mathbf{G}}$ is the approximate geometry. Thus, ΔE contains errors both due to the approximate geometry and approximate energy. To decompose this error into the GEO and "non-GEO" part, we add and subtract $E(\tilde{\mathbf{G}})$ to the r.h.s of Eq. S1

$$\Delta E = \underbrace{E(\tilde{\mathbf{G}}) - E(\mathbf{G}_0)}_{E_{geo}\geq 0} + \underbrace{\tilde{E}(\tilde{\mathbf{G}}) - E(\tilde{\mathbf{G}})}_{E_{\mathrm{p}}}. \tag{S2}$$

We can also add and subtract $\tilde{E}(\mathbf{G}_0)$ to the r.h.s of Eq. S1 to obtain an alternative form of Eq. S2

$$\Delta E = \underbrace{\tilde{E}(\tilde{\mathbf{G}}) - \tilde{E}(\mathbf{G}_0)}_{E'_{geo}\leq 0} + \underbrace{\tilde{E}(\mathbf{G}_0) - E(\mathbf{G}_0)}_{E'_{\mathrm{p}}}. \tag{S3}$$

The signs of E_{geo} and E'_{geo} also dictate the following chain of inequalities:

$$E_{\mathrm{p}} \leq \Delta E \leq E'_{\mathrm{p}}. \tag{S4}$$

For an illustration of the error decomposition by means of Eqs. S1 and S3 see Fig. S1. As discussed throughout the text, E_{geo} is typically accurately approximated by $-E'_{geo}$, and thus : $E_{geo} \approx \frac{1}{2}(E'_{\mathrm{p}} - E_p)$ (see Fig. S1).

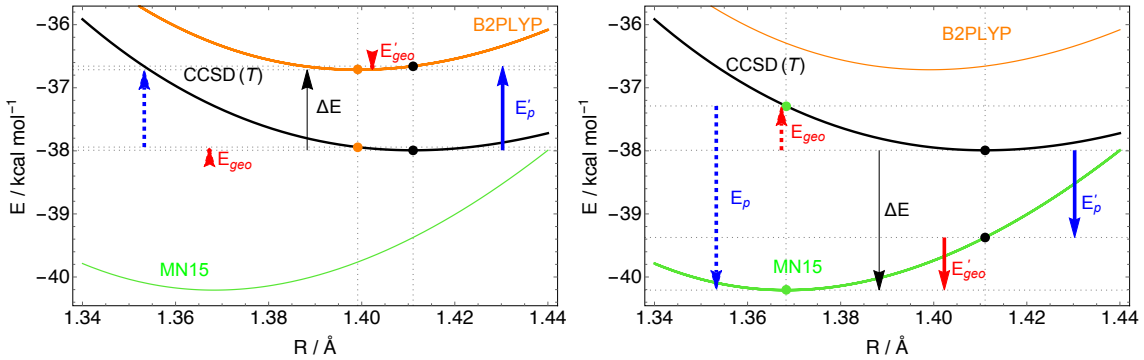


Figure S1: *The decomposition of ΔE error by Eqs. S1 and S3) for the atomization energy of the F_2 molecules obtained from the B2PLYP (E_{geo} and E'_{geo} are negligible) and MN15 functionals (E_{geo} and E'_{geo} are non-negligible)*

For equilibrium structures (minima of potential energy surfaces), E_{geo} is always positive. On the other hand, in the case of transition states (the first order saddle points of potential energy surfaces), the sign of E_{geo} is not definite. In that case, the $\mathbf{H}_0$ Hessian will have one negative eigenvalue. Thus, one term on the r.h.s of Eq 4 of the underlying E_{geo}^{harm} is negative, and all other are positive.

S2. Additional details on the results in Figure 1a (top panel)

Note: In some figures in the SI, 'ωB97XD' is used for the 'ωB97X-D' functional. Also, 'DFT [D3]' and 'DFT-D3' notations are interchangeably used in the SI figures.

The CCSD(T)/A'V5Z (see Ref. [1] for the basis set description) $E(\mathbf{G})$ values and $\mathbf{G}$ geometries have been taken from Ref. [1]. All other results have been obtained from the Gaussian16 (G16) calculations by always employing the same (A'V5Z) basis set. Unless otherwise stated, the D3 empirical dispersion correction of Grimme and co-workers is used with the original D3 damping function[2], invoked by the 'EmpiricalDispersion=GD3' keyword in G16. The E'_{geo} and E_{geo} errors for individual molecules and employed approximate methods are given in Tables S2 and S1, respectively. In Figure S2, we show the MAE for the atomization energies for the same set of molecules. The plot In Figure S5 shows the data of Table S1.

	HF	PBE	BLYP	TPSS	B3LYP	MN15	ωB97	ωB97X-D	LSDA	PBE-D3	MP2	B2PLYP	PBE0
acetaldehyde	0.957	0.203	0.203	0.067	0.049	0.067	0.039	0.112	0.412	0.207	0.032	0.018	0.122
acetic	1.431	0.322	0.567	0.227	0.038	0.093	0.024	0.091	0.375	0.324	0.030	0.016	0.077
c2h3f	0.698	0.165	0.292	0.109	0.059	0.084	0.072	0.107	0.263	0.169	0.024	0.037	0.094
c2h5f	0.391	0.277	0.495	0.192	0.036	0.045	0.023	0.017	0.331	0.287	0.033	0.024	0.035
ch2ch	0.191	0.150	0.118	0.042	0.085	0.092	0.063	0.083	0.402	0.152	0.866	0.058	0.074
ch2nh2	0.276	0.166	0.182	0.073	0.056	0.223	0.040	0.053	0.577	0.171	0.054	0.036	0.065
ch3nh	0.284	0.228	0.168	0.064	0.029	0.074	0.030	0.049	0.829	0.234	0.041	0.022	0.074
ch3nh2	0.384	0.188	0.225	0.080	0.035	0.116	0.039	0.042	0.411	0.196	0.030	0.025	0.058
ethanol	0.622	0.270	0.440	0.168	0.036	0.084	0.019	0.049	0.445	0.278	0.036	0.020	0.056
formic	1.448	0.265	0.428	0.174	0.038	0.084	0.029	0.099	0.269	0.278	0.009	0.009	0.088
methanol	0.555	0.207	0.311	0.126	0.018	0.081	0.017	0.043	0.319	0.221	0.016	0.011	0.050
propane	0.212	0.225	0.282	0.085	0.034	0.039	0.009	0.016	0.432	0.225	0.058	0.027	0.038
propene	0.468	0.193	0.156	0.055	0.076	0.082	0.074	0.115	0.461	0.195	0.048	0.048	0.109
propyne	0.991	0.171	0.090	0.045	0.138	0.135	0.136	0.200	0.499	0.173	0.047	0.051	0.170
mean	0.636	0.216	0.283	0.108	0.052	0.093	0.044	0.077	0.430	0.222	0.094	0.029	0.079
RMS	0.755	0.222	0.317	0.122	0.060	0.102	0.054	0.090	0.452	0.228	0.234	0.032	0.087

Table S1: *E_{geo} values of Figure 1a (top panel) in kcal/mol for individual molecules.*

	HF	PBE	BLYP	TPSS	B3LYP	MN15	ωB97	ωB97X-D	LSDA	PBE-D3	MP2	B2PLYP	PBE0
acetaldehyde	-1.055	-0.196	-0.196	-0.065	-0.050	-0.071	-0.042	-0.117	-0.393	-0.199	-0.032	-0.018	-0.126
acetic	-1.583	-0.311	-0.539	-0.219	-0.037	-0.095	-0.025	-0.094	-0.361	-0.314	-0.030	-0.016	-0.080
c2h3f	-0.761	-0.160	-0.274	-0.104	-0.058	-0.087	-0.075	-0.110	-0.260	-0.164	-0.024	-0.037	-0.097
c2h5f	-0.419	-0.267	-0.462	-0.182	-0.034	-0.048	-0.023	-0.017	-0.319	-0.277	-0.034	-0.023	-0.035
ch2ch	-0.192	-0.157	-0.124	-0.057	-0.127	-0.137	-0.105	-0.133	-0.443	-0.159	-1.224	-0.100	-0.118
ch2nh2	-0.297	-0.160	-0.174	-0.071	-0.055	-0.203	-0.039	-0.053	-0.548	-0.165	-0.053	-0.036	-0.064
ch3nh	-0.308	-0.213	-0.159	-0.061	-0.027	-0.074	-0.032	-0.050	-0.749	-0.219	-0.042	-0.022	-0.073
ch3nh2	-0.426	-0.183	-0.217	-0.078	-0.034	-0.109	-0.039	-0.042	-0.378	-0.190	-0.031	-0.025	-0.057
ethanol	-0.685	-0.263	-0.419	-0.162	-0.036	-0.084	-0.020	-0.050	-0.429	-0.271	-0.037	-0.020	-0.056
formic	-1.603	-0.256	-0.407	-0.167	-0.038	-0.086	-0.030	-0.103	-0.258	-0.269	-0.009	-0.009	-0.090
methanol	-0.613	-0.201	-0.295	-0.121	-0.018	-0.080	-0.018	-0.044	-0.304	-0.215	-0.017	-0.011	-0.050
propane	-0.226	-0.221	-0.274	-0.082	-0.034	-0.040	-0.009	-0.017	-0.415	-0.221	-0.059	-0.027	-0.038
propene	-0.511	-0.187	-0.152	-0.054	-0.077	-0.085	-0.079	-0.119	-0.448	-0.190	-0.048	-0.052	-0.112
propyne	-1.093	-0.164	-0.086	-0.044	-0.141	-0.139	-0.145	-0.208	-0.484	-0.165	-0.046	-0.052	-0.175
mean	-0.698	-0.210	-0.270	-0.105	-0.055	-0.095	-0.049	-0.083	-0.413	-0.216	-0.120	-0.032	-0.084
RMS	0.832	0.215	0.301	0.118	0.065	0.104	0.061	0.097	0.431	0.221	0.329	0.039	0.092

Table S2: *E'_{geo} values of Figure 1a (top panel) in kcal/mol for individual molecules.*

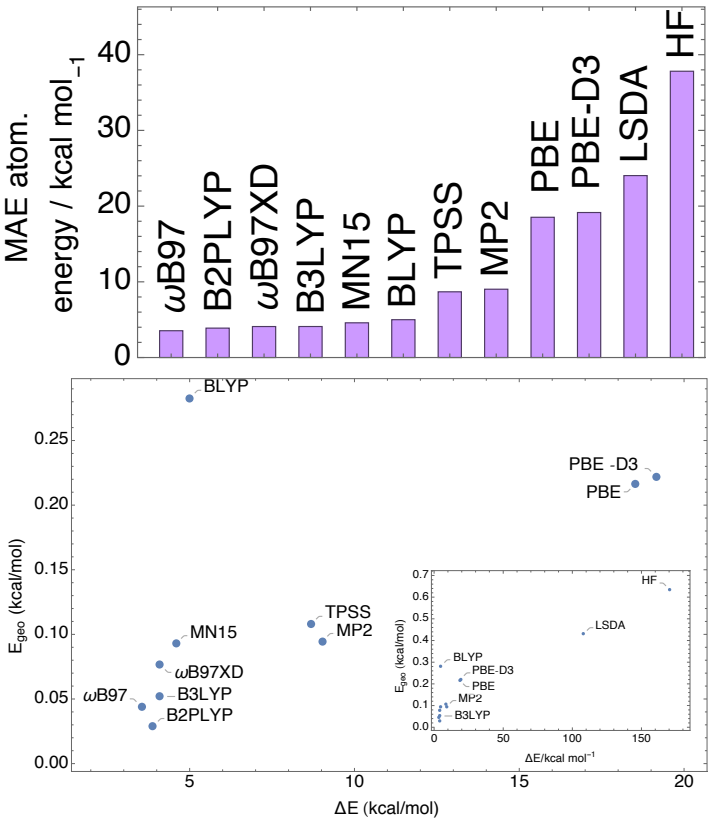


Figure S2: *Top panel: MAEs for the atomization energies for the set of molecules (listed in Tables S2 and S1) [Fig. 1a (top panel)] obtained from different methods. The LSDA and HF MAEs are divided by the factor of* 4.5*. Bottom panel: MAEs from the top panel vs. average* E_{geo} *errors for the same set of molecules.*

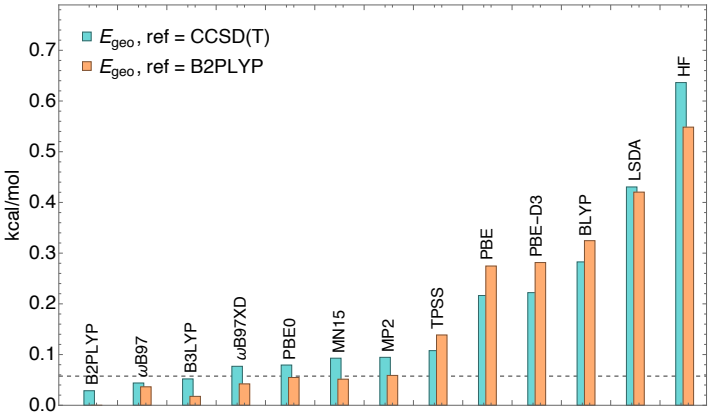


Figure S3: *Averaged GEOs with B2PLYP as a proxy reference vs. GEOs with CCSD(T) as a reference (same as in the top panel of Fig. 1a). B2PLYP is not a reliable reference below the horizontal dashed line which goes through twice the GEO of B2PLYP.*

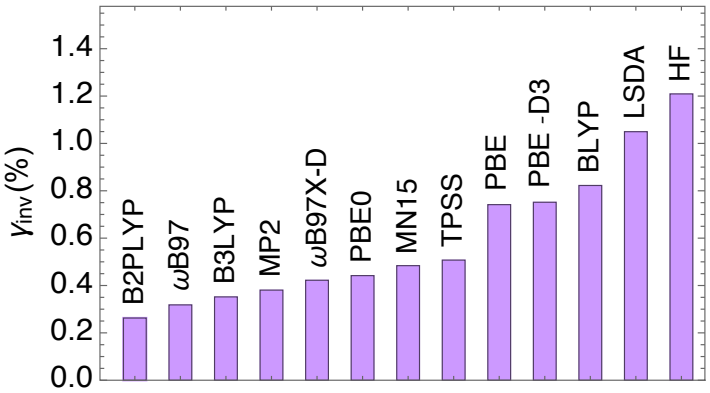


Figure S4: *The* inverted γ, $\gamma_{inv} = \sqrt{\frac{2E_{geo}}{D}}$, *averaged over the molecular dataset considered in 1a (top panel). The underlying* E_{geo} *values are given in Table S1, and the* D *values are given in Table S6.*

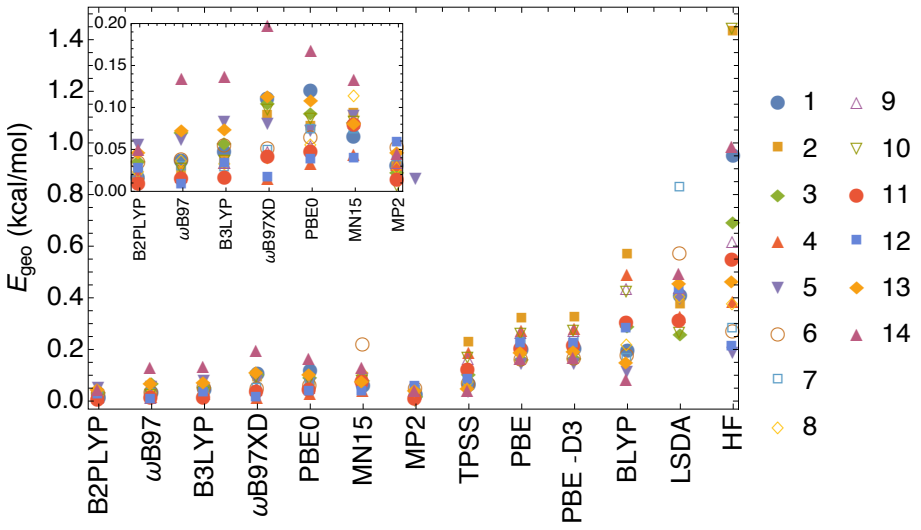


Figure S5: *Plots showing data of Table S1. The order of molecules is the same as in Table S1.*

S3. Additional details on the results in Figure 1a (bottom panel)

The dataset of molecules used in Figure 1a (bottom panel) is shown in Figure S6. The PBEh-3c and HF-3c calculations have been performed with the ORCA package[3]. The GFN1-xTB and DFTB3-D3 calculations have been performed with the ADF package[4]. All other calculations have been performed with the G16 package. We used aug-cc-pVTZ[5] basis set for all DFT and HF calculations, except for PBEh-3c and HF-3c, which come with their own basis set. For DFTB3-D3 calculations, we have used the set of parameters of Ref. [6] in tandem with the D3 correction with the Becke-Johnson damping function[7] . The set of parameters for PM6 and PM7 are those implemented in G16 when the keywords *"PM6"* and *"PM7"*, respectively, are used. The E_{geo} errors for individual molecules (Figure S6) and employed approximate methods are shown in Figure S7, and the underlying data values are given in Table S4.

