THE JOURNAL OF PHYSICAL CHEMISTRY Letters

pubs.acs.org/JPCL Letter

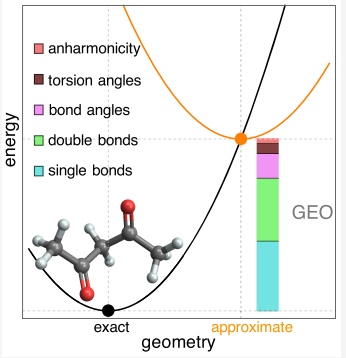
Quantifying and Understanding Errors in Molecular Geometries

Stefan Vuckovic and Kieron Burke*

Cite This: J. Phys. Chem. Lett. 2020, 11, 9957−9964 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Electronic structure calculations are ubiquitous in most branches of chemistry, but all have errors in both energies and equilibrium geometries. Quantifying errors in possibly dozens of bond angles and bond lengths is a Herculean task. A single natural measure of geometric error is introduced, the geometry energy offset (GEO). GEO links many disparate aspects of geometry errors: a new ranking of different methods, quantitative insight into errors in specific geometric parameters, and insight into trends with different methods. GEO can also reduce the cost of high-level geometry optimizations and shows when geometric errors distort the overall error of a method. Results, including some surprises, are given for both covalent and weak interactions.



Downloaded via VRIJE UNIV AMSTERDAM on November 24, 2020 at 11:59:00 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

Whenever one runs an electronic structure calculation within the Born−Oppenheimer approximation, whether it is a density functional calculation, *ab initio*, or semiempirical, of a molecule or a material, one must always answer the question: Which geometry should I use? Whatever the limitations of your method are, they will show up in giving an approximate energy at any given geometry which will minimize at some approximate geometry. Sometimes the differences between the exact geometry and the approximation are so slight that it does not matter. Whenever it does matter, common sense often dictates a choice: When comparing different methods, the requirement of apples-to-apples comparison means comparing several methods with a fixed geometry.[1−3] Other times, when the cost of a single calculation is severe, geometry optimization is prohibitively expensive, and one resorts to using geometries from a cheaper method.

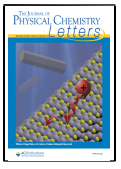
This problem is compounded when comparing geometric parameters computed with different methods. As a molecule grows in size, there are $3N - 6$ distinct degrees of freedom for the equilibrium structure, with errors in bond lengths, angles, etc. Some are more accurate in one method, some are better in another (see, e.g., ref 4). Should one average over all such parameters? But what if one method is better for bond lengths, and another for angles? And how do such errors in geometry correlate with other energetic errors?

We define the geometry energy offset (GEO) of a given method as

$$E_{\mathrm{geo}} = E(\tilde{\mathbf{G}}) - E(\mathbf{G}_0) \tag{1}$$

where $E(\mathbf{G})$ is the exact ground-state energy at geometry $\mathbf{G}$, $\mathbf{G}_0$ is the exact geometry, and $\tilde{\mathbf{G}}$ is an approximate geometry. We define the theoretically exact geometry as the exact minimum of the exact ground-state potential energy surface. For main-group chemistry molecules and approximations that we consider here, CCSD(T)/CBS provides sufficiently accurate reference both in terms of ground-state energies and geometries.[5−11] The simple definition of eq 1 leads to all the analysis and results contained in the paper. Figure 1 summarizes some of our most important results with GEO, with more details within this paper and the Supporting Information. On the left, we plot GEO energies averaged over a data set of small organic molecules (top). Here we use CCSD(T)/A′V5Z[12] as a reference, given its high accuracy for geometries of main-group molecules.[12] Accurate calculation of GEO using CCSD(T) is expensive (see Computational Details), but using the approximations themselves is less expensive and yields nearly identical results (E'_{geo}). We can see that different approximations perform characteristically well or poorly. Thus, every method can be ranked by its GEO value, and some perform much better for geometries than they do for, e.g., atomization energies. This is crucial information for understanding the accuracy of different methods for geometries. Directly below, we evaluate a much greater variety of methods for a data set of medium-sized organic molecules. Here, we use B2PLYP as a proxy reference, since it is the winner in the top panel and CCSD(T) is already too expensive. Thus, strictly speaking, this panel measures closest to B2PLYP geometries, which approximates GEO well for all but the best four methods (see Figure S3, where we analyze in

Received: October 4, 2020
Accepted: October 28, 2020
Published: November 10, 2020



© 2020 American Chemical Society

https://dx.doi.org/10.1021/acs.jpclett.0c03034
J. Phys. Chem. Lett. 2020, 11, 9957−9964

