Supporting Information for

"Quantification of Geometric Errors Made Simple: Application to Main-group Molecular Structures"

Stefan Vuckovic*,†,‡

†*Institute for Microelectronics and Microsystems (CNR-IMM), Via Monteroni, Campus Unisalento, 73100 Lecce, Italy*

‡*Department of Chemistry & Pharmaceutical Sciences and Amsterdam Institute of Molecular and Life Sciences (AIMMS), Faculty of Science, Vrije Universiteit, De Boelelaan 1083, 1081HV Amsterdam, The Netherlands*

E-mail: stefanvuckovic1@gmail.com

Contents

List of Tables

List of Figures

S1 Further details on GEO and GEO'

For an approximate electronic structure method, the total error is given by:

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

and thus contains errors both due to the approximate geometry and approximate energy (see Section 1 (*Introduction*) for the definition of the quantities in Eq S1). To decompose this error into GEO and non-geometric parts (*'purely energetic components'*, denoted by P and P' below), we add and subtract $E\left(\tilde{\mathbf{G}}_0\right)$ to the r.h.s of Eq. S1:

$$\Delta E = \underbrace{E\left(\tilde{\mathbf{G}}_0\right) - E\left(\mathbf{G}_0\right)}_{GEO \geq 0} + \underbrace{\tilde{E}\left(\tilde{\mathbf{G}}_0\right) - E\left(\tilde{\mathbf{G}}_0\right)}_{P}. \tag{S2}$$

Adding and subtracting $\tilde{E}\left(\mathbf{G}_0\right)$ to the r.h.s of Eq. S1 we obtain an alternative form of Eq. S2:

$$\Delta E = \underbrace{\tilde{E}\left(\tilde{\mathbf{G}}_0\right) - \tilde{E}\left(\mathbf{G}_0\right)}_{-GEO' \leq 0} + \underbrace{\tilde{E}\left(\mathbf{G}_0\right) - E\left(\mathbf{G}_0\right)}_{P'}. \tag{S3}$$

The signs of GEO and GEO' also dictate the following chain of inequalities:

$$P \leq \Delta E \leq P'. \tag{S4}$$

Since GEO is typically very accurately approximated by GEO', then we also have: $GEO \approx \frac{1}{2}\left(P' - P\right)$. As discussed in Ref. 1, for equilibrium structures (minima of potential energy surfaces), GEO and GEO' are always positive, whereas their signs for transition states (the first order saddle points of potential energy surfaces) are not definite.

S2 Computational details

All calculations except for ones stated in the next paragraph have been performed with the G16 package,[2] with *'tight'* convergence criteria for geometry optimizations, and with *'ultrafine'* grids for DFT calculations. For the LDA functional, the Slater exchange was combined with VWN correlation (SVWN).[3]

Calculations using 3c methods (Section 3.5), and using different version of MP2 for the phenyl radical, have been performed with the ORCA 5.0.0. package.[4] These MP2 calculations have been performed within the resolution of the identity (RI) approximation.

S3 Set of accurate semixperimental B2se structures used in the present work

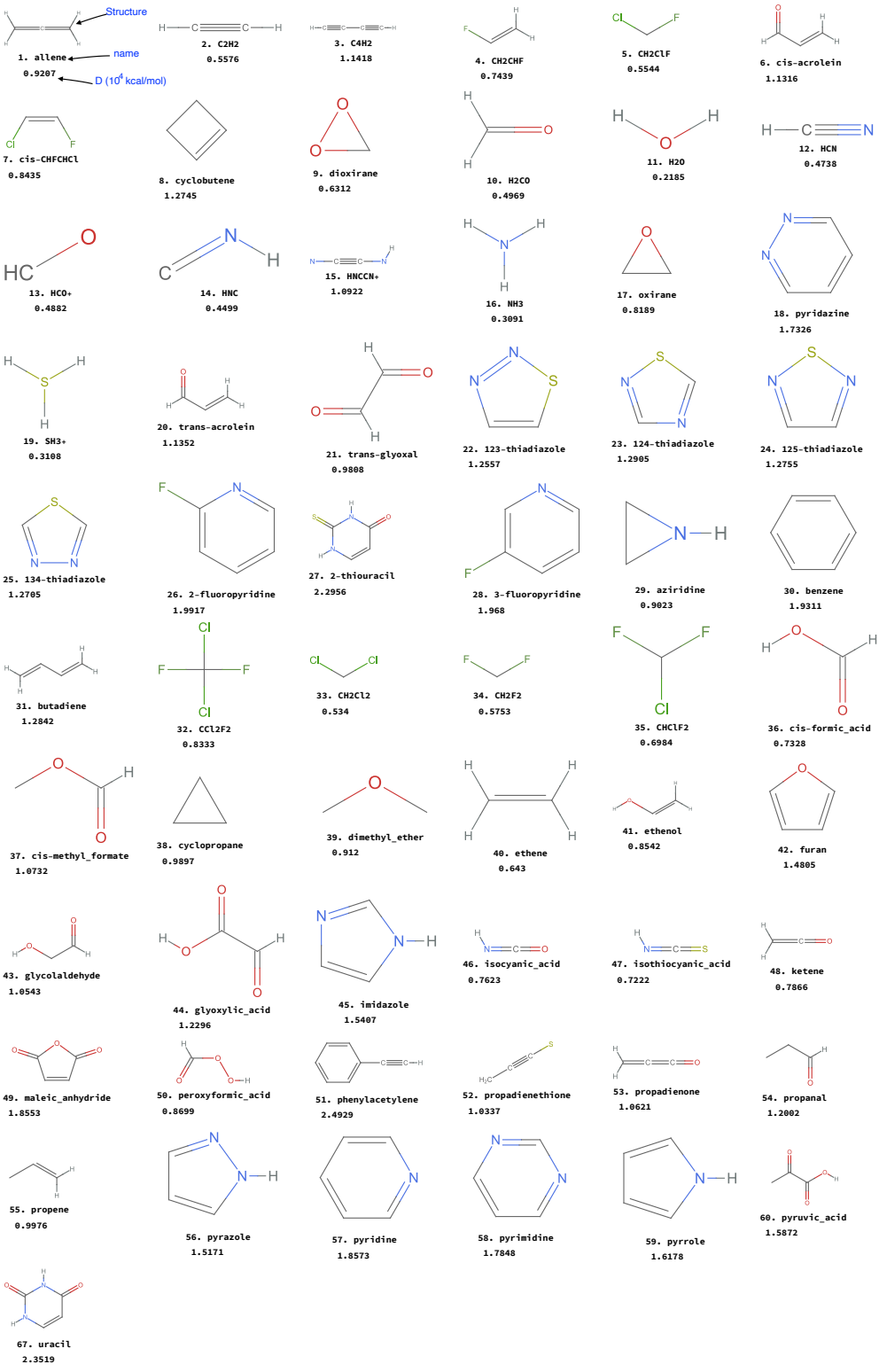


Figure S1: B2se set of molecules with their D values in 10^4 kcal/mol. The phenyl radical, as the only open-shell B2se species is excluded here and is analysed separately in Section 4.4.

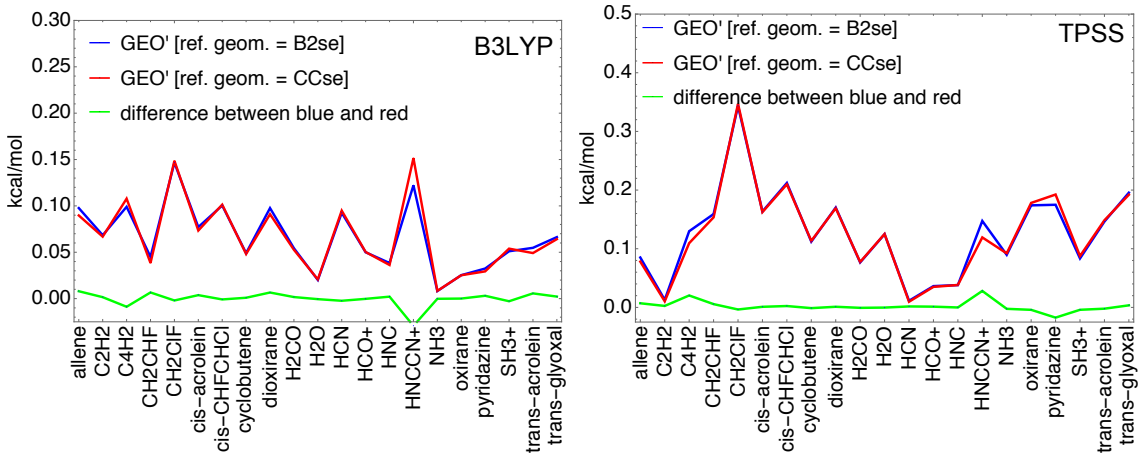


Figure S2: *GEO′* for B3LYP/AVQZ (left) and TPSS/AVQZ (right) w.r.t. to CCse reference geometries (blue) and B2se reference geometries (red) for selected molecules

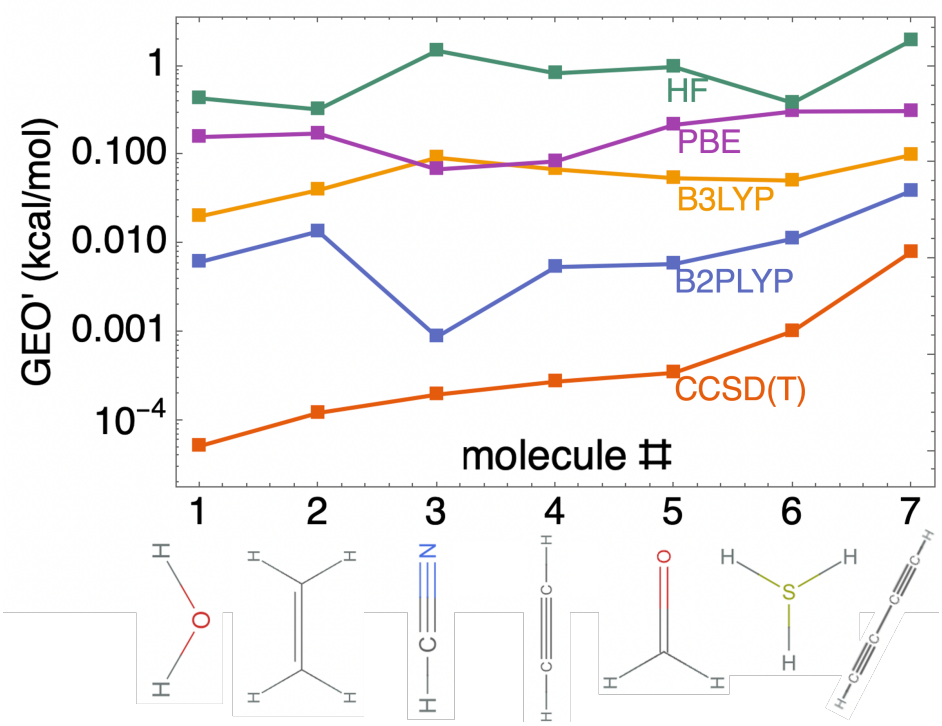


Figure S3: GEO' values on the *log* scale measured against the B2se structures of small molecules. For HF and CCSD(T) calculations, we use the cc-pVnZ basis set for the H atom and aug-cc-pCVnZ basis set for all other atoms, with $n = 5$ for molecules #1 to #5, and $n = Q$ for molecules #6 and #7. For PBE, B2PLYP and B3PLYP calculations, aug-cc-pVQZ was used.

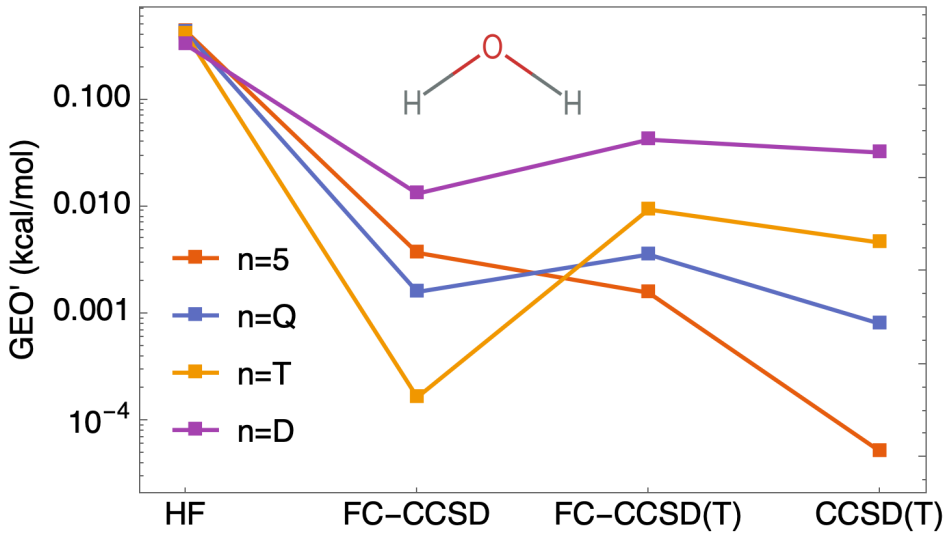


Figure S4: Plots showing how GEO' for the water molecule measured against its B2se structure changes as we go from HF to CCSD(T) via FC-CCSD and FC-CCSD(T), where FC stands for the frozen core approximation. Aug-cc-pCVnZ basis set was used for the O atom and cc-pVnZ for the H atom (note the *log*-scale on the y-axis.