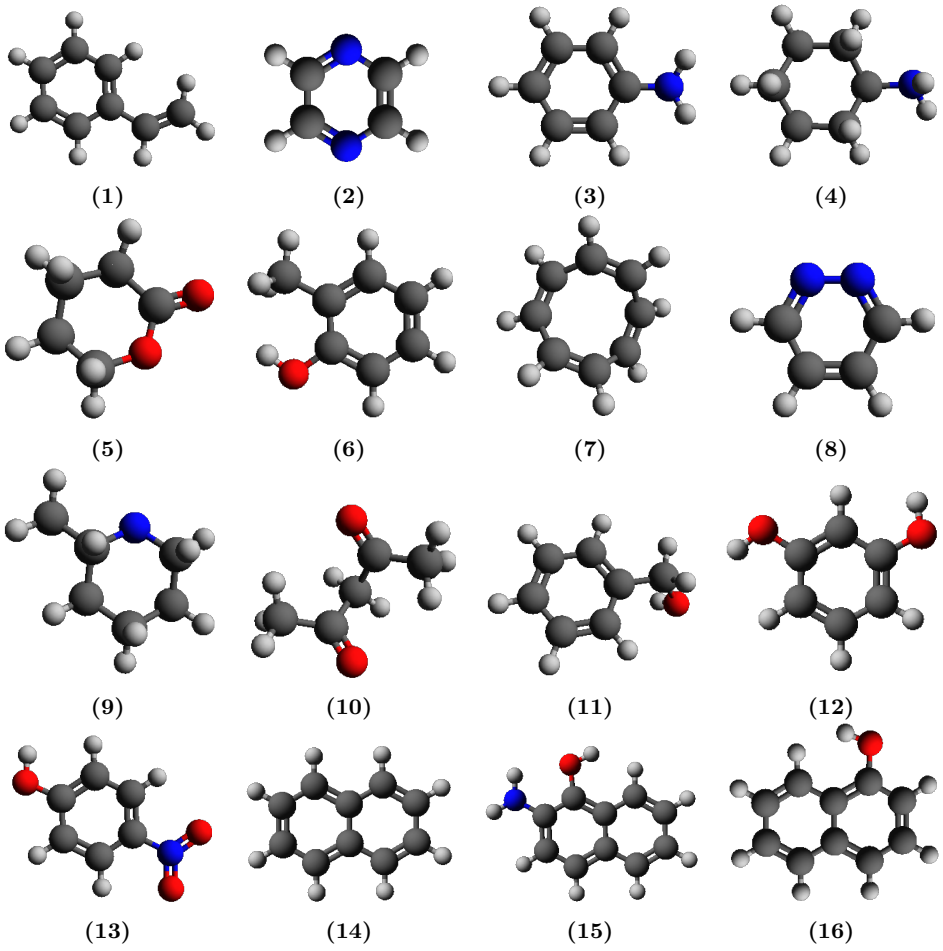


Figure S6: *Molecules used in error plots of Figure 1a (bottom panel).*

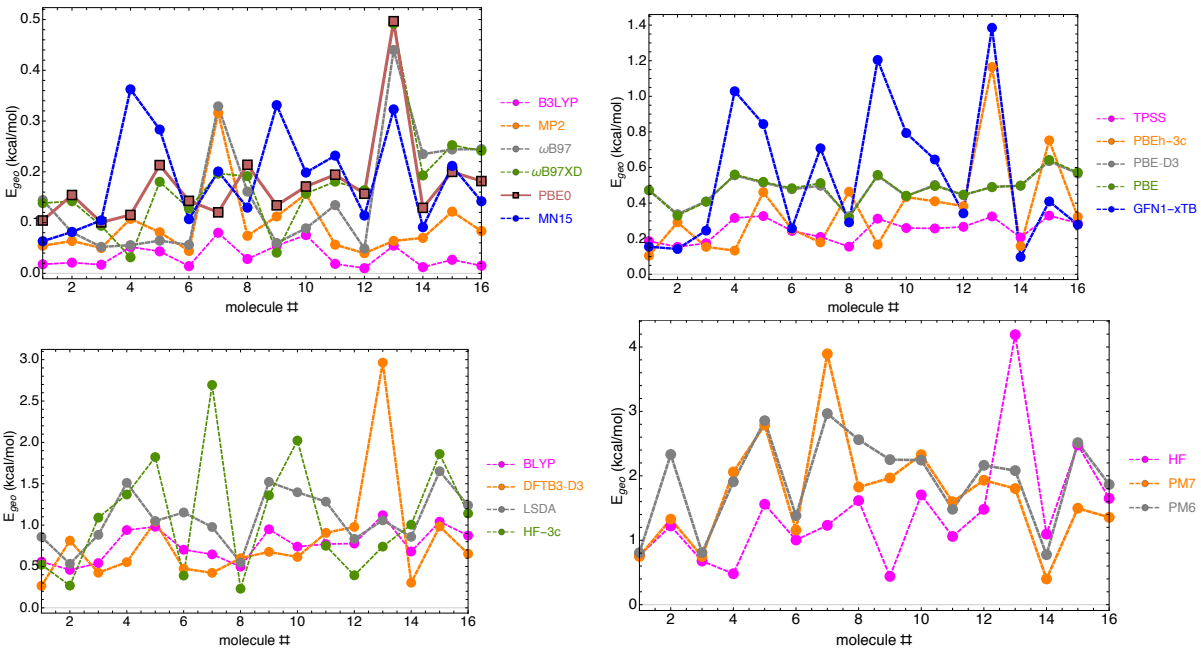


Figure S7: *Plots showing E_{geo} errors for different approximate methods for the set of molecules shown in Figure S6. Averaged error bars are shown in Figure 1a (bottom panel). Note the difference in the y-axes scales.*

	ωB97	B3LYP	ωB97X-D	MN15	MP2	TPSS	PBE	PBE-D3	BLYP	LSDA	HF	DFTB3-D3	GFN1-xTB	HF-3c	PBEh-3c	PM6	PM7	PBE0
1	0.144	0.018	0.138	0.063	0.054	0.186	0.472	0.473	0.557	0.858	0.763	0.266	0.155	0.521	0.105	0.806	0.749	0.104
2	0.081	0.021	0.142	0.082	0.064	0.154	0.330	0.338	0.459	0.532	1.222	0.812	0.142	0.269	0.292	2.332	1.327	0.154
3	0.052	0.017	0.094	0.104	0.050	0.175	0.407	0.410	0.541	0.882	0.677	0.424	0.245	1.090	0.154	0.808	0.747	0.100
4	0.055	0.052	0.032	0.363	0.107	0.317	0.561	0.556	0.940	1.511	0.480	0.552	1.028	1.370	0.133	1.908	2.059	0.115
5	0.064	0.043	0.180	0.283	0.081	0.327	0.520	0.514	0.980	1.050	1.559	1.042	0.844	1.823	0.461	2.856	2.788	0.213
6	0.056	0.014	0.128	0.107	0.044	0.243	0.484	0.477	0.703	1.153	1.004	0.474	0.258	0.390	0.253	1.381	1.158	0.143
7	0.329	0.080	0.196	0.201	0.316	0.210	0.512	0.497	0.646	0.977	1.232	0.423	0.708	2.694	0.179	2.966	3.893	0.120
8	0.161	0.028	0.192	0.129	0.074	0.155	0.316	0.326	0.498	0.548	1.616	0.600	0.292	0.231	0.464	2.560	1.825	0.214
9	0.060	0.054	0.041	0.331	0.112	0.312	0.558	0.554	0.949	1.523	0.440	0.676	1.204	1.360	0.167	2.251	1.966	0.134
10	0.089	0.075	0.157	0.199	0.156	0.260	0.442	0.439	0.739	1.397	1.704	0.615	0.794	2.022	0.434	2.244	2.328	0.171
11	0.134	0.019	0.181	0.232	0.056	0.258	0.496	0.502	0.770	1.283	1.059	0.906	0.645	0.750	0.410	1.481	1.598	0.194
12	0.049	0.010	0.164	0.114	0.040	0.267	0.447	0.444	0.777	0.833	1.479	0.978	0.342	0.393	0.383	2.164	1.931	0.157
13	0.440	0.055	0.492	0.323	0.064	0.325	0.489	0.491	1.119	1.058	4.191	2.964	1.385	0.740	1.165	2.081	1.802	0.497
14	0.235	0.012	0.193	0.091	0.070	0.207	0.499	0.497	0.681	0.859	1.095	0.304	0.097	1.005	0.158	0.775	0.400	0.129
15	0.244	0.027	0.253	0.212	0.122	0.330	0.643	0.636	1.041	1.651	2.478	0.984	0.409	1.860	0.752	2.515	1.497	0.200
16	0.245	0.015	0.241	0.142	0.083	0.287	0.573	0.566	0.874	1.239	1.651	0.653	0.279	1.141	0.323	1.866	1.356	0.182
mean	0.152	0.034	0.176	0.186	0.093	0.251	0.484	0.483	0.767	1.085	1.416	0.792	0.552	1.104	0.365	1.937	1.714	0.177
RMS	0.190	0.040	0.203	0.208	0.114	0.258	0.491	0.489	0.792	1.132	1.665	0.998	0.676	1.303	0.450	2.054	1.900	0.198

Table S3: *E_{geo} values Figure 1a (bottom panel) for individual molecules and approximate methods, all values are in kcal/mol. The molecules are ordered in Figure S6*

S4. Details on the results of Figure 1b

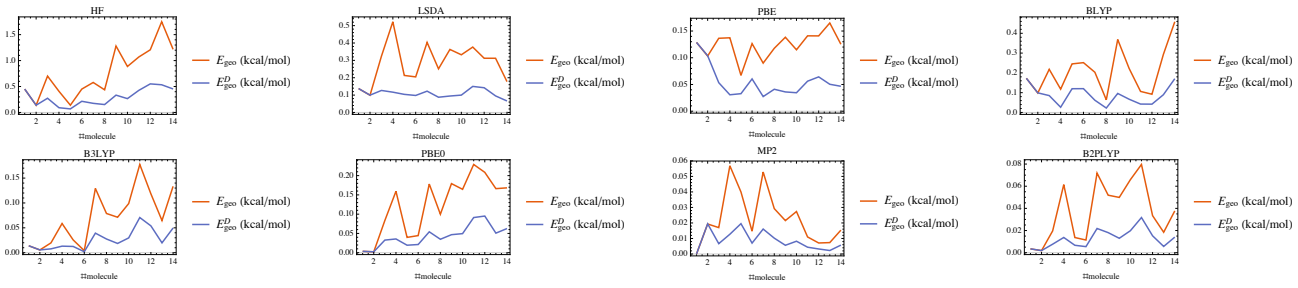


Figure S8: *Plots comparing E_{geo} and E^D_{geo} for selected methods and for a set of molecules shown in Fig. 1b.*

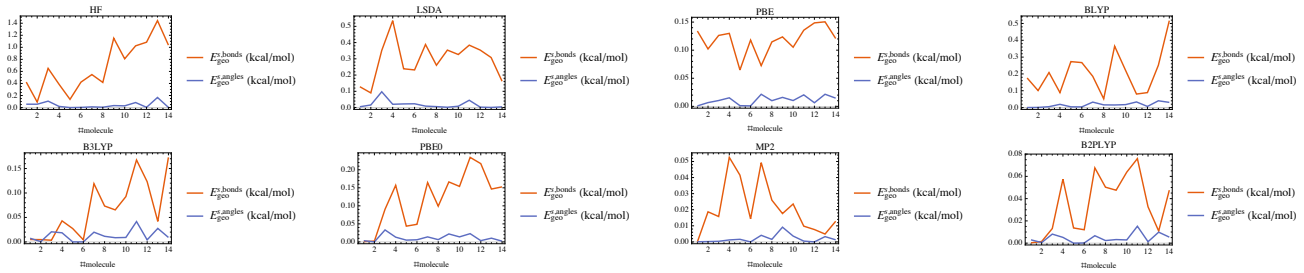


Figure S9: *Plots comparing total bond length and angle errors contribution to* E_{geo}^{simple} *for selected methods and for a set of molecules shown in Fig. 1b.*

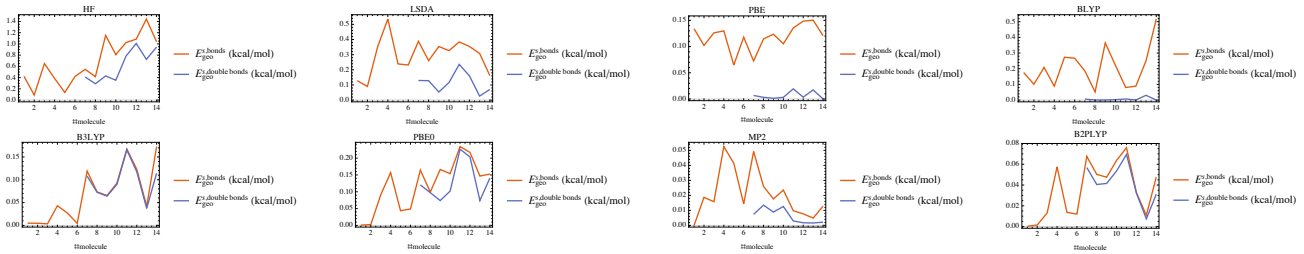


Figure S10: *Plots comparing total bond and double bond length errors contribution to* E_{geo}^{simple} *for selected methods and for a set of molecules shown in Fig. 1b.*

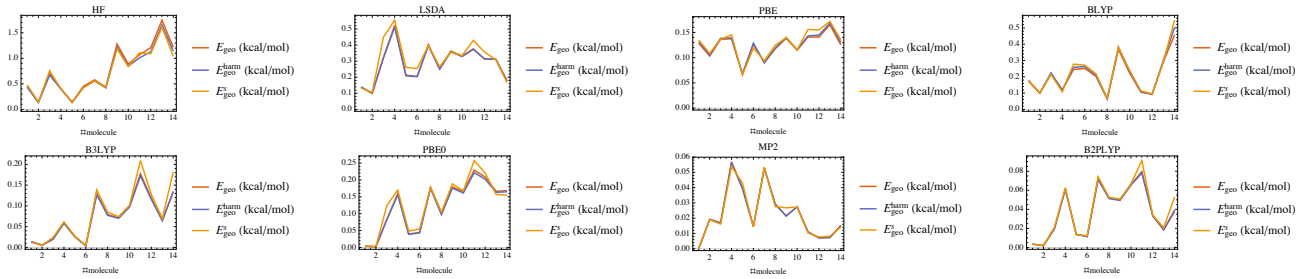


Figure S11: *Plots comparing* E_{geo}, E_{geo}^{harm} *and* E_{geo}^{simple} *for selected methods and for a set of molecules shown in Fig. 1(b).*

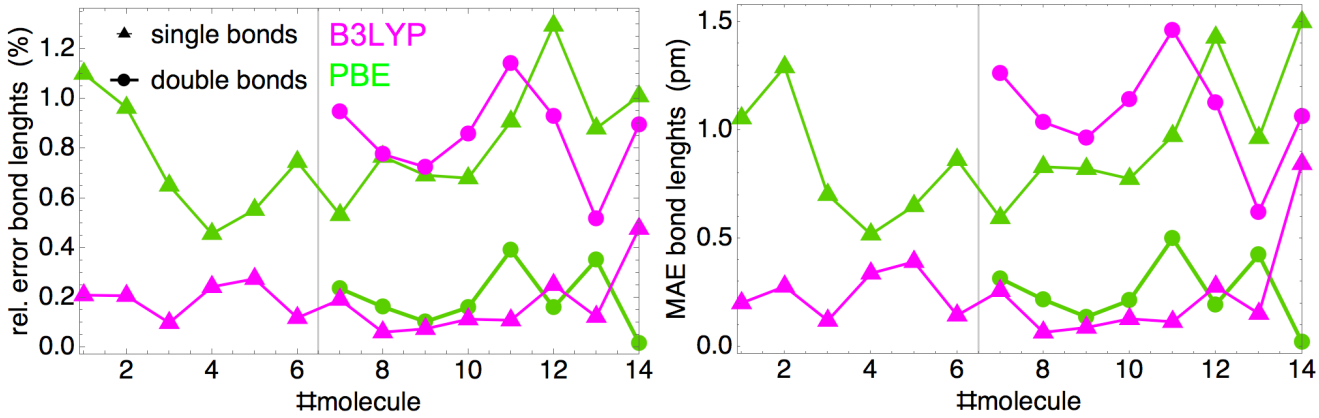


Figure S12: *PBE and B3LYP errors in bond lengths for a set of molecules shown in Fig. 1b. Left panel: relative errors; middle panel: mean absolute errors.*

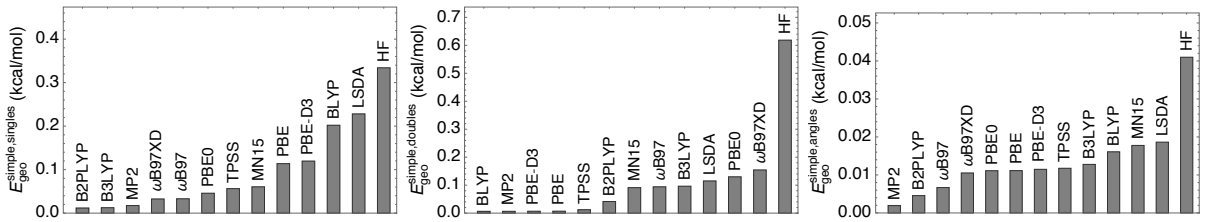


Figure S13: *Rankings of approximations for a set of molecules shown in Fig. 1b based on averaged-E_{geo}^{simple} (over a number of molecules) contributions from: single bonds (left panel), double bonds (middle panel), angles (right panel).*

S5. The GEO analysis for the water molecule

We use CCSD(T)/A'V6Z [1] as a reference method in all calculations performed in this section. The same basis set is employed in calculations involving approximate electronic structure methods and all calculations have been performed with the Gaussian16 package.