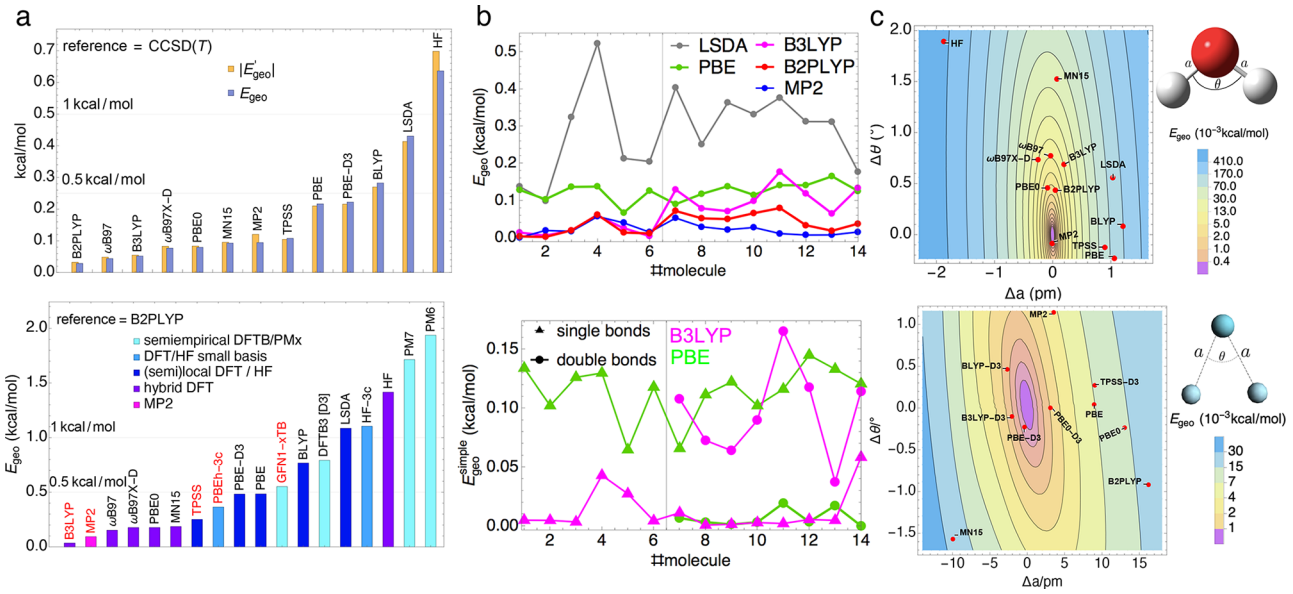


Figure 1. Why GEO is useful. Left: exact and approximate GEO rankings for quantum chemical methods on small molecules (top) and GEO of many different methods over medium-sized organic molecules (bottom) with B2PLYP as a proxy reference. For the lists of molecules and further details, see Figures S2−S7 and Tables S1−S3. Center: E_{geo} for a few methods and a few molecules (top) and single- and double-bond contributions to E_{geo}^{simple} for two popular methods, showing a huge difference in their accuracies (bottom). For the list of molecules, see Figure 3. Right: GEO contours as a function of errors in the bond angle and length for the water molecule (top panel) and weakly bonded Ne_2Ar (lower panel) and positions of different approximations.

more detail the accuracy of B2PLYP as a GEO reference). This shows some surprises: lower level (and less costly) methods can outperform higher level methods because they have been trained empirically. For example, the semiempirical GFN1-xTB method of Grimme and co-workers competes with DFT with the PBE functional[13] and outperforms DFT with BLYP.[14,15]

In the center, we show the very disparate behavior of two popular representative density functionals for single bonds and for double bonds. PBE, as a generalized gradient approximation (GGA) is far more accurate for double bonds than for singles. But a (global) hybrid, B3LYP, totally reverses this trend: more accurate for single bonds, but surprisingly far less accurate for double bonds. Explanations of the accuracy of hybrids[16,17] typically center on atomization energies, not bond lengths, and do not explain these trends. The top right panel shows a trade-off between angle- and bond-length errors for water. There are clear behaviors of different levels of density functionals. The local density approximation has noticeable errors in both the bond length and angle (but far smaller than those of HF). One can clearly see how GGAs like PBE and BLYP and the meta-GGA TPSS[18] greatly reduce the angular error, but have almost no effect on the bond length. Finally, by mixing some fraction (about 1/4) of exact exchange, PBE0 and B3LYP lie along a line joining their parent GGA to HF (see Figure S16 for more details on how the position of the approximation in the GEO contour plot depends on the exact exchange fraction). Furthermore, the mixing fraction of exact exchange almost perfectly cancels the bond-length error, while increasing the angle-error. Better functionals have about the same accuracy, while MP2 has almost perfect geometry.

All these results and trends are for strong covalent bonds. But GEO is even more important for weak interactions, where GEO energies can be comparable to the binding energy itself. To illustrate this, in the right, we contrast contours of GEO for two A_2B molecules, one covalently bonded and the other a noncovalent interaction: Water and the van der Waals trimer, Ne_2Ar, with the different methods from the left figures plotted as points in the plane. The noncovalent case is strikingly different. First, its binding energy is only 0.37 kcal/mol, so GEO errors are now more than relevant on this scale. This is accompanied by huge errors in bond length, related to the softness of the potential. Finally, the performance of different electronic structure methods is very different from the covalent case. For covalent bonds, MP2 is exceptionally good; for NCIs, density functional approximations are much better. These effects seem to have largely been ignored when ranking functionals for such complexes, which is usually done at a fixed geometry.[1−3] We expect improved performance for weak interaction methods once GEO errors are accounted for.

The rest of this paper explains how GEO works and shows how useful it can be. We focus on just three immediate applications: (i) obtaining insight into geometric errors in molecular benchmark energy sets; (ii) establishing an energetic scale comparing the quality of geometries from different approximate quantum-mechanical (QM) solvers; (iii) how this scale can be used for choosing a geometry optimization solver that has a good accuracy to cost ratio. By truncating E_{geo} to second order, we define E_{geo}^{harm} and E_{geo}^{simple} (see below), which we use to gain insight into how errors in different structural parameters translate to E_{geo} contributions. We apply our logic first to covalent bonds, where GEO is typically negligible relative to atomization energies, and then to noncovalent bonds, where GEO is often comparable to binding energies, and so is even more important.

Performance of Approximations. One of the most valuable uses of GEO is to rank different electronic structure methods for their geometric accuracy, as illustrated in the upper left panel of Figure 1. We stress that this ranking is quite different from traditional rankings by purely energetic performance, such as for atomization energies (AE). Roughly, GEO is typically about 1/80-th of the mean absolute error in AE (see Figure S2, where we show the correlation plot between GEO

https://dx.doi.org/10.1021/acs.jpclett.0c03034
J. Phys. Chem. Lett. 2020, 11, 9957−9964

