Assessing the domain-based local pair natural orbital (DLPNO) approximation for non-covalent interactions in sizable supramolecular complexes

Montgomery Gray and John M. Herbert*
Department of Chemistry & Biochemistry,
The Ohio State University, Columbus, Ohio 43210 USA

Abstract

The titular DLPNO approximation has become a *de facto* standard for extending correlated wave function models to large molecular systems yet its fidelity for intermolecular interaction energies has not been thoroughly vetted. Non-covalent interaction energies are sensitive to tails of the electron density, involving parts of the wave function that are far from the nuclei and may be discarded in some local correlation treatments. Meanwhile, the accuracy of the DLPNO approximation is known to deteriorate as molecular size increases and questions have been raised regarding the accuracy of benchmark calculations for large van der Waals complexes. Here, we test the DLPNO approximation at the level of second-order Møller-Plesset perturbation theory (MP2) in systems with up to 240 atoms, for which canonical MP2 calculations can be performed for comparison. For small dimers, we find that DLPNO-MP2 interaction energies are within 3% of canonical values but the approximation is quite poor for larger systems, unless the results are extrapolated to to the limit where the threshold for discarding PNOs is taken to zero. For a sequence of nanoscale graphene dimers up to $(C_{96}H_{24})_2$, extrapolated DLPNO-MP2 interaction energies agree with canonical values to within 1%, independent of system size, provided that the basis set does not contain diffuse functions. The presence of diffuse basis functions causes oscillatory behavior as a function of the PNO threshold, making it impossible to extrapolate the results in a meaningful way.

I. INTRODUCTION

The coupled-cluster method using singles, doubles, and perturbative triples [CCSD(T)] is the gold standard of single-reference quantum chemistry, for both thermochemistry[1,2] and non-covalent interactions.[3–5] However, this benchmark-level accuracy comes at near-intractable cost for large systems. Efforts to develop localized-orbital approximations to canonical CCSD(T) are experiencing a renaissance,[6–18] and among several competing approaches the domain-based local pair natural orbital (DLPNO) approximation[19–21] has emerged as the most widely used variant,[6–11] owing to its implementation in the ORCA program.[22]

The fidelity of the DLPNO approximation has been exhaustively evaluated for small systems, including for non-covalent interactions,[11,23,24] but there have been few systematic tests for non-covalent interactions in large systems.[25–27] Neglect or approximation of distant-pair interactions is known to degrade the accuracy in such cases.[28] As a result, the accuracy of the DLPNO approximation decreases with system size,[25,29] at least for thermochemical calculations. For benchmark-quality accuracy, results must be extrapolated to the canonical limit, meaning that the threshold for neglecting weak pairs is extrapolated to zero.[24,30]

Even so, recent work has cast doubt on the fidelity of localized-orbital approximations to canonical CCSD(T), by demonstrating discrepancies exceeding 10 kcal/mol as compared to fixed-node diffusion Monte Carlo (FN-DMC) calculations;[31,32] see Ref. 32 for a summary of current state-of-the-art benchmarks. Some of the relevant CCSD(T) benchmarks were obtained with the DLPNO approach while others used an alternative localized natural orbital (LNO) implementation of CCSD(T),[15–18] which is available in the MRCC program.[33] All of them have been extrapolated to the complete basis set (CBS) limit, yet sizable discrepancies remain in multiple cases. Establishing the veracity of the benchmarks is vitally important for the development of low-cost methods including density functional theory (DFT). Notably, a variety of dispersion-inclusive functionals and dispersion-corrected DFT approaches afford results of wildly varying quality for supramolecular complexes with $\gtrsim 100$ atoms.[34] This is despite providing rather accurate interaction energies for small van der Waals dimers, consistent across a variety of DFT methods.[34,35] First-principles dispersion models[36–38] also need to be tested against high-quality *ab initio* data in large systems.

As such, the purpose of this work is to examine the accuracy of the DLPNO approximation for large van der Waals complexes. The most thorough previous effort along these lines is a recent study of double-hybrid DFT for transition metal compounds,[41] where the localized-orbital technique was applied to the second-order Møller-Plesset (MP2) energy. As compared to what is typical in thermochemical DLPNO-CCSD(T) calculations, tighter PNO thresholds were found to be required in order that the DLPNO-MP2 interaction energy remain faithful to the canonical MP2 result. Not addressed in

---

*herbert@chemistry.ohio-state.edu

https://doi.org/10.26434/chemrxiv-2024-9cvcx ORCID: https://orcid.org/0000-0002-1663-2278

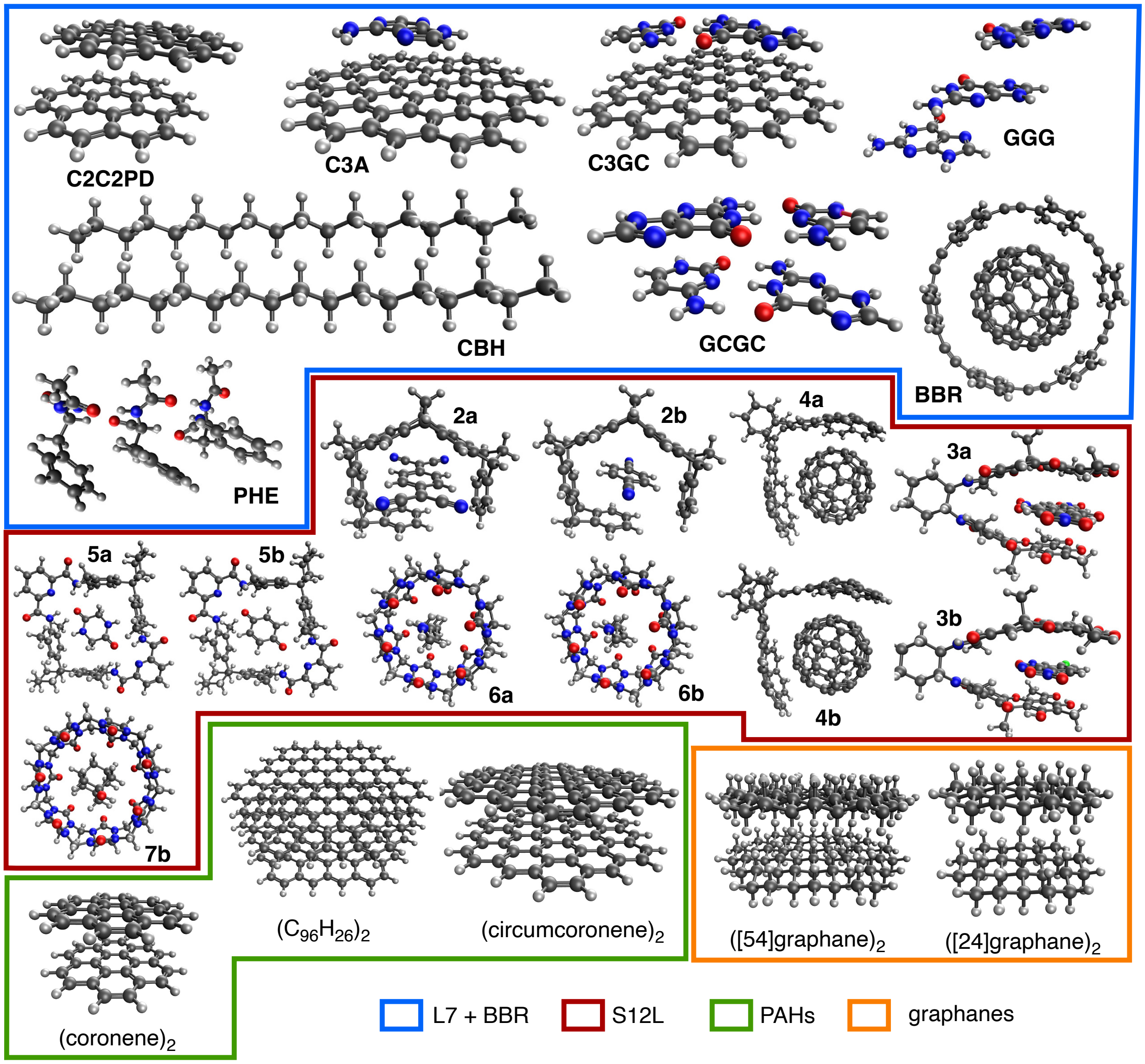


Fig. 1: Large supramolecular complexes investigated in this work, including the L7 complexes,[39] S12 complexes,[40] a buckyball-in-a-ring complex (**BBR**),[31] graphene dimers from the circumcoronene family, and the corresponding graphane (perhydrocircumcoronene) dimers.

that study is the performance of the DLPNO approximation in the presence of diffuse basis functions, which are the bane of linear-scaling approximations but are sometimes needed to obtain converged intermolecular interaction energies.[42–45]

Systematic tests of the DLPNO approximation in large van der Waals complexes are missing from the literature and the present work aims to fill this gap, examining the accuracy of the DLPNO approximation as a function of molecular size for nanoscale complexes as large as $(C_{96}H_{24})_2$. Because we need the canonical result for comparison, we perform these calculations at the MP2 level rather than the CCSD(T) level. Our results establish the accuracy that can be expected when the PNO thresholds are extrapolated to the canonical limit, and we examine the impact of diffuse basis functions on that extrapolation.