To show how GEO depends on errors in individual bond lengths and angles, we consider the water molecule in internal coordinates. It has three degrees of freedom, but to analyse the E_{geo} error we need just two: a and θ (see Figure S14) since the lengths of the two O−H bonds are the same within any approximation. In Figure S14, we also show the errors of different methods in the bond length, angle, together with the E_{geo} errors.

From a series of CCSD(T) calculations at fixed geometries, we can see how E_{geo} depends on Δa and $\Delta\theta$, the error in the bond length and bond angle, respectively. The resulting contour plot is shown in Figure S15a. From this plot, we can see that the E_{geo} error increases much more quickly with the error in the bond length than in the bond angle. To better understand this observation, we consider the harmonic approximation to E_{geo} (Eq. 3). For clarity, we write this equation again:

$$E_{geo} \approx E_{geo}^{\text{harm}} = \frac{1}{2}\Delta\tilde{\mathbf{G}}^{\intercal}\mathbf{H}\left(\mathbf{G}\right)\Delta\tilde{\mathbf{G}}. \tag{S5}$$

For the water molecule in internal coordinates, we can now write E_{geo}^{harm} as a function of:

$$\Delta\tilde{\mathbf{G}} = \left(2\Delta a, \Delta\theta\right). \tag{S6}$$

The underlying $\mathbf{H}\left(\mathbf{G}\right)$ matrix in internal coordinates is given by:

$$\mathbf{H}\left(\mathbf{G}\right) = \begin{pmatrix} k_{aa} & k_{a\theta} \\ k_{\theta a} & k_{\theta\theta} \end{pmatrix} \sim \begin{pmatrix} 0.122\text{pm}^{-2} & 0.006^{\circ-1}\text{pm}^{-1} \\ 0.006^{\circ-1}\text{pm}^{-1} & 0.031^{\circ-2} \end{pmatrix} \text{kcal/mol} \tag{S7}$$

where k-s are the force constants at the highly accurate reference [CCSD(T)] geometry. Plugging Eqs. S6 and S7 into Eq. S5, we obtain:

$$E_{geo}^{\text{harm}}\left(\Delta a, \Delta\theta\right) = k_{aa}\Delta a^2 + \frac{1}{2}k_{\theta\theta}\Delta\theta^2 + 2k_{a\theta}\Delta a\Delta\theta, \tag{S8}$$

where we also used the Hessian symmetry: $k_{a\theta} = k_{\theta a}$. In Eq. S7, the units were chosen such that Δa errors in pm and $\Delta\tilde{\theta}$ errors in degrees yield E_{geo}^{harm} in kcal/mol. In Figure S15b, we show the E_{geo}^{harm} contour plot as a function of Δa and $\Delta\tilde{\theta}$. We can see that it is in a very close agreement with its *exact* counterpart (Figure S15a). A further simplification of E_{geo}^{harm} (Eq. S8) obtained by setting $k_{a\theta}$ to 0:

$$E_{geo}^{\text{harm}}\left(\Delta a, \Delta\theta\right) \approx E_{geo}^{\text{simple}}\left(\Delta a, \Delta\theta\right) = k_{aa}\Delta a^2 + \frac{1}{2}k_{\theta\theta}\Delta\theta^2, \tag{S9}$$

gives a sensible approximation to $\Delta E_g\left(\Delta a, \Delta\theta\right)$, as it can be seen from Figure S15c. In Figure S15d, we compare the size of the three errors from different methods and we can see that both E_{geo}^{harm} and E_{geo}^{simple} are in a close agreement with E_{geo}. Thus, both Eq. S8 and S9 give us a simple yet very accurate representation of the GEO for the water molecule. From Eq. S8 and values of the force constants (Eq. S7), we can see that an error of 2pm in the bond length results in the GEO of ∼0.5kcal/mol.

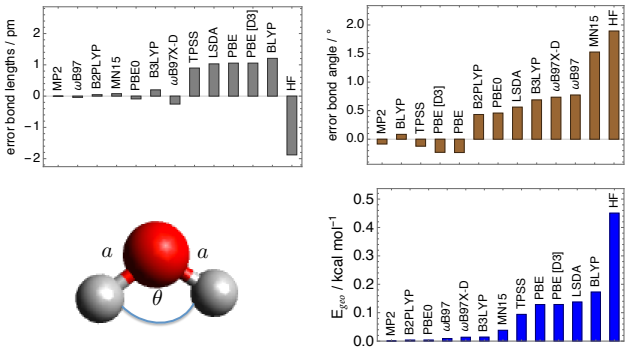


Figure S14: *The error in the bond length, bond angle, and the E_{geo} errors for the water molecule obtained from various approximations.*

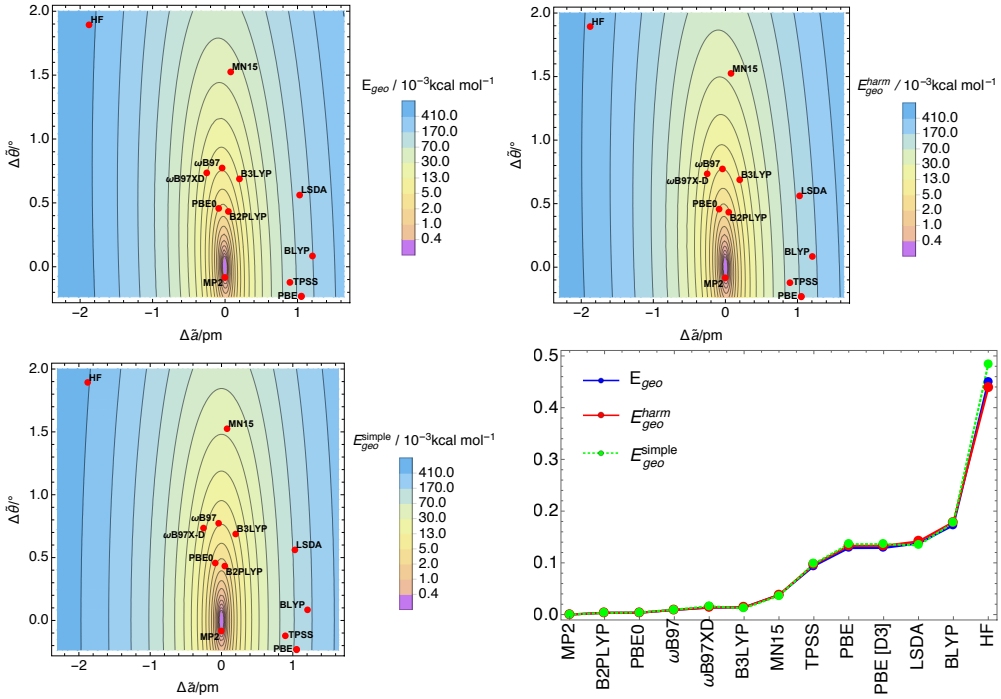


Figure S15: *(a) Contour plot of GEO as a function of errors in the bond lengths for the water molecule (also shown in the top panel of Fig. 1a); (b) same plot from the harmonic approximation (Eq. S8); (c) same plot from the* simple *quadratic approximation (Eq. S9); (d) Comparison of the* E_{geo} , E_{geo}^{harm}, *and* E_{geo}^{simple} *obtained from different methods.*

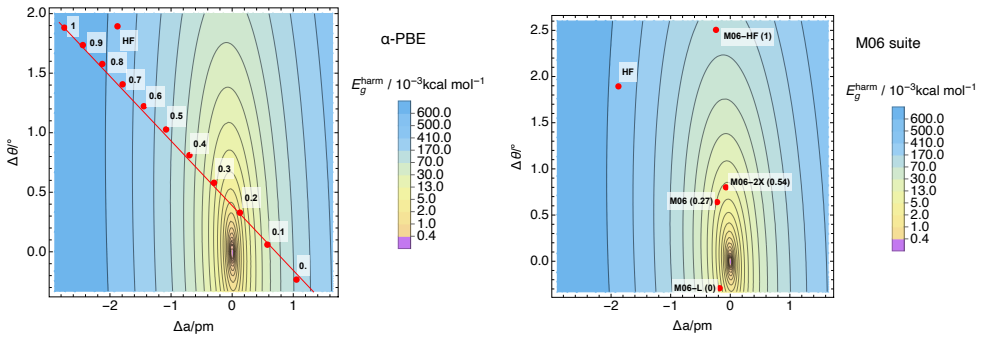


Figure S16: *The contour* E_{geo}^{harm} *plot for the water molecule (same as in Fig. S15) with the positions of different approximations. Left panel: The positions of approximations at different* α *values obtained from the* α*-PBE exchange-correlation functional defined as:* $E_{\text{XC}}^{\alpha} = \alpha \left(E_{\text{X}}^{\text{HF}} - E_{\text{X}}^{\text{PBE}}\right) + E_{\text{C}}^{\text{PBE}}$*, where* E_{X}^{HF} *is the exact exchange, and* $E_{\text{X}}^{\text{PBE}}$ *and* $E_{\text{c}}^{\text{PBE}}$ *are the PBE exchange and correlation, respectively [8]. Right panel: The position of approximations from the M06 suite [9, 10, 11], with the amount of the exact exchange for each of functional shown in the brackets. While (almost) all* α*-PBE functionals lie on the same line, this is not the case with the functionals from the M06 suite, since the parameters of their semilocal part vary across the M06 suite.*

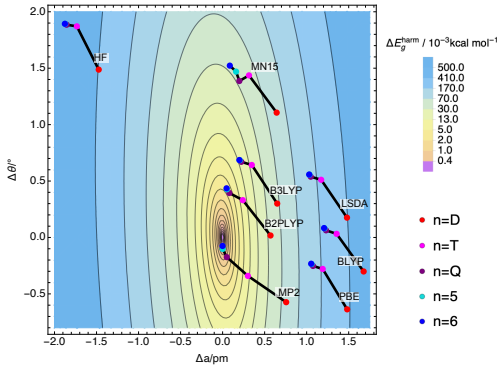


Figure S17: *The contour E_{geo}^{harm} plot for the water molecule (same as in Fig. S15) with the positions of different approximations with the aug-cc-pVnZ basis set, where $n = \{D,\ T,\ Q,\ 5,\ 6\ \}$.*

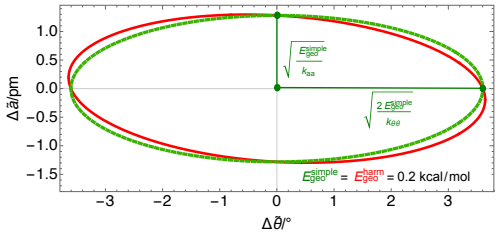


Figure S18: *The ellipses satisfying Eqs. S10 and S11 for the water molecule.*

On the other hand, because of the smaller size of $k_{\theta\theta}$, we can see that the GEO for the water molecule is much more forgiving of the bond angle error, since an error of 2°, results in the GEO of only ∼0.06kcal/mol. Interestingly, Eq. S9 can be rearranged to be in a form of an ellipse equation:

$$\frac{k_{aa}}{E_{geo}^{\mathrm{harm}}}\Delta a^2 + \frac{1}{2E_{geo}^{\mathrm{harm}}}k_{\theta\theta}\Delta\theta^2 + 2\frac{k_{a\theta}}{E_{geo}^{\mathrm{harm}}}\Delta a\Delta\theta = 1. \tag{S10}$$

Similarly, rearranging Eq. S9, we obtain an axis-aligned ellipse equation:

$$\frac{k_{aa}}{E_{geo}^{\mathrm{simple}}}\Delta a^2 + \frac{k_{\theta\theta}}{2E_{geo}^{\mathrm{simple}}}\Delta\theta^2 = 1. \tag{S11}$$

Since $k_{a\theta}$ is small, the two ellipses of Eq. S10 and S11 are in a close agreement, as it can be seen from Figure S18.

S6. GEO FOR SMALL MOLECULES IN INTERNAL COORDINATES

In this section we show detailed GEO analysis for a set of small molecules. For all molecules, CCSD(T) has been used as a reference and all calculation have been obtained with the G16 package. The following basis set have ben used: A'V6Z for the water molecule (Figure S20), aug-cc-pVQZ for H_2S (Figure S21) and ethene (Figure S23), and aug-cc-pVTZ for H_2CO (Figure S22), methanol (Figure S24), cyclopropane(Figure S24).

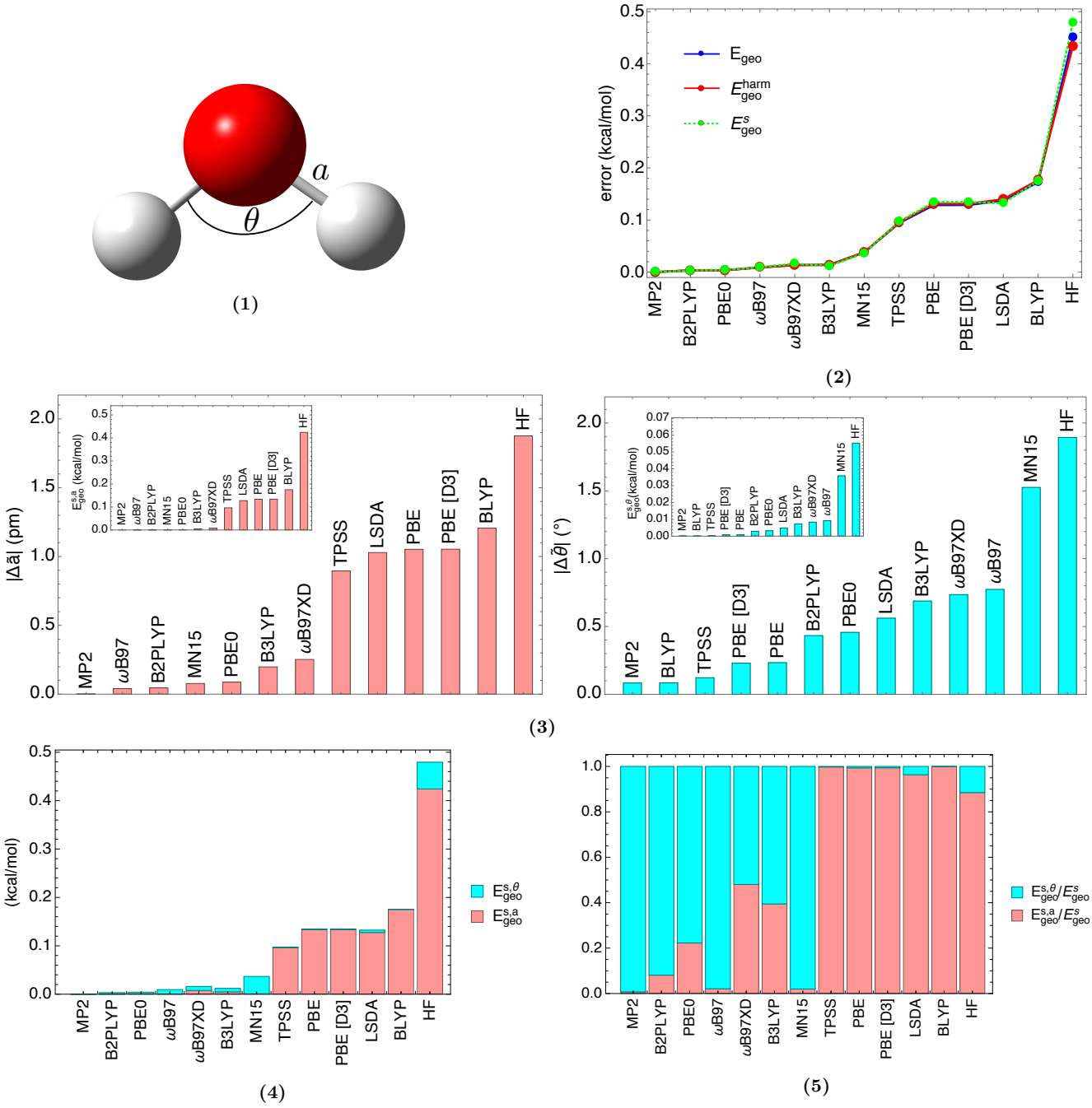


Figure S20: *The GEO analysis for the water molecule. Panel (a): considered degrees of freedom; (b): Comparison of GEO with its harm and simple (labeled as "s" in the plot for brevity) approximations obtained from various methods; (c): errors in specific i-th degree of freedom specified in (a), with insets showing the resulting* $E_{geo}^{simple,i}$ *(see Eq. 5); (d) different* $E_{geo}^{simple,i}$ *contributions to the* total E_{geo}^{simple}*; (e) the resulting* weights, *i.e.* $E_{geo}^{simple,i}/E_{geo}^{\text{simple}}$.