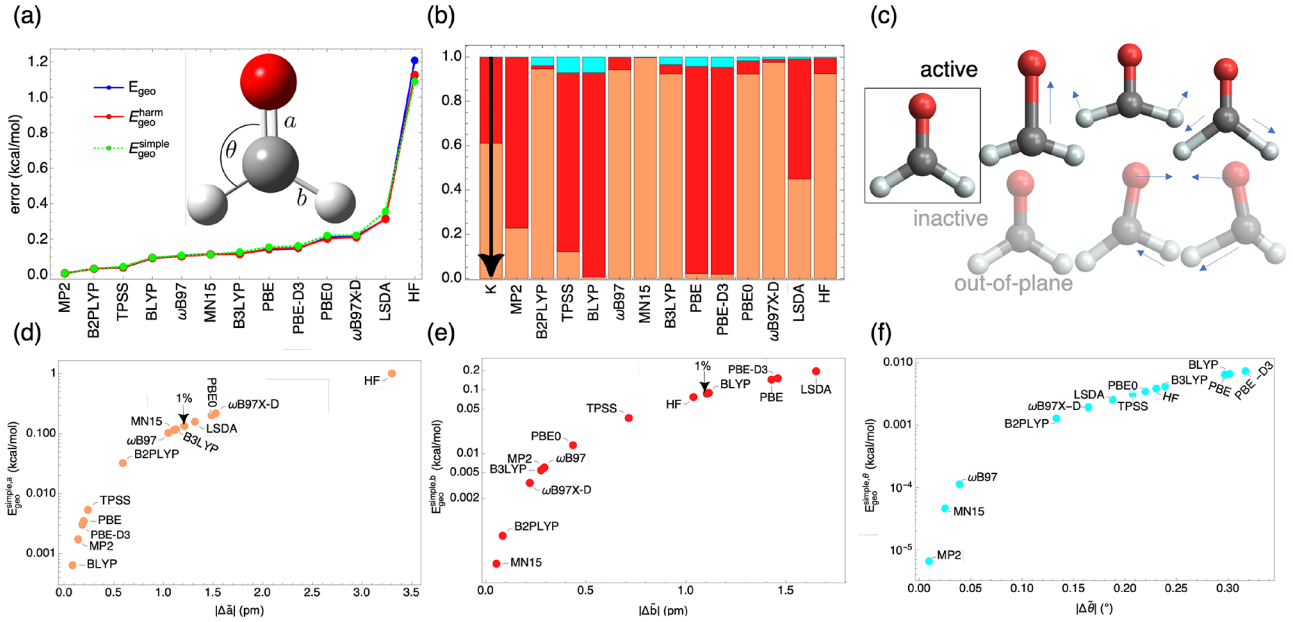


Figure 2. GEO analysis for formaldehyde. Top: (a) GEO rankings of approximations. The plots also show that E_{geo} is accurately approximated by E_{geo}^{harm} (eq 3) and E_{geo}^{simple} (eq 5). (b) E_{geo}^{simple} weights, $E_{geo}^{simple,i}/E_{geo}^{simple}$ (see lower panels for color legends). When all bonds are stretched by the same γ factor, GEO becomes E_{geo}^{γ} (see the text), and the underlying weights are shown in the first bar and labeled by "*K*". (c) GEO-active and GEO-inactive modes (those that have no contribution to the right-hand side of eq 4). For the E_{geo}^{harm} weights in normal modes, analogous to panel b, see Figure S32. (d)−(f) Plots showing errors in individual geometric parameters (*x*-axis) and how these errors translate to E_{geo}^{simple} terms by virtue of eq 5 (*y*-axis). The points marked by the arrows, show GEO when the bond lengths are stretched by 1%.

and MAE for AE). Relative to their poor performance for AE, LDA, and HF are surprisingly good for geometries, because their MAE for atomization is so poor. Nonempirical functionals (PBE, TPSS) are close to this line, so improvements in AEs are reflected in improvements in GEO. Empirical functionals typically perform very well, especially the global hybrid B3LYP, and BLYP, as a GGA, yields surprisingly poor geometries. Also, the addition of D3 corrections[19] to any functional makes little difference to its geometric performance for covalent bonds.

If we want to calculate GEO errors for larger main-group molecules, for which CCSD(T) is too expensive, we can use B2PLYP as a reference in place of CCSD(T). Multireference systems, such as transition metal dimers, are more delicate and would need a better reference than CCSD(T)[20,21] to calculate GEO.

The bigger picture is shown in the lower-left panel of Figure 1, which includes many different kinds of methods. Here, we had to use B2PLYP[22] as a reference (see above). Besides the QM solvers considered in Figure 1a(top), we also include semiempirical QM solvers, such as DFTB[23,24] and PM*x*,[25,26] and the highly practical HF-3c[27] and PBEh-3c,[28] both of which use a small basis set and contain empirical parameters. The results are summarized in the lower panel of Figure 1a, where different error bar colors indicate methods at different levels of theory, with the best method for each level of theory shown in red. Trends are similar to the panel above, but overall GEOs are larger as the molecules are bigger. This plot is not accurate below about 0.1 kcal/mol, because of the B2PLYP reference. Thus, B3LYP does not really rank as No. 1, its errors are simply correlated with the reference. As we discuss later, since GFN1-xTB[24] has the best performance among all semiempirical methods shown, it can serve as an excellent starting point in optimization schemes.

Geometry Optimization. On the basis of GEO, one can establish the following *sequence* composed of the best method for each level of complexity: GFN1-xTB → TPSS (or PBEh-3c) → B3LYP → B2PLYP. This sequence can be used in automated explorations of chemical space and molecular screenings assisted by QM solvers,[29−31] which are powerful tools for the discovery of new molecules with desired properties. In these procedures, on the basis of energetic criteria (e.g., their binding energies with a specific enzyme), molecules are filtered out, and QM geometry optimizations of a large number of molecules make the procedure computationally demanding. As the number of molecules in the screening decreases, more expensive and more accurate methods are typically used. Thus, on the basis of our sequence determined by the GEO criterion, in the first step of the screening one can employ GFN1-xTB for optimizing geometries of all initial molecular candidates. After the first cycle of filtering out molecules, TPSS[18] can be employed as an optimizer, and so on. In the last round, B2PLYP geometries can be confidently used, given they are energetically very close to the CCSD(T) geometries (∼0.03 kcal/mol for the testset considered in Figure 1a). Even if one only wants the CCSD(T) geometries for small molecules, one can use the same *sequence* to preoptimize the molecular geometry, before the CCSD(T) optimizer is turned on, and thereby save computational time.

