RESEARCH ARTICLE | APRIL 09 2025

Systematic discrepancies between reference methods for noncovalent interactions within the S66 dataset ⊘

Benjamin X. Shi 💿 ; Flaviano Della Pia 💿 ; Yasmine S. Al-Hamdani 💿 ; Angelos Michaelides 💿 ; Dario Alfè 💿 ; Andrea Zen 🎞 💿

Check for updates

J. Chem. Phys. 162, 144107 (2025) https://doi.org/10.1063/5.0254021



Articles You May Be Interested In

Coupled cluster benchmarks of large noncovalent complexes: The L7 dataset as well as DNA–ellipticine and buckycatcher–fullerene

J. Chem. Phys. (April 2021)

Canonical coupled cluster binding benchmark for nanoscale noncovalent complexes at the hundred-atom scale

J. Chem. Phys. (December 2024)

A comparison of geometric parameters from PBE-based doubly hybrid density functionals PBE0-DH, PBE0-2, and xDH-PBE0

J. Chem. Phys. (November 2013)



14 April 2025 14:50:46



r٦٦

Export Citation

Systematic discrepancies between reference methods for noncovalent interactions within the S66 dataset

Cite as: J. Chem. Phys. 162, 144107 (2025); doi: 10.1063/5.0254021 Submitted: 20 December 2024 • Accepted: 20 March 2025 • **Published Online: 9 April 2025**

Benjamin X. Shi, 1 🔟 Flaviano Della Pia, 1 🔟 Yasmine S. Al-Hamdani, 2,3,4 🔟 Angelos Michaelides, 1 🔟 Dario Alfè.^{2,5,3,4} (D) and Andrea Zen^{2,4,a)} (D)

AFFILIATIONS

¹Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

²Dipartimento di Fisica Ettore Pancini, Universita di Napoli Federico II, Monte Sant'Angelo, I-80126 Napoli, Italy

³Thomas Young Centre, University College London, London WC1E 6BT, United Kingdom

⁴Department of Earth Sciences, University College London, London WC1E 6BT, United Kingdom

⁵London Centre for Nanotechnology, University College London, London WCIE 6BT, United Kingdom

a) Author to whom correspondence should be addressed: and rea.zen@unina.it

ABSTRACT

The accurate treatment of noncovalent interactions is necessary to model a wide range of applications, from molecular crystals to surface catalysts to aqueous solutions and many more. Quantum diffusion Monte Carlo (DMC) and coupled cluster theory with single, double, and perturbative triple excitations [CCSD(T)] are considered two widely trusted methods for treating noncovalent interactions. However, while they have been well-validated for small molecules, recent work has indicated that these two methods can disagree by more than 7.5 kcal/mol for larger systems. The origin of this discrepancy remains unknown. Moreover, the lack of systematic comparisons, particularly for medium-sized complexes, has made it difficult to identify which systems may be prone to such disagreements and the potential scale of these differences. In this work, we leverage the latest developments in DMC to compute interaction energies for the entire S66 dataset, containing 66 medium-sized complexes with a balanced representation of dispersion and electrostatic interactions. Comparison to previous CCSD(T) references reveals systematic trends, with DMC predicting stronger binding than CCSD(T) for electrostatic-dominated systems, while the binding becomes weaker for dispersion-dominated systems. We show that the relative strength of this discrepancy is correlated to the ratio of electrostatic and dispersion interactions, as obtained from energy decomposition analysis methods. Finally, we have pinpointed model systems: the hydrogen-bonded acetic acid dimer (ID 20) and dispersion-dominated uracil-cyclopentane dimer (ID 42), where these discrepancies are particularly prominent. These systems offer cost-effective benchmarks to guide future developments in DMC, CCSD(T), as well as the wider electronic structure theory community.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0254021

I. INTRODUCTION

Noncovalent interactions play a crucial role in many areas of science. These interactions govern the structure of molecular crystals¹ (e.g., in pharmaceutical drugs) and biomolecules² such as DNA and proteins and are relevant to supramolecular³ science and nanotechnology.^{4,5} They also underlie important processes across chemistry and biology, from protein-ligand binding⁶ to catalytic reactions, both on the surface⁷ and in solution.³ Understanding and unlocking new processes for these applications will increasingly rely on accurate computational modeling tools that can treat noncovalent interactions.8

Two methods of choice for modeling noncovalent interactions are quantum diffusion Monte Carlo⁹ (DMC) and coupled cluster theory¹⁰ with single, double, and perturbative triple excitations [CCSD(T)]. While these methods may not be as affordable as density functional theory¹¹ (DFT), the reference data they provide are pivotal for benchmarking and parametrizing the density functional approximations (DFAs) necessary for practical routine simulations. For example, the local density approximation (LDA) and many extensions build upon a DMC-based parametrization of the correlation energy,¹² while CCSD(T) interaction energy datasets have helped aid in the development of many modern dispersion corrections.^{13–16} In particular, the applicability of these methods to larger systems has rapidly expanded in recent years, arising from computer hardware improvements and, more importantly, algorithmic/methodological developments to both DMC^{17–26} and CCSD(T).^{27–34}

DMC and CCSD(T) solve the Schrödinger equation to model the systems with distinct approaches and corresponding approximations. Despite these differences, there are many examples where DMC and CCSD(T) have come into alignment. For example, in addition to small molecules,^{35–37} agreement has been obtained for graphene bilayer binding energies,³⁸ molecular crystal lattice energies,^{39–41} molecule–surface interactions,^{42–47} and vacancy formation energies.⁴⁸ Recently, this agreement has been shown to start to falter^{49–52} for large dispersion-bound molecules, with differences as large as 7.5 kcal/mol for a buckyball-ring ($C_{60}@[6]$ CPPA) complex.