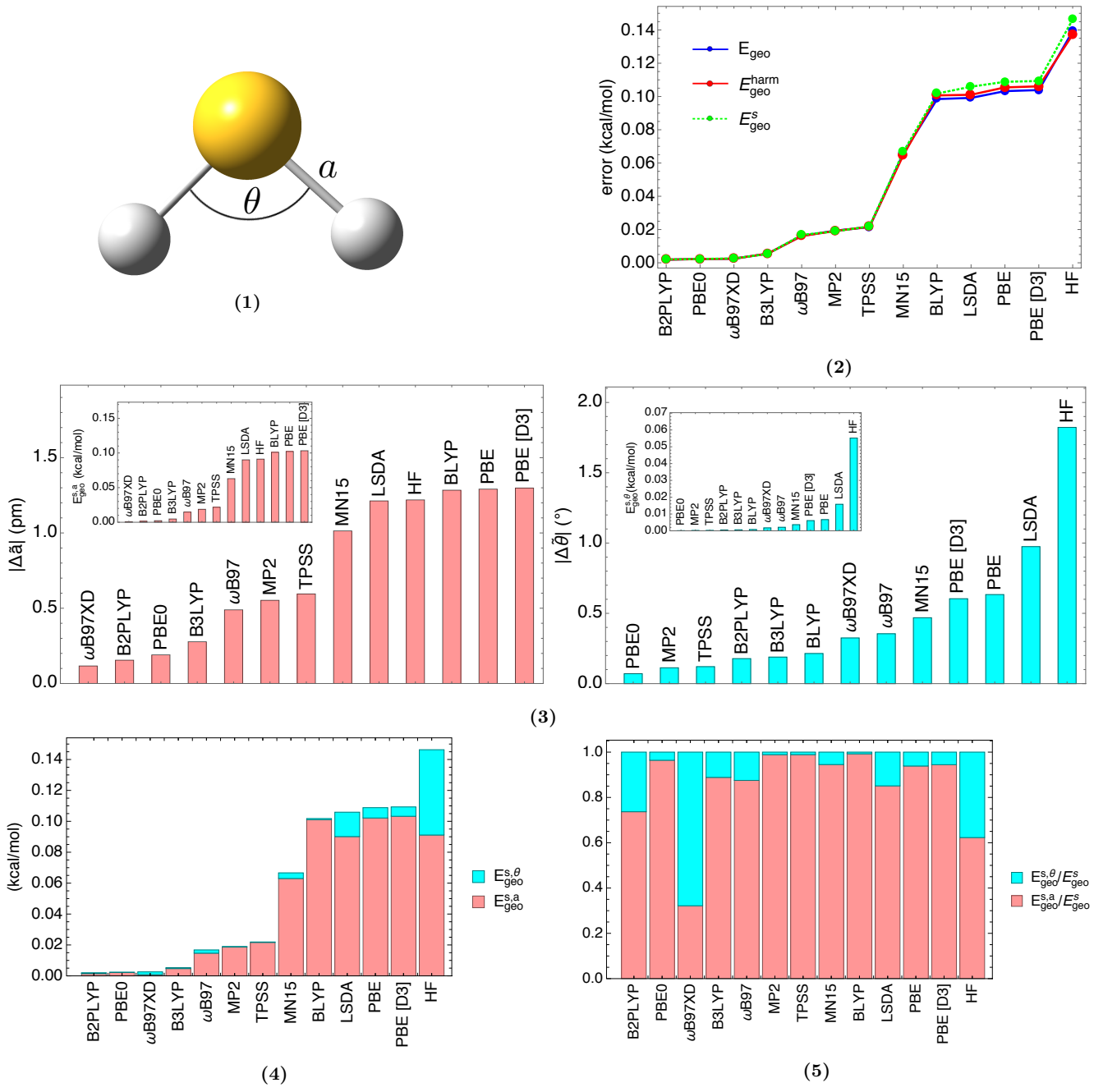


Figure S21: *The same plots as in Figure S20, but for the H_2S molecule.*

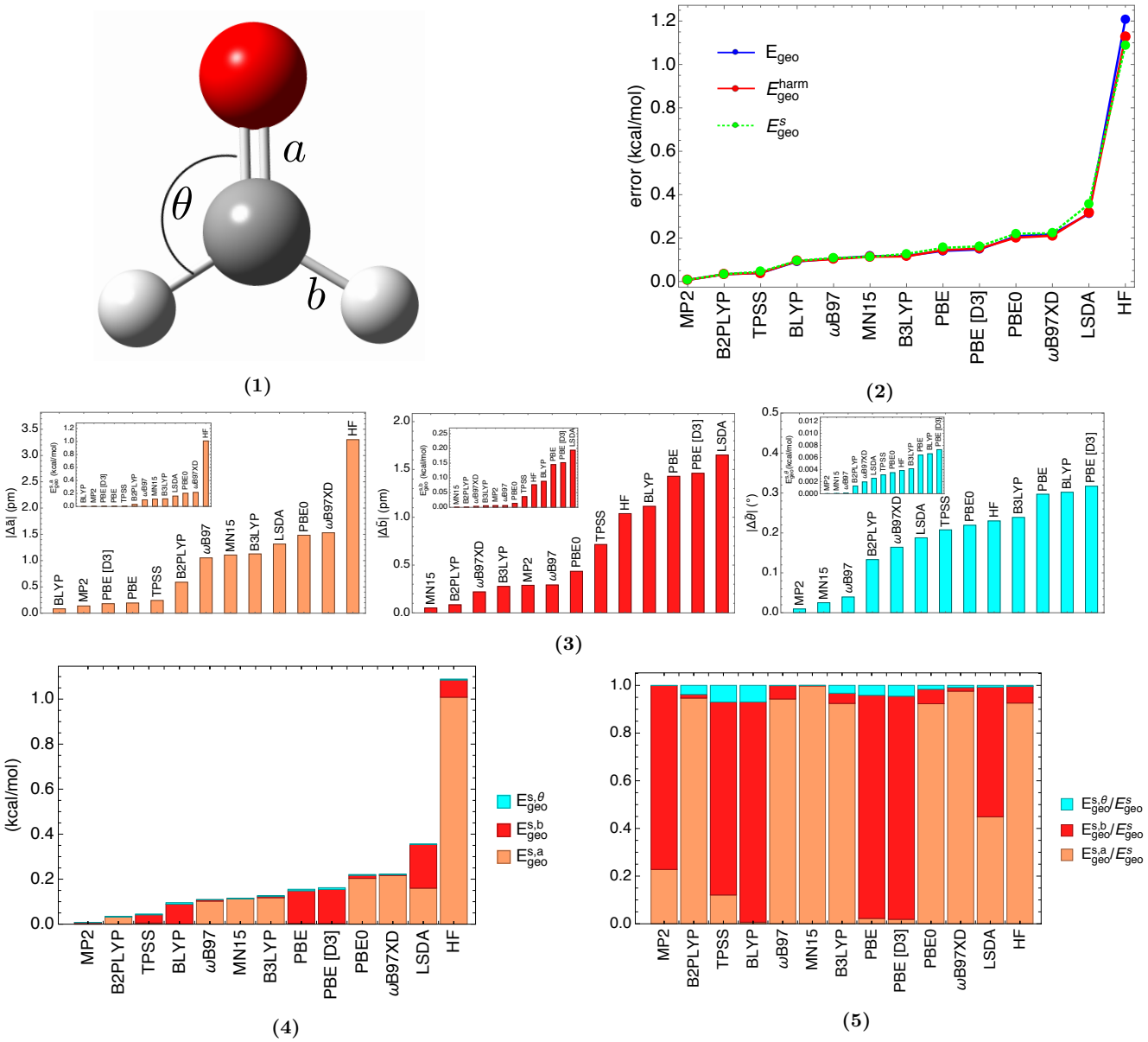


Figure S22: *The same plots as in Figure S20, but for the H_2CO molecule.*

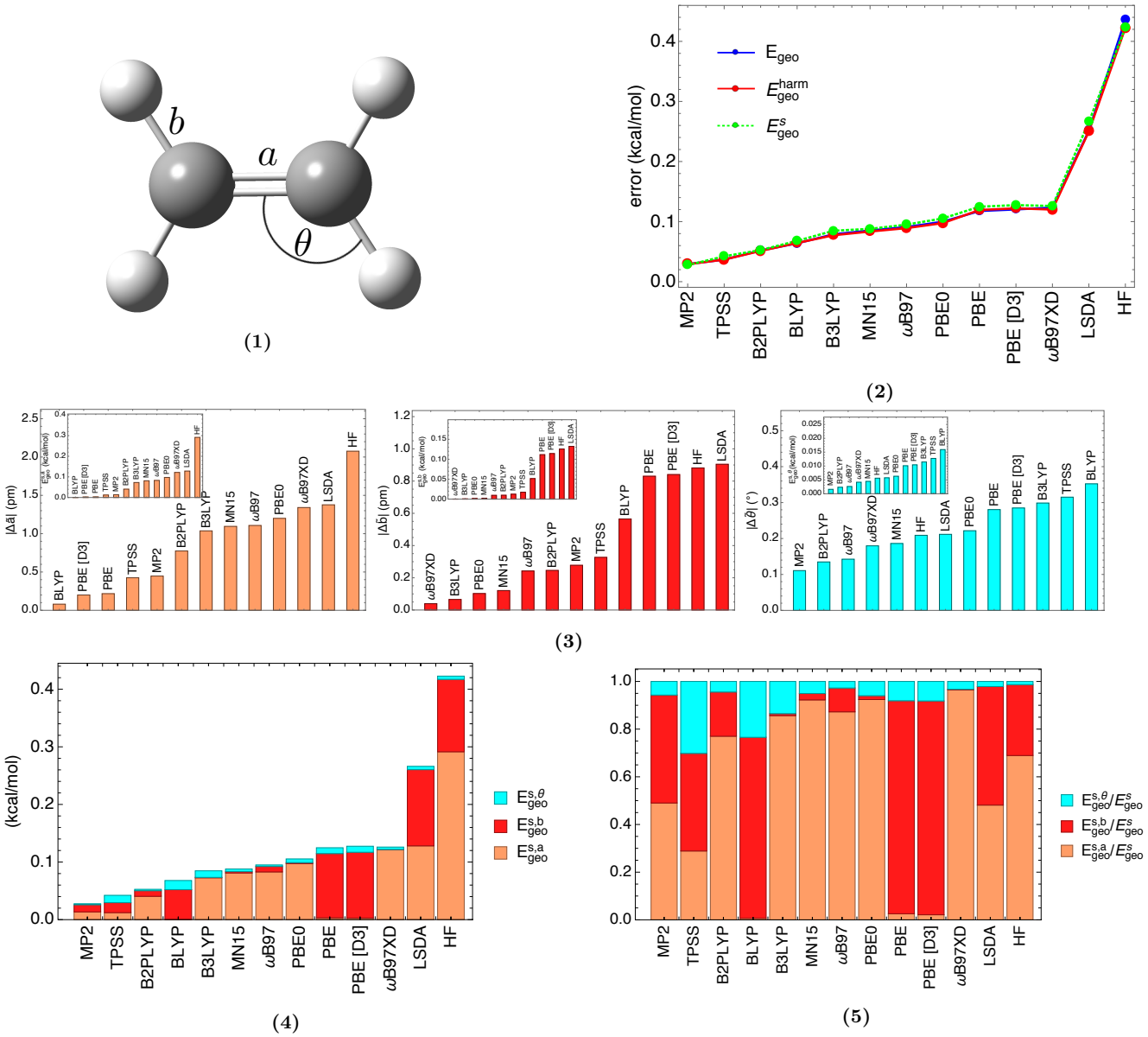


Figure S23: *The same plots as in Figure S20, but for the ethene molecule.*

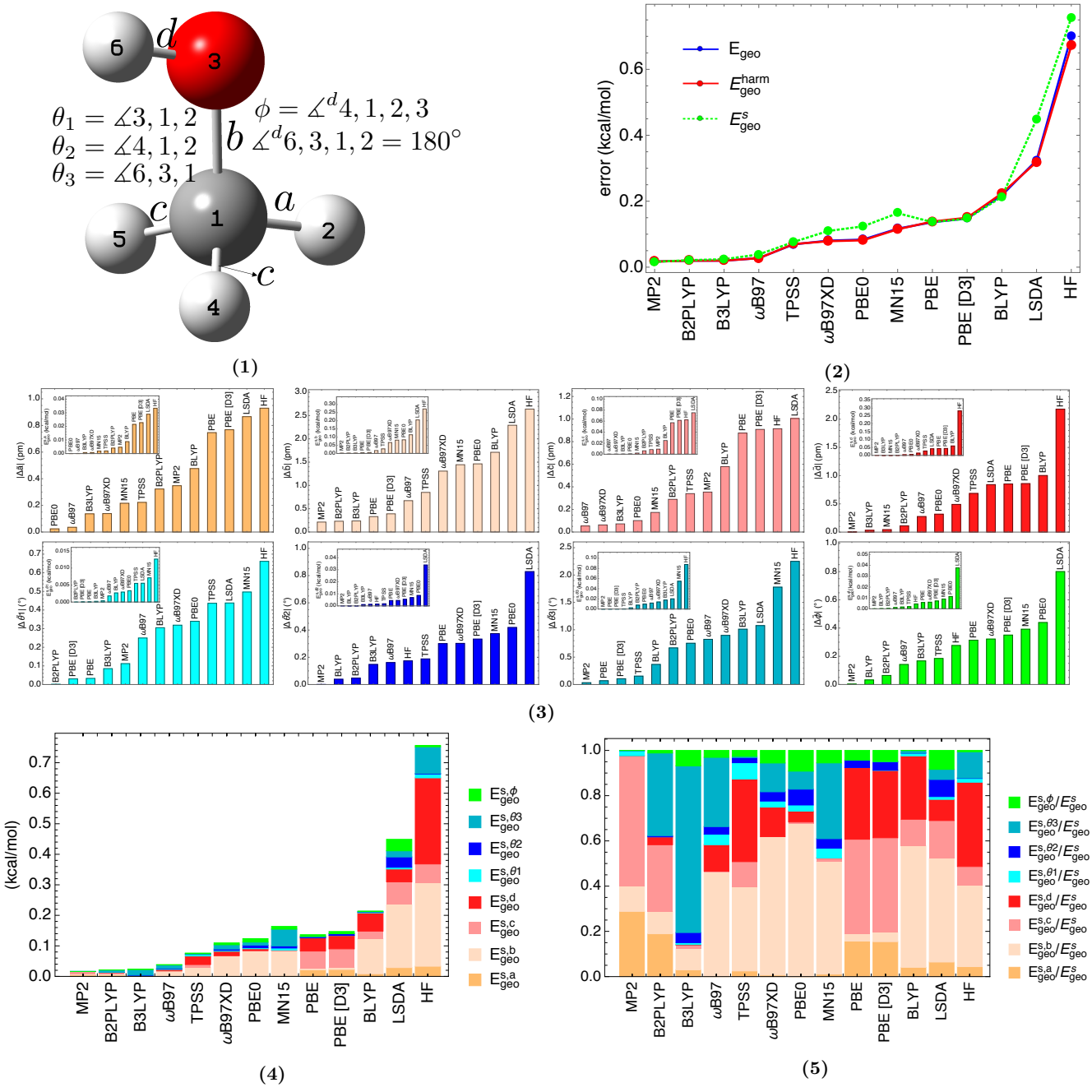


Figure S24: *The same plots as in Figure S20, but for the methanol molecule.*

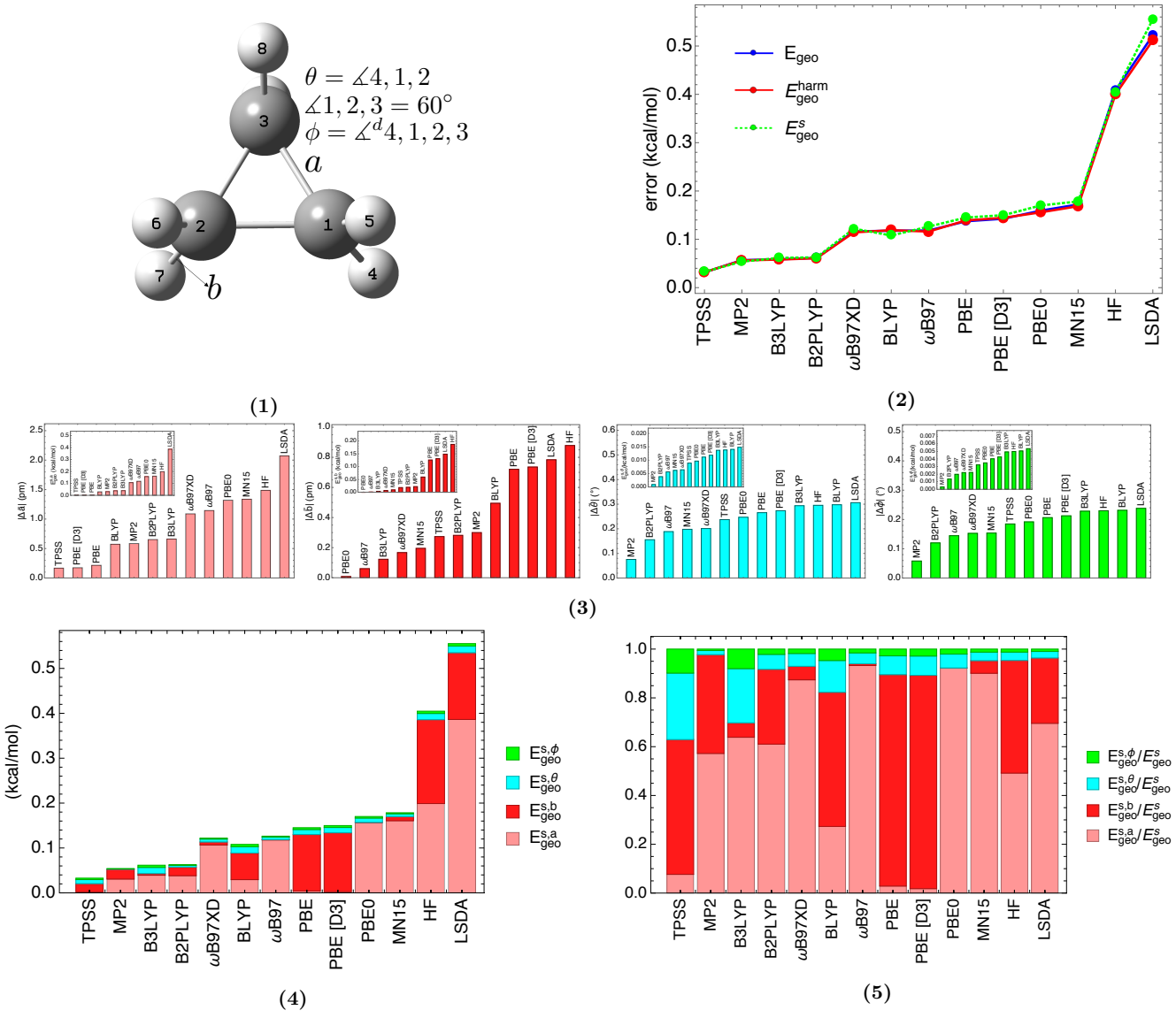
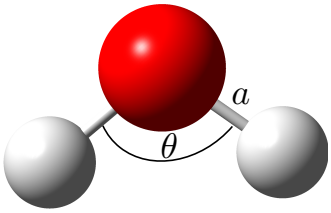
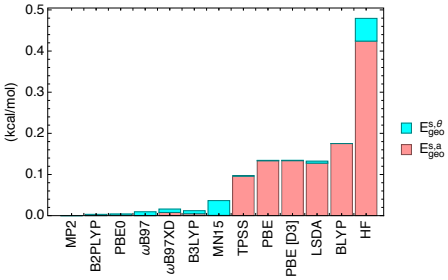
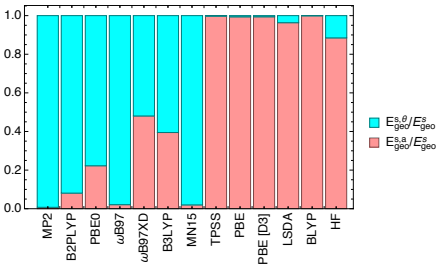