II. COMPUTATIONAL PROCEDURES

The large supramolecular complexes examined in this work are presented in Fig. 1. These include several complexes that are now standard benchmark systems for quantum chemistry, including the L7[39] and S12L[40] data sets and a "buckyball-in-a-ring" complex (**BBR**, C_{60}@[6]-cycloparaphenyleneacetylene).[31,46] Notably, several of these complexes exhibit differences exceeding 1 kcal/mol between FN-DMC and CCSD(T)/CBS benchmarks, even upon accounting for the uncertainties in either benchmark. These include **BBR**, **C2C2PD**, **C3A**, **C3GC**, **2a**, **2b**, **4a**, **5a**, and **6a**.[31,32]

In addition ,we consider dimers of polybenzenoids from the circumcoronene sequence: the coronene dimer $(C_{24}H_{12})_2$, the circumcoronene dimer $(C_{54}H_{18})_2$, and the circumcircumcoronene dimer $(C_{96}H_{24})_2$. These are models of graphene nano-flakes and we place them in a cofacial arrangement ("AA stacking"),[47–49] at 3.8 Å face-to-face separation, equivalent to the separation in the cofacial or "sandwich" isomer of benzene dimer.[50] Some previous benchmarks exist for these complexes,[51–53] which have often been considered alongside the corresponding graphane models,[54] namely, the perhydrocircumcoronenes that are also shown in Fig. 1.

A variety of basis sets are examined including (aug-)cc-pVDZ and (aug-)cc-pVTZ of the Dunning variety,[55,56] and def2-SVP, def2-SVPD, and def2-TZVP of Karlsruhe extraction.[57,58] The DLPNO approximation exhibits similar trends in both cc-pVDZ and cc-pVTZ so we judged it unnecessary to test cc-pVQZ, which would substantially limit the system sizes for which we could obtain canonical results. Although it is well-established that MP2 overestimates dispersion interactions,[37,59–62] the primary purpose of this work is to test the accuracy of the DLPNO approximation, not the accuracy of MP2. That said, a CBS extrapolation of the MP2 energy often serves as the foundation upon which a smaller-basis CCSD(T) correction is added,[25,63] and with that in mind we examine two-point MP2/CBS extrapolations using cc-pVXZ and aug-cc-pVXZ with X = D and T. Counterpoise correction is applied to all interaction energies, as it has a sizable effect even in the higher-quality basis sets examined here.[44,64,65]

All calculations were performed using ORCA v. 5.0.3,[22] except for calculations to establish the MP2/CBS limit, which were performed using Q-Chem v. 6.1.0.[66] MP2 calculations were performed within the resolution-of-identity (RI) approximation, with appropriate auxiliary basis sets matched to the target basis,[67] however no RI approximation was applied at the self-consistent field (SCF) level. The integral screening thresholds was set to $\tau_{\text{ints}} = 10^{-16}$ a.u. and the SCF convergence criterion to $\tau_{\text{SCF}} = 10^{-8}$ E_{h}. Following previous recommendations,[8] three different PNO-related settings were tested. A "loose" setting corresponds to a domain threshold $T_{\text{cutDO}} = 2 \times 10^{-2}$ and a PNO threshold $T_{\text{cutPNO}} = 10^{-7}$; the "normal" setting means $T_{\text{cutDO}} = 1 \times 10^{-2}$ and $T_{\text{cutPNO}} = 10^{-8}$; and finally, "tight" thresholds are $T_{\text{cutDO}} = 5 \times 10^{-3}$ and $T_{\text{cutPNO}} = 10^{-9}$.[68]

Extrapolation of the DLPNO-MP2 energy to the canonical MP2 result is performed using both a loose/normal (L/N) and a normal/tight (N/T) scheme. These extrapolations make use of the the formula[24]

$$E = E_x + F(E_y - E_x) \,, \tag{1}$$

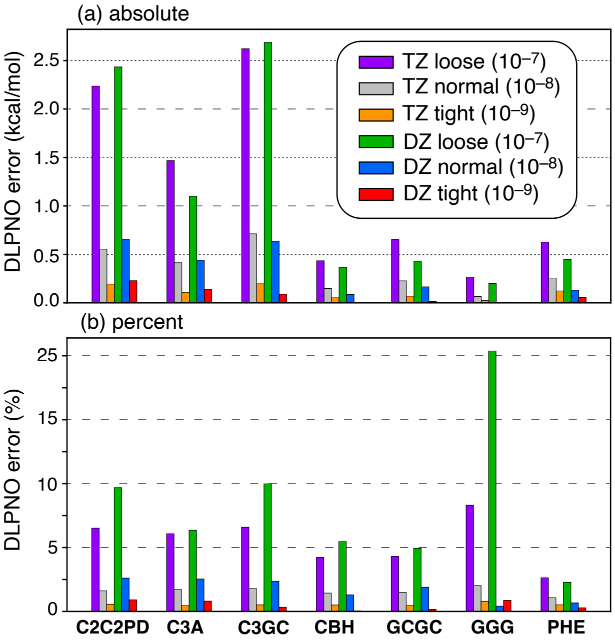


Fig. 2: Errors in the DLPNO approximation for L7 interaction energies in (a) absolute and (b) percentage terms, as a function of the PNO threshold T_{cutPNO}, for basis sets cc-pVDZ and cc-pVTZ. Counterpoise-corrected interaction energies were computed at the RI-MP2 level and error is defined as the difference between the DLPNO and the canonical RI-MP2 result.

where $F = 1.5$ is an empirical parameter and E_x denotes the energy obtained using a threshold $T_{\text{cutPNO}} = 10^{-x}$. The L/N extrapolation scheme corresponds to $x = 7$ and $y = 8$ while N/T extrapolation means $x = 8$ and $y = 9$.

III. RESULTS AND DISCUSSION

A. Basis-set comparison

The primary purpose of this work is to examine the efficacy of the DLPNO approximation as a function of molecular size, rather than to obtain converged benchmarks, so we first investigate whether extremely large basis sets are necessary to address that particular issue. To do this, we compare counterpoise-corrected RI-MP2/cc-pVDZ and RI-MP2/cc-pVTZ results for the L7 data set,[39] in Fig. 2. These data characterize the DLPNO error, defined as the difference in the interaction energy relative to the canonical MP2 result. We find that DLPNO-MP2 interaction energies are systematically less attractive than canonical MP2 values but the error decreases systematically as the threshold T_{cutPNO} is tightened, in both the cc-pVDZ and cc-pVTZ basis sets. In two cases, the error exceeds 1 kcal/mol for the loose setting ($T_{\text{cutPNO}} = 10^{-7}$), and for the parallel-displaced coronene dimer the error exceeds 2 kcal/mol. That said, both basis sets afford very similar DLPNO errors, with

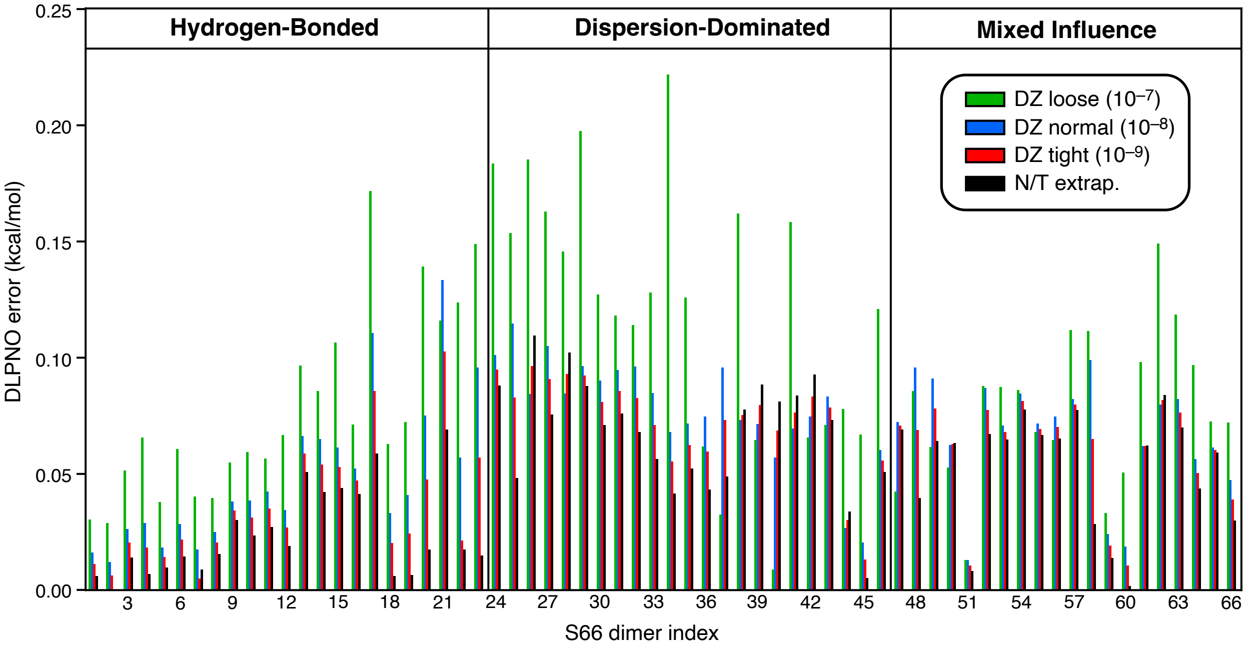


Fig. 3: DLPNO errors for S66 dimers at the RI-MP2/cc-pVDZ level.

differences no larger than 0.4 kcal/mol. (For tight PNO thresholds, the errors differ by $\lesssim$ 0.1 kcal/mol.) This suggests that we can use cc-pVDZ as a representative basis set to examine the DLPNO error in larger systems than would otherwise be feasible, given available computing resources.