S4 More basis set dependence results

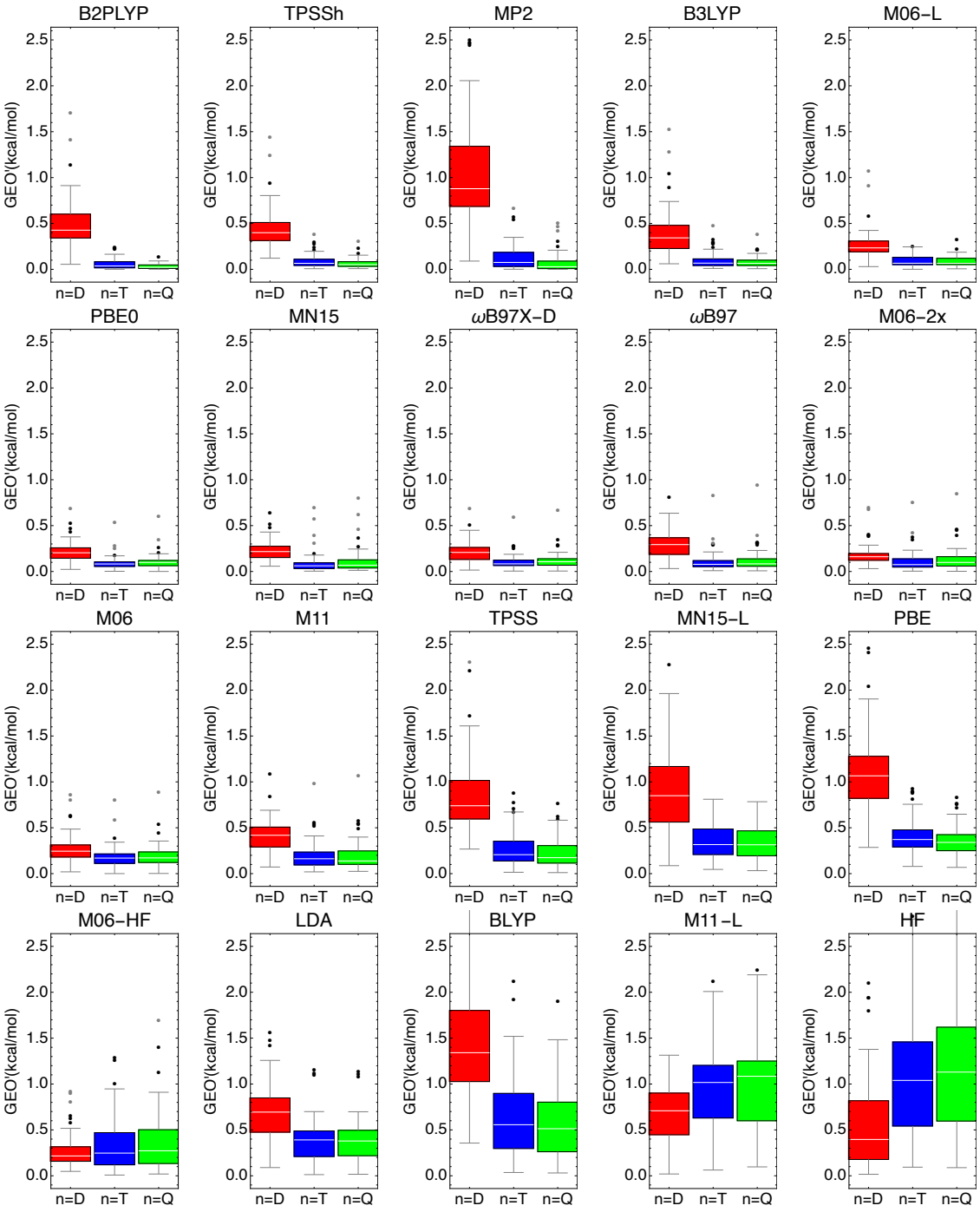


Figure S5: Same as Figure 3, but for more approximations.

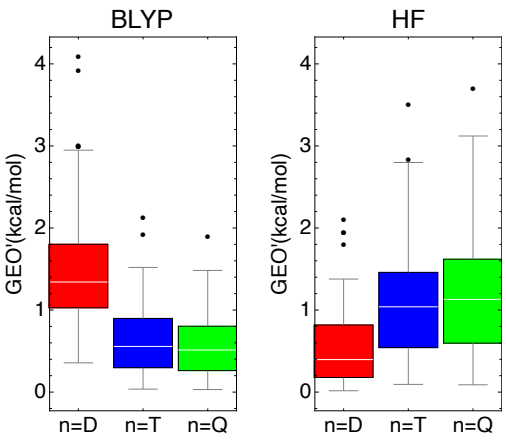


Figure S6: Same as Fig. S5, but for a full range of GEO' values for HF and BLYP.

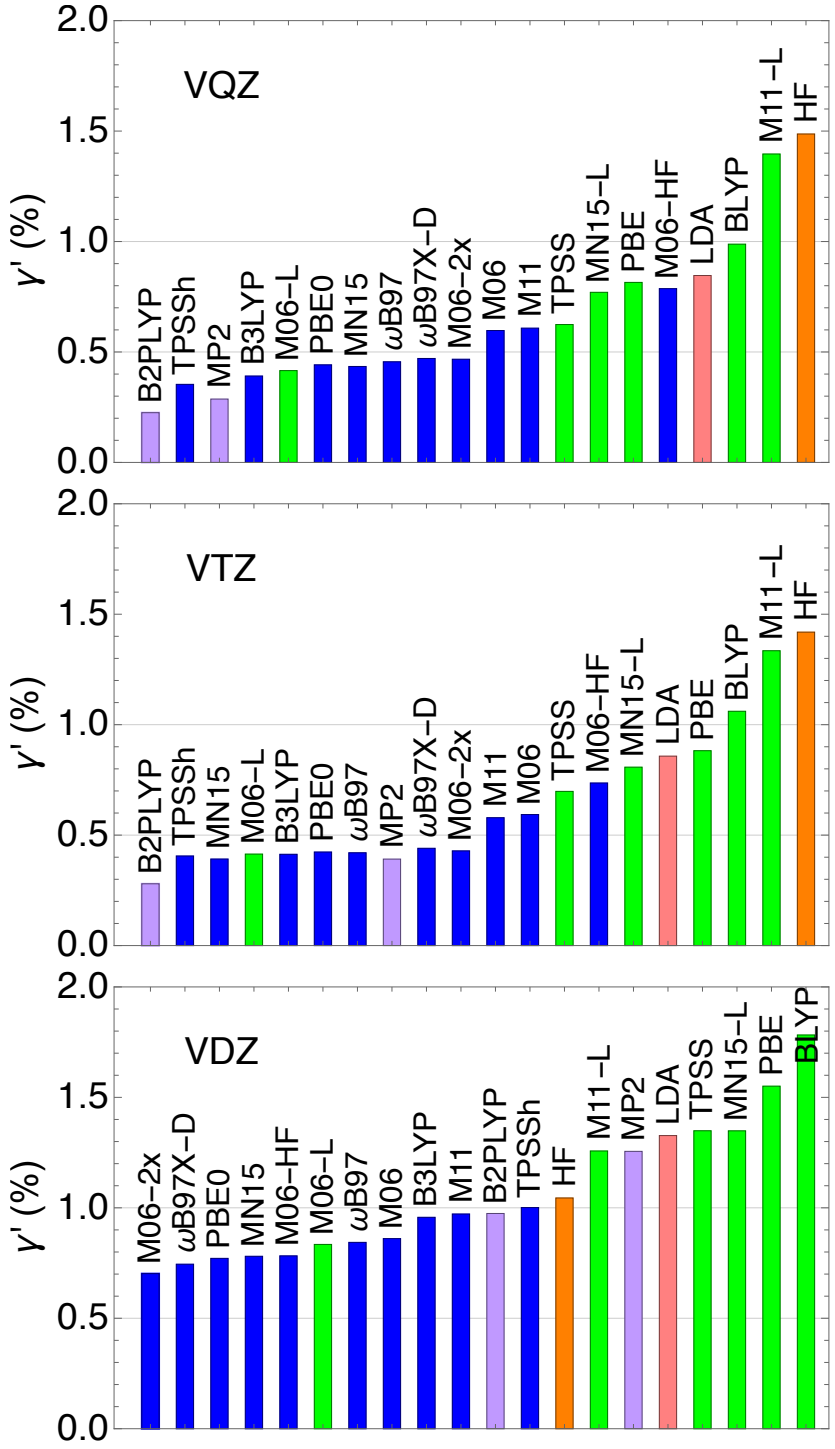


Figure S7: γ' bars, same as Fig. 6, but within the VnZ basis set (no diffuse functions).

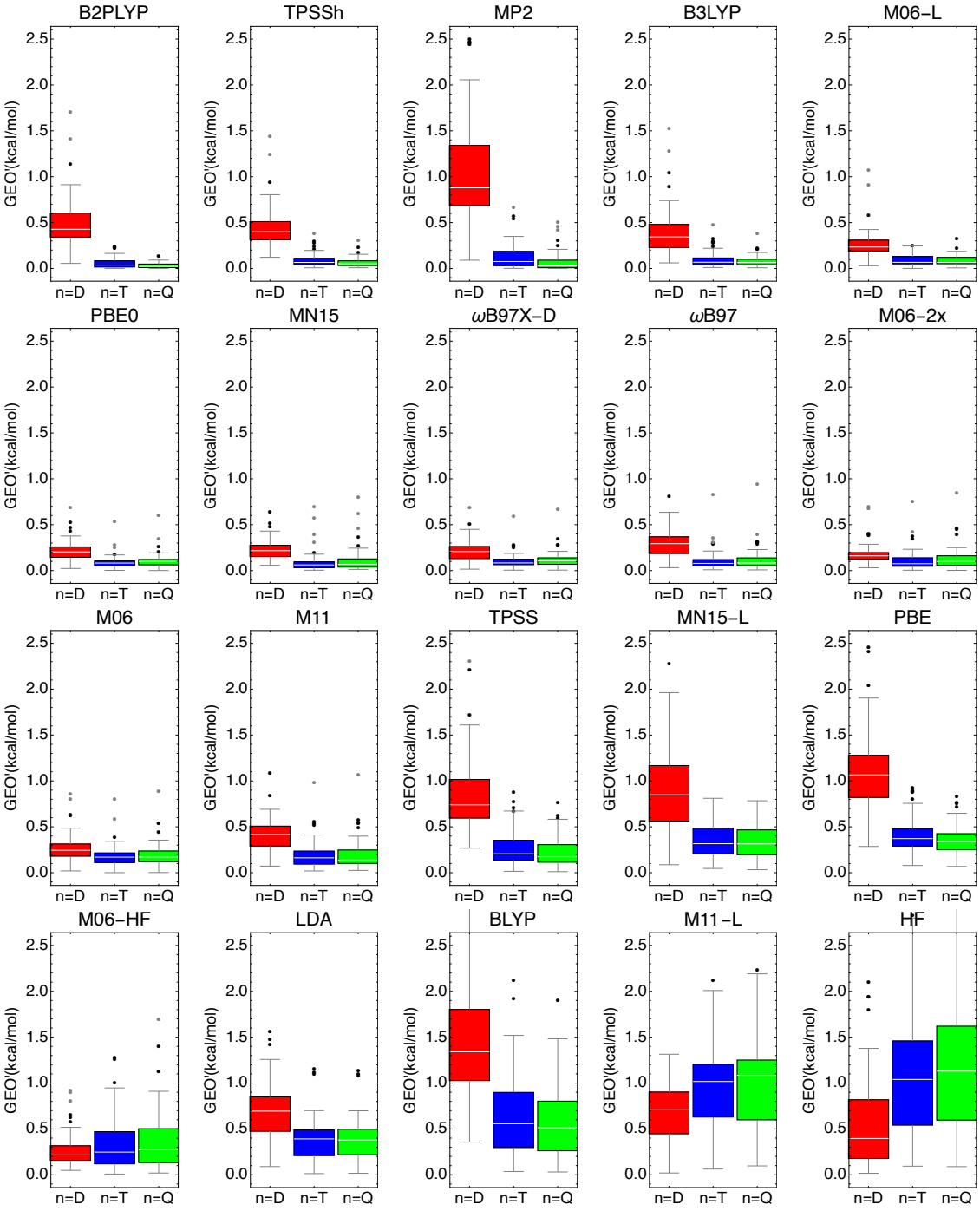


Figure S8: Same as Fig. S5, but showing the γ' results in place of GEO'.