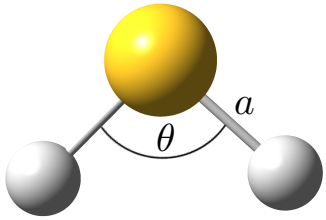
Figure S25: *The same plots as in Figure S20, but for the cyclopropane molecule.*

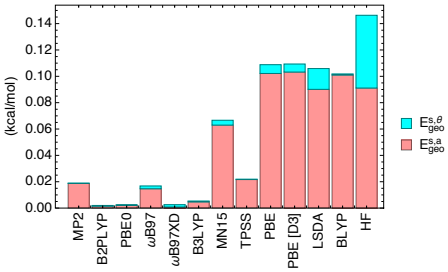

a
θ

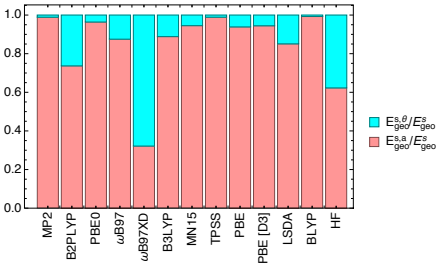

(kcal/mol)
MP2
B2PLYP
PBE0
ωB97
ωB97XD
B3LYP
MN15
TPSS
PBE
PBE [D3]
LSDA
BLYP
HF


MP2
B2PLYP
PBE0
ωB97
ωB97XD
B3LYP
MN15
TPSS
PBE
PBE [D3]
LSDA
BLYP
HF

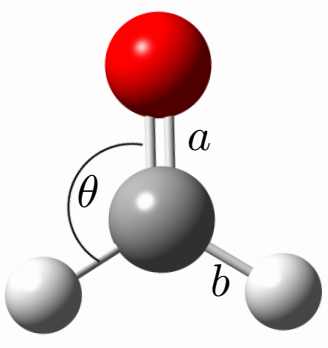
(1)

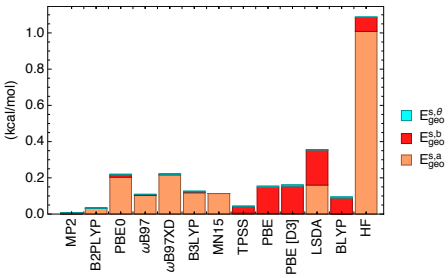

a
θ

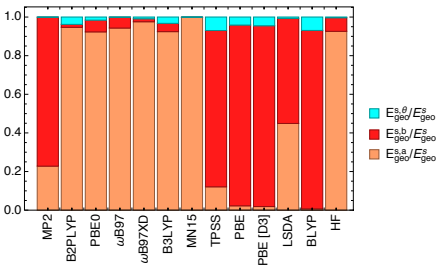

(kcal/mol)
MP2
B2PLYP
PBE0
ωB97
ωB97XD
B3LYP
MN15
TPSS
PBE
PBE [D3]
LSDA
BLYP
HF


MP2
B2PLYP
PBE0
ωB97
ωB97XD
B3LYP
MN15
TPSS
PBE
PBE [D3]
LSDA
BLYP
HF

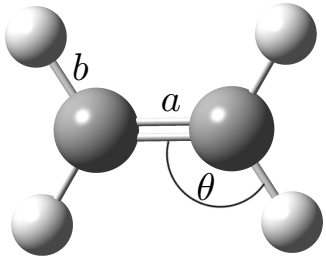
(2)

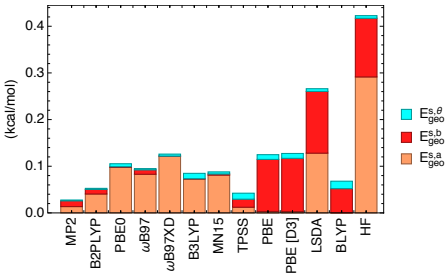

a
θ
b

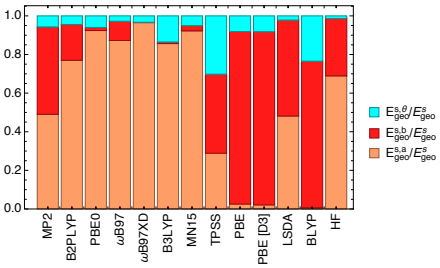

(kcal/mol)
MP2
B2PLYP
PBE0
ωB97
ωB97XD
B3LYP
MN15
TPSS
PBE
PBE [D3]
LSDA
BLYP
HF


MP2
B2PLYP
PBE0
ωB97
ωB97XD
B3LYP
MN15
TPSS
PBE
PBE [D3]
LSDA
BLYP
HF

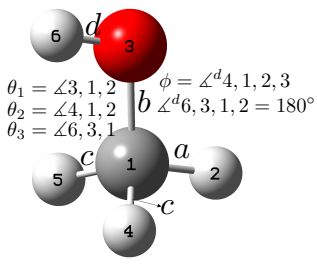
(3)

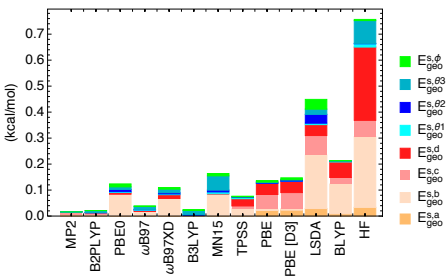

b
a
θ

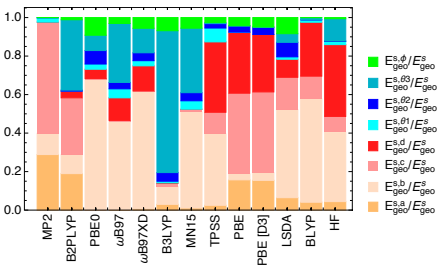

(kcal/mol)
MP2
B2PLYP
PBE0
ωB97
ωB97XD
B3LYP
MN15
TPSS
PBE
PBE [D3]
LSDA
BLYP
HF


MP2
B2PLYP
PBE0
ωB97
ωB97XD
B3LYP
MN15
TPSS
PBE
PBE [D3]
LSDA
BLYP
HF

(4)


θ1 = ∠3, 1, 2
θ2 = ∠4, 1, 2
θ3 = ∠6, 3, 1
ϕ = ∠d4, 1, 2, 3
∠d6, 3, 1, 2 = 180°
d
b
c
a
c
6
3
5
1
2
4


(kcal/mol)
MP2
B2PLYP
PBE0
ωB97
ωB97XD
B3LYP
MN15
TPSS
PBE
PBE [D3]
LSDA
BLYP
HF


MP2
B2PLYP
PBE0
ωB97
ωB97XD
B3LYP
MN15
TPSS
PBE
PBE [D3]
LSDA
BLYP
HF

(5)

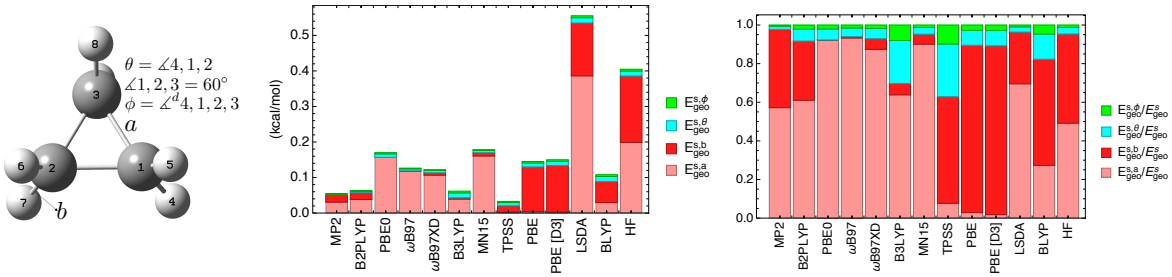
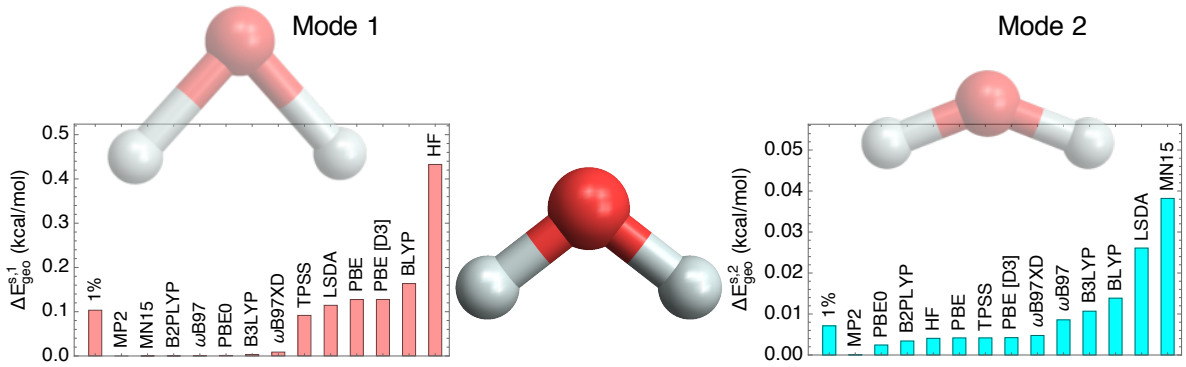


Figure S27: *Summary of Figs. S20-S25.*

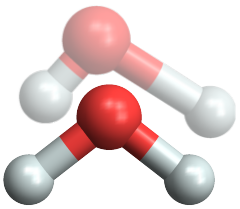
	HF	PBE	BLYP	TPSS	B3LYP	MN15	ωB97	ωB97X-D	LSDA	PBE-D3	MP2	B2PLYP	PBE0
H_2O	0.450	0.129	0.173	0.094	0.014	0.038	0.009	0.013	0.138	0.129	0.000	0.003	0.004
H_2S	0.139	0.103	0.098	0.021	0.006	0.066	0.016	0.002	0.099	0.104	0.019	0.002	0.002
H_2CO	1.208	0.141	0.092	0.039	0.118	0.117	0.105	0.218	0.312	0.147	0.007	0.033	0.209
ethene	0.436	0.118	0.064	0.037	0.079	0.085	0.091	0.123	0.251	0.120	0.029	0.052	0.100
methanol	0.701	0.136	0.218	0.069	0.020	0.117	0.027	0.080	0.324	0.148	0.017	0.020	0.084
cyclopropane	0.408	0.137	0.118	0.032	0.059	0.172	0.118	0.116	0.522	0.143	0.057	0.061	0.159
mean	0.557	0.127	0.127	0.049	0.049	0.099	0.061	0.092	0.274	0.132	0.022	0.029	0.093
RMS	0.649	0.128	0.137	0.055	0.064	0.108	0.076	0.118	0.307	0.133	0.028	0.036	0.120

Table S4: *The E_{geo} values for molecules in Figs. S20- S25 in kcal/mol. CCSD(T) has been used as a reference in all calculations. The following basis set have ben used: A'V6Z for the water molecule aug-cc-pVQZ for H_2S and ethene, and aug-cc-pVTZ for H_2CO, methanol, cyclopropane.*

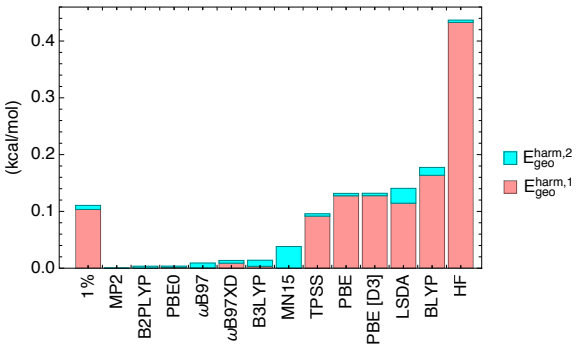
S7. GEO analysis in normal modes (Hessian eigenvectors)



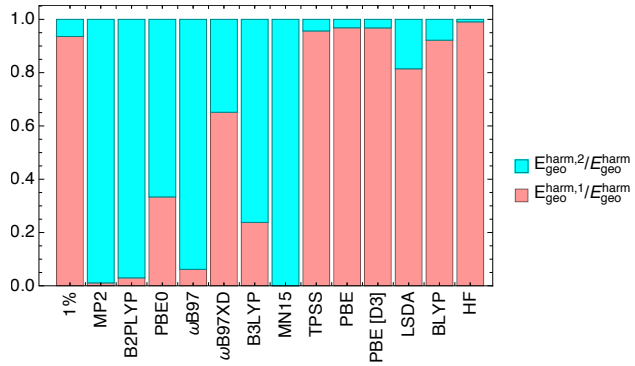
(1) *modes that do have contributions to GEO*



(2) *modes that do not have contributions to GEO*



(3) *Contribution of different normal modes to GEO*



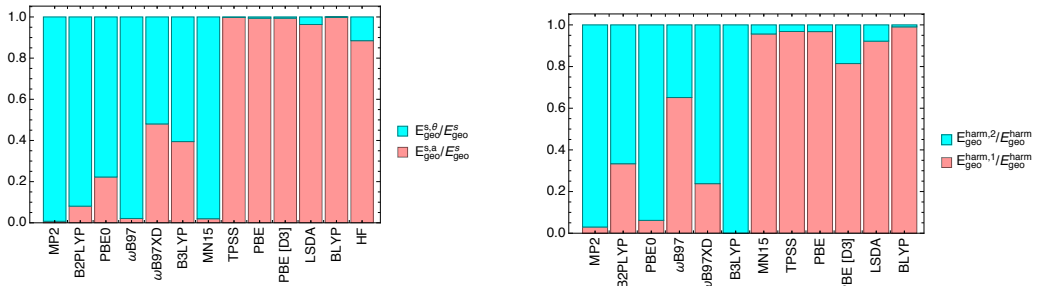
(4) *Weights of individual normal mode contributions to GEO*

Figure S29: *Normal-mode GEO analysis for the H_2O molecule. The "1%" label denotes the GEO errors (or weights) when $\gamma = 1\%$ (see the main text).*

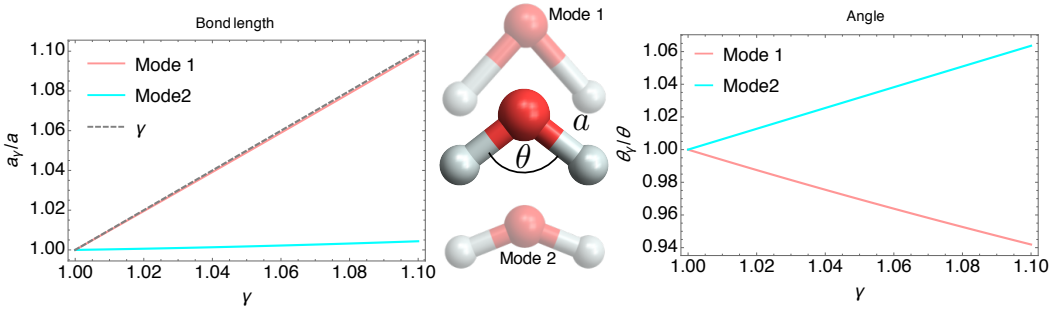
The water has 3 normal modes (note again that by normal modes we refer to the eigenvectors of $\mathbf{H}_0$), and two of these are GEO-active. In the left panel of Fig. S30(1) we show the E_{geo}^{simple} weights in internal coordinates (see Eq. 5), i.e. the E_{geo} contributions from the errors in θ (bond angle) and a (OH bond length). In the right panel of the same figure, we show the E_{geo} weights but given in terms of the errors in normal coordinates (modes), p_1 and p_2, which correspond to the eigenvectors of the Hessian (see Eq. 4). The third normal mode that leads to an asymmetric OH stretch is GEO-inactive (see Fig. S29). By the similarity of the patterns in the two panels of Fig. S30(1), we can expect that the changes in p_1 mostly affect a and that the changes in p_2 mostly affect θ (see also the middle panel of Fig. S30(2), where we plot the normal modes.) But, to show explicitly how the errors in p_1 and p_2 translate into the errors in a and θ, we consider again the γ-scaling, and we scale the individual normal coordinates as: $p_i^{\gamma} = (1+\gamma)\, p_i$. From the left panel of Fig. S30(2), we can see that upon scaling of p_1 by $1+\gamma$, a is scaled by the (almost) factor of γ. At the same time, scaling p_2 by $1+\gamma$ affects very little a. Note that when all coordinates are scaled by γ:

$$\tilde{\mathbf{G}} = (1+\gamma)\, \mathbf{G}_0 \quad \text{all angles are unchanged and all bond lengths scaled by } \gamma. \tag{S12}$$

Thus, when p_1 and p_2 are scaled by $1+\gamma$, the angle changes at the same speed in both cases, but in opposite directions, since the two changes in the angles need to cancel each other out [Eq S12]. This is shown in the right panel of Fig. S30(2). Overall, mode 1 contains contributions from both R and a, whereas, mode 2 contains mostly contribution from a. That is why the weights in the two panels of Fig. S30(1) are similar, but not exactly the same.