Percentage errors [Fig. 2(b)] confirm the similarity of the cc-pVDZ and cc-pVTZ results. For both the normal and tight thresholds, differences between basis sets are $< 1\%$ except for the guanine trimer **GGG**, which has the smallest interaction energy amongst the L7 complexes ($\approx$ 2 kcal/mol). For $T_{\text{cutPNO}} < 10^{-7}$ (*i.e.*, normal or tight settings), the percentage error introduced by the DLPNO approximation changes by less than 1% between basis sets. Results for the def2-SVP and def2-TZVP basis sets (Fig. S1) confirm similar trends as T_{cutPNO} is tightened, with comparable errors in the double- and triple-ζ basis set. The DLPNO errors are slightly larger than those observed for cc-pVXZ basis sets, and approach or exceed 10% of the interaction energy in several cases, when loose thresholds are employed. For the **BBR** complex (Fig. 1), which we did not attempt using cc-pVXZ basis sets, the DLPNO errors are 6–7 kcal/mol for def2-SVP and def2-TZVP using loose thresholds, although these errors are reduced below 2 kcal/mol for normal thresholds.

These tests establish that the cc-pVDZ basis set affords DLPNO errors that are representative of those obtained using either the cc-pVTZ, def2-SVP, or def2-TZVP, and can be used for more comprehensive testing to follow. The role of diffuse functions will be considered separately, in Section III E.

B. DLPNO approximation in small systems

We now proceed to establish how the DLPNO errors scale with molecular size, using the cc-pVDZ basis set, and in order to establish a baseline we first present results for small systems at the RI-MP2/cc-pVDZ level. For these tests we use the S66 data set,[69] and individual results for each S66 dimer are presented in Fig. 3. Three different values of T_{cutPNO} are compared along with a N/T extrapolation, the latter of which will be called DLPNO[N/T] in the discussion that follows. These data are partitioned into the three standard subsets of S66,[69] namely, hydrogen-bonded dimers (for which electrostatics dominates dispersion), dispersion-dominated complexes, and dimers where dispersion and electrostatics are comparable in magnitude.

For the normal threshold ($T_{\text{cutPNO}} = 10^{-8}$), errors with respect to the canonical RI-MP2 interaction energy range up up to 0.11 kcal/mol with an average of 0.06 kcal/mol, whereas for $T_{\text{cutPNO}} = 10^{-9}$ they range up to 0.10 kcal/mol with an average of 0.06 kcal/mol. Extrapolated values afford errors up to 0.11 kcal/mol with an average of 0.05 kcal/mol. While these energy differences seem small, and close inspection of Fig. 3 illustrates that the DLPNO error vanishes systematically as T_{cutPNO} is tightened for the hydrogen-bonded complexes, that is not always the case for the other two subsets of S66, for which dispersion is more important. Nevertheless, the normal and tight thresholds (meaning $T_{\text{cutPNO}} \leq 10^{-8}$) afford errors $\lesssim$ 0.1 kcal/mol for all of these small dimers. Even for the loose-threshold case, the maximum error is 0.24 kcal/mol.

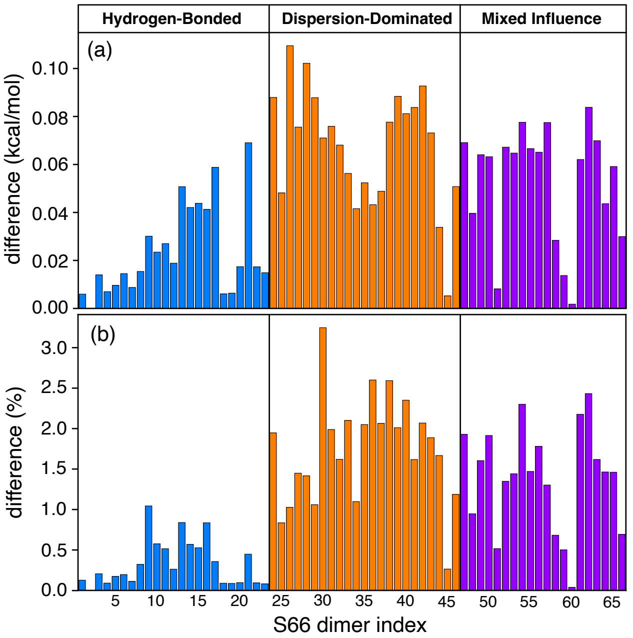
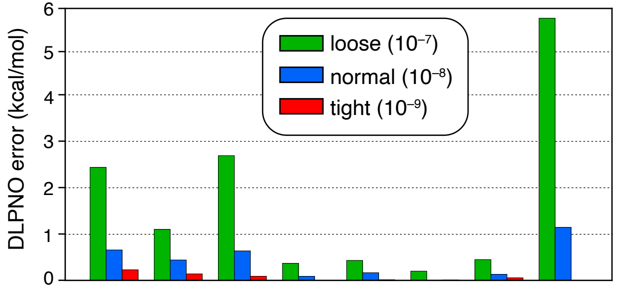


Fig. 4: (a) Absolute and (b) percent differences between the DLPNO-MP2[N/T]/cc-pVDZ interaction energy and the canonical MP2/cc-pVDZ value, for the S66 dimers.

(a) absolute



(b) percent

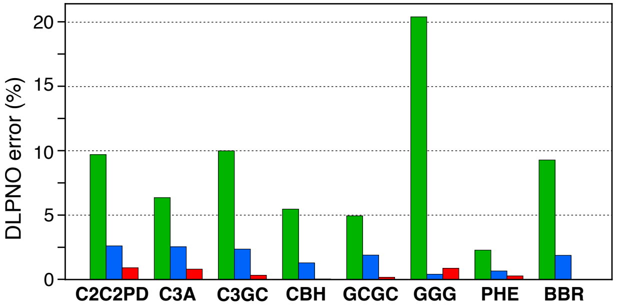


Fig. 5: (a) Absolute and (b) percent DLPNO errors in RI-MP2/cc-pVDZ calculations for the L7 and **BBR** complexes.

Figure 4 provides the DLPNO[N/T]-MP2 error, relative to the canonical MP2 result, for each of the S66 dimers. All errors are $\lesssim 0.1$ kcal/mol (or $\lesssim 3\%$), comparable to what is obtainable using the "normal" threshold $T_{\text{cutPNO}} = 10^{-8}$, with no need to use the tighter threshold or to extrapolate. For these small systems, then, one has reached a point of diminishing returns already at the level of the normal threshold. Errors are largest for the dispersion-bound subset of S66 and smallest for the hydrogen-bonded subset. This is consistent with the idea that dispersion is sensitive to tails of the density and "weak pairs" (in many-body language). Even for the dispersion-dominated complexes, however, DLPNO errors amount to $\lesssim 3\%$ of the total interaction energy. The DLPNO approximation thus works exceedingly well for these small non-covalent complexes, with normal thresholds or anything tighter, and is almost negligibly different from the canonical result upon extrapolation. Larger errors for the dispersion-bound subset do portend some problems in large van der Waals complexes, however, including π-stacked PAHs. Larger systems are considered next.

C. DLPNO approximation in larger systems

Although the DLPNO errors are quite small for the S66 complexes, interaction energies E_{int} for S66 range only from $|E_{\text{int}}| = 1$ 19 kcal/mol,[69] whereas van der Waals complexes with $\gtrsim 100$ atoms may have $|E_{\text{int}}| > 100$ kcal/mol.[34] We next examine the L7[39] and S12L[40] complexes in detail, as examples of larger systems, to which we add **BBR** from Ref. [31], as an example where LNO-CCSD(T)/CBS benchmarks differ markedly from FN-DMC results.