Simplifications and Analysis Tools. As mentioned above, a much less costly calculation is

$$E'_{geo} = \tilde{E}(\tilde{\mathbf{G}}) - \tilde{E}(\mathbf{G}_0) \quad (2)$$

where $\tilde{E}$ is the approximate energy (see Supplementary Section 1 for more mathematical details). At minima, the signs of E_{geo} and E'_{geo} are definite: $E'_{geo} \leq 0$ and $E_{geo} \geq 0$. From Figure 1a, the mean $|E'_{geo}|$ is in excellent agreement with its E_{geo}

https://dx.doi.org/10.1021/acs.jpclett.0c03034
J. Phys. Chem. Lett. 2020, 11, 9957−9964

counterpart for all approximations (see also Tables S1 and S2 comparing E_{geo} and E'_{geo} values for individual molecules).

The next step is to approximate GEO by expanding E_0 around its minimum to second order:

$$E_{geo}^{harm} = \frac{1}{2}\Delta\tilde{\mathbf{G}}^{T}\mathbf{H}_0\Delta\tilde{\mathbf{G}} \tag{3}$$

where $\mathbf{H}_0$ is the Hessian at the minimum, composed of force constants and $\Delta\tilde{\mathbf{G}} = \tilde{\mathbf{G}} - \mathbf{G}_0$ is the error in specific geometric parameters (degrees of freedom) that determine the relative positions of nuclei. These could be simple Cartesians or any other choice of coordinates. Again, this is extremely accurate when approximated with most electronic structure methods (see Figures S20−S25). Thus, we can use eq 3 for further analysis and decomposition of GEO. One can easily diagonalize $\mathbf{H}_0$ and obtain GEO modes, p. In these coordinates, eq 3 becomes

$$E_{geo}^{harm} = \frac{1}{2}\sum_{i=1}^{N_g} f_i^p(\Delta p_i)^2 \tag{4}$$

where f_i^p are the underlying force constants ($\mathbf{H}_0$ eigenvalues) and $\Delta p_i = \tilde{p}_i - p_i$ represent $\Delta\tilde{G}$ written in terms of the errors in the GEO normal modes ($\mathbf{H}_0$ eigenvectors) and $N_g \leq 3N - 6$ is the number of GEO-active modes. A highly appealing feature of eq 4 is that at the minimum each term contributes positively to E_{geo}^{harm} which, in turn, allows us to obtain weights of each modes' contribution to the total GEO (see Figures S31−S33, where for a set of small molecules we show the weights of different GEO-active modes). GEO-inactive modes are those that do not contribute to the total E_{geo}^{harm}. For example, by symmetry, no (sensible) electronic structure approximation gives unequal OH bond lengths in the water molecule, so the asymmetric stretch of the OH bond is GEO-inactive (see Figure S29). The higher the symmetry of a molecule, the fewer modes are GEO-active. For ethene, all modes that distort its D_{2h} symmetry (asymmetric and out-of-plane vibrations) are GEO-inactive, so only 3 of its 12 modes are GEO-active (see Figure S33).

Besides the GEO modes, a more chemically intuitive analysis in terms of bond lengths, angles, and torsion angles of E_{geo} can be obtained by considering eq 3 in internal coordinates. Considering only the underlying diagonal elements of $\mathbf{H}_0^q$ (the Hessian in internal coordinates), we find the following *simple* approximation to E_{geo}^{harm}:

$$E_{geo}^{harm} \approx E_{geo}^{simple} = \frac{1}{2}\sum_{i=1}^{3N-6} f_{i,i}^q(\Delta q_i)^2 \tag{5}$$

where $\Delta q_i = \tilde{q}_i - q$ are the errors in internal coordinates. While the right-hand side of eq 4 is exactly equal to E_{geo}^{harm}, this is not so for E_{geo}^{simple}, since the off-diagonal $\mathbf{H}_0^q$ are typically small but nonzero. For the organic molecules we consider here, E_{geo}^{simple} is typically in good agreement with both E_{geo}^{harm} and the "exact" E_{geo} (see Figures S20−S25). This, in turn, allows us to *safely* use eq 5 to decompose E_{geo}^{harm} into its positive contributions arising from errors in specific geometric parameters.

In Figure 2, we illustrate how a GEO analysis works for a simple case, formaldehyde. In panel a, we give GEO rankings of different approximations for this molecule, which somewhat align with the database averages of Figure 1a. As with all covalent cases we studied, E_{geo}^{harm} and E_{geo}^{simple} are in excellent agreement with GEO, which allows us to use eq 5 to decompose contributions from different structural parameters. The fractional contributions of each coordinate are shown in panel b. Angle errors give only a minor contribution to GEO for all methods. For the hybrids, the GEO error comes nearly entirely from the error in the double bond, while in the case of semilocal functionals, nearly the entire GEO error comes from the error in the single-bond lengths, consistent with the trends in Figure 1b. The rankings for the single- and double-bond lengths and the bond angle are shown in the lower panels and how they correlate with E_{geo}^{simple}. Our GEO axes are logarithmic, but the actual curves are parabolic. The rankings differ substantially from those for the total GEO. The leftmost is the double bond, and here the semilocal functionals (no mixing of exact exchange) do best, outperforming even B2PLYP! The hybrids do no better than simple LDA. But, the roles are reversed in the middle panel, showing that hybrids greatly improve single-bond length error. Finally, the angle-error is shown, with a variety of results, but no clear trends. For the set of molecules considered in Figure 1b, MP2 and B2PLYP yield the best angles on average (see the right panel of Figure S13). All three curves in the lower panels of Figure 2 are monotonic, so rankings by a E_{geo}^{simple} contribution correspond directly to rankings by the error in the underlying geometric parameter. Finally, in Figure 2c, we make a distinction between the GEO-active and -inactive modes of formaldehyde.