S5 Accuracy of hybrids for molecular geometries as a function of the amount of exact exchange

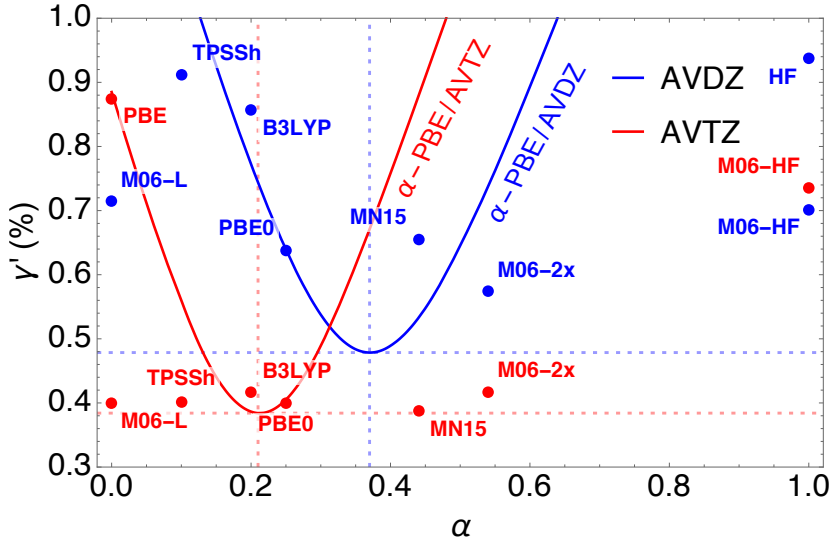


Figure S9: Same as the bottom panel of Fig. 7, but showing the γ' results in place of GEO'.

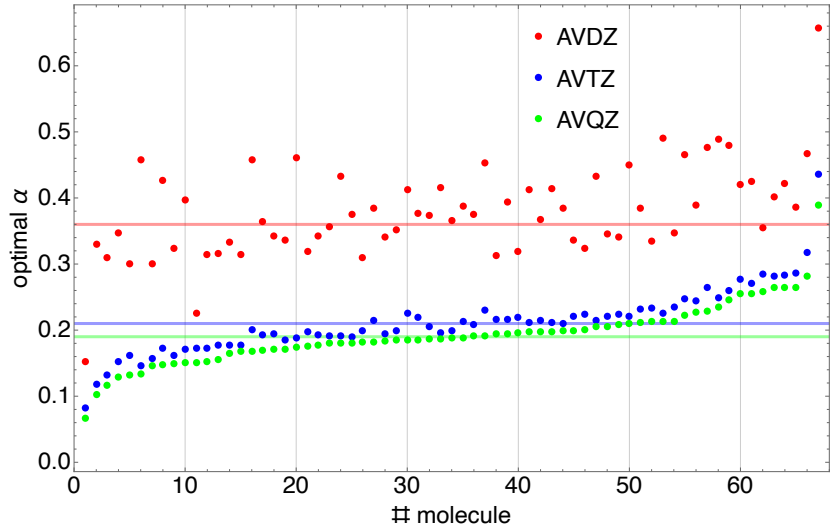


Figure S10: Optimal α values (ones giving the lowest GEO') for α-PBE at the three basis sets for each of the B2se molecules. The dashed horizontal lines go through the α values that minimize mean GEO' for B2se. The molecules are ordered such that their optimal α within the AVQZ basis set increases as we go from the left to the right. The underlying molecular structures ordered in this way are given in Figure S13.

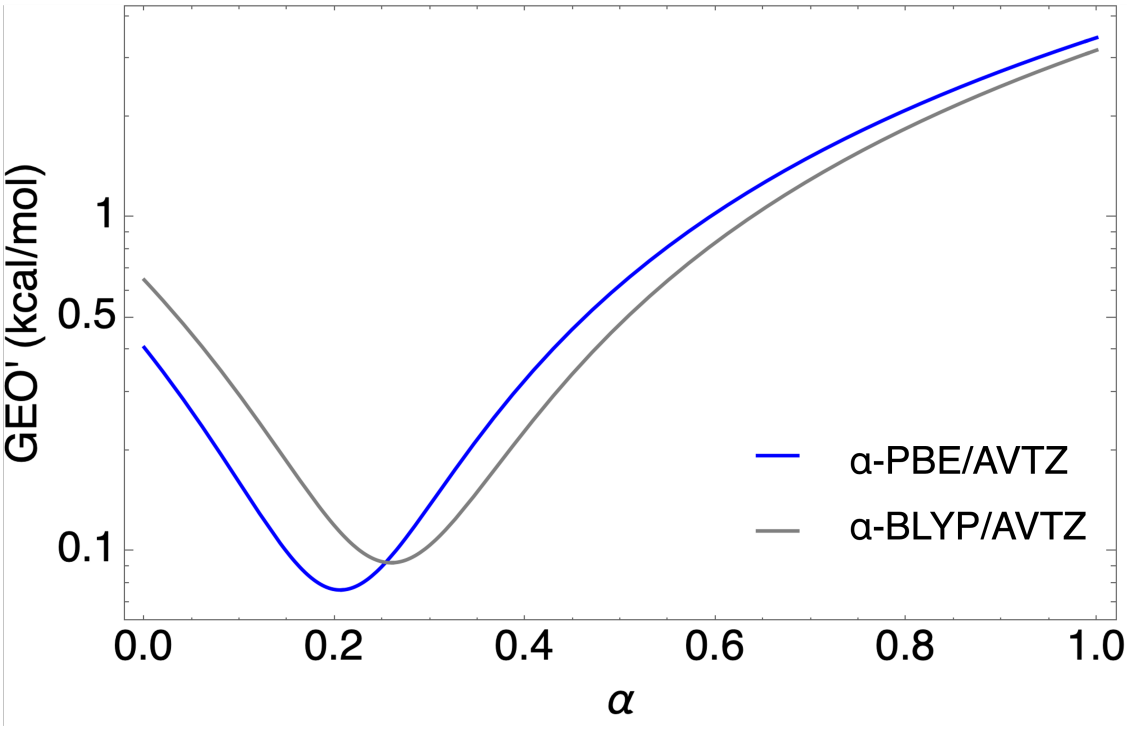


Figure S11: Mean GEO' values of the α-PBE and α-BLYP as a function of α (amount of the exact exchange mixing). Aug-cc-pVTZ basis set was used in all calculations. α-PBE and α-BLYP stand for exchange-correlation functionals where the α-amount of exact exchange replaces the semilocal exchange used in PBE and BLYP, respectively. BLYP employs B88 exchange[5] and LYP correlation,[6] whereas PBE employs PBE exchange and PBE correlation.[7] Note the *log*-scale on the y-axis.

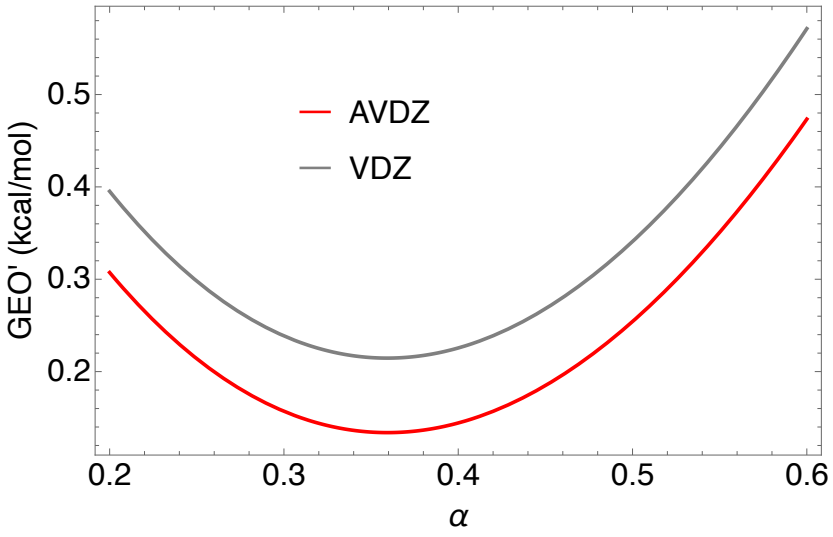


Figure S12: Plots comparing AVDZ and VDZ mean GEO' for α-PBE as a function of α for the B2se dataset.

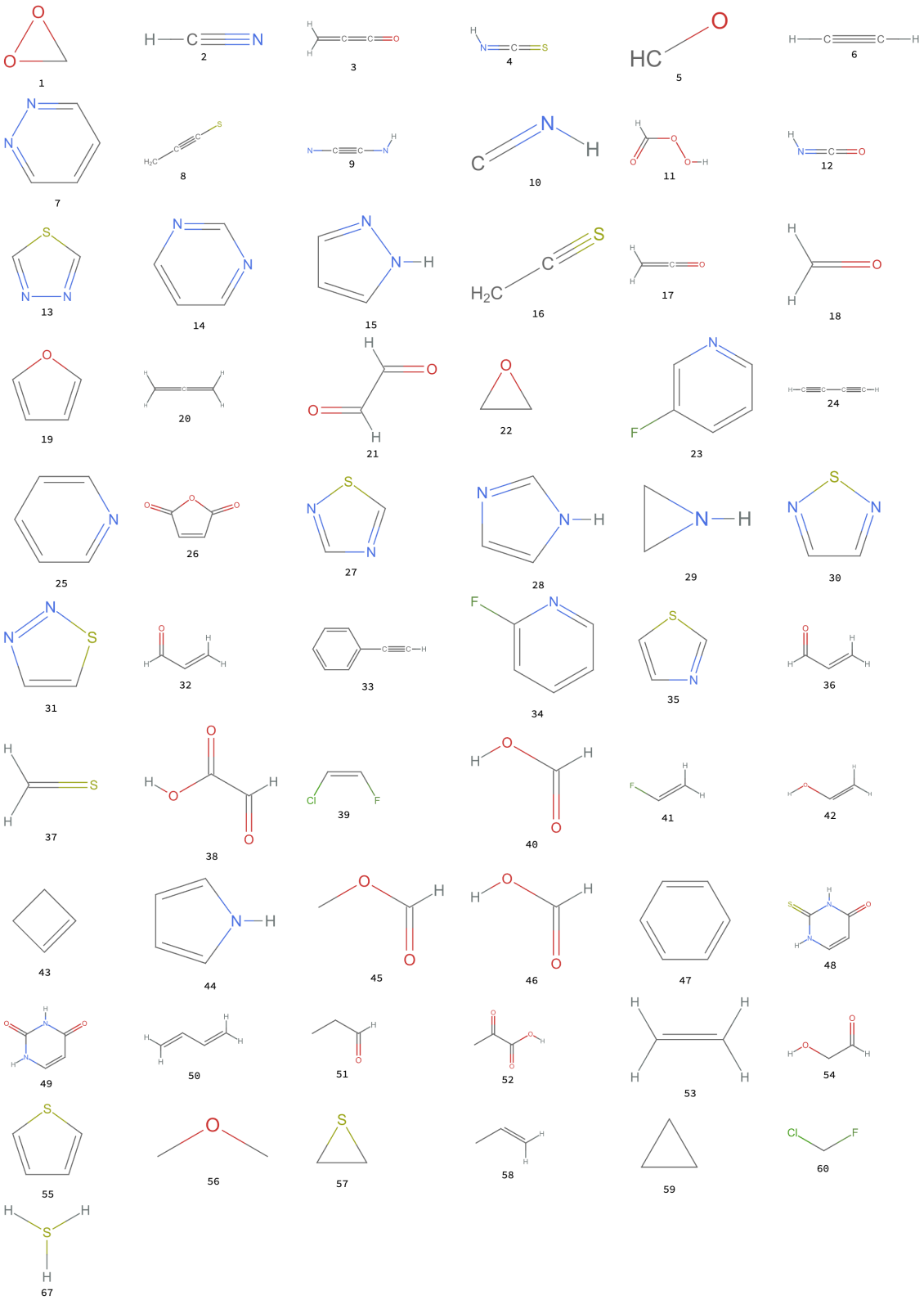


Figure S13: The order of molecules of the plot in Fig. S10.