DLPNO errors for the L7 and **BBR** complexes are plotted in Fig. 5, computed at the RI-MP2/cc-pVDZ level, and they are much larger than the corresponding errors for the S66 complexes, except (arguably) in the case of tight PNO thresholds. For the loose, normal, and tight thresholds the average DLPNO error is 1.7 kcal/mol, 0.4 kcal/mol, and 0.1 kcal/mol, respectively. Maximum errors are 5.6 kcal/mol, 1.1 kcal/mol, and 0.2 kcal/mol, respectively. Loose thresholds are thus unacceptable for benchmark calculations in systems of this size, with percentage errors ranging up to nearly 20%. However, normal thresholds afford DLPNO errors $\lesssim 1$ kcal/mol and reduce the percentage errors below 10%. Tight thresholds are practically indistinguishable from canonical results.

The accuracy can be further improved via extrapolation, and both L/N and N/T results are shown in Fig. 6. Errors are reduced to $\lesssim 1$ kcal/mol by either extrapolation. For each of the L7 complexes, errors are $<$ 0.5 kcal/mol even for L/N extrapolation, whereas for **BBR** the N/T extrapolation is required to reach 0.5 kcal/mol DLPNO error, which is about 1% of the total interaction energy in that case. Note that achieving $\sim 1\%$ error requires N/T extrapolation for both the most strongly-bound complex (**BBR**) and for the most weakly-bound complex (**GGG**).

DLPNO-MP2/cc-pVDZ errors for the S12L complexes are shown in Fig. 7. (Similar errors are obtained at

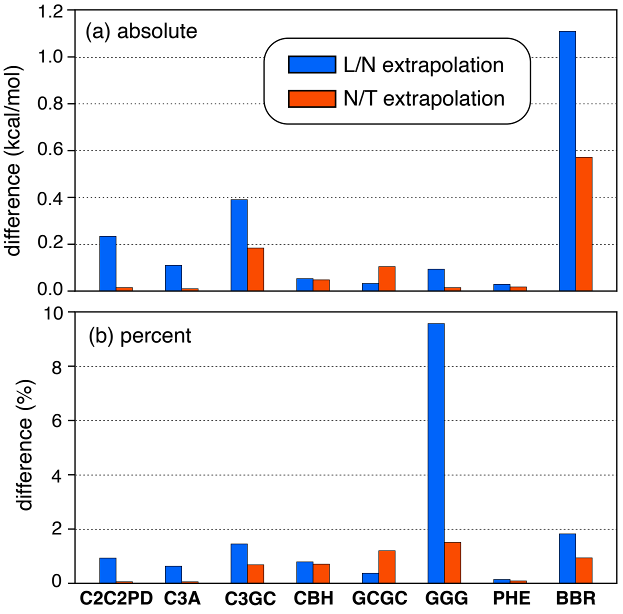


Fig. 6: (a) Absolute and (b) percent differences between L/N and N/T extrapolated values and canonical RI-MP2/cc-pVDZ interaction energies, for the L7 complexes along with **BBR**.

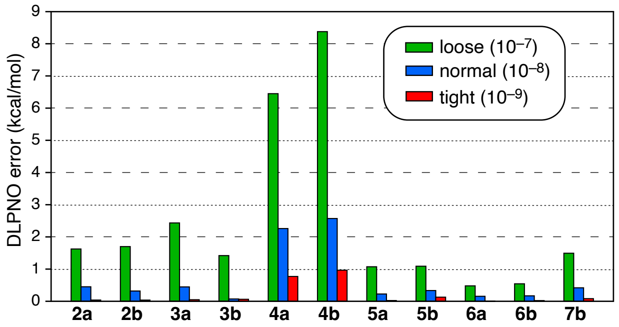


Fig. 7: DLPNO errors for S12L interaction energies computed at the RI-MP2/cc-pVDZ level.

the MP2/def2-SVP level; see Fig. S2.) Errors are large for $T_{\text{cutPNO}} = 10^{-7}$, ranging from 0.5 kcal/mol to 8.4 kcal/mol with an average of 2.4 kcal/mol. Tighter thresholds afford acceptable errors, ranging up to 2.4 kcal/mol (with an average of 0.7 kcal/mol) for $T_{\text{cutPNO}} = 10^{-8}$ versus a maximum error of 1.0 kcal/mol for $T_{\text{cutPNO}} = 10^{-9}$, with an average of 0.2 kcal/mol in that case. The largest errors are for the dispersion-bound systems **4a** and **4b**, which are complexes of C_{60} and C_{70} with the corannulene-based "buckycatcher" molecule, $C_{60}H_{28}$.[70] Even in those cases, however, setting $T_{\text{cutPNO}} \leq 10^{-8}$ affords an interaction energy that is very close to the canonical RI-MP2 value. DLPNO errors for the S12L complexes can be further reduced by extrapolation (Fig. S3), affording errors below 0.6 kcal/mol (or $\lesssim$ 2%) for L/N extrapolation, and $<$ 0.3 kcal/mol (or $<$ 1%) for N/T extrapolation.

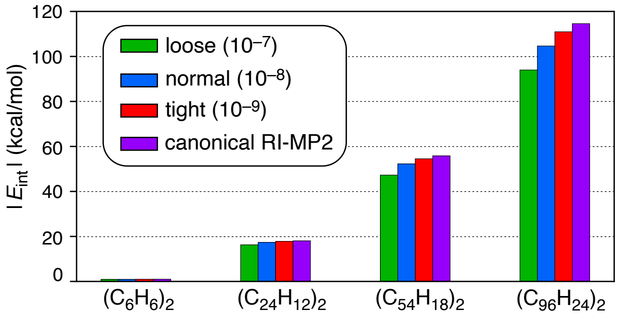


Fig. 8: Absolute interaction energies $|E_{\text{int}}|$ for polybenzenoid graphene dimers of increasing size, computed at the RI-MP2/cc-pVDZ level and comparing DLPNO-MP2 results with various values of T_{cutPNO}.

It is worth noting that the tight threshold needed for N/T extrapolation adds significantly to the cost of the DLPNO calculations. We find that the $(C_{96}H_{24})_2$ system that is explored in Section III D represents an approximate crossover point when $T_{\text{cutDO}} = 5 \times 10^{-3}$ and $T_{\text{cutPNO}} = 10^{-7}$. That calculation requires 614 min for DLPNO-MP2/cc-pVDZ, running on 40 processors with 4.5 Gb of memory per processor, which should be compared to 617 min for canonical RI-MP2/cc-pVDZ on 48 processors with 5.0 Gb per processor. A single-point DLPNO-MP2/cc-pVDZ calculation with $T_{\text{cutPNO}} = 10^{-8}$ requires 4,825 min on 40 processors, and $T_{\text{cutPNO}} = 10^{-9}$ requires 6,531 min. Although the timing comparisons will improve significantly (in favor of the DLPNO approach) when CCSD(T) replaces MP2, these data make it clear that the need for tight thresholds comes at a significant increase in cost.

D. Errors versus system size

Prompted by the fact that dispersion-bound complexes in S66 afforded the largest errors, along with the documented dependence of the DLPNO error on molecular size,[25,29] we decided to examine a sequence of increasingly large polybenzenoid dimers in the circumcoronene family. Absolute interaction energies $|E_{\text{int}}|$ for these complexes are plotted in Fig. 8, comparing DLPNO-MP2 results with different thresholds alongside the canonical MP2 interaction energy. Especially for the $(C_{54}H_{18})_2$ and $(C_{96}H_{24})_2$ systems, there is evident accumulation of error that increases with system size, which is systematically reduced (though not eliminated) as T_{cutPNO} is tightened. For the largest of these systems, the DLPNO error reaches 30 kcal/mol for $T_{\text{cutPNO}} = 10^{-7}$ and is several kcal/mol even in the coronene and circumcoronene dimers. Use of $T_{\text{cutPNO}} = 10^{-9}$ pushes the error below about 1 kcal/mol except in the case of $(C_{96}H_{24})_2$, where it remains nearly 5 kcal/mol even with this tightest threshold. As such, DLPNO results without extrapola-

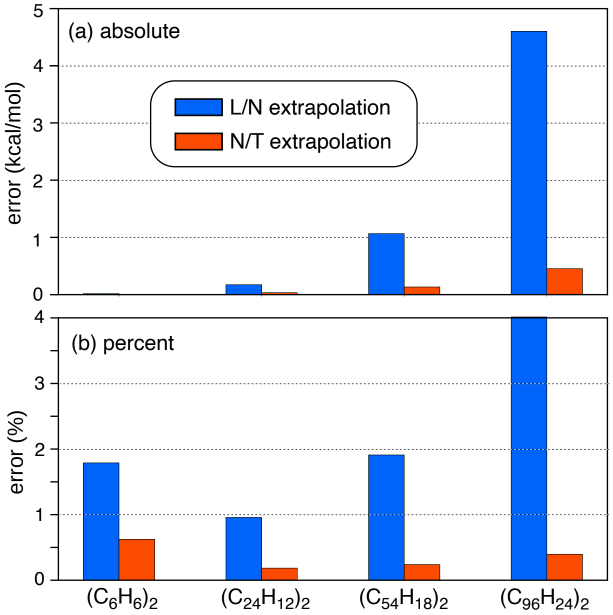


Fig. 9: (a) Absolute and (b) percent errors in DLPNO[L/N]- and DLPNO[N/T]-MP2/cc-pVDZ interaction energies for polybenzenoid dimers, as compared to the canonical RI-MP2/cc-pVDZ result.

tion do not provide benchmark-quality results for systems of this size.