Absolute GEO Scale. A problem that bedevils benchmarking of atomization energies is whether to consider total atomization energy errors or errors per bond. Here we show that there exists a universal GEO scale, independent of any method, that overcomes this problem for geometric errors. Consider a small expansion of all coordinates, $\Delta\mathbf{G} = \gamma\mathbf{G}_0$, producing $E_{geo}^{\gamma} = \gamma^2 D/2$, where $D = \mathbf{G}_0^{T}\mathbf{H}_0\mathbf{G}_0$. Thus, E_{geo}^{γ} is the GEO value for a very specific geometric error, that of expansion (or compression) of the exact geometry. For our small molecules, if $\gamma = 1\%$, E_{geo}^{γ} is a fraction of a kcal/mol. Thus, any calculation of GEO by any method for any molecule can be compared to this intrinsic property of the molecule. Moreover, E_{geo}^{γ} scales with the size of the molecule (compare, e.g., D values for small molecules shown in Table S6 with those for medium-sized molecules shown in Table S7), so that GEOs measured relative to it do not grow with molecular size. We can even decompose E_{geo}^{γ} in terms of Hessian eigenvectors or simple internal coordinates, giving an internally defined distribution of contributions. This only includes bond lengths, as no angle changes when molecules are uniformly expanded.

To compare GEO values on the D scale across different molecules, we define $E_{geo}^{D} = E_{geo}(D_{H_2O}/D)$, which contains D for the molecule in its denominator, ensuring it does not grow with molecular size, and D_{H_2O} in its numerator to normalize it. In Figure 3, we repeat the plot in the top panel of Figure 1b, but with E_{geo}^{D}, which varies much less with molecular size. For the same set of molecules, we report the underlying D values in Table 1, where we also show that approximate calculations of D typically yield highly accurate estimates (even HF is not too bad). In Figure S34, we show the decomposition of D for each bond in each molecule, with double bonds being about 0.25 and most singles being about 0.1 if to an H atom, and about 0.16 if between heavier atoms. The units here are 10^4 kcal/mol. For rare gas dimers (bonded by weak interactions), D values are several orders of magnitude smaller (see Table S8).

Returning to Figure 2, the leftmost column of (b) is the D decomposition of the single versus double bond, showing that

https://dx.doi.org/10.1021/acs.jpclett.0c03034
J. Phys. Chem. Lett. 2020, 11, 9957−9964

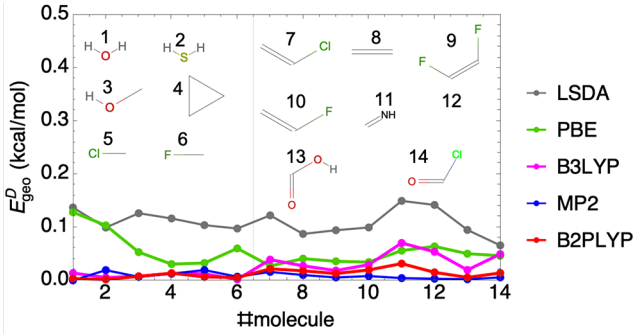


Figure 3. Same plot as in the top panel of Figure 1b but with E^D_{geo} (see the absolute GEO scale section) on the *y*-axis. For more plots comparing E_{geo} and E^D_{geo}, see Figure S8.

Table 1. Accurate *D* Values [CCSD(T)] in 10^4 kcal/mol, and Approximate Values from Selected Methods, for the Set of Molecules Considered in Figure 1b[a]

molecule	CCSD(T)	B2PLYP	B3LYP	PBE	HF
1	0.221	0.220	0.219	0.211	0.247
2	0.220	0.221	0.215	0.209	0.239
3	0.566	0.567	0.563	0.545	0.629
4	0.989	0.998	0.990	0.966	1.076
5	0.452	0.452	0.443	0.435	0.488
6	0.463	0.463	0.458	0.443	0.511
7	0.727	0.736	0.733	0.714	0.810
8	0.634	0.643	0.642	0.623	0.699
9	0.848	0.856	0.856	0.823	0.971
10	0.735	0.744	0.744	0.719	0.828
11	0.555	0.564	0.569	0.547	0.636
12	0.485	0.489	0.497	0.476	0.563
13	0.725	0.723	0.731	0.697	0.842
14	0.594	0.592	0.599	0.573	0.696
MAE		0.005	0.007	0.017	0.073
(%)		0.7	1.3	3.1	12.2

[a]In the last row, we report the mean absolute percentage errors of approximations. For the list of molecules, see Figure 3. Tables S5−S8 report *D* values for other molecules or other approximations considered in this work.

under expansion, 60% of the energy cost is to stretch the double bond, 40% to stretch the single. Then BLYP clearly makes an unusually small error in the double, and a relatively large error in the single.

Noncovalent Interactions. The rest of this paper is devoted to weak interactions. We re-examine all aspects of GEO for these cases, as GEO energies can be a more significant fraction of the binding energies here. Force constants for weak bonds are so much weaker that even very small GEO values can lead to large errors in bond lengths. For weak interactions, we include D3 corrections[19] to the DFT approximations, which typically greatly improve energetic accuracy. We also only allow the weak bond length to vary in complexes, i.e., one degree of freedom.