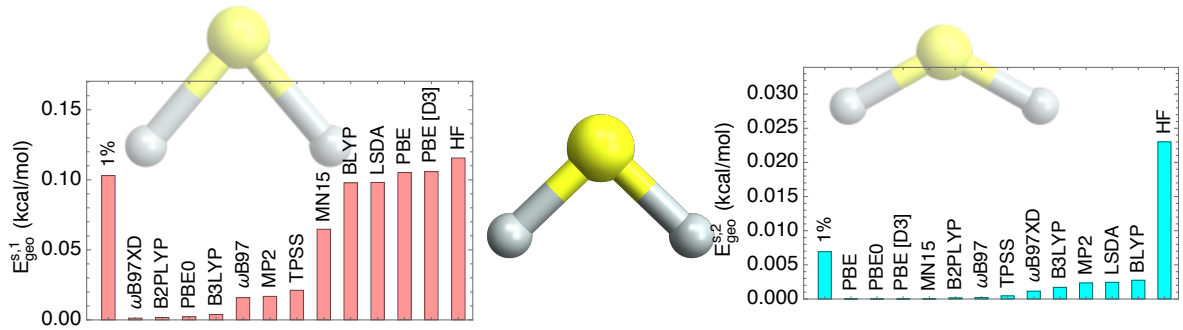


(1) E^{simple}_{geo} *weights for the water molecule in internal coordinates (left), GEO-active normal modes (right).*

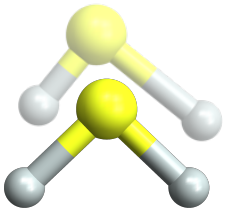


(2) *The change in the bond lengths(left) and bond angle (right) upon scaling GEO active normal modes:* $p_{i,\gamma} = (1+\gamma)p_i$ *where* p_i *are* accurate *[CCSD(T)] normal modes. The GEO active normal modes are shown in the middle panel.*

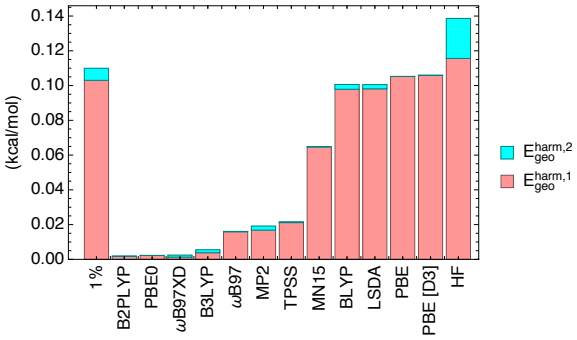
Figure S30: *Normal-mode vs. internal coordinates GEO analysis for the* H_2O *molecule.*



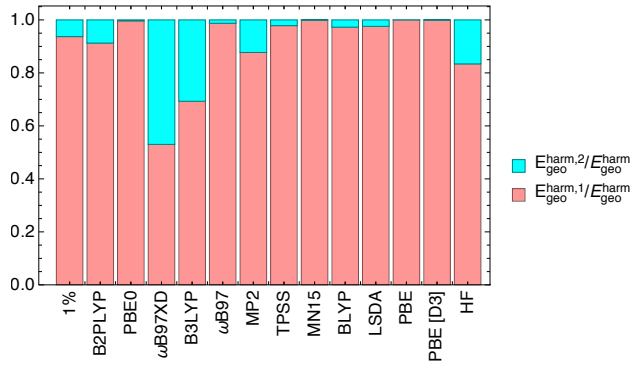
(1) *modes that do have contributions to GEO*



(2) *modes that do not have contributions to GEO*

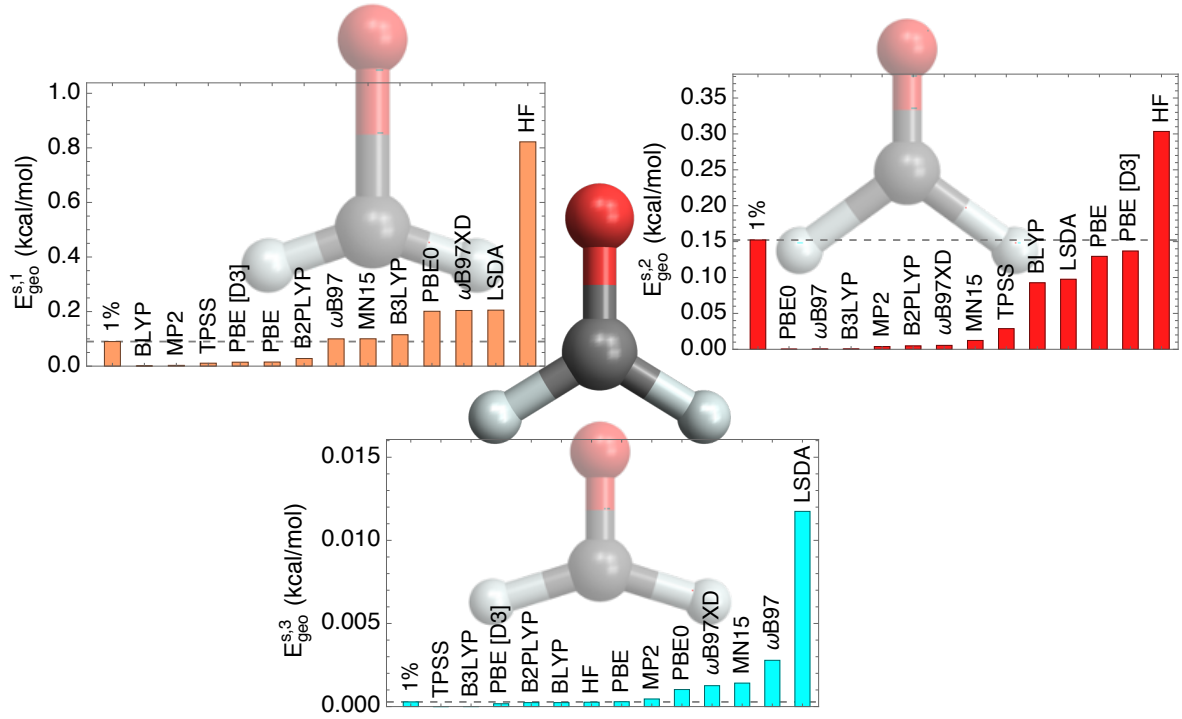


(3) *Contribution of different normal modes to GEO*

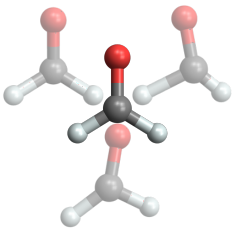


(4) *Weights of individual normal mode contributions to GEO*

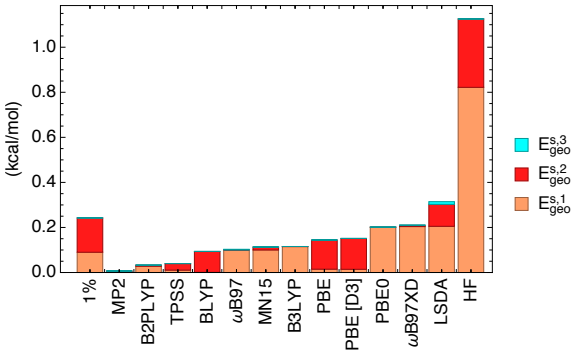
Figure S31: *Normal-mode GEO analysis for the H_2S molecule.*



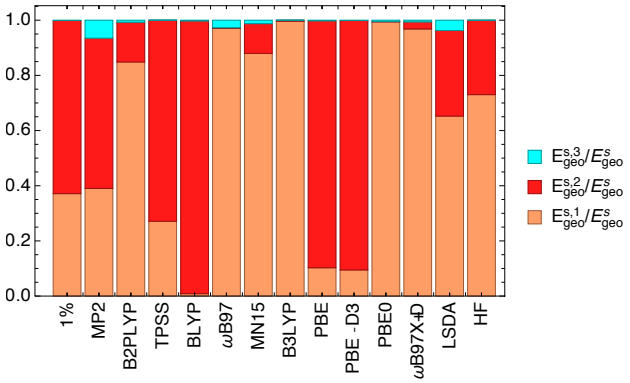
(1) *modes that do have contributions to GEO*



(2) *modes that do not have contributions to GEO*

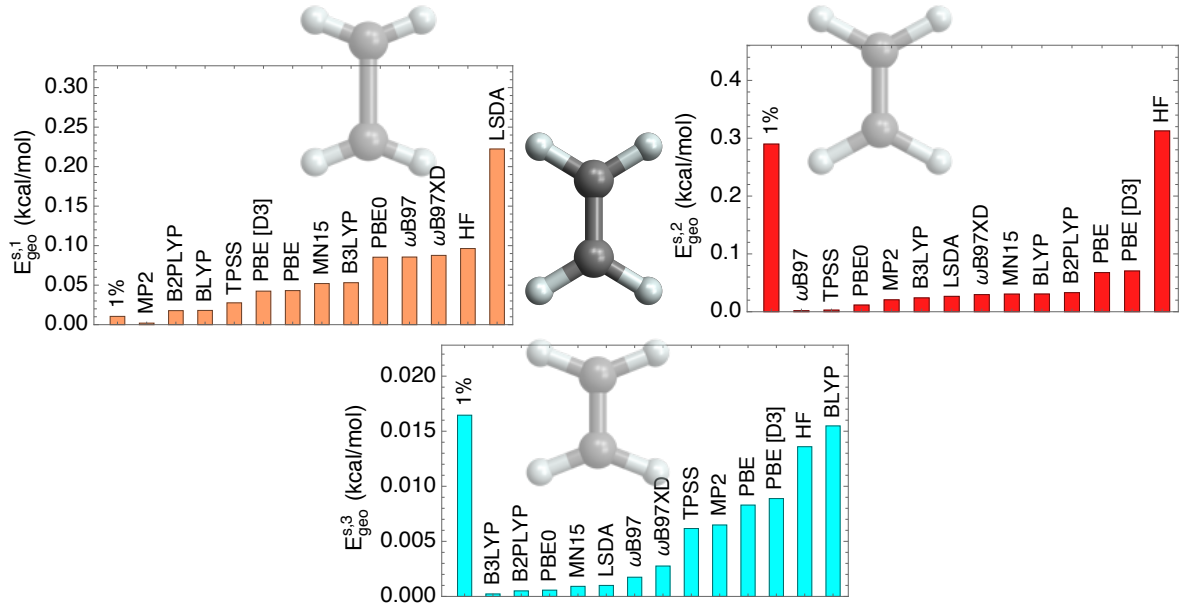


(3) *Contribution of different normal modes to GEO*

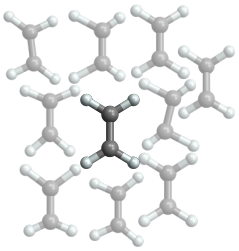


(4) *Weights of individual normal mode contributions to GEO*

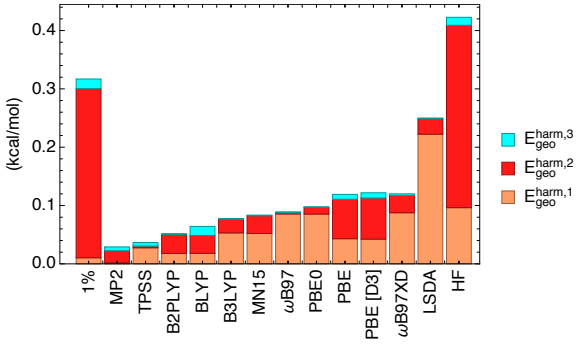
Figure S32: *Normal-mode GEO analysis for the H_2CO molecule.*



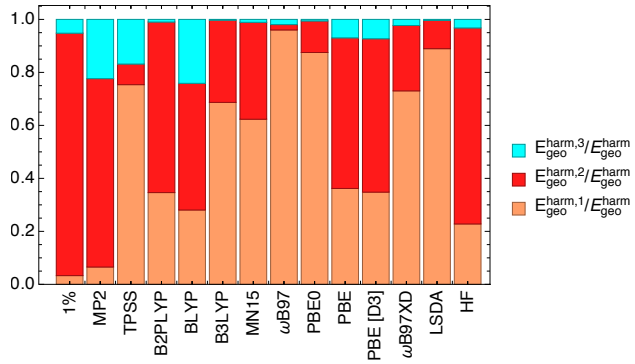
(1) *modes that do have contributions to GEO*



(2) *modes that do not have contributions to GEO*



(3) *Contribution of different normal modes to GEO*



(4) *Weights of individual normal mode contributions to GEO*

Figure S33: *Normal-mode GEO analysis for the ethene molecule.*

S8. Details on the γ expansion and D values for molecular sets

The D molecule-specific constant (absolute GEO scale) is given by:

$$D = \mathbf{G}_0^{\intercal}\mathbf{H}_0\mathbf{G}_0, \tag{S13}$$

and thereby can be easily calculated once the force constants (in e.g. Cartesian coordinates) are available. In analogy to approximating E_{geo} by E'_{geo}, we can calculate D more cheaply from approximate minima and force constants:

$$D' = \tilde{\mathbf{G}}^{\intercal}\tilde{\mathbf{H}}\tilde{\mathbf{G}}. \tag{S14}$$

Furthermore, in analogy to E_{geo}^{simple} approximation to E_{geo}, we can also define D^{simple} as:

$$D \approx D^{simple} = \sum_{i}^{\text{all bonds}} K^{simple,i} = \sum_{i}^{\text{all bonds}} R_i^2\, f_{ii}, \tag{S15}$$

where R_i-s are the bond lengths and f_{ii} are the force constants for each of the bonds. Note that there is no angle contribution to the r.h.s. of Eq S15, since by the γ-scaling (see the main text), angles remain unchanged.

In Table S6, we show the accurate D and approximate D' values. From this table, we can see that even HF D' are decently accurate and already B3LYP ones are in an excellent agreement with D. For the same set of molecules, E^{simple} and contributions from individual bonds, we show in Fig. S34. For other organic molecules in the paper, we give the D values in Tables S6 and S8. Finally, in Table S8, we show the D values for rare gas dimers, and we can see that these D values are by an order of magnitude smaller than those for covalently bonded molecules.

	CCSD(T)	MP2	B2PLYP	B3LYP	PBE0	TPSS	PBE	HF
h2o	0.221	0.222	0.220	0.219	0.224	0.211	0.211	0.247
h2s	0.220	0.228	0.221	0.215	0.221	0.213	0.209	0.239
methanol	0.566	0.572	0.567	0.563	0.579	0.547	0.545	0.629
cyclopropane	0.989	1.007	0.998	0.99	1.012	0.972	0.966	1.076
ch3cl	0.452	0.465	0.452	0.443	0.459	0.436	0.435	0.488
ch3f	0.463	0.469	0.463	0.458	0.47	0.447	0.443	0.511
c2h3cl	0.727	0.742	0.736	0.733	0.754	0.717	0.714	0.81
c2h4	0.634	0.643	0.643	0.642	0.651	0.629	0.623	0.699
c-c2h2f2	0.848	0.858	0.856	0.856	0.881	0.828	0.823	0.971
c2h3f	0.735	0.745	0.744	0.744	0.761	0.724	0.719	0.828
imine	0.555	0.557	0.564	0.569	0.58	0.551	0.547	0.636
h2co	0.485	0.485	0.489	0.497	0.508	0.48	0.476	0.563
formic	0.725	0.724	0.723	0.731	0.753	0.698	0.697	0.842
hcocl	0.594	0.595	0.592	0.599	0.623	0.573	0.573	0.696
MAE	/	0.007	0.005	0.007	0.019	0.013	0.017	0.073
RMS	/	0.009	0.006	0.008	0.021	0.015	0.018	0.079
MARE (%)	/	1.3	0.7	1.3	3.	2.5	3.1	12.2

Table S5: *Accurate D values [CCSD(T)] in 10^4 kcal/mol, and approximate D' values computed from Eqs. S13 and S14, respectively, for the set of molecules considered in Figs. 1b.*