S6 Geometric performance of Grimme's '3-c' composite methods

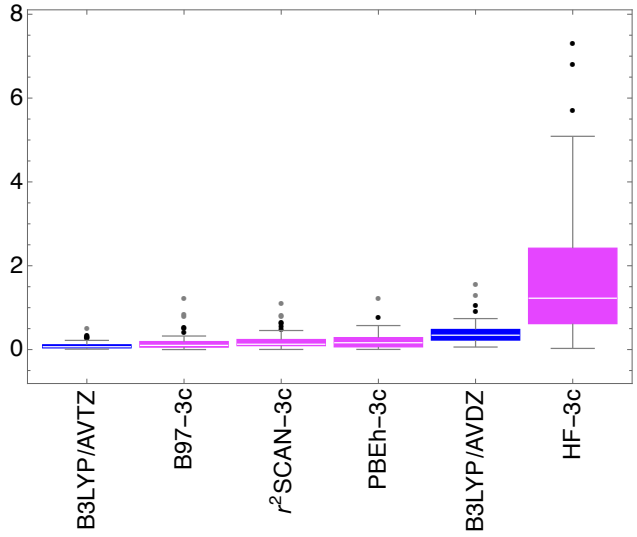
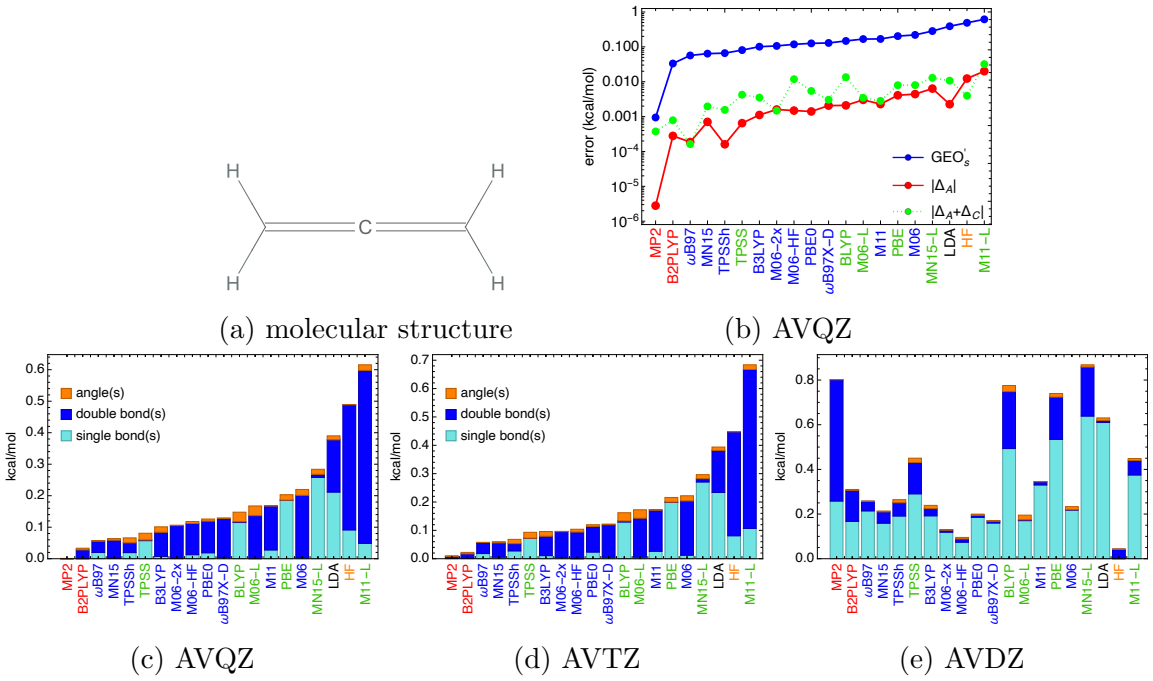


Figure S14: Same as Fig. 9, but for a full range on the y-axes.

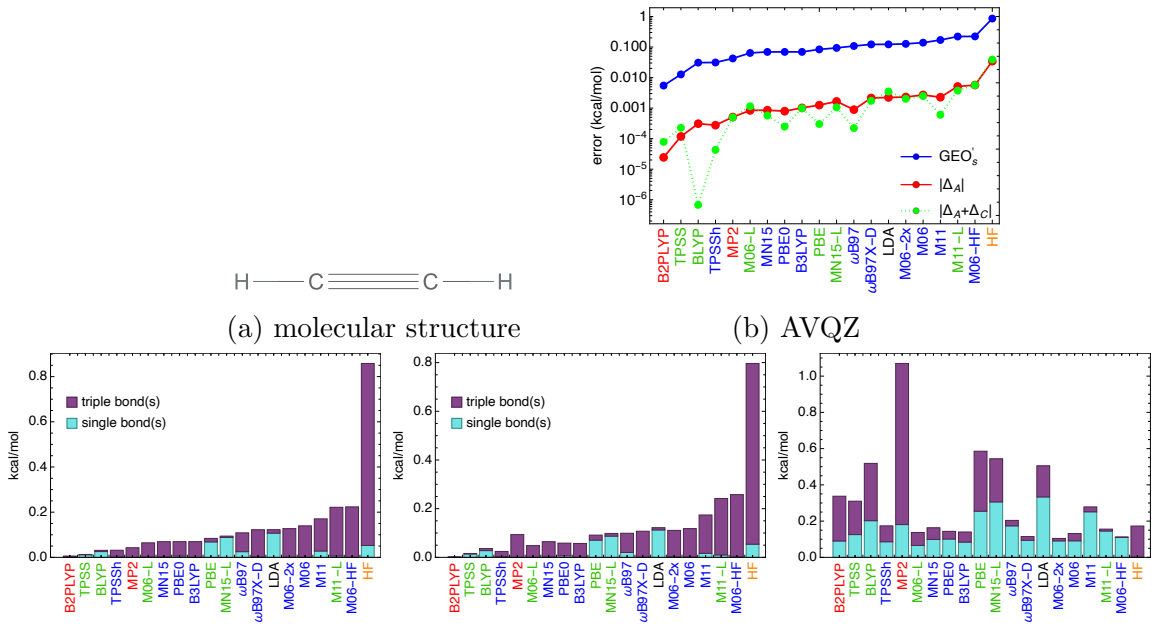
S7 Breakdown of GEO' into contributions due to errors in specific geometric parameters for selected B2se molecules



(a) molecular structure (b) AVQZ

(c) AVQZ (d) AVTZ (e) AVDZ

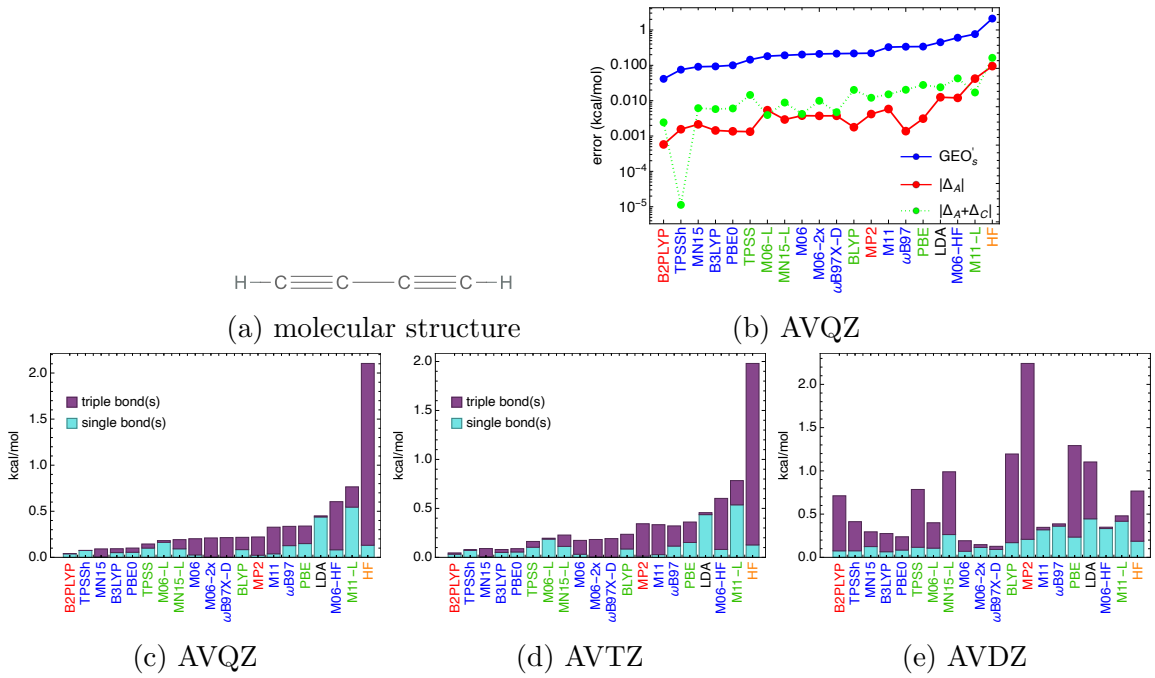
Figure S16: GEO'_{s} analysis for allene: (a) 2D structure; (b) GEO'_{s}, Δ_A, and Δ_C values from approximations within the AVQZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_{s} into different components from approximate methods within: (c) AVQZ; (d) AVTZ; (e) AVDZ basis set. The approximations in all panels are ranked by the GEO'_{s} values in (c).



(a) molecular structure (b) AVQZ

(c) AVQZ (d) AVTZ (e) AVDZ

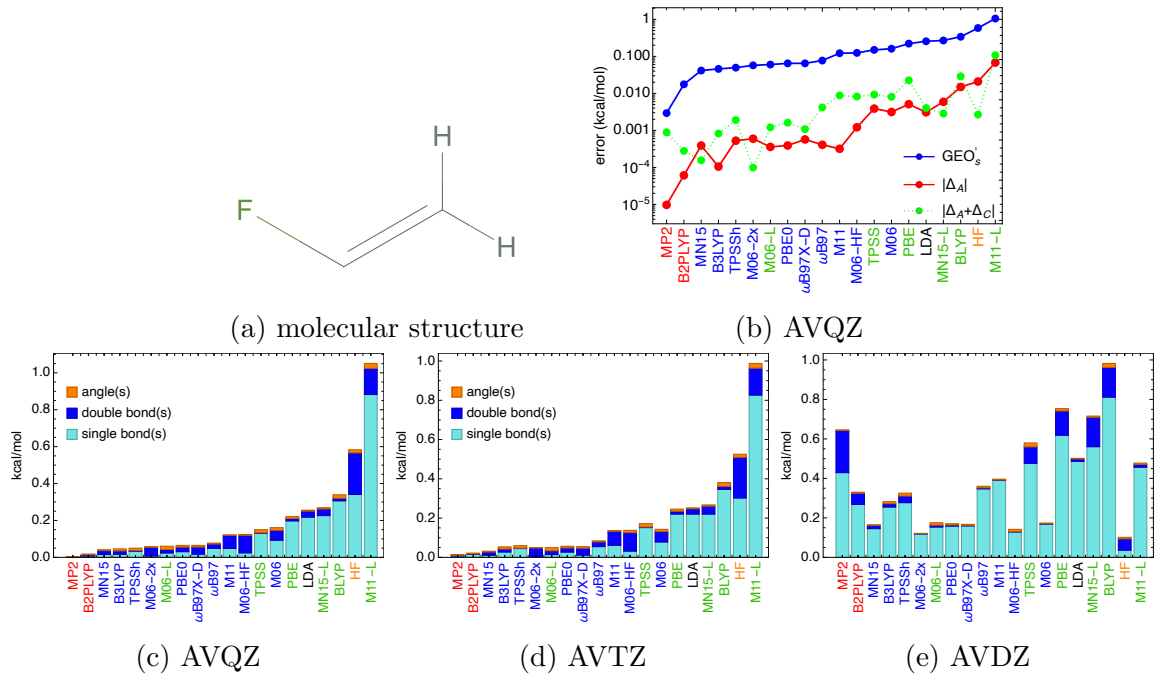
Figure S18: GEO'_s analysis for acetylene: (a) 2D structure; (b) GEO'_s, Δ_A, and Δ_C values from approximations within the AVQZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_s into different components from approximate methods within: (c) AVQZ; (d) AVTZ; (e) AVDZ basis set. The approximations in all panels are ranked by the GEO'_s values in (c).



(a) molecular structure (b) AVQZ

(c) AVQZ (d) AVTZ (e) AVDZ

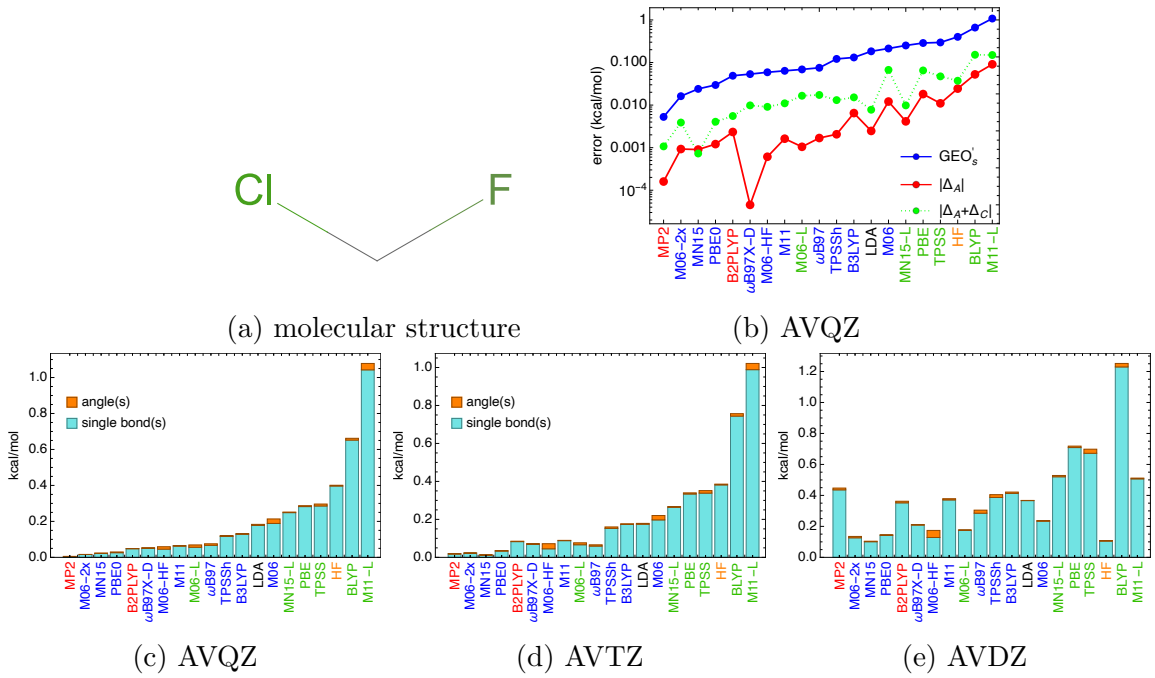
Figure S20: GEO'_s analysis for buta-1,3-diyne: (a) 2D structure; (b) GEO'_s, Δ_A, and Δ_C values from approximations within the AVQZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_s into different components from approximate methods within: (c) AVQZ; (d) AVTZ; (e) AVDZ basis set. The approximations in all panels are ranked by the GEO'_s values in (c).