The upper panel of Figure 4 shows a prototypical van der Waals system, Ne_2, with the exact curve and various approximations. In most cases, the geometries are quite accurate (minima are marked by beads) so that GEO energies are very small. The ωB97X-D functional[34] has a large geometric error, which is significantly reduced by the new functionals built upon it (see Figure S41).[35−37] Nevertheless,

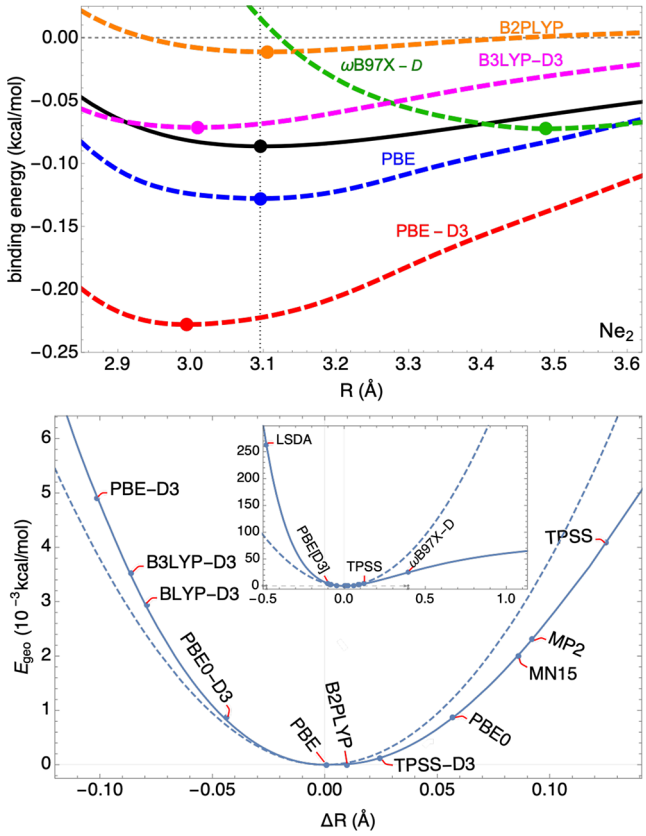


Figure 4. GEO analysis for the Ne_2 binding energies. Top panel: binding curves of Ne_2 with various methods. Lower panel: GEO for different approximations. For more details, see Figures S40−S44.

ωB97X-D is interesting here because of its highly accurate energy minimum, resulting from a cancellation of geometric and nongeometric errors. Since functionals are applied to cases where neither accurate geometries nor energies are known, the primary concern is to predict an accurate energy at the approximate minimum. By this criterion, ωB97X-D is the most accurate approximation shown in the lower panel of Figure 4! Note that the D3 correction worsens PBE here. Furthermore, B2PLYP gives an excellent geometry.

The approximate E'_{geo} works well when the geometry is reasonably good, and the harmonic approximation is largely still valid (dashed line in the lower panel of Figure 4) but is less accurate than for typical covalent cases. Decomposition into Hessian eigenvectors is irrelevant here, as for now we only adjust one geometry parameter, the length of the weak bond. The analysis is thus the same as for a simple diatomic, where the Cartesian coordinate difference is the internal coordinate.

Typically, databases are established using a fixed reference geometry, which may or may not be very close to the exact minimum (as measured by GEO). Optimization of parameters in an approximation will then miss the trade-off between geometry and energy that can occur in applications beyond the training database, where presumably a geometry-optimized calculation should be designed to yield the best energy.

Implications for Benchmarking Molecular Energies. In quantum chemistry, the performance of approximate QM solvers is usually assessed by single-point calculations at reference geometries.[1−3,38] Now we show the importance of GEO in such comparisons using the S66×8 data set.

https://dx.doi.org/10.1021/acs.jpclett.0c03034
J. Phys. Chem. Lett. 2020, 11, 9957−9964

In Figure 5, we show plots comparing GEO with the total |Δ*E*| errors of MP2 (with a complete basis set) and the PBE0

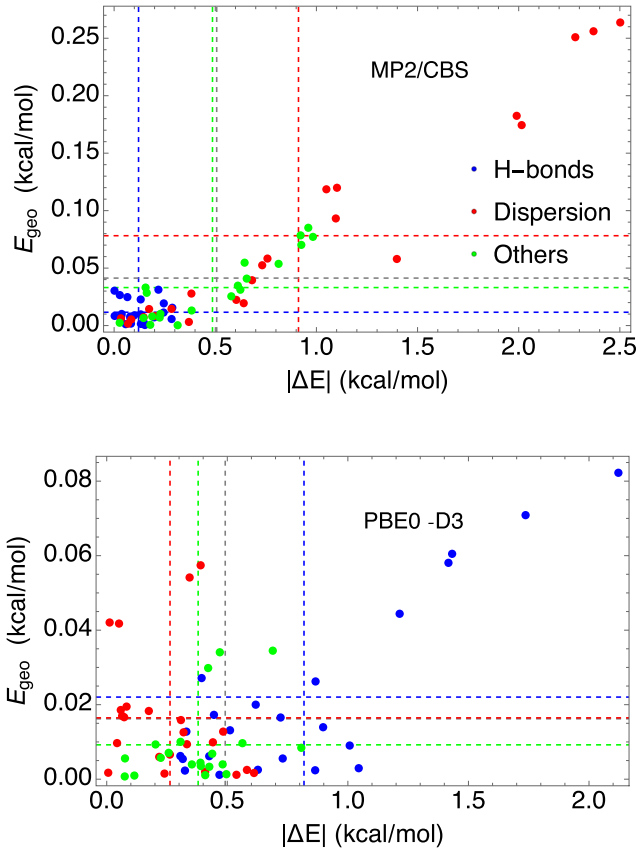


Figure 5. $E_X[G_X] - E_0[G_0]$ vs GEO errors for the binding energies of the S66 complexes of MP2/CBS (top panel), PBE0-D3 (lower panel). The complexes are classified into "H-bonds", "Dispersion", and "Others" as in the original S66 publication.[32] Colored dashed lines represent MAEs for different S66 categories, and the gray dash line represents the overall S66 MAEs.

hybrid with the D3 correction for the S66 binding energies (other methods are in SI in Figures S45−S47). Vertical and horizontal dashed lines represent the averaged GEO and |Δ*E*|, respectively, separated into H-bonds, dispersion dominated interactions, and others. Both methods have overall mean average energy errors of about 0.5 kcal/mol, but MP2 is worst for dispersion complexes, while PBE0-D3 is worst for H-bonds. But clearly, MP2 yields much worse geometries with GEO values about 4 times larger. Also, for MP2, there is a strong correlation between GEO and errors in binding energies. But for PBE0-D3, GEO errors do not correlate with such errors. For the geometry of weak bonded complexes, PBE0-D3 is clearly superior to MP2, but not (overall) for binding energies. Here we employ the S66 data set, but the same methodology can be applied to other noncovalent data sets of interest.[39,40]