The origin of the discrepancy between DMC and CCSD(T) for large dispersion-bound molecules is a topic of current debate, ${}^{50,53-55}$ particularly on the validity of the perturbative triples (T) contribution in CCSD(T). Schäfer *et al.*⁵⁰ have suggested that part of this discrepancy arises from missing contributions in (T) that can be accounted for by the (cT) approach. In addition, Semidalas *et al.*⁵⁶ have reported non-trivial discrepancies between CCSD(T) and post-CCSD(T) methods such as CCSDT(Q). Conversely, Lambie *et al.*⁵⁴ have found that CCSD(T) does not differ significantly from CCSDT(Q) using the Pariser–Parr–Pople (PPP) model^{57–59} for large conjugated systems. Similarly, Fishman *et al.*⁵³ and Lao⁵⁵ report only a slight overbinding of CCSD(T) against CCSDT(Q) that cannot explain the discrepancy with DMC.

Understanding these discrepancies between DMC and CCSD(T) for large molecules requires cross-validating these methods across systematic datasets, particularly those involving medium- to large-sized molecules,^{51,60,61} which sample a range of noncovalent interactions. While DMC and CCSD(T) have both been compared (to great agreement) for the A2462 and S2263 datasets of small molecular complexes, DMC has not been frequently applied to study medium-sized datasets. In particular, it has not been used to study the S66 dataset,⁶⁴ a compilation of 66 dimers that probes the two major types of noncovalent interactions: dispersion and hydrogen-bonding together with those of mixed character. As well as covering a range of interactions, many of the molecules considered form the building blocks for larger biomolecules along different binding configurations. Furthermore, the parallel-displaced benzene dimer⁶⁵ is included in this set of complexes, making it an interesting modeling challenge. Such a dataset has been pivotal toward benchmarking^{66–74} DFAs in DFT as well as lower-level approximations to wave-function methods⁷⁵⁻ and even machine-learning models.82

In this work, we leverage the latest developments in DMC to compute interaction energies for the entire S66 dataset. When compared to CCSD(T) estimates (taken from the literature), we reveal a consistent weaker binding of dispersion interactions and consistent stronger binding of electrostatic interactions in DMC.

In particular, we show that their differences are correlated to the ratio of electrostatic and dispersive interactions within the system. The discrepancies in dispersion-dominated systems are shown to be reduced when utilizing an (empirically fitted) CCSD(cT) formulation,⁵⁰ although notable differences remain. We identify specific systems with well-defined differences between DMC and CCSD(T) that can serve as model systems for testing future developments in both methods, setting the stage toward resolving their discrepancies.

II. METHODS

A. Diffusion Monte Carlo

The DMC interaction energies of the S66⁶⁴ dataset are computed as

$$\Delta E_{\rm int.} = E_{\rm dimer} - E_{\rm mon.\,1} - E_{\rm mon.\,2},\tag{1}$$

where E_{dimer} is the total energy of the dimer and $E_{\text{mon. 1}}$, $E_{\text{mon. 2}}$ are the total energies of the constituent monomers. In the S66 dataset, these monomers are kept fixed to their geometry in the dimer, which is in general different from their equilibrium geometry. In this work, we first computed the energies of the monomers with DMC at a chosen reference geometry. Subsequently, we added the deformation energy, that is, the energy difference between the geometry of the monomer in the dimer and against this reference geometry using CCSD(T). We provide further details on these calculations in Sec. S2.1 of the supplementary material and show for a subset of the S66 complexes that differences between DMC and CCSD(T) predictions of the deformation energies are within ~0.12 kcal/mol.

A detailed description of the DMC method can be found in Ref. 9. In this work, we compute fixed-node DMC interaction energies by using the CASINO code.²³ We use energy-consistent correlated electron pseudopotentials⁸⁴ (eCEPP) with the determinant locality approximation (DLA).²² The trial wavefunctions were of the Slater-Jastrow type with single Slater determinants, and the single-particle orbitals obtained from DFT local-density approximation (LDA) plane-wave calculations performed with PWscf^{85,8} using an energy cutoff of 600 Ry and re-expanded in terms of Bsplines.⁸⁷ The Jastrow factor included a two-body electron-electron (e-e) term, two-body electron-nucleus (e-n) terms, and three-body electron-electron-nucleus (e-e-n) terms. The variational parameters of the Jastrow have been optimized by minimizing the variance of each system. The final DMC estimates of $\Delta E_{int.}$ were extrapolated toward the zero time-step limit $(\tau \rightarrow 0)$ by making a cubic fit to a series of time-step estimates from 0.1 au down to 0.003 au. We estimate errors that capture both the stochastic errors in the fit as well as the errors in the cubic fit due to the changing behavior near the zero time-step