(a) molecular structure (b) AVQZ

(c) AVQZ (d) AVTZ (e) AVDZ

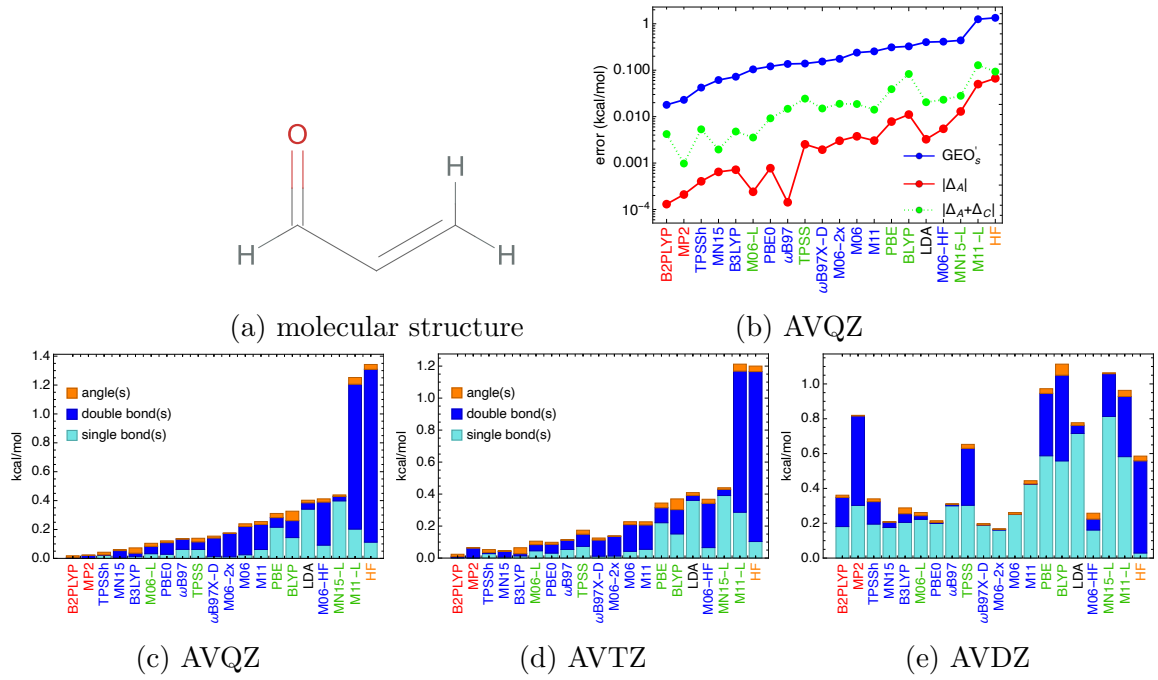
Figure S22: GEO'_s analysis for fluoroethene: (a) 2D structure; (b) GEO'_s, Δ_A, and Δ_C values from approximations within the AVQZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_s into different components from approximate methods within: (c) AVQZ; (d) AVTZ; (e) AVDZ basis set. The approximations in all panels are ranked by the GEO'_s values in (c).



(a) molecular structure (b) AVQZ

(c) AVQZ (d) AVTZ (e) AVDZ

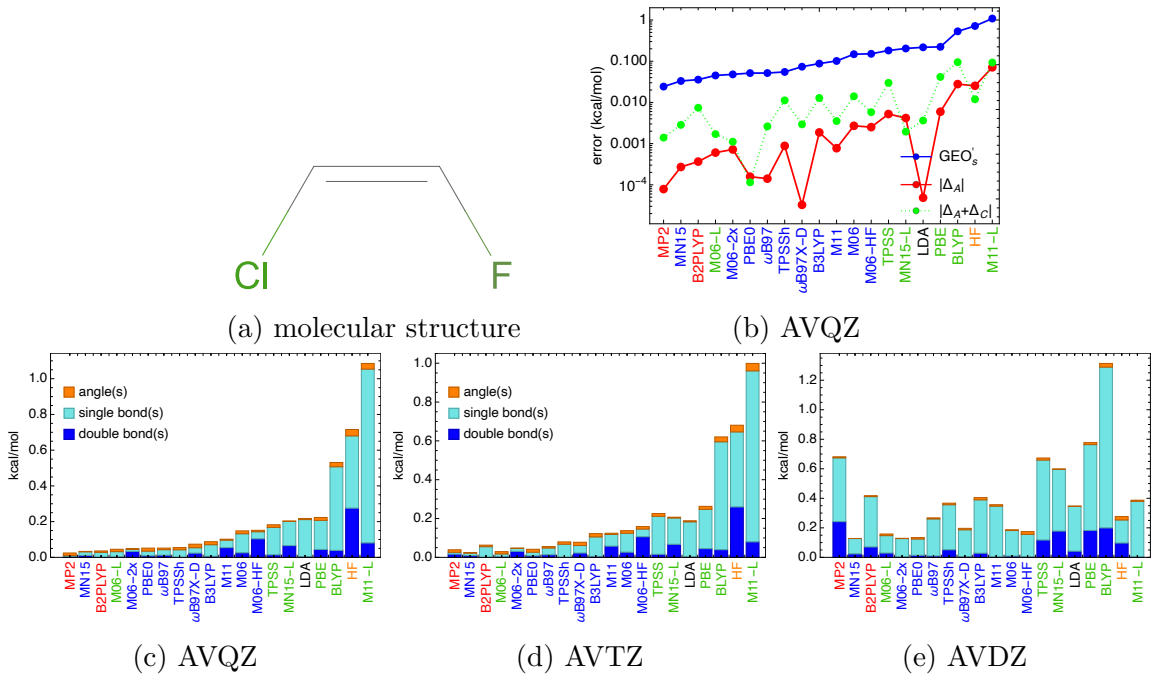
Figure S24: GEO'_s analysis for chlorofluoromethane: (a) 2D structure; (b) GEO'_s, Δ_A, and Δ_C values from approximations within the AVQZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_s into different components from approximate methods within: (c) AVQZ; (d) AVTZ; (e) AVDZ basis set. The approximations in all panels are ranked by the GEO'_s values in (c).



(a) molecular structure (b) AVQZ

(c) AVQZ (d) AVTZ (e) AVDZ

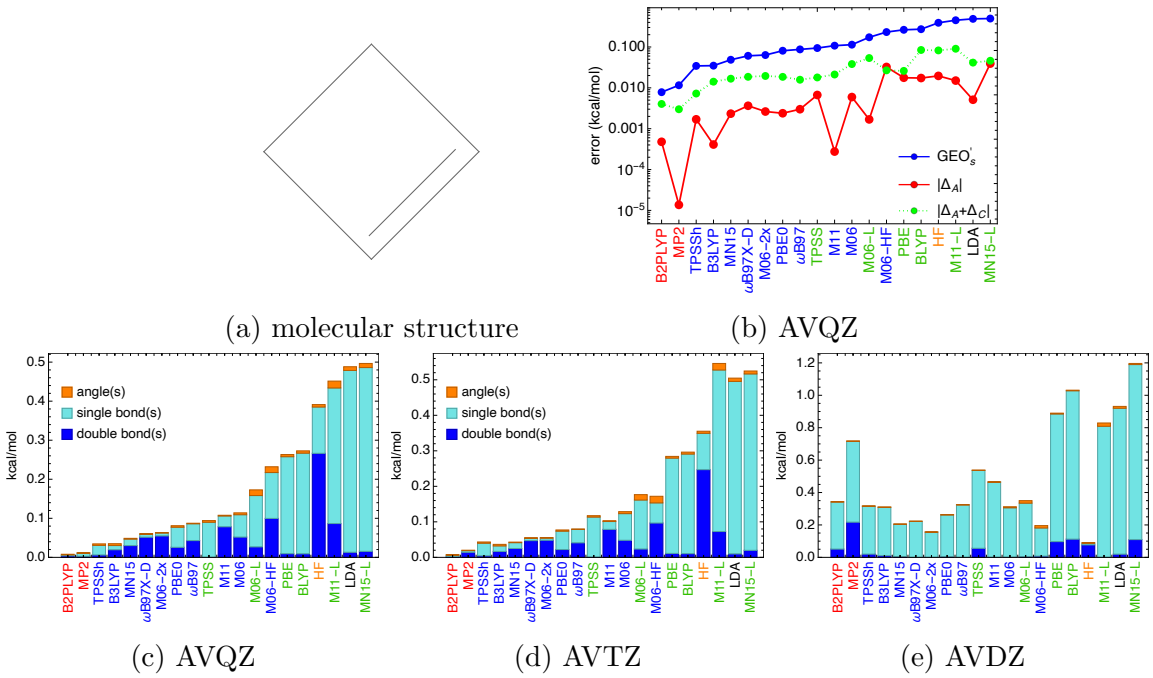
Figure S26: GEO'_{s} analysis for prop-2-enal: (a) 2D structure; (b) GEO'_{s}, Δ_A, and Δ_C values from approximations within the AVQZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_{s} into different components from approximate methods within: (c) AVQZ; (d) AVTZ; (e) AVDZ basis set. The approximations in all panels are ranked by the GEO'_{s} values in (c).



(a) molecular structure (b) AVQZ

(c) AVQZ (d) AVTZ (e) AVDZ

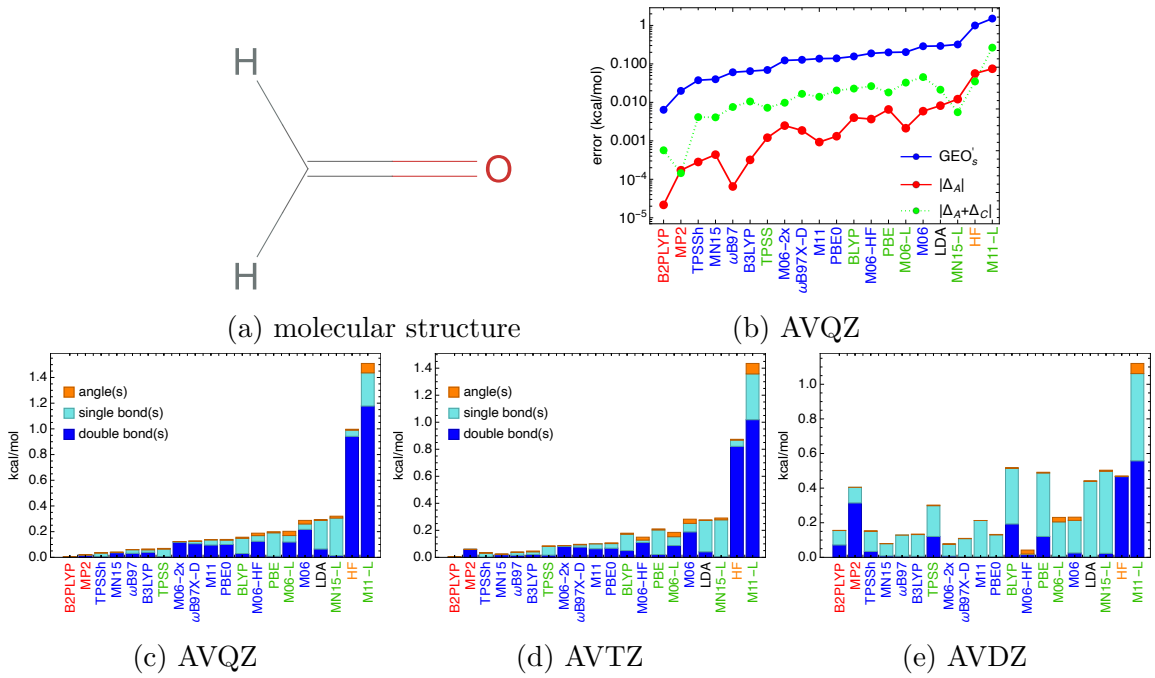
Figure S28: GEO'_s analysis for (Z)-1-chloro-2-fluoroethene: (a) 2D structure; (b) GEO'_s, Δ_A, and Δ_C values from approximations within the AVQZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_s into different components from approximate methods within: (c) AVQZ; (d) AVTZ; (e) AVDZ basis set. The approximations in all panels are ranked by the GEO'_s values in (c).