Finally, we consider the effect of varying geometries on binding energy errors in Table 2. Each column is a method for finding a geometry, each row is the method used for finding the energy, all averaged over the 66 complexes. The bolded entries along the diagonal are the errors of each method at its own geometry. The last column is the errors on accurate geometries, given by CCSD(T), i.e., the best estimate of the exact geometry here. The numbers are very similar in all cases, suggesting that improvements in geometry do not matter. However, this makes the B2PLYP column even more surprising: For all methods, the errors are smaller at the B2PLYP geometries than at their own geometries, sometimes by almost a factor of 2, and always better than on the accurate geometries! How can this be? We explored and found that most approximations overbind the S66 complexes (particularly those bonded by dispersion), and B2PLYP typically overestimates the bond lengths. For this reason, energies of approximate methods at the B2PLYP minimum are more accurate than at their own minimum (see the benzene−uracil binding curves shown in Figure S48.)

Here we have covered only the most obvious topics that the GEO concept brings into focus. For main group chemistry and weak interactions, GEO calculations and analysis yield an ideal tool for understanding geometric errors and for ranking different approximations, one that is very different from tables of errors in atomization/binding energies and could easily be applied to transition state geometries. An advantage of the GEO concept is that it uses energetic units to assess qualities of approximate geometries, and thus it can be easily coupled with the standard energetic scores[1−3] used to rank different approximations. Specifically, the GEO scores can be included in the calculation of *the weighted total mean absolute deviation* (WTMAD)[1] that will pertain to new benchmark databases. Here we analyze errors in geometries obtained from electronic structure methods, but the very same tools are also applicable to molecular mechanics methods. Here we study GEO at fixed basis sets, but in future work, we will also study how the GEO

Table 2. MAE of Different Methods at Different Minima of S66×8 Binding Curves for the S66 Dataset[a]

	MP2/CBS	B2PLYP	B3LYP-D3	PBE0-D3	PBE-D3	MN15	ωB97X-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	0.40	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	**0.60**	0.60	0.34	0.59
ωB97X-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

[a]The error is defined as $E_X[G_Y] - E_0[G_0]$, where X is a method in rows and Y is the method in columns, and $E_0[G_0]$ is the binding energy at the CCSD(T) minimum. The S66×8 binding curves from CCSD(T)/CBS (used as a reference here) and MP2/CBS have been taken from the original S66×8 dataset.[32,33] 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 data points for each of the S66 complex (see details above).

https://dx.doi.org/10.1021/acs.jpclett.0c03034
J. Phys. Chem. Lett. 2020, 11, 9957−9964

varies across different basis set and different approximations (as hinted in Figure S17).

COMPUTATIONAL DETAILS

Now we go back to the top panel of Figure 1a, where we calculate mean GEO for a data set of 14 small organic molecules, which is the AV5Z subset of the W4-11-GEOM set produced by Karton and co-workers.[12] The geometries from this data set have been optimized at the CCSD(T) level, which is the level of theory that we use for this data set as a reference. GEOs for approximate methods that we consider here are obtained as follows. The $\tilde{E}(\mathbf{G}_0)$ quantity is obtained from the total energies of each of the approximate method at CCSD(T) geometries. Then we relax CCSD(T) geometries by using each of the approximate methods, and this allows us to calculate the $\tilde{E}(\tilde{\mathbf{G}})$ term. Finally, we obtain the total energies of each of the approximate method at CCSD(T) geometries to compute $E_0(\tilde{\mathbf{G}})$. These energies are obtained from single-point CCSD(T)/A′V5Z calculations on the $\tilde{\mathbf{G}}$ geometries. In this way, we ensure that the level of theory used for the reference single-point energy calculations matches the level of theory used for obtaining reference geometries by Karton and others.[12,41] Other GEOs in the paper are computed in the same manner and in the Supporting Information we provide further computational details.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.0c03034.

GEO mathematical details, details on the γ expansion and D values, additional details on the results of figures in the main text, tables of E_{geo} and MAE values, additional GEO figures (PDF)

AUTHOR INFORMATION

Corresponding Author

Kieron Burke − *Department of Chemistry and Department of Physics, University of California, Irvine, California 92697, United States;* orcid.org/0000-0002-6159-0054; Email: kieron@uci.edu

Author

Stefan Vuckovic − *Department of Chemistry, University of California, Irvine, California 92697, United States;* orcid.org/0000-0002-0768-9176

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.0c03034

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank J. Rezác for providing us the S66×8 binding energies for a selection of wavefunction methods. S.V. acknowledges funding from the Rubicon project (019.181EN.026), which is financed by The Netherlands Organisation for Scientific Research (NWO). KB acknowledges funding from NSF (CHE 1856165).

REFERENCES

(1) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184−32215.

(2) Yu, H. S.; Zhang, W.; Verma, P.; He, X.; Truhlar, D. G. Nonseparable exchange−correlation functional for molecules, including homogeneous catalysis involving transition metals. *Phys. Chem. Chem. Phys.* **2015**, *17*, 12146−12160.

(3) Mardirossian, N.; Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **2017**, *115*, 2315−2372.

(4) Su, N. Q.; Xu, X. Beyond energies: Geometry predictions with the XYG3 type of doubly hybrid density functionals. *Chem. Commun.* **2016**, *52*, 13840−13860.

(5) Ruden, T. A.; Helgaker, T.; Jørgensen, P.; Olsen, J. Coupled-cluster connected quadruples and quintuples corrections to the harmonic vibrational frequencies and equilibrium bond distances of HF, N2, F2, and CO. *J. Chem. Phys.* **2004**, *121*, 5874−5884.