Extrapolated DLPNO-MP2 errors for these polybenzenoid dimers are presented in Fig. 9. The L/N extrapolation performs poorly for the two largest systems, with errors that exceed 1 kcal/mol for $(C_{54}H_{18})_2$ and 4 kcal/mol for $(C_{96}H_{24})_2$. (Due to the magnitude of the interaction energies in these large systems, even the latter error amounts to only 4% of $|E_{\text{int}}|$.) On the other hand, the N/T extrapolation performs exceptionally well and reduces errors to < 0.5 kcal/mol even for these two systems, or about 0.5% of $|E_{\text{int}}|$. This suggest that the N/T extrapolation works well, even for systems where the accuracy of the DLPNO approximation itself is seemingly degraded, namely, nanoscale dispersion-bound complexes with $\gtrsim 100$ atoms. Absent extrapolation, however, the DLPNO results are unacceptable in these systems.

To examine whether the error accumulation in these graphene nano-flakes is merely a function of molecular size, or whether extended conjugation may also play a role, we next examine analogous dimers of the perhydrocircumcoronenes, which are the structures whose basic building block is cyclohexane rather than benzene. A structure for the dimer of [24]graphane (perhydrocoronene,[71] $C_{24}H_{36}$) was obtained from Ref. 52 and larger structures were constructed around it, maintaining the interlayer separation at 4.7 Å, as in the $(C_{24}H_{36})_2$ structure in Ref. 52. DLPNO errors for these graphane dimers, are given in Table 1, again computed at the RI-MP2/cc-pVDZ level.

Table 1: DLPNO errors at the MP2/cc-pVDZ level for dimers of polycyclic hydrocarbons

Complex	DLPNO Error (kcal/mol)		
	loose	normal	tight
$(\text{benzene})_2$	0.08	0.04	0.01
$(\text{cyclohexane})_2$	0.06	0.03	0.01
difference[a]	−0.02(25%)	−0.01(25%)	0.00(0%)
$(\text{coronene})_2$	1.79	0.71	0.26
$([24]\text{graphane})_2$	0.81	0.51	0.15
difference[b]	−0.98(55%)	−0.20(28%)	−0.11(42%)
$(\text{circumcoronene})_2$	8.52	3.55	1.27
$([54]\text{graphane})_2$	3.27	1.95	0.78
difference[b]	−5.26(62%)	−1.60(45%)	−0.49(39%)

[a] Difference between the graphene and graphane values (given in parenthesis as a percentage of graphane value).

Although the DLPNO errors do increase with system size for both the circumcoronene (or [n]graphene) and perhydrocircumcoronene ([n]graphane) dimers, the effect is much larger for the conjugated hydrocarbons, up to 5 kcal/mol larger for the C_{54} dimers with loose thresholds. This may result from many-body dispersion screening effects that arise in the conjugated systems, or what Dobson has termed "type B" nonadditivity,[72,73] although this effect is ordinarily considered not to be captured at the MP2 level.[72] It could simply be an effect of the more delocalized molecular orbitals in graphene as compared to graphane, leading to larger errors in a localized-orbital implementation. Whatever the origin, these data suggest that systems with extensive conjugation may represent some of the most problematic cases for the DLPNO approximation, for which the use of tight thresholds should be considered mandatory.

E. Effect of diffuse functions

We next wish to examine the role of diffuse basis functions. However, our attempts to perform MP2/aug-cc-pVXZ calculations for the L7 complexes were stymied by SCF convergence problems in ORCA, even with aug-cc-pVDZ. (The usual solution to such problems is to tighten the shell-pair drop tolerance,[45] but we have set that threshold to 10^{-16} a.u. for all calculations reported here.) We were able to converge these calculations using jun-cc-pVDZ for some of the L7 and S12L complexes (see Fig. S4), and more consistently using def2-SVPD. DLPNO errors for RI-MP2/def2-SVPD interaction energies are shown in Fig. 10 for the L7 complexes and in Fig. S5 for the S12L complexes.

For L7, these errors range up to 2.8 kcal/mol for **BBR**, using normal thresholds, although they are smaller than 0.1 kcal/mol for some of the other complexes and percentage errors are on par with, or even smaller than, those observed using cc-pVDZ (Fig. 5). The S12L complexes **4a** and **4b** exhibit DLPNO errors larger than 3 kcal/mol.

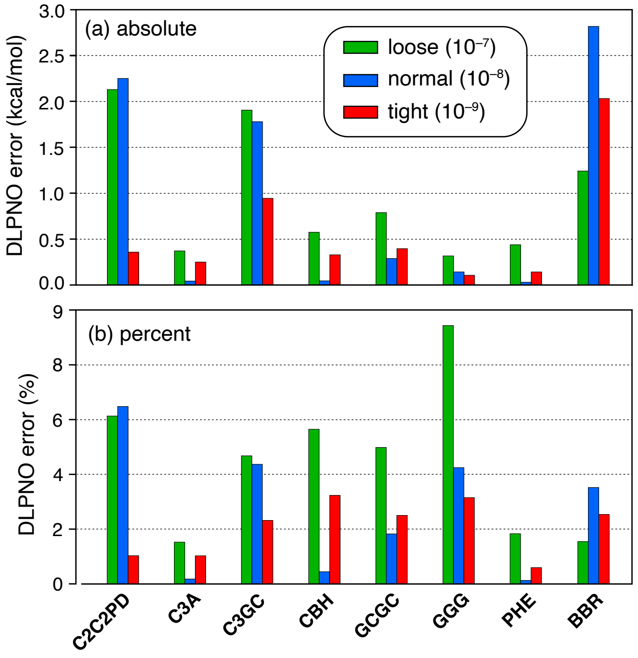


Fig. 10: (a) Absolute and (b) percent differences between the DLPNO-MP2/def2-SVPD and canonical MP2/def2-SVPD interaction energies for the L7 complexes.

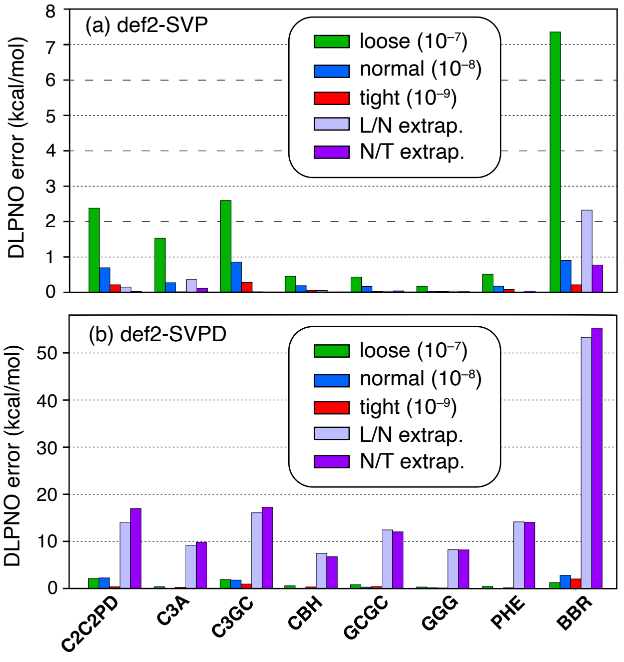


Fig. 11: DLPNO errors as a function of T_{cutPNO} for RI-MP2 calculations on L7 complexes using (a) the def2-SVP basis set and (b) the def2-SVPD basis set. The L/N and N/T extrapolations use Eq. (1).

More importantly, the DLPNO errors for both sets of complexes oscillate (as a function of T_{cutPNO}) in the presence of diffuse basis functions and thus do not converge in a manner that lends itself to extrapolation. This not an artifact of Karlsruhe basis sets, since calculations using def2-SVP do not exhibit such oscillations (Fig. 2), but seems to be inherent to the use of diffuse basis functions. DLPNO-MP2/jun-cc-pVDZ calculations also fail to converge monotonically as a function of T_{cutPNO} (Fig. S4).