(a) molecular structure (b) AVQZ

(c) AVQZ (d) AVTZ (e) AVDZ

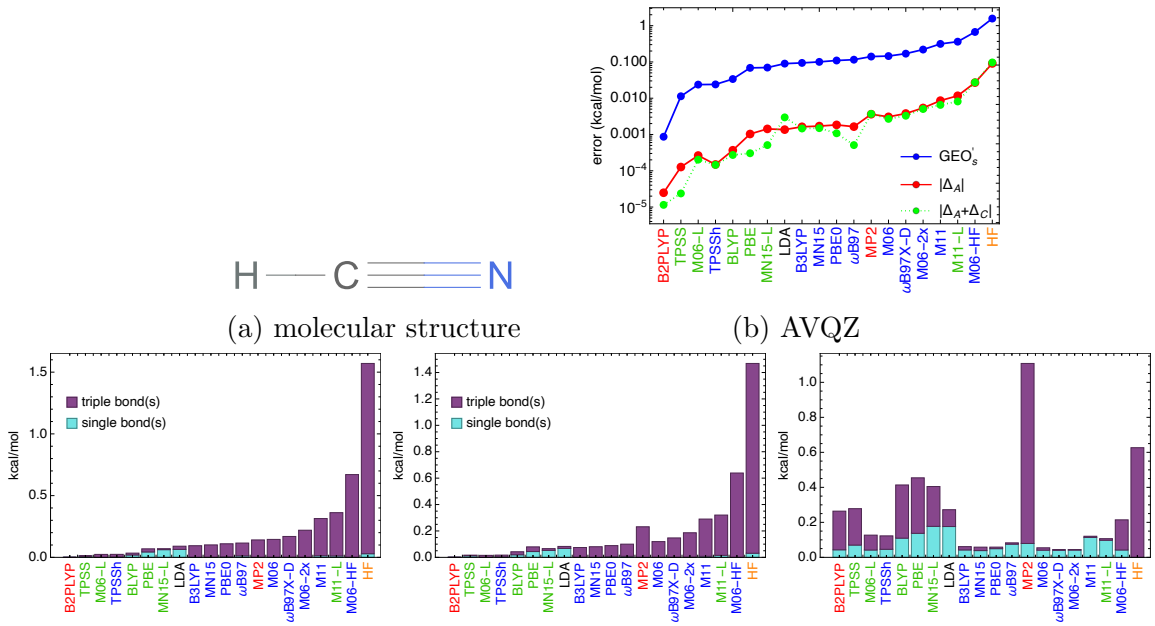
Figure S30: GEO'_s analysis for cyclobutene: (a) 2D structure; (b) GEO'_s, Δ_A, and Δ_C values from approximations within the AVQZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_s into different components from approximate methods within: (c) AVQZ; (d) AVTZ; (e) AVDZ basis set. The approximations in all panels are ranked by the GEO'_s values in (c).



(a) molecular structure (b) AVQZ

(c) AVQZ (d) AVTZ (e) AVDZ

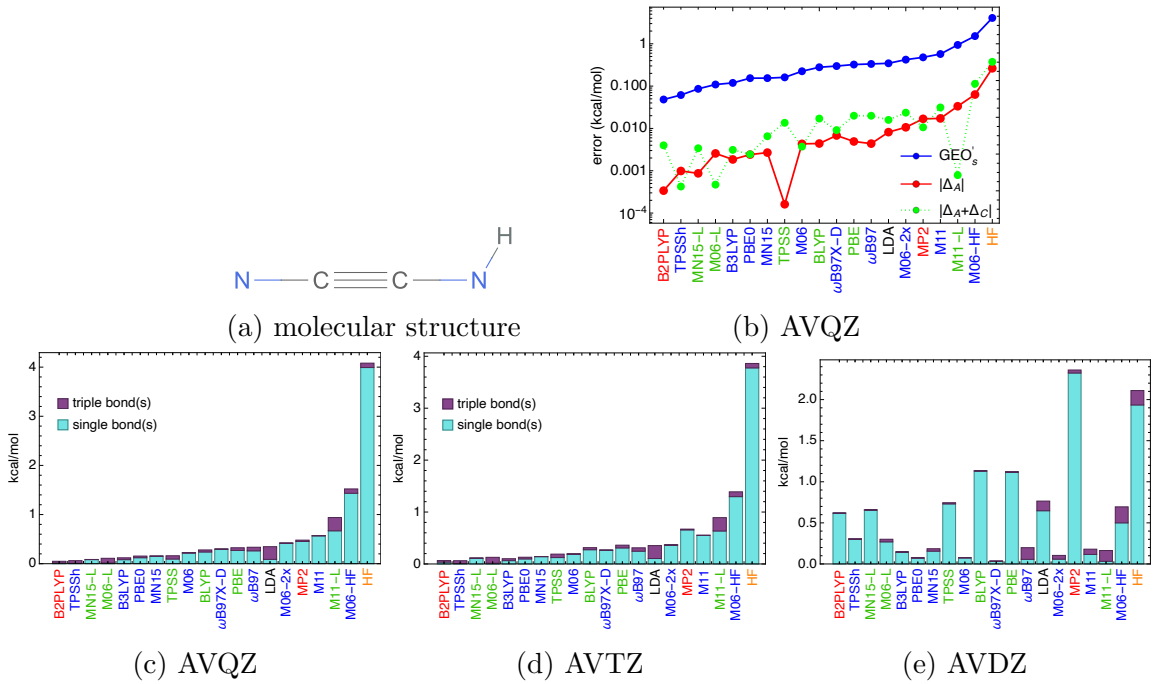
Figure S32: GEO'_s analysis for formaldehyde: (a) 2D structure; (b) GEO'_s, Δ_A, and Δ_C values from approximations within the AVQZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_s into different components from approximate methods within: (c) AVQZ; (d) AVTZ; (e) AVDZ basis set. The approximations in all panels are ranked by the GEO'_s values in (c).



(a) molecular structure (b) AVQZ

(c) AVQZ (d) AVTZ (e) AVDZ

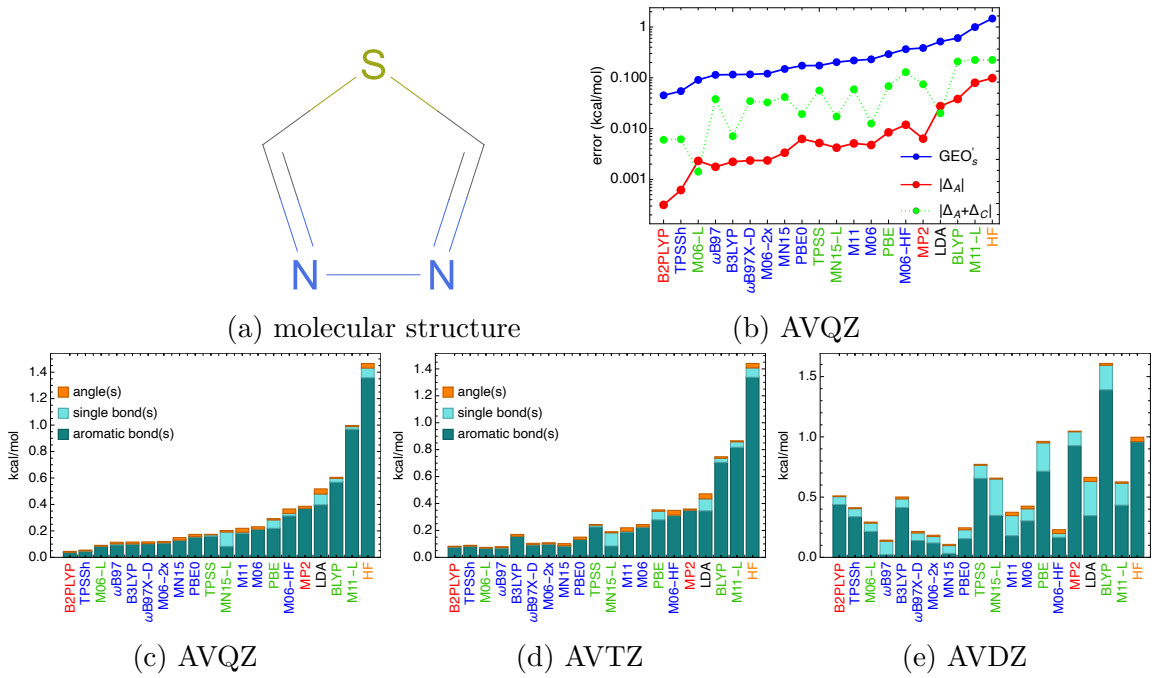
Figure S34: GEO'_s analysis for formonitrile: (a) 2D structure; (b) GEO'_s, Δ_A, and Δ_C values from approximations within the AVQZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_s into different components from approximate methods within: (c) AVQZ; (d) AVTZ; (e) AVDZ basis set. The approximations in all panels are ranked by the GEO'_s values in (c).



(a) molecular structure (b) AVQZ

(c) AVQZ (d) AVTZ (e) AVDZ

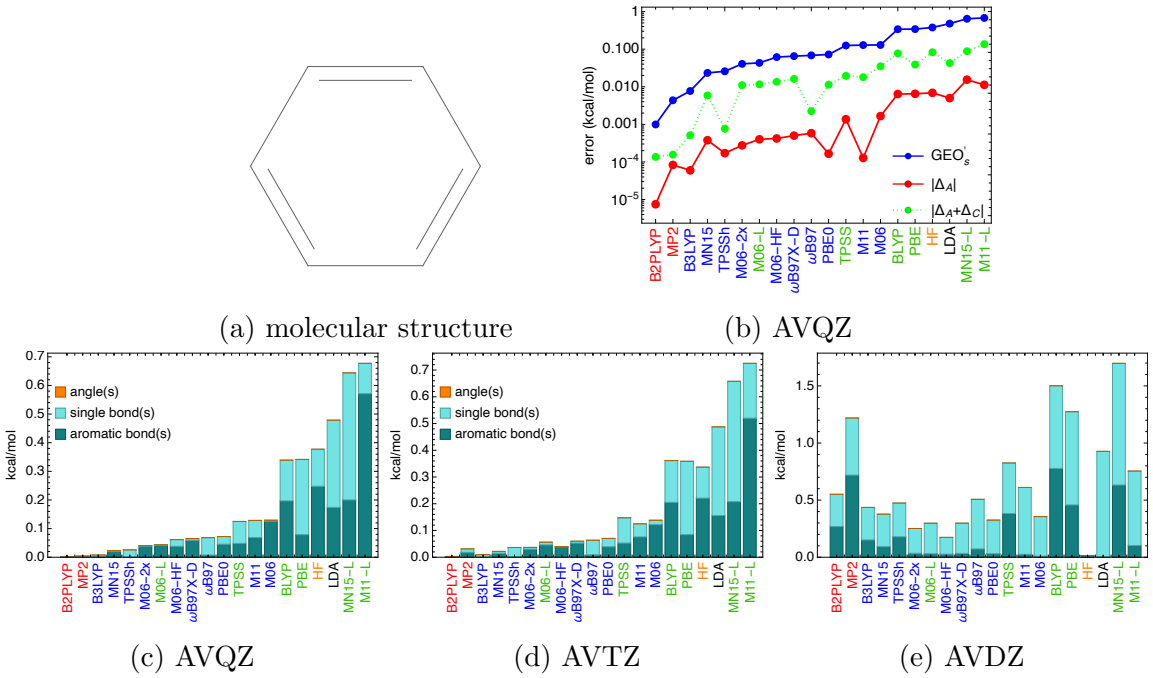
Figure S36: GEO'_s analysis for $HNCCN^+$: (a) 2D structure; (b) GEO'_s, Δ_A, and Δ_C values from approximations within the AVQZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_s into different components from approximate methods within: (c) AVQZ; (d) AVTZ; (e) AVDZ basis set. The approximations in all panels are ranked by the GEO'_s values in (c).