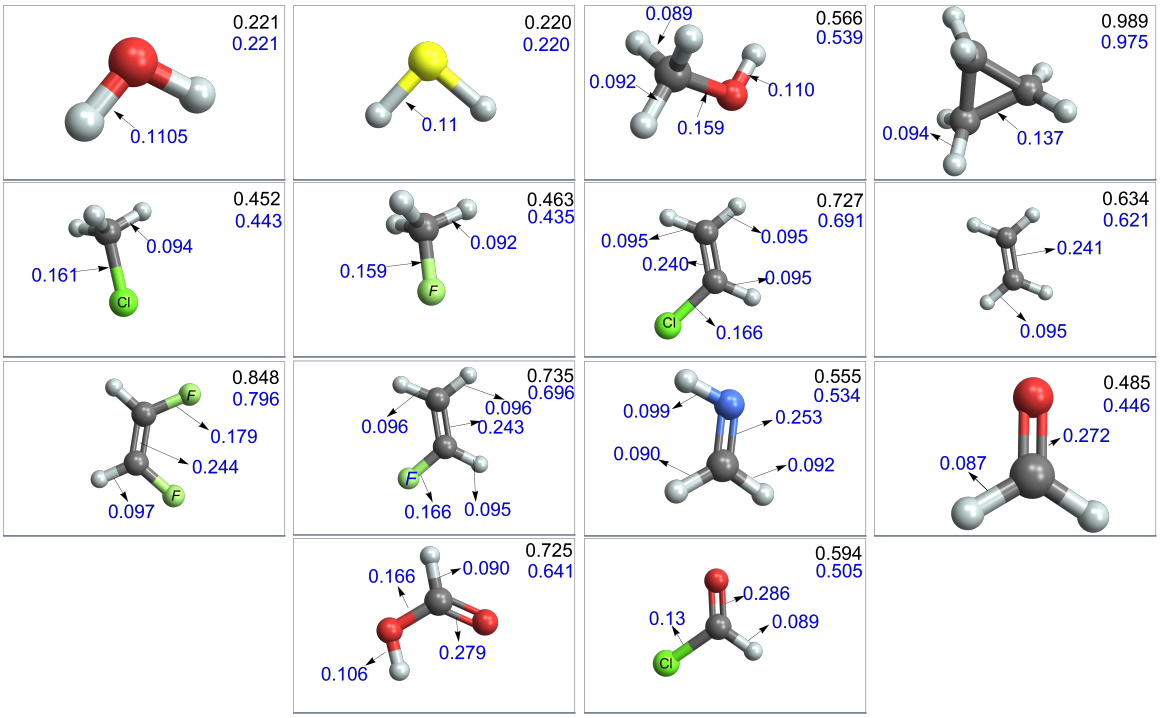


Figure S34: *D values in 10^4 kcal/mol (a value in black in the top-right corner of each panel), D^{simple} values (a value in blue the top-right corner of each panel), and D^{simple} bond-decomposition of Eq. S15 for the set of molecules considered in Fig. 1b.*

acetaldehyde	0.842
acetic	1.086
c2h3f	0.735
c2h5f	0.819
ch2ch	0.749
ch2nh2	0.598
ch3nh	0.543
ch3nh2	0.660
ethanol	0.917
formic	0.73
methanol	0.568
propane	1.066
propene	0.982
propyne	0.877

Table S6: *D values in 10^4 kcal/mol for individual molecules for molecules considered in Fig. 1a(top) at the CCSD(T) level of theory. For sufficiently small γ, D can be computed by finite difference from: $D = \frac{2E^{\gamma}_{geo}}{\gamma^2}$, and we have used here $\gamma = 0.5\%$. We have found that this γ value is small enough to stay within the harmonic approximation, but still not too small to lose numerical accuracy.*

mol. #	K
1	2.544
2	1.695
3	2.229
4	2.386
5	2.083
6	2.483
7	2.505
8	1.693
9	2.387
10	2.009
11	2.457
12	2.339
13	2.738
14	3.208
15	3.715
16	3.404

Table S7: *D values in* 10^4 *kcal/mol obtained as described in the caption Table S6 but with B2PLYP/aug-cc-pVTZ as a reference. The order of molecules is shown in Fig. S6.*

He_2	0.000154
Ne_2	0.000724
Ar_2	0.002164
HeNe	0.000301
NeAr	0.001079

Table S8: *D values in* 10^4 *kcal/mol computed from the CCSD(T) force constants for the noble gas dimers (Eq. S13). For the basis set informations, see Fig. S44.*

S9. Statistical measures for the quality of approximate geometries of the methylamine molecule

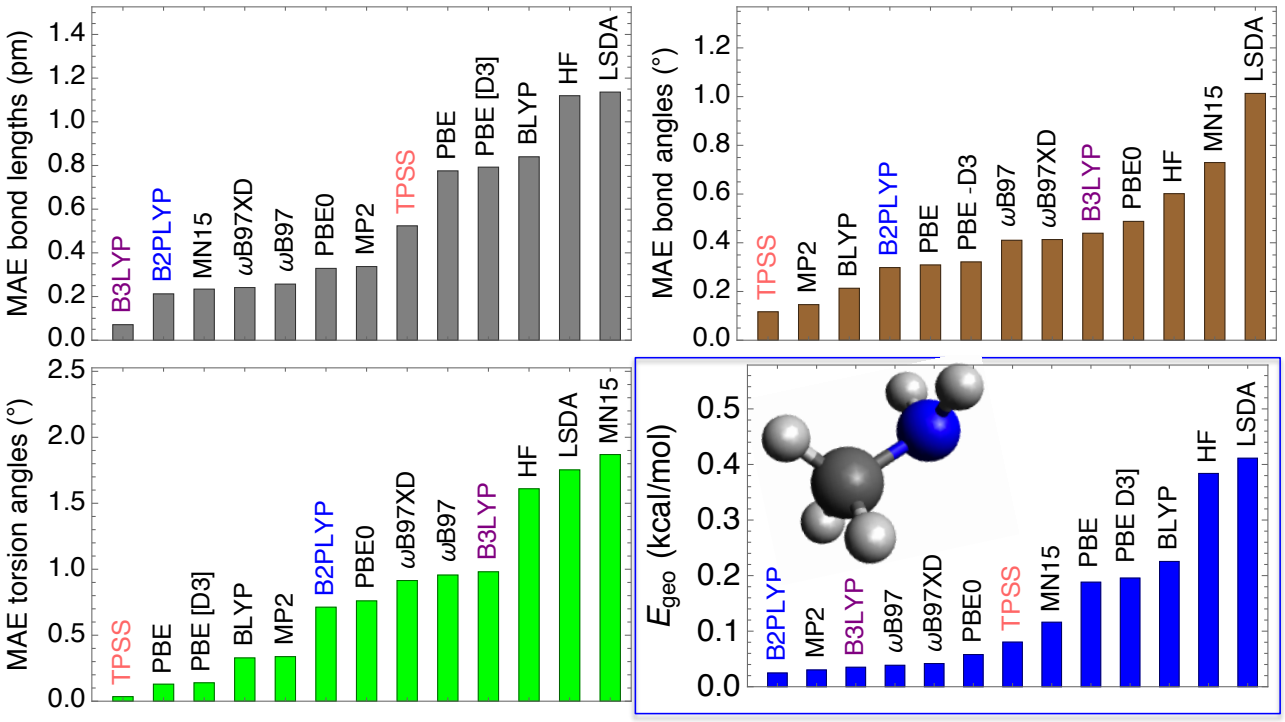


Figure S35: *Mean absolute errors (MAE) in (i) calculated bond lengths; (ii) bond angles; (iii) torsion angles (iv) GEO errors obtained from various approximate methods for the methylamine molecule.*

In addition to the measures in Figure S35, in Figure S37 we show other approximate measures for the geometry of the methylamine molecule. For comparison, we also show in Figure S38 the errors in atomization energies obtained with the same methods.

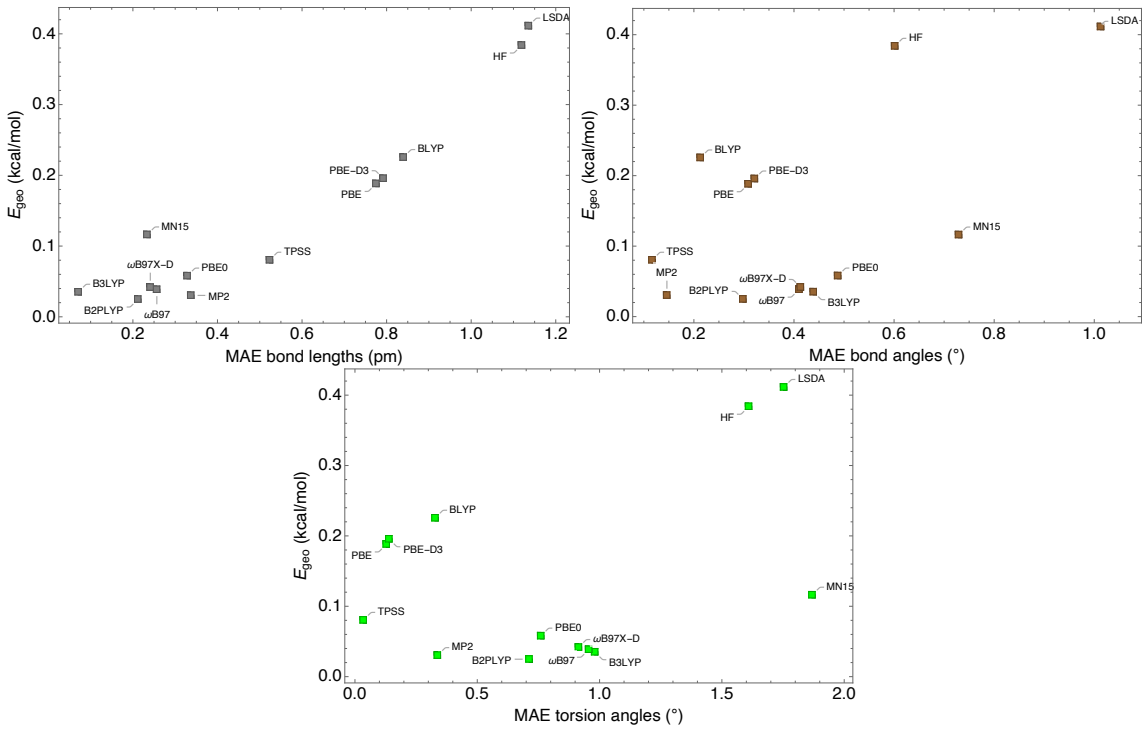


Figure S36: *(i) GEO vs MAEs in bond lengths; (ii) GEO vs MAEs in bond angles; (iii) GEO vs MAEs in torsion angles; for the methylamine molecule. The data are the same as in Fig. S35.*

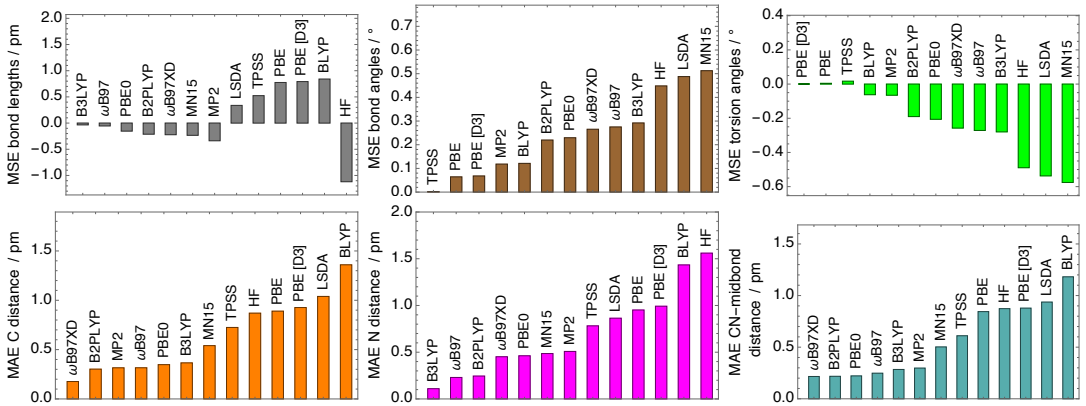


Figure S37: *Other statistical measures for the quality of the approximate geometries of the methylamine molecule (see Figure 3 of the main text): (i) means signed error (MSE) for the bond lengths; (ii): MSE for the bond angles; (iii) MSE for the torsion angles (iv) mean absolute error (MAE) for the distances between the C atom and the other atoms in the molecule (v) MAE for the distances between the N atom and the other atoms in the molecule (vi) MAE for the distances between the CN midbond point and the atoms in the molecule.*

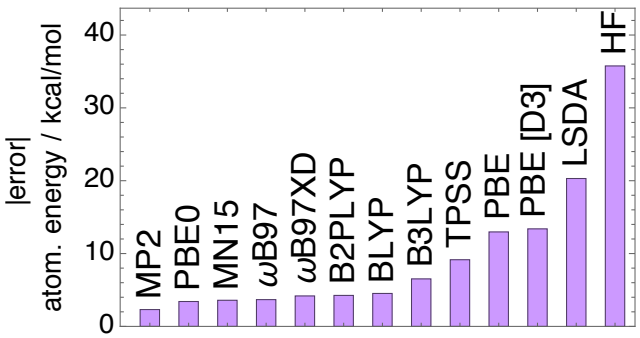
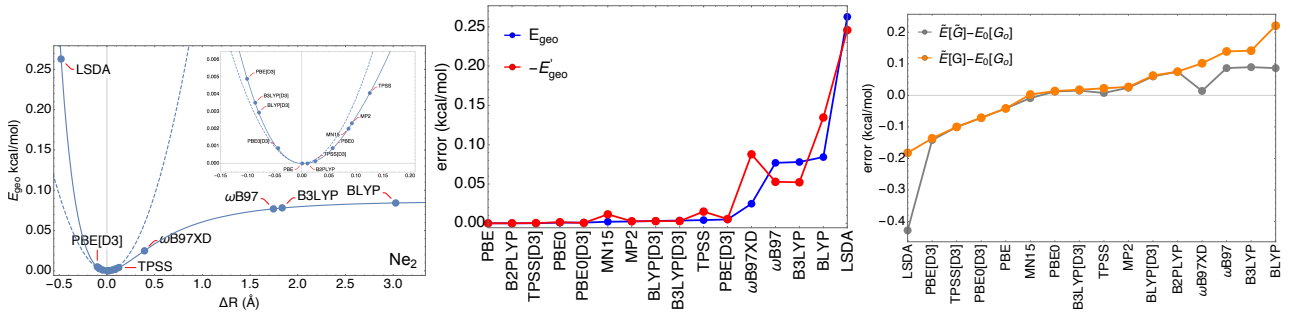


Figure S38: *Absolute errors for the atomization energies of the methylamine molecule obtained from different methods. The LSDA and HF errors are divided by the factor of* 4.5

S10. GEO ANALYSIS FOR NOBLE GAS DIMERS



(1) *Ne2 [0.09 kcal/mol]*

Figure S40: *Geo analysis for the neon dimer. For the basis set information, see Figure S44. Accurate binding energy of the dimer is shown in the square brackets in the subcaption. The aug-cc-pV5Z basis set has been used in all calculations. In the left top panel, the ΔR position of approximations are not accurate when $\Delta R > 1$Å. In these cases, the approximations typically give unphysical binding curves with no minimum and only infimum at $R \to \infty$.*

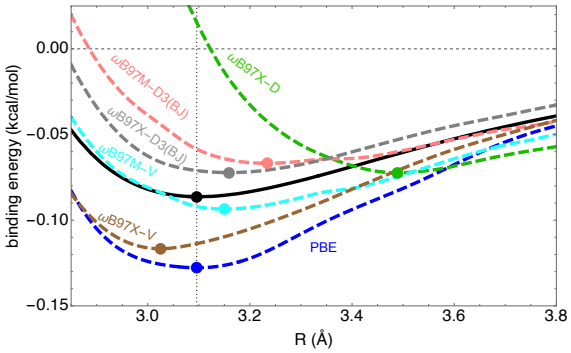


Figure S41: *Binding curves of Ne_2 with functionals built upon ωB97X-D. The ωB97X-V [12] and ωB97M-V [13] functionals have been developed by Mardirossian and Martin Head-Gordon and include VV10 non-local correlation (NLC) terms [14]. Najibi and Goerigk developed the versions of these two functionals, ωB97X-D3(BJ) and ωB97M-D3(BJ), respectively, by removing the NLC terms and adding optimized D3(BJ) terms [15]. The binding curves of these four functionals have been obtained from ORCA 4.2.1.[3] The reference CCSD(T) (black), PBE and ωB97X-D curves are shown for comparison. The aug-cc-pV5Z basis set is used in all calculations.*

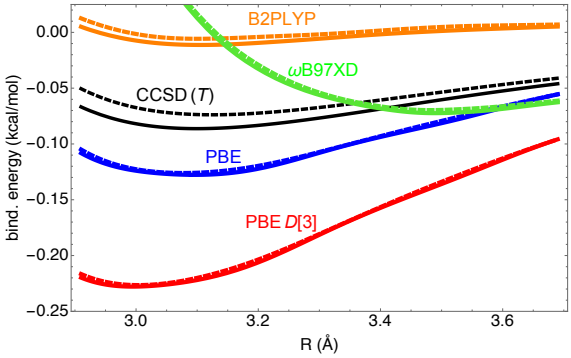


Figure S42: *Counterpoise (CP) corrected (dashed lines) and counterpoise uncorrected binding curves for the neon dimer (solid lines) obtained from various approximations. Detailed analysis of Baerends and co-workers have shown that for small dimers bonded by dispersion, CP-uncorrected curves with a finite basis set are in a better agreement with the "exact" (i.e. no errors due to a finite basis set) curve curve than its CP-corrected counterpart (see Refs. [16, 17]). Thus, in the main text and in Figure S40, we use CP-uncorrected curves. Nonetheless, due to a large basis set used here [aug-cc-pV5Z], we can see from this figure that the energy difference between curves with and without the CP is small, even in the CCSD(T) case and nearly negligible in the case of DFT curves.*