For the **C2C2PD** and **BBR** complexes, setting $T_{\text{cutPNO}} = 10^{-8}$ affords a larger DLPNO error as compared to $T_{\text{cutPNO}} = 10^{-7}$, and for **C3A**, **CBH**, **GCGC**, and **PHE** the tight value $T_{\text{cutPNO}} = 10^{-9}$ affords a larger DLPNO error than the normal value $T_{\text{cutPNO}} = 10^{-8}$. Although the absolute differences are < 3 kcal/mol (or $< 10\%$), they are significant on the scale of benchmark-quality work. For the S12L systems (Fig. S5), DLPNO errors obtained with tight thresholds are larger than those computed using loose thresholds for complexes **2a**, **2b**, **6a**, **6b**, and **7a**, while for **4a** and **7b** the normal-threshold error is larger than the loose-threshold error.

This oscillatory behavior has a dramatic and detrimental effect on the extrapolation to the canonical limit, as shown for the L7 complexes and **BBR** in Fig. 11, which presents the extrapolations using both def2-SVP and def2-SVPD data. (For the corresponding percentage errors, see Figs. S5 and S6.) For the latter basis set, extrapolated results are significantly less accurate as compared to interaction energies computed even using loose thresholds, to the point that the extrapolation is simply unusable. In contrast, for def2-SVP the L/N and N/T extrapolations are typically more accurate than the normal- and tight-threshold results, respectively. This proves *not* to be the case for the **BBR** and **C3A** complexes, even using def2-SVP, although even in those cases the extrapolations still improve upon the looser of the two thresholds that was used as input, which is no longer the case once diffuse basis functions are added. Again, this behavior is not limited to Karlsruhe basis sets; extrapolations using jun-cc-pVDZ frequently afford larger errors (with respect to the canonical MP2 calculation) than the input data, as shown in Fig. S7.

One possible (or partial) explanation for these issues with extrapolation, especially in the case of def2-SVP, is that the extrapolation formula in Eq. (1) was parameterized using cc-pVXZ basis sets, for which an empirical parameter $F = 1.5 \pm 0.2$ was determined.[24] To examine this issue, we used RI-MP2/def2-SVPD calculations on the S66 data set to redetermine the value of F in Eq. (1). We find that the aggregate, extrapolated DLPNO error for S66 is minimized when $F = 1.097$ for L/N extrapolation or $F = 0.980$ for N/T extrapolation, leading to mean absolute DLPNO errors of 0.04 kcal/mol (L/N) and 0.02 kcal/mol (N/T). We then used the reparameterized extrapolations to examine the L7 and **BBR** complexes at the RI-MP2/def2-SVPD level.

Table 2 compares these new extrapolations alongside the values obtained using $F = 1.5$, as used above and in Ref. 24. The re-optimized extrapolations provide very little improvement ($\lesssim 0.2$ kcal/mol), as compared to extrapolated errors that exceed 8 kcal/mol and range to

Table 2: DLPNO-MP2/def2-SVPD errors (in kcal/mol) upon extrapolation.

Complex	L/N		N/T	
	$F = 1.5$	$F = 1.097$	$F = 1.5$	$F = 0.980$
C2C2PD	14.1	14.1	17.0	16.0
C3A	9.2	9.3	9.8	9.7
C3GC	16.1	16.0	17.3	16.8
BBR	53.3	53.9	55.3	54.9
CBH	7.4	7.2	6.8	6.9
GCGC	12.4	12.2	12.0	12.1
GGG	8.2	8.2	8.2	8.2
PHE	14.1	14.0	14.1	14.0

> 50 kcal/mol for **BBR**. This supports the idea that it is the diffuse functions themselves that are the source of the large errors, not the particular choice of basis-set family or the empirical extrapolation scheme. This is also supported by the fact that the reparameterized extrapolations work about equally well as the original one in the def2-SVP basis set, affording DLPNO errors that average 0.5 kcal for the L7 systems with L/N extrapolation, or 0.1 kcal/mol for the N/T scheme. These are essentially identical to the extrapolated errors obtained using $F = 1.5$.

These data strongly suggest that diffuse functions should be avoided in DLPNO calculations, which is potentially problematic for intermolecular interaction energies where diffuse functions often play an important role in obtaining converged results,[42–45] although counterpoise correction can often serve the same purpose.[34,44] To provide a quantitative perspective on this, we performed canonical RI-MP2 calculations for the S66 data set using (aug-)cc-pVDZ and (aug-)cc-pVTZ, combined with a two-point (aug-)cc-p[D/T]Z extrapolation to the CBS limit. That extrapolation uses the formula[74]

$$E_{\text{corr}}(\infty) = \frac{X^3 E_{\text{corr}}(X) - Y^3 E_{\text{corr}}(Y)}{X^3 - Y^3} \tag{2}$$

for the correlation energy (with $X = 2$ and $Y = 3$), and is combined with the corresponding Hartree-Fock/(aug-)cc-pVQZ energy to estimate the interaction energy at the MP2/CBS level. We omit quadruple-ζ MP2 calculations on the grounds that these may not be feasible for the nanoscale van der Waals complexes of interest in this work.

Figure 12 shows the difference between putative MP2/CBS interaction energies for the S66 complexes, comparing cc-pV[D/T]Z extrapolation to aug-cc-pV[D/T]Z extrapolation. Differences range up to up to 0.7 kcal/mol (6%), with an average absolute difference of 0.2 kcal/mol (4%) and a maximum percentage difference of 10%. Note that calculations with diffuse basis sets are significantly more expensive. For the largest S66 system (pentane dimer), the RI-MP2/aug-cc-pVDZ calculation took 1.2 h on 12 processors, as compared to only 0.2 h for RI-MP2/cc-pVDZ on the same hardware. For small systems, this additional effort may not be justified to capture < 1 kcal/mol interaction energy, but the difference between diffuse and non-augmented basis sets is likely extensive with system size, so it remains to be determined whether diffuse functions are important in sizable van der Waals complexes. Our own experience with DFT calculations of protein–ligand interaction energies suggests this may be the case.[34,44]

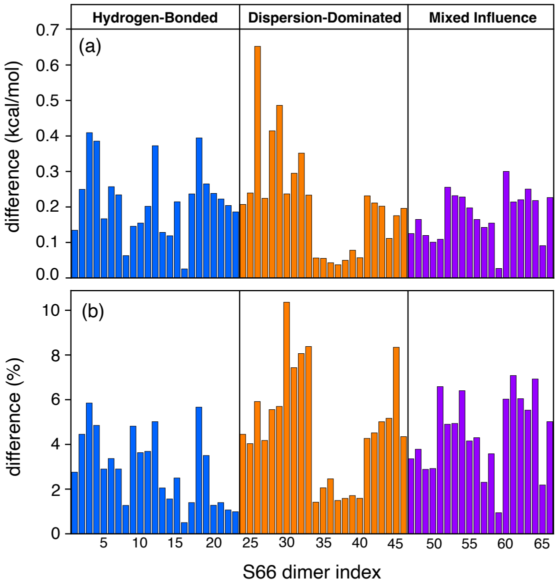


Fig. 12: (a) Absolute and (b) percent differences between CBS-extrapolated RI-MP2 interaction energies for the S66 dimers, comparing results in cc-pV[D/T]Z versus aug-cc-pV[D/T]Z basis sets.

IV. CONCLUSIONS

The DLPNO approximation is an excellent tool for high-quality *ab initio* calculation of interaction energies for large supramolecular complexes, which would not be accessible in canonical implementations of the MP2 or CCSD(T) methods. However, whereas "loose" thresholds[68] ($T_{\text{cutPNO}} = 10^{-7}$) are faithful to within 3% percent of the canonical result in small systems, the DLPNO error (measured with respect to canonical results) is largest for dispersion-bound complexes and grows with molecular size. Tight thresholds ($T_{\text{cutPNO}} = 10^{-9}$) are mandatory in complexes with $\gtrsim 100$ atoms. Alternatively, extrapolation to the canonical limit ($T_{\text{cutPNO}} \to 0$) can afford good results even when only the "normal" threshold value ($T_{\text{cutPNO}} = 10^{-7}$) is feasible, by means of an extrapolation scheme that is accurate to within 1% of the interaction energy. This is true even for large complexes involving C_{60} and C_{70} with the equally large "buckycatcher" partner, $C_{60}H_{28}$. For systems such as $(C_{96}H_{24})_2$, which is a model of π-stacking in graphene, the L/N extrapolation procedure is insufficient (affording a DLPNO error of 5 kcal/mol) and a N/T extrapolation must be used instead, requiring a tight PNO threshold of $T_{\mathrm{cutPNO}} = 10^{-9}$. This reduces the DLPNO error below 1 kcal/mol, to about 0.5% for $(C_{96}H_{24})_2$.

We have also shown that the success of these extrapolations is predicated on the use of basis sets that do not contain diffuse functions. When diffuse functions are included, the results are no longer systematically convergent results as T_{cutPNO} is reduced, and extrapolation to the canonical limit becomes impossible. DLPNO errors > 50 kcal/mol are documented herein for interaction energies involving C_{60}. For small complexes, such as those in the S66 data set, addition of diffuse functions modifies interaction energies by $< 5\%$ upon CBS extrapolation, although this may not be the case for larger complexes.[44] More work is needed on this front, to decide whether diffuse functions can sensibly be avoided in benchmark-quality calculations involving large supramolecular complexes.