(a) molecular structure (b) AVQZ

(c) AVQZ (d) AVTZ (e) AVDZ

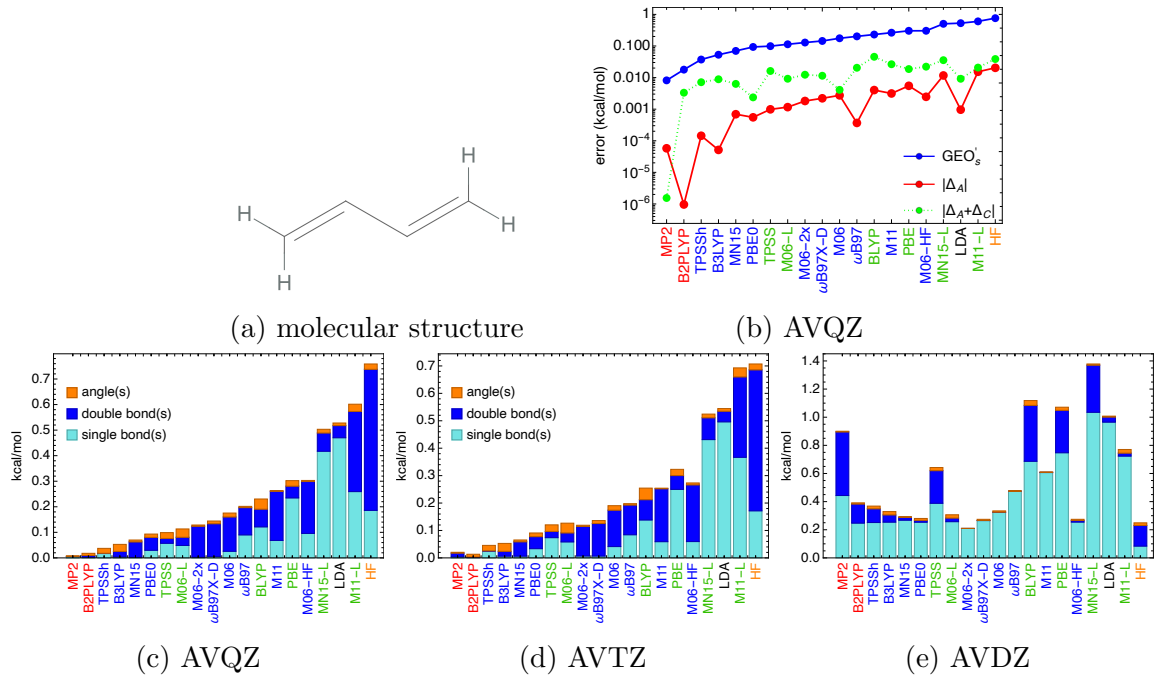
Figure S38: GEO'_s analysis for 1,3,4-thiadiazole: (a) 2D structure; (b) GEO'_s, Δ_A, and Δ_C values from approximations within the AVQZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_s into different components from approximate methods within: (c) AVQZ; (d) AVTZ; (e) AVDZ basis set. The approximations in all panels are ranked by the GEO'_s values in (c).



(a) molecular structure (b) AVQZ

(c) AVQZ (d) AVTZ (e) AVDZ

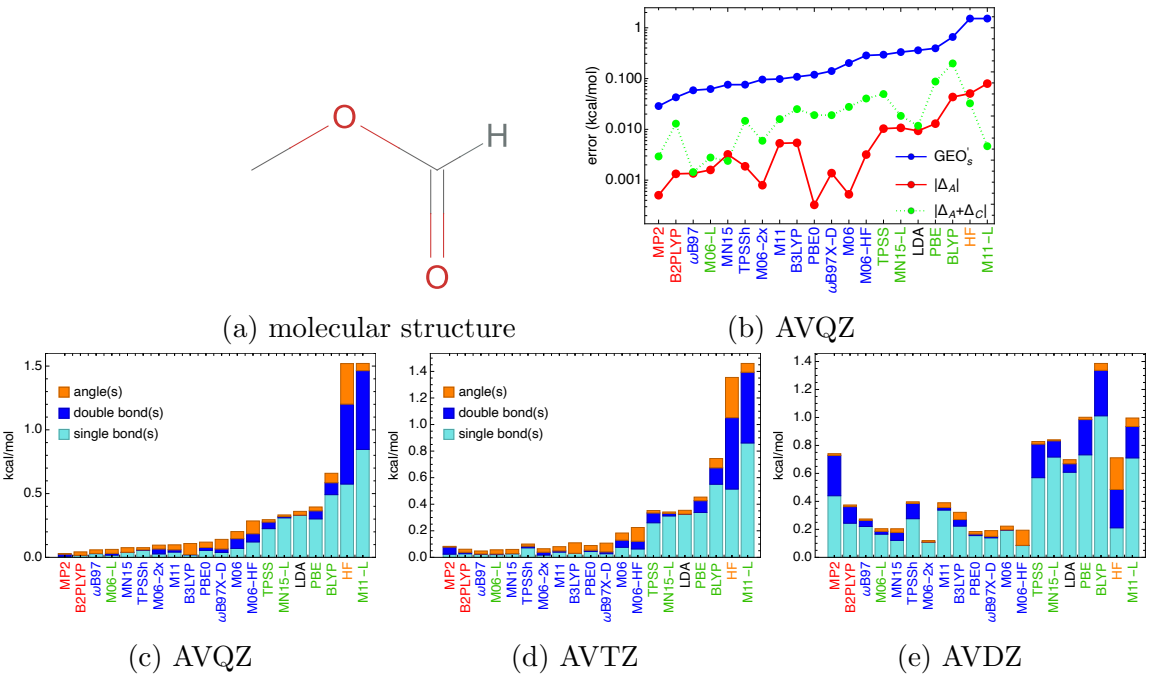
Figure S40: GEO'_{s} analysis for benzene: (a) 2D structure; (b) GEO'_{s}, Δ_A, and Δ_C values from approximations within the AVQZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_{s} into different components from approximate methods within: (c) AVQZ; (d) AVTZ; (e) AVDZ basis set. The approximations in all panels are ranked by the GEO'_{s} values in (c).



(a) molecular structure (b) AVQZ

(c) AVQZ (d) AVTZ (e) AVDZ

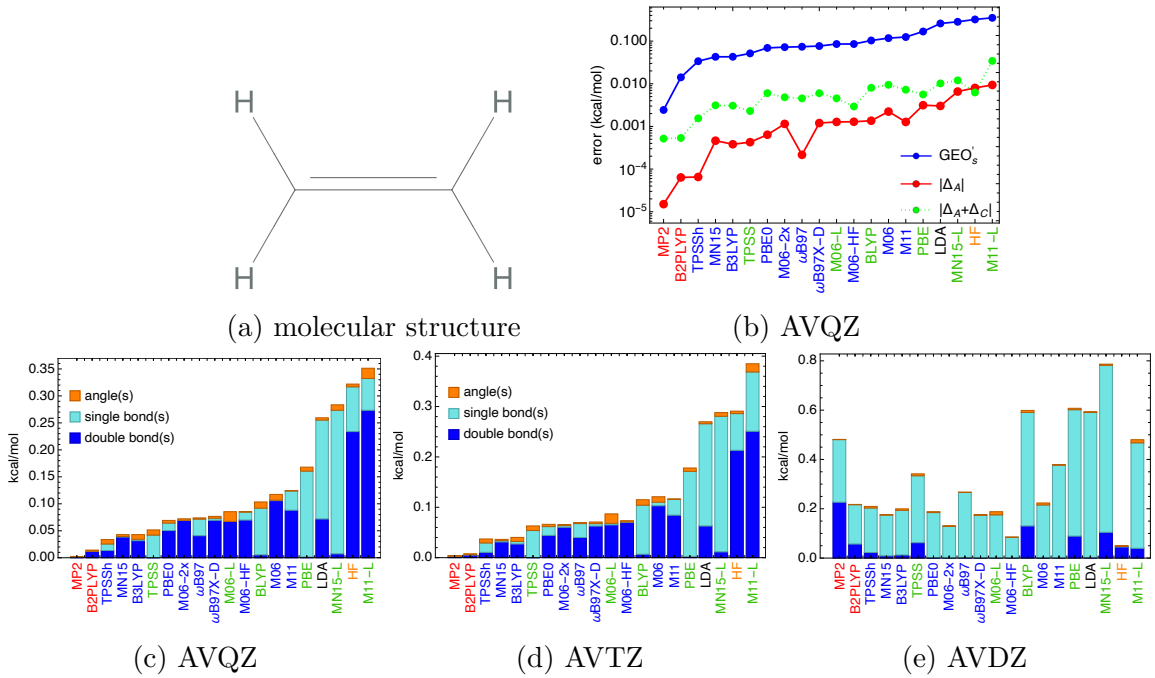
Figure S42: GEO'_s analysis for buta-1,3-diene: (a) 2D structure; (b) GEO'_s, Δ_A, and Δ_C values from approximations within the AVQZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_s into different components from approximate methods within: (c) AVQZ; (d) AVTZ; (e) AVDZ basis set. The approximations in all panels are ranked by the GEO'_s values in (c).



(a) molecular structure (b) AVQZ

(c) AVQZ (d) AVTZ (e) AVDZ

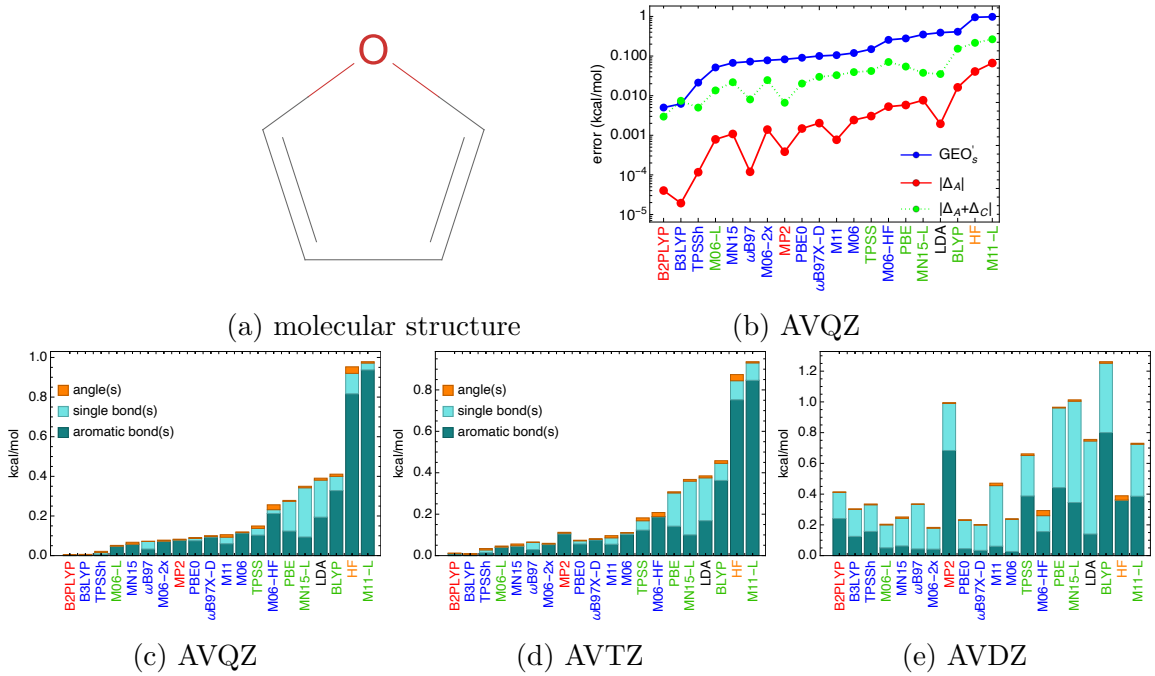
Figure S44: GEO'_{s} analysis for methyl formate: (a) 2D structure; (b) GEO'_{s}, Δ_A, and Δ_C values from approximations within the AVQZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_{s} into different components from approximate methods within: (c) AVQZ; (d) AVTZ; (e) AVDZ basis set. The approximations in all panels are ranked by the GEO'_{s} values in (c).



(a) molecular structure (b) AVQZ

(c) AVQZ (d) AVTZ (e) AVDZ

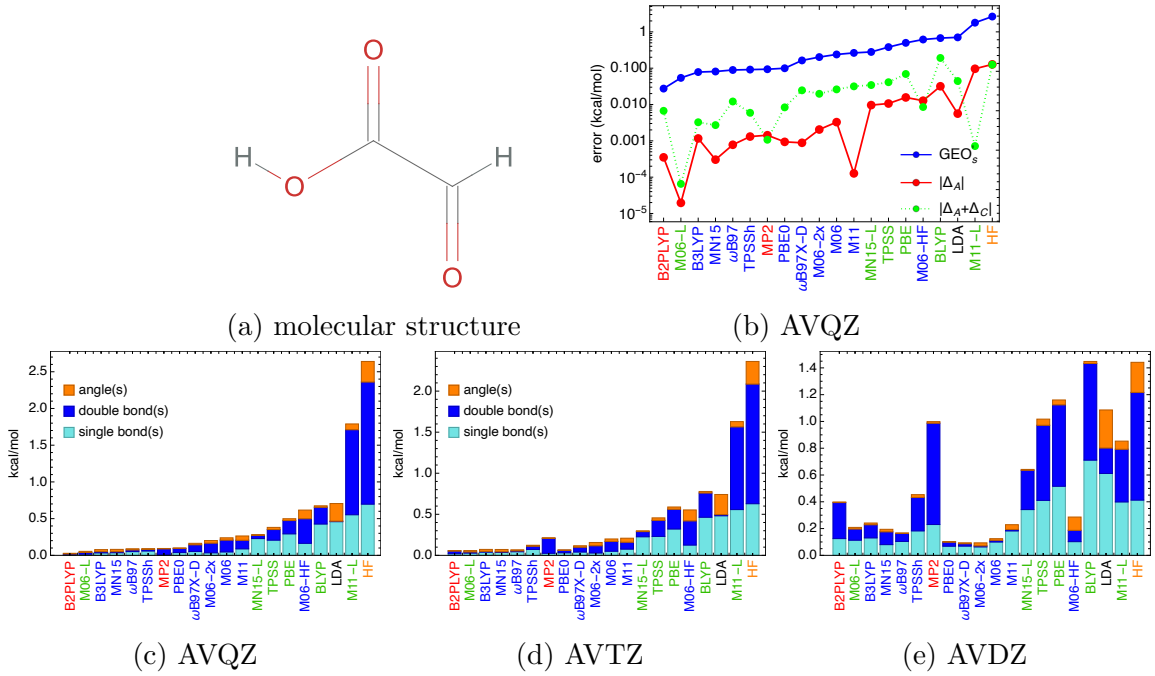
Figure S46: GEO'_s analysis for ethylene: (a) 2D structure; (b) GEO'_s, Δ_A, and Δ_C values from approximations within the AVQZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_s into different components from approximate methods within: (c) AVQZ; (d) AVTZ; (e) AVDZ basis set. The approximations in all panels are ranked by the GEO'_s values in (c).