(6) Feller, D.; Peterson, K. A.; Dixon, D. A. A survey of factors contributing to accurate theoretical predictions of atomization energies and molecular structures. *J. Chem. Phys.* **2008**, *129*, 204105.

(7) Feller, D.; Peterson, K. A.; Dixon, D. A. Further benchmarks of a composite, convergent, statistically calibrated coupled-cluster-based approach for thermochemical and spectroscopic studies. *Mol. Phys.* **2012**, *110*, 2381−2399.

(8) Heckert, M.; Kallay, M.; Gauss, J. Molecular equilibrium geometries based on coupled-cluster calculations including quadruple excitations. *Mol. Phys.* **2005**, *103*, 2109−2115.

(9) Feller, D.; Peterson, K. A.; de Jong, W. A.; Dixon, D. A. Performance of coupled cluster theory in thermochemical calculations of small halogenated compounds. *J. Chem. Phys.* **2003**, *118*, 3510−3522.

(10) Helgaker, T.; Gauss, J.; Jorgensen, P.; Olsen, J. The prediction of molecular equilibrium structures by the standard electronic wave functions. *J. Chem. Phys.* **1997**, *106*, 6430−6440.

(11) Piccardo, M.; Penocchio, E.; Puzzarini, C.; Biczysko, M.; Barone, V. Semi-Experimental Equilibrium Structure Determinations by Employing B3LYP/SNSD Anharmonic Force Fields: Validation and Application to Semirigid Organic Molecules. *J. Phys. Chem. A* **2015**, *119*, 2058−2082.

(12) Spackman, P. R.; Jayatilaka, D.; Karton, A. Basis set convergence of CCSD (T) equilibrium geometries using a large and diverse set of molecular structures. *J. Chem. Phys.* **2016**, *145*, 104101.

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

(14) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098.

(15) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785.

(16) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(17) Burke, K.; Ernzerhof, M.; Perdew, J. P. The adiabatic connection method: A non-empirical hybrid. *Chem. Phys. Lett.* **1997**, *265*, 115.

(18) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the density functional ladder: Nonempirical meta−generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.

(19) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(20) Xu, X.; Zhang, W.; Tang, M.; Truhlar, D. G. Do Practical Standard Coupled Cluster Calculations Agree Better than Kohn−

Sham Calculations with Currently Available Functionals When Compared to the Best Available Experimental Data for Dissociation Energies of Bonds to 3 d Transition Metals? *J. Chem. Theory Comput.* **2015**, *11*, 2036−2052.
(21) Hait, D.; Tubman, N. M.; Levine, D. S.; Whaley, K. B.; Head-Gordon, M. What Levels of Coupled Cluster Theory Are Appropriate for Transition Metal Systems? A Study Using Near-Exact Quantum Chemical Values for 3d Transition Metal Binary Compounds. *J. Chem. Theory Comput.* **2019**, *15*, 5370−5385.
(22) Grimme, S. Semiempirical hybrid density functional with perturbative second-order correlation. *J. Chem. Phys.* **2006**, *124*, 034108.
(23) Gaus, M.; Goez, A.; Elstner, M. Parametrization and benchmark of DFTB3 for organic molecules. *J. Chem. Theory Comput.* **2013**, *9*, 338−354.
(24) Grimme, S.; Bannwarth, C.; Shushkov, P. A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements (Z= 1−86). *J. Chem. Theory Comput.* **2017**, *13*, 1989−2009.
(25) Stewart, J. J. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173−1213.
(26) Stewart, J. J. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **2013**, *19*, 1−32.
(27) Sure, R.; Grimme, S. Corrected small basis set Hartree-Fock method for large systems. *J. Comput. Chem.* **2013**, *34*, 1672−1685.
(28) Grimme, S.; Brandenburg, J. G.; Bannwarth, C.; Hansen, A. Consistent structures and interactions by density functional theory with small atomic orbital basis sets. *J. Chem. Phys.* **2015**, *143*, 054107.
(29) Grimme, S. Exploration of chemical compound, conformer, and reaction space with meta-dynamics simulations based on tight-binding quantum chemical calculations. *J. Chem. Theory Comput.* **2019**, *15*, 2847−2862.
(30) Wang, L.-P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martínez, T. J. Discovering chemistry with an ab initio nanoreactor. *Nat. Chem.* **2014**, *6*, 1044.
(31) Pracht, P.; Bohle, F.; Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169.
(32) Rezác, J.; Riley, K. E.; Hobza, P. S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427−2438.
(33) We thank J. Rezác for giving us access to the S66×8 binding energies for these wave function methods.
(34) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom−atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615−6620.
(35) Mardirossian, N.; Head-Gordon, M. ωB97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9904.
(36) Mardirossian, N.; Head-Gordon, M. ωB97M-V: A combinatorially optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **2016**, *144*, 214110.
(37) Najibi, A.; Goerigk, L. 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. *J. Chem. Theory Comput.* **2018**, *14*, 5725−5738.
(38) Gould, T. Diet GMTKN55' offers accelerated benchmarking through a representative subset approach. *Phys. Chem. Chem. Phys.* **2018**, *20*, 27735−27739.
(39) Řezáč, J.; Hobza, P. Describing Noncovalent Interactions beyond the Common Approximations: How Accurate Is the "Gold Standard, " CCSD(T) at the Complete Basis Set Limit? *J. Chem. Theory Comput.* **2013**, *9*, 2151−2155.
(40) Kraus, P.; Frank, I. Density Functional Theory for Microwave Spectroscopy of Noncovalent Complexes: A Benchmark Study. *J. Phys. Chem. A* **2018**, *122*, 4894−4901.
(41) Dunning, T. H.; Peterson, K. A.; Wilson, A. K. Gaussian basis sets for use in correlated molecular calculations. X. The atoms aluminum through argon revisited. *J. Chem. Phys.* **2001**, *114*, 9244−9253.

https://dx.doi.org/10.1021/acs.jpclett.0c03034
J. Phys. Chem. Lett. 2020, 11, 9957−9964