SUPPLEMENTARY MATERIAL

Coordinates for the graphene and graphane tests systems created for this work.

ACKNOWLEDGEMENTS

This work was support by the National Institute of General Medical Sciences of the National Institutes of Health under award no. R43GM148095-01A1. Calculations were performed at the Ohio Supercomputer Center.[75]

Author Contributions

Montgomery Gray: Conceptualization (contributing); Data curation (lead); Formal analysis (lead); Investigation (lead); Methodology (lead); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review and editing (equal). **John M. Herbert:** Conceptualization (lead); Formal analysis (contributing); Funding acquisition (lead); Methodology (contributing); Project administration (lead); Supervision (lead); Visualization (contributing); Writing – review and editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

---

[1] J. D. Watts, "Chemical reactions: Thermochemical calculations", in *Handbook of Computational Chemistry*, J. Leszczynski, Ed.; Springer: Dordrecht, 2012; chapter 13, pages 467–482.

[2] A. Karton, "Quantum mechanical thermochemical predictions 100 years after the Schrödinger equation", *Annu. Rep. Comput. Chem.*, **18**, 123–166 (2022).

[3] J. Řezáč and P. Hobza, "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.*, **9**, 2151–2155 (2013).

[4] K. Patkowski, "Benchmark databases of intermolecular interaction energies: Design, construction, and significance", *Annu. Rep. Comput. Chem.*, **13**, 3–91 (2017).

[5] M. Kodrycka and K. Patkowski, "Platinum, gold, and silver standards of intermolecular interaction energy calculations", *J. Chem. Phys.*, **151**, 070901:1–17 (2019).

[6] C. Riplinger, B. Sandhoefer, A. Hansen, and F. Neese, "Natural triple excitations in local coupled cluster calculations with pair natural orbitals", *J. Chem. Phys.*, **139**, 134101:1–13 (2013).

[7] M. Sparta and F. Neese, "Chemical applications carried out by local pair natural orbital based coupled-cluster methods", *Chem. Soc. Rev.*, **43**, 5032–5041 (2014).

[8] D. G. Liakos, M. Sparta, M. K. Kesharwani, J. M. L. Martin, and F. Neese, "Exploring the accuracy limits of local pair natural orbital coupled-cluster theory", *J. Chem. Theory Comput.*, **11**, 1525–1539 (2015).

[9] Y. Guo, C. Riplinger, U. Becker, D. G. Liakos, Y. Minenkov, L. Cavallo, and F. Neese, "Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [DLPNO-CCSD(T)]", *J. Chem. Phys.*, **148**, 011101:1–5 (2018).

[10] Y. Guo, C. Riplinger, D. G. Liakos, U. Becker, M. Saitow, and F. Neese, "Linear scaling perturbative triples correction approximations for open-shell domain-based local pair natural orbital coupled cluster singles and doubles theory [DLPNO-CCSD(T_0/T)]", *J. Chem. Phys.*, **152**, 024116:1–13 (2020).

[11] D. G. Liakos and F. Neese, "Comprehensive benchmark results for the domain based local pair natural orbital coupled cluster method (DLPNO-CCSD(T)) for closed- and open-shell systems", *J. Phys. Chem. A*, **124**, 90–100 (2020).

[12] W. Li, Z. Ni, and S. Li, "Cluster-in-molecule local correlation method for post-Hartree–Fock calculations of large systems", *Mol. Phys.*, **114**, 1447–1460 (2016).

[13] Y. Guo, U. Becker, and F. Neese, "Comparison and combination of "direct" and fragment based local correlation methods: Cluster in molecules and domain based local pair natural orbital perturbation and coupled cluster theories", *J. Chem. Phys.*, **148**, 124117:1–11 (2018).

[14] Q. Ma and H.-J. Werner, "Explicitly correlated local

coupled-cluster methods using pair natural orbitals", *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, **8**, e1371:1–32 (2018).

[15] P. R. Nagy and M. Kállay, "Optimization of the linear-scaling local natural orbital CCSD(T) method: Redundancy-free triples correction using Laplace transform", *J. Chem. Phys.*, **146**, 214106:1–15 (2017).

[16] P. R. Nagy, G. Samu, and M. Kállay, "Optimization of the linear-scaling local natural orbital CCSD(T) method: Improved algorithm and benchmark applications", *J. Chem. Theory Comput.*, **14**, 4193–4215 (2018).

[17] P. R. Nagy and M. Kállay, "Approaching the basis set limit of CCSD(T) energies for large molecules with local natural orbital coupled-cluster methods", *J. Chem. Theory Comput.*, **15**, 5275–5298 (2019).

[18] P. R. Nagy, L. Gyevi-Nagy, B. D. Lörincz, and M. Kállay, "Pursuing the basis set limit of CCSD(T) non-covalent interaction energies for medium-sized complexes: Case study on the S66 compilation", *Mol. Phys.*, **121**, e2109526:1–26 (2022).

[19] F. Neese, A. Hansen, and D. G. Liakos, "Efficient and accurate approximations to the local coupled cluster singles doubles method using a truncated pair natural orbital basis", *J. Chem. Phys.*, **131**, 064103:1–15 (2009).

[20] P. Pinski, C. Riplinger, E. F. Valeev, and F. Neese, "Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. I. An efficient and simple linear scaling local MP2 method that uses an intermediate basis of pair natural orbitals", *J. Chem. Phys.*, **143**, 034108:1–17 (2015).

[21] C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, and F. Neese, "Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory", *J. Chem. Phys.*, **144**, 024109:1–10 (2016).

[22] F. Neese, F. Wennmohs, U. Becker, and C. Riplinger, "The ORCA quantum chemistry program package", *J. Chem. Phys.*, **152**, 224108:1–18 (2020).

[23] D. G. Liakos, A. Hansen, and F. Neese, "Weak molecular interactions studied with parallel implementations of the local pair natural orbital coupled pair and coupled cluster methods", *J. Chem. Theory Comput.*, **7**, 76–87 (2011).

[24] A. Altun, F. Neese, and G. Bistoni, "Extrapolation to the limit of a complete pair natural orbital space in local coupled-cluster calculations", *J. Chem. Theory Comput.*, **16**, 6142–6149 (2020).

[25] J. Chen, T. Sun, Y. Want, and W. Wang, "Toward a less costly but accurate calculation of the CCSD(T)/CBS noncovalent interaction energy", *J. Comput. Chem.*, **41**, 1252–1260 (2019).

[26] E. Brémond, H. Li, J. C. Sancho-García, and C. Adamo, "Double hybrids and noncovalent interactions: How far can we go?", *J. Phys. Chem. A*, **126**, 2590–2599 (2022).

[27] G. Santra and J. M. L. Martin, "Pure and hybrid SCAN, rSCAN, and r^{2}SCAN: Which one is preferred in KS- and HF-DFT calculations, and how does D4 dispersion correction affect this ranking?", *Molecules*, **27**, 141:1–11 (2022).

[28] Z. Ni, Y. Guo, F. Neese, W. Li, and S. Li, "Cluster-in-molecule local correlation method with an accurate distant pair correction for large systems", *J. Chem. Theory Comput.*, **17**, 756–766 (2021).

[29] I. Sandler, J. Chen, M. Taylor, S. Sharma, and J. Ho, "Accuracy of DLPNO-CCSD(T): Effect of basis set and system size", *J. Phys. Chem. A*, **125**, 1553–1563 (2021).

[30] K. Sorathia and D. P. Tew, "Basis set extrapolation in pair natural orbital theories", *J. Chem. Phys.*, **153**, 174112:1–8 (2020).

[31] Y. S. Al-Hamdani, P. R. Nagy, A. Zen, D. Barton, M. Kállay, J. G. Brandenburg, and A. Tkatchenko, "Interactions between large molecules pose a puzzle for reference quantum mechanical methods", *Nat. Commun.*, **12**, 3927:1–12 (2021).

[32] C. Villot, F. Ballesteros, D. Wang, and K. U. Lao, "Coupled cluster benchmarking of large noncovalent complexes in L7 and S12L as well as the C_{60} dimer, DNA–ellipticine, and HIV–indinavir", *J. Phys. Chem. A*, **126**, 4326–4341 (2022).

[33] M. Kállay, P. R. Nagy, D. Mester, Z. Rolik, G. Samu, J. Csontos, J. Csóka, P. B. Szabó, L. Gyevi-Nagy, I. Ladjánszki, L. Szegedy, B. Ladóczki, B. Ladóczki, K. P. andd M. Farkas, P. D. Mezei, and A. Ganyecz, "The MRCC program system: Accurate quantum chemistry from water to proteins", *J. Chem. Phys.*, **152**, 074107:1–18 (2020).

[34] M. Gray and J. M. Herbert, "Density functional theory for van der Waals complexes: Size matters", *Annu. Rep. Comput. Chem.* (2024). (accepted for publication; preprint available at DOI: 10.26434/chemrxiv-2023-xgr3q-v3).