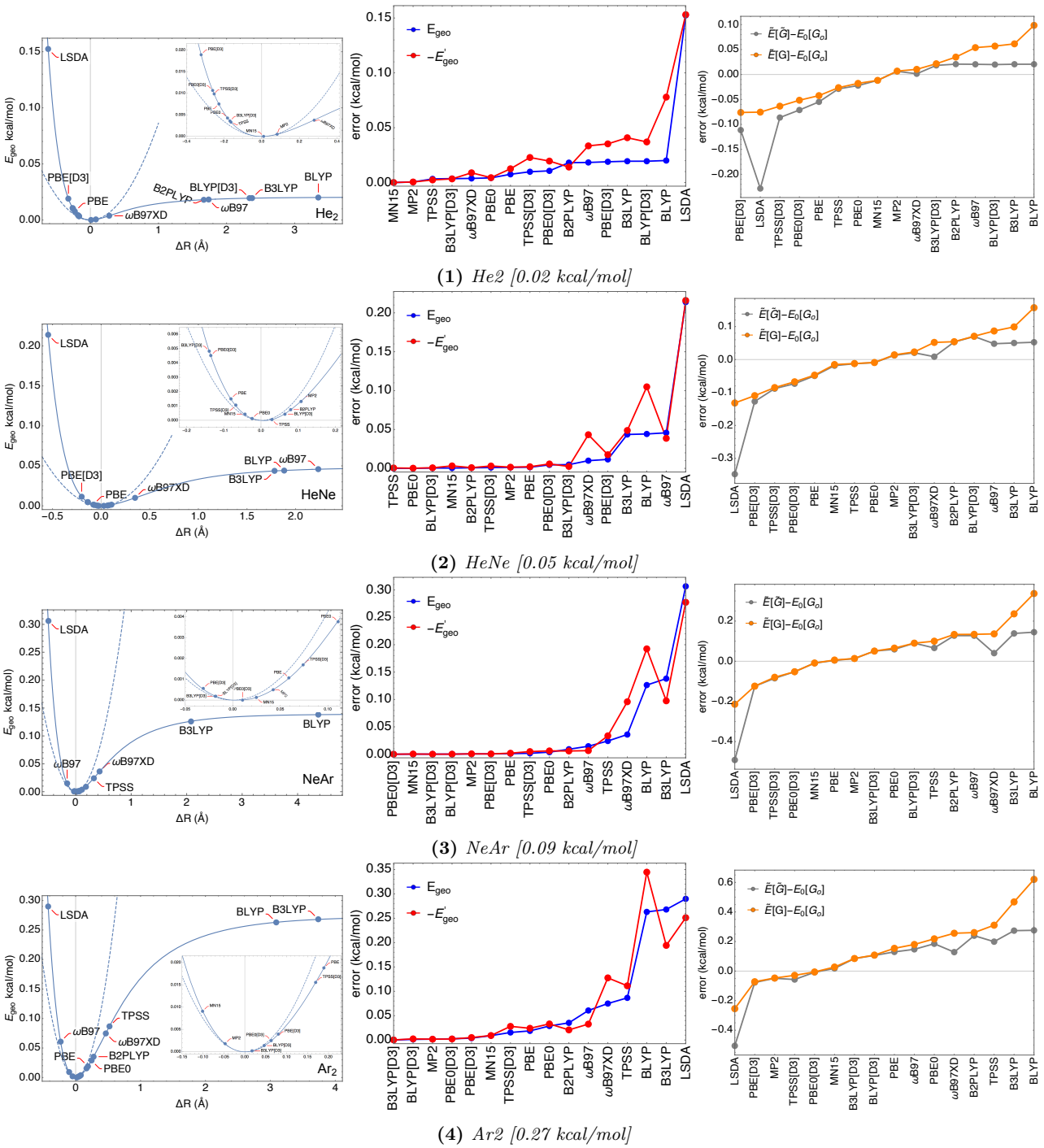


Figure S44: *The GEO analysis for noble gas dimers. The aug-cc-pV5Z basis has been used for Ne2 and He2, and aug-cc-pVQZ has been used for all other dimers. CCSD(T) has been used as a reference for all dimers. In the left panels, the ΔR position of approximations are not accurate when $\Delta R > 1$Å. In these cases, the approximations typically give unphysical binding curves with no minimum and only infimum at $R \to \infty$.*

S11. Details on the results of Figure 5 [S66 results]

The S66x8 set contains 66 dimers bonded by noncovalent interactions.[18] For each complex and each approximate method we do single point calculations to obtain 8 datapoints for different separations between fragments along the dissociation curve of a given complex. The geometries at different separations between fragments in a complex have been taken from the S66x8 dataset.[18] The CCSD(T)/CBS (used as a reference here), MP2/CBS, CCSD/CBS and SCS-CCSD/CBS binding energies have been taken from the original S66x8 dataset.[18, 19] All other binding energy we obtain from the G16, by performing counterpoise corrected calculations within the aug-cc-pVTZ basis set. From these 8 datapoints (for each complex and for each approximate method) we use a spline interpolation to construct a dissociation curve. The E_{geo} error for a given approximate method is then calculated by the difference between the CCSD(T) binding energies at the approximate minimum and at the CCSD(T) minimum (see Eq. 1). In principle, in addition to finding a minimum intermolecular separation between fragments, we should have also optimized the fragments with a given approximate method. However, we assume here that this has a minor effect on the E_{geo} contribution to the binding energies errors of the S66 complexes.

The same plots shown in Figure 5 in the main text are shown here with other approximate methods. Namely, in Figure S45 we show the $|E_{geo}|$ vs. $|\Delta E|$ plots (as in Figure 5 of the main text) for selected DFT functionals with the D3 empirical correction. The same plots are shown in Figure S46 for CCSD/CBS and SCS-CCSD/CBS. The same plots obtained from the PBE functional with and without the D3 correction is shown in Figure S47.

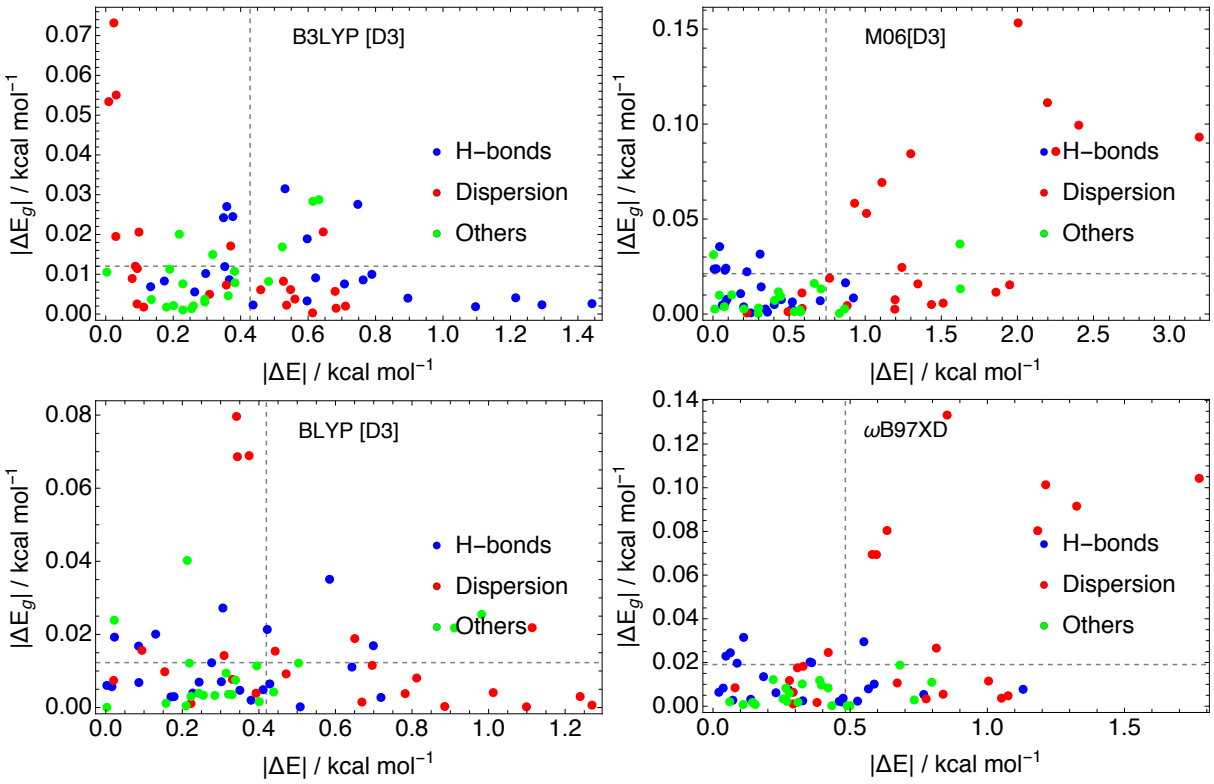


Figure S45: *E_{geo} (y-axis) vs. $|\Delta E|$ (x-axis) errors in kcal/mol for different methods and for the S66 dataset.*

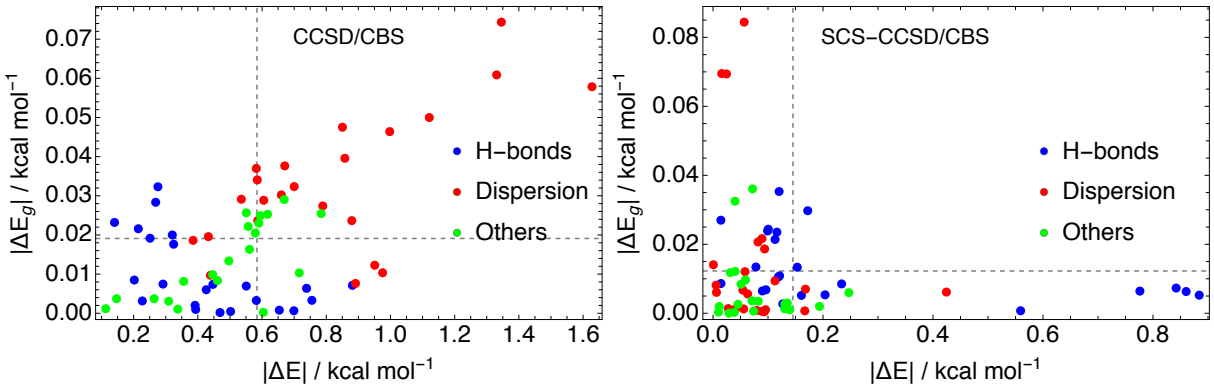


Figure S46: *E_{geo} (y-axis) vs. $|\Delta E|$ (x-axis) errors in kcal/mol for different methods and for the S66 dataset.*

	MP2/CBS	B2PLYP	B3LYP-D3	PBE0-D3	PBE-D3	MN15	wB97X-D	PBE	CCSD(T)
MP2/CBS	0.51	0.26	0.45	0.43	0.38	0.47	0.46	0.28	0.47
B2PLYP	1.63	1.34	1.45	1.43	1.38	1.52	1.51	1.40	1.46
B3LYP-D3	0.44	0.34	0.43	0.42	0.41	0.41	0.42	0.53	0.43
PBE0-D3	0.51	0.43	0.48	0.49	0.48	0.45	0.47	0.54	0.48
PBE-D3	0.48	0.4	0.44	0.45	0.46	0.40	0.41	0.44	0.44
MN15	0.56	0.34	0.59	0.58	0.53	0.6	0.60	0.34	0.59
wB97X-D	0.42	0.25	0.47	0.45	0.41	0.47	0.48	0.38	0.46
PBE	2.29	1.77	2.04	1.98	1.90	2.14	2.12	1.71	2.06

Table S9: *MAE of different methods (rows) at different minima of S66x8 binding curves (columns) for the S66 dataset. The error is defined as: $E_X[R_Y] - E_0$, where X is a method in rows and Y is the method in columns, and E_0 is the binding energy at the CCSD(T) minimum. The S66x8 binding curves from CCSD(T)/CBS (used as a reference here) and MP2/CBS have been taken from the original S66x8 dataset.[18, 19] All other binding curves have been obtained from counterpoise corrected calculations within the aug-cc-pVTZ basis set. The minimum of each binding curve has been found numerically after the interpolation of 8 datapoints for each of the S66 complex (see details above).*

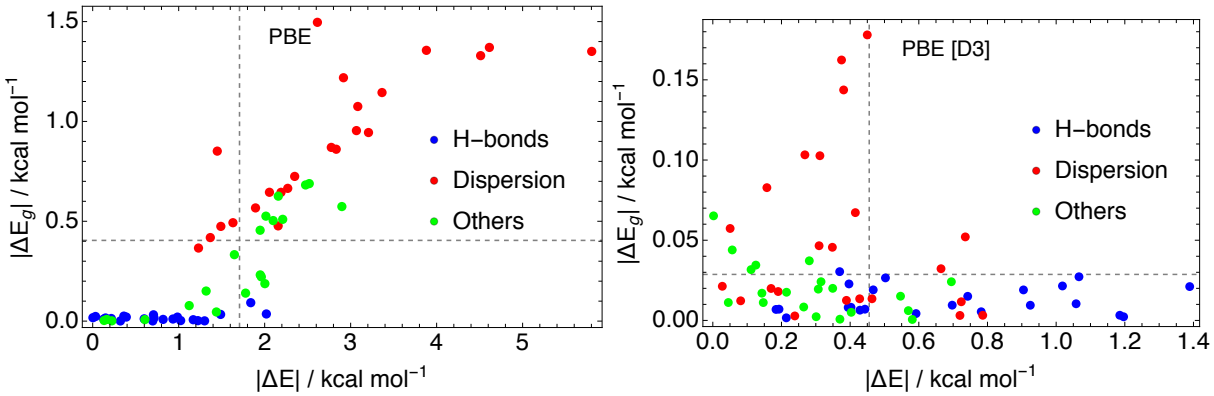


Figure S47: *E_{geo} (y-axis) vs. $|\Delta E|$ (x-axis) errors in kcal/mol for different methods and for the S66 dataset.*

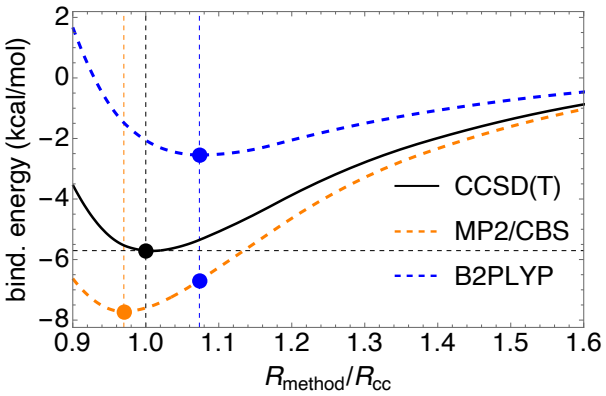


Figure S48: *Benzene-uracil binding curves (interpolated from the 8 datapoints along the dissociation curve).*

References

[1] Peter R Spackman, Dylan Jayatilaka, and Amir Karton. Basis set convergence of ccsd (t) equilibrium geometries using a large and diverse set of molecular structures. *The Journal of chemical physics*, 145(10):104101, 2016.

[2] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *The Journal of chemical physics*, 132(15):154104, 2010.

[3] Frank Neese. Software update: the orca program system, version 4.0. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(1):e1327, 2018.

[4] G t Te Velde, F Matthias Bickelhaupt, Evert Jan Baerends, C Fonseca Guerra, Stan JA van Gisbergen, Jaap G Snijders, and Tom Ziegler. Chemistry with adf. *Journal of Computational Chemistry*, 22(9):931–967, 2001.

[5] Thom H. Dunning. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90:1007, 1989.

[6] Michael Gaus, Albrecht Goez, and Marcus Elstner. Parametrization and benchmark of dftb3 for organic molecules. *Journal of Chemical Theory and Computation*, 9(1):338–354, 2013.

[7] Axel D Becke and Erin R Johnson. A density-functional model of the dispersion interaction. *The Journal of chemical physics*, 123(15):154101, 2005.

[8] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865, 1996.

[9] Yan Zhao and Donald G. Truhlar. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *The Journal of Chemical Physics*, 125(19):194101, November 2006.

[10] Yan Zhao and Donald G. Truhlar. The m06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four m06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts*, 120(1-3):215–241, July 2007.

[11] Yan Zhao and Donald G. Truhlar. Density functional for spectroscopy: no long-range self-interaction error, good performance for rydberg and charge-transfer states, and better performance on average than b3lyp for ground states. *The Journal of Physical Chemistry A*, 110(49):13126–13130, December 2006.

[12] Narbe Mardirossian and Martin Head-Gordon. ωb97x-v: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Physical Chemistry Chemical Physics*, 16(21):9904, 2014.

[13] Narbe Mardirossian and Martin Head-Gordon. ωb97m-v: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *The Journal of Chemical Physics*, 144(21):214110, June 2016.

[14] Oleg A. Vydrov and Troy Van Voorhis. Nonlocal van der waals density functional: The simpler the better. *The Journal of Chemical Physics*, 133(24):244103, December 2010.

[15] Asim Najibi and Lars Goerigk. The nonlocal kernel in van der waals density functionals as an additive correction: An extensive analysis with special emphasis on the b97m-v and ωb97m-v approaches. *Journal of Chemical Theory and Computation*, 14(11):5725–5738, October 2018.

[16] Xiao Wei Sheng, Lukasz Mentel, Oleg V Gritsenko, and Evert Jan Baerends. Counterpoise correction is not useful for short and van der waals distances but may be useful at long range. *Journal of computational chemistry*, 32(13):2896–2901, 2011.

[17] ŁM Mentel and EJ Baerends. Can the counterpoise correction for basis set superposition effect be justified? *Journal of Chemical Theory and Computation*, 10(1):252–267, 2014.

[18] Jan Rezác, Kevin E Riley, and Pavel Hobza. S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures. *Journal of chemical theory and computation*, 7(8):2427–2438, 2011.

[19] We thank J. Rezác for giving us access to the S66x8 binding energies for these wavefunction methods.