(a) molecular structure (b) AVQZ

(c) AVQZ (d) AVTZ (e) AVDZ

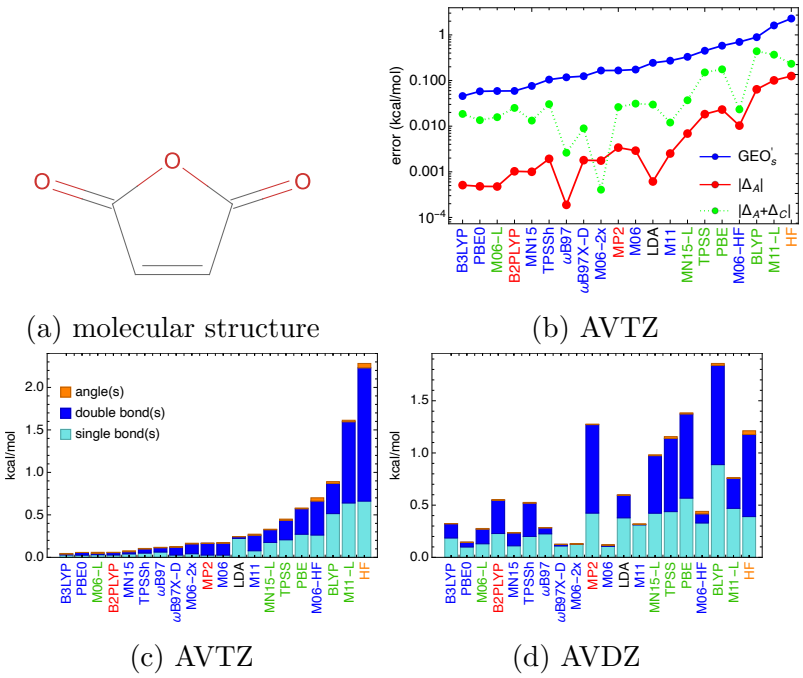
Figure S48: GEO'_{s} analysis for furan: (a) 2D structure; (b) GEO'_{s}, Δ_A, and Δ_C values from approximations within the AVQZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_{s} into different components from approximate methods within: (c) AVQZ; (d) AVTZ; (e) AVDZ basis set. The approximations in all panels are ranked by the GEO'_{s} values in (c).



(a) molecular structure (b) AVQZ

(c) AVQZ (d) AVTZ (e) AVDZ

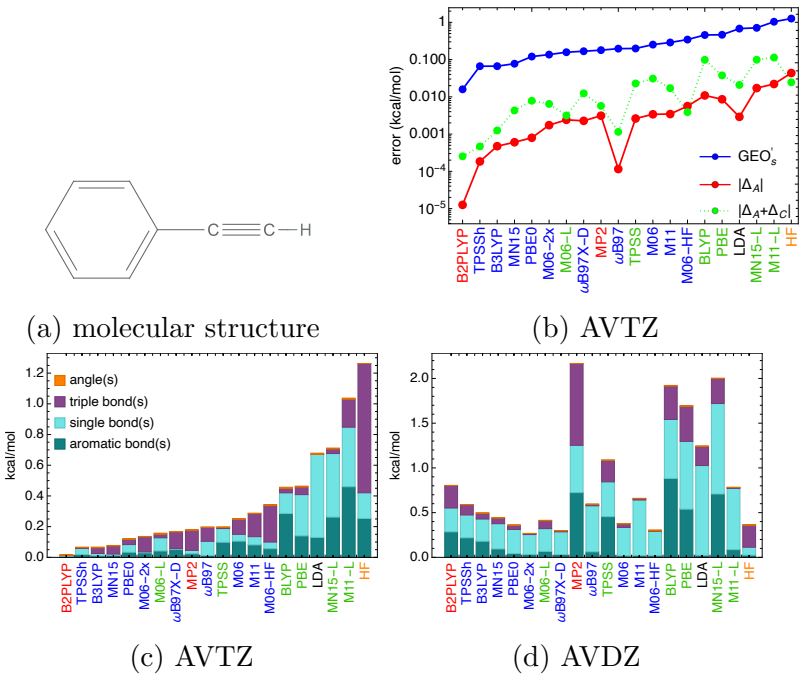
Figure S50: GEO'_s analysis for glyoxylic acid: (a) 2D structure; (b) GEO'_s, Δ_A, and Δ_C values from approximations within the AVQZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_s into different components from approximate methods within: (c) AVQZ; (d) AVTZ; (e) AVDZ basis set. The approximations in all panels are ranked by the GEO'_s values in (c).



(a) molecular structure (b) AVTZ

(c) AVTZ (d) AVDZ

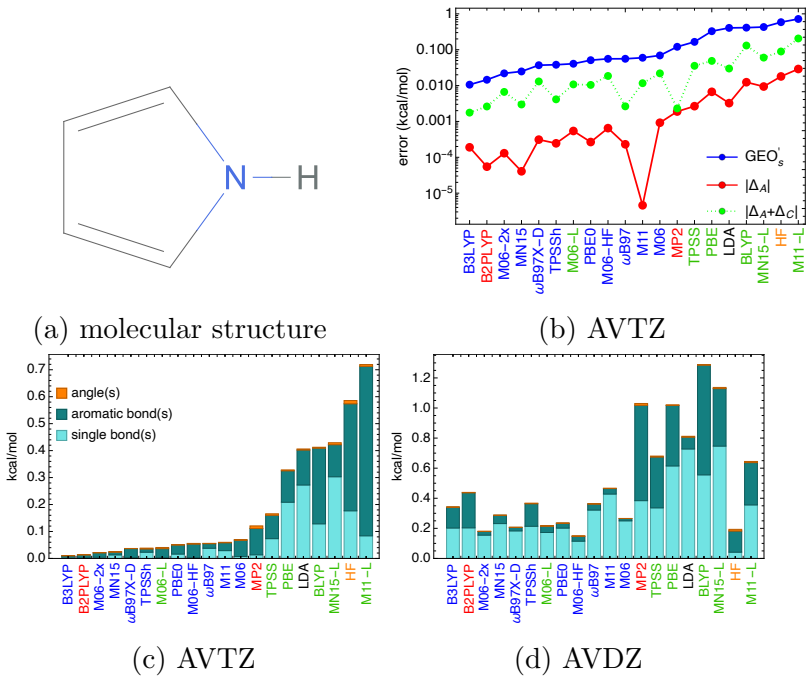
Figure S52: GEO'_s analysis for furan-2,5-dione: (a) 2D structure; (b) GEO'_s, Δ_A, and Δ_C values from approximations within the AVTZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_s into different components from approximate methods within: (c) AVTZ; (d) AVDZ basis set. The approximations in both panels are ranked by the GEO'_s values in (c).



(a) molecular structure (b) AVTZ

(c) AVTZ (d) AVDZ

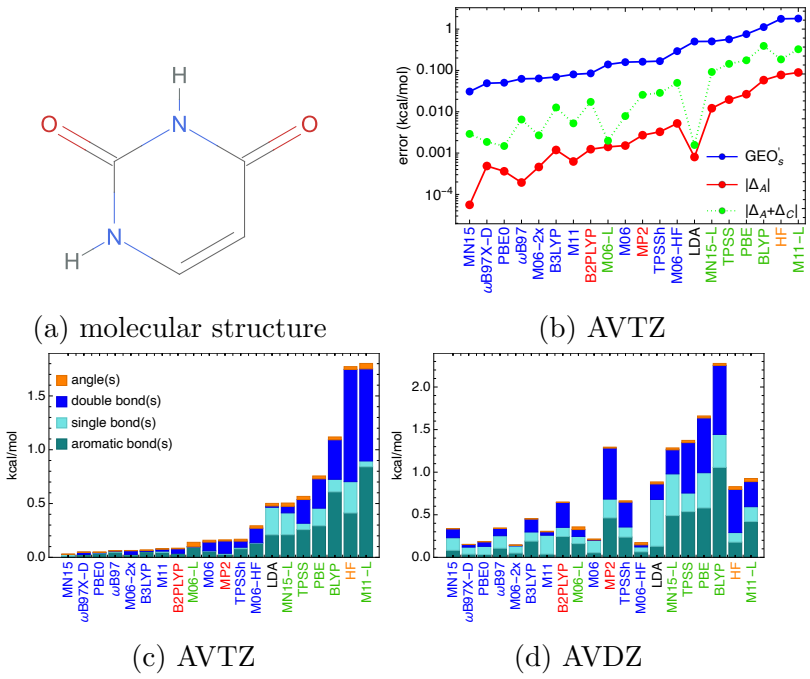
Figure S54: GEO'_s analysis for ethynylbenzene: (a) 2D structure; (b) GEO'_s, Δ_A, and Δ_C values from approximations within the AVTZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_s into different components from approximate methods within: (c) AVTZ; (d) AVDZ basis set. The approximations in both panels are ranked by the GEO'_s values in (c).



(a) molecular structure (b) AVTZ

(c) AVTZ (d) AVDZ

Figure S56: GEO'_s analysis for 1H-pyrrole: (a) 2D structure; (b) GEO'_s, Δ_A, and Δ_C values from approximations within the AVTZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_s into different components from approximate methods within: (c) AVTZ; (d) AVDZ basis set. The approximations in both panels are ranked by the GEO'_s values in (c).



(a) molecular structure (b) AVTZ

(c) AVTZ (d) AVDZ

Figure S58: GEO'_s analysis for 1H-pyrimidine-2,4-dione: (a) 2D structure; (b) GEO'_s, Δ_A, and Δ_C values from approximations within the AVTZ basis set (note the *log*-scale on the y axes). The breakdown of GEO'_s into different components from approximate methods within: (c) AVTZ; (d) AVDZ basis set. The approximations in both panels are ranked by the GEO'_s values in (c).

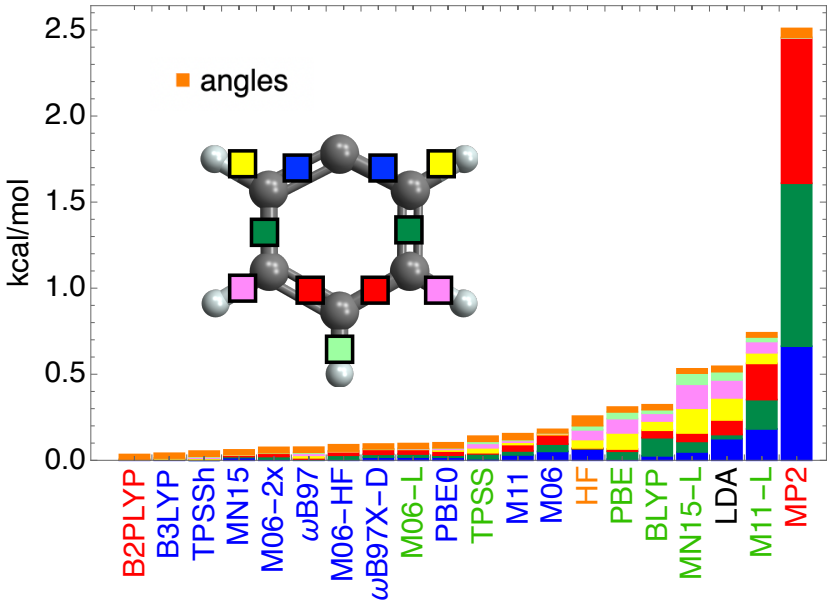


Figure S59: GEO' contributions from errors in each of the bond length for the phenyl radical. AVTZ used in all calculations.

References

(1) Vuckovic, S.; Burke, K. Quantifying and Understanding Errors in Molecular Geometries. *The Journal of Physical Chemistry Letters* **2020**, *11*, 9957–9964.

(2) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian?16 Revision C.01. 2016; Gaussian Inc. Wallingford CT.

(3) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics* **1980**, *58*, 1200–1211.

(4) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA quantum chemistry program package. *The Journal of Chemical Physics* **2020**, *152*, 224108.

(5) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098.

(6) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical review B* **1988**, *37*, 785.

(7) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.