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

[36] J. Hermann, R. A. DiStasio Jr., and A. Tkatchenko, "First-principles models for van der Waals interactions in molecules and materials: Concepts, theory, and applications", *Chem. Rev.*, **117**, 4714–4758 (2017).

[37] K. U. Lao and J. M. Herbert, "Accurate and efficient quantum chemistry calculations of noncovalent interactions in many-body systems: The XSAPT family of methods", *J. Phys. Chem. A*, **119**, 235–253 (2015).

[38] K. Carter-Fenk, K. U. Lao, K.-Y. Liu, and J. M. Herbert, "Accurate and efficient *ab initio* calculations for supramolecular complexes: Symmetry-adapted perturbation theory with many-body dispersion", *J. Phys. Chem. Lett.*, **10**, 2706–2714 (2019).

[39] R. Sedlak, T. Janowski, M. Pitoňák, J. Řezáč, P. Pulay, and P. Hobza, "Accuracy of quantum chemical methods for large noncovalent complexes", *J. Chem. Theory Comput.*, **9**, 3364–3374 (2013).

[40] T. Risthaus and S. Grimme, "Benchmarking of London dispersion-accounting density functional theory methods on very large molecular complexes", *J. Chem. Theory Comput.*, **9**, 1580–1591 (2013).

[41] H. Neugebaur, P. Pinski, S. Grimme, F. Neese, and M. Bursch, "Assessment of DLPNO-MP2 approximations in double-hybrid DFT", *J. Chem. Theory Comput.*, **19**, 7695–7703 (2023).

[42] J. Witte, J. B. Neaton, and M. Head-Gordon, "Push it to the limit: Characterizing the convergence of common sequences of basis sets for intermolecular interactions as described by density functional theory", *J. Chem. Phys.*, **144**, 194306:1–12 (2016).

[43] M. Gray and J. M. Herbert, "Comprehensive basis-set testing of extended symmetry-adapted perturbation theory and assessment of mixed-basis combinations to reduce cost", *J. Chem. Theory Comput.*, **18**, 2308–2330 (2022).

[44] M. Gray, P. E. Bowling, and J. M. Herbert, "Systematic examination of counterpoise correction in density functional

theory", *J. Chem. Theory Comput.*, **18**, 6742–6756 (2022).

[45] M. Gray, P. E. Bowling, and J. M. Herbert, "In defense of (certain) Pople-type basis sets", *ChemRxiv* (2024). (DOI: 10.26434/chemrxiv-2024-jz473).

[46] T. Kawase, K. Tanaka, N. Fujiwara, H. R. Darabi, and M. Oda, "Complexation of a carbon nanoring with fullerenes", *Angew. Chem. Int. Ed. Engl.*, **115**, 1662–1666 (2003).

[47] M. Birowska, K. Milowska, and J. A. Majewski, "Van der Waals density functionals for graphene layers and graphite", *Acta Phys. Pol. A*, **120**, 845–848 (2011).

[48] A. V. Rozhkov, A. O. Sboychakov, A. L. Rakhmanov, and F. Nori, "Electronic properties of graphene-based bilayer systems", *Phys. Rep.*, **648**, 1–104 (2016).

[49] M. Gray and J. M. Herbert, "Origins of offset-stacking in porous frameworks", *J. Phys. Chem. C*, **127**, 2675–2686 (2023).

[50] K. Carter-Fenk and J. M. Herbert, "Electrostatics does not dictate the slip-stacked arrangement of aromatic π–π interactions", *Chem. Sci.*, **11**, 6758–6765 (2020).

[51] A. A. Fokin, D. Gerbig, and P. R. Schreiner, "σ/σ- and π/π-interactions are equally important: Multilayered graphanes", *J. Am. Chem. Soc.*, **133**, 20036–20039 (2011).

[52] T. Janowski and P. Pulay, "A benchmark comparison of σ/σ and π/π dispersion: The dimers of naphthalene and decalin, and coronene and perhydrocoronene", *J. Am. Chem. Soc.*, **134**, 17520–17525 (2012).

[53] K. Carter-Fenk and J. M. Herbert, "Reinterpreting π-stacking", *Phys. Chem. Chem. Phys.*, **22**, 24870–24886 (2020).

[54] J. O. Sofo, A. S. Chaudhari, and G. D. Barber, "Graphane: A two-dimensional hydrocarbon", *Phys. Rev. B*, **75**, 153401:1–4 (2007).

[55] T. H. Dunning, Jr., "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen", *J. Chem. Phys.*, **90**, 1007–1023 (1989).

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

[57] F. Weigend and R. Ahlrichs, "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy", *Phys. Chem. Chem. Phys.*, **7**, 3297–3305 (2005).

[58] D. Rappoport and F. Furche, "Property-optimized Gaussian basis sets for molecular response calculations", *J. Chem. Phys.*, **133**, 134105:1–11 (2010).

[59] P. Jurečka, J. Šponer, J. Černý, and P. Hobza, "Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs", *Phys. Chem. Chem. Phys.*, **8**, 1985–1993 (2006).

[60] A. Tkatchenko, R. A. DiStasio Jr., M. Head-Gordon, and M. Scheffler, "Dispersion-corrected Møller–Plesset second-order perturbation theory", *J. Chem. Phys.*, **131**, 094106:1–7 (2009).

[61] D. Cremer, "Møller–Plesset perturbation theory: From small molecule methods to methods for thousands of atoms", *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, **1**, 509–530 (2011).

[62] B. Nguyen, G. P. Chen, M. M. Agee, A. M. Burow, M. Tang, and F. Furche, "Divergence of many-body perturbation theory for noncovalent interactions of large molecules", *J. Chem. Theory Comput.*, **16**, 2258–2273 (2020).

[63] J. Pogrebetsky, A. Siklitskaya, and A. Kubas, "MP2-based correction scheme to approach the limit of a complete pair natural orbitals space in DLPNO-CCSD(T) calculations", *J. Chem. Theory Comput.*, **19**, 4023–4032 (2023).

[64] R. M. Richard, K. U. Lao, and J. M. Herbert, "Achieving the CCSD(T) basis-set limit in sizable molecular clusters: Counterpoise corrections for the many-body expansion", *J. Phys. Chem. Lett.*, **4**, 2674–2680 (2013).

[65] L. A. Burns, M. S. Marshall, and C. D. Sherrill, "Comparing counterpoise-corrected, uncorrected, and averaged binding energies for benchmarking noncovalent interactions", *J. Chem. Theory Comput.*, **10**, 49–57 (2014).

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

Wesolowski, K. B. Whaley, H. L. Woodcock III, P. M. Zimmerman, S. Faraji, P. M. W. Gill, M. Head-Gordon, J. M. Herbert, and A. I. Krylov, “Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package”, *J. Chem. Phys.*, **155**, 084801:1–59 (2021).

[67] F. Weigend, A. Köhn, and C. Hättig, “Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations”, *J. Chem. Phys.*, **116**, 3175–3183 (2002).

[68] Philosophically, we dislike the idea of ascribing names like “tight” or “loose” to well-defined numerical thresholds. In our view, such terminology unduly insulates users from numerical details of the algorithm, leading to overreliance upon (and overconfidence in) software-specific default values for various thresholds, which might change without warning from one version of the software to the next. Unfortunately, the loose/normal/tight language for T_{cutDO} and T_{cutPNO} is ubiquitous in the literature, so we shall use it here, although our discussion is liberally interspersed with the actual numerical value of T_{cutPNO}.

[69] J. Řezác, K. E. Riley, and P. Hobza, “S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures”, *J. Chem. Theory Comput.*, **7**, 2427–2438 (2011). Erratum: *J. Chem. Theory Comput.* **10**, 1359–1360 (2014).

[70] A. Sygula, F. R. Fronczek, R. Sygula, P. W. Ragideau, and M. M. Olmstead, “A double concave hydrocarbon buckycatcher”, *J. Am. Chem. Soc.*, **129**, 3842–3843 (2007).

[71] J. C. Fetzer, W. R. Biggs, and M. Zander, “The preparation and purification of perhydrocoronene for use as a spectroscopic matrix”, *Z. Naturforsch. A*, **46**, 291–292 (1991).

[72] J. F. Dobson, “Beyond pairwise additivity in London dispersion interactions”, *Int. J. Quantum Chem.*, **114**, 1157–1161 (2014).

[73] K. U. Lao and J. M. Herbert, “A simple correction for nonadditive dispersion within extended symmetry-adapted perturbation theory (XSAPT)”, *J. Chem. Theory Comput.*, **14**, 5128–5142 (2018).

[74] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, “Basis-set convergence in correlated calculations on Ne, N_2, and H_2O”, *Chem. Phys. Lett.*, **286**, 243–252 (1998).

[75] Ohio Supercomputer Center, http://osc.edu/ark:/19495/f5s1ph73 (accessed 2023-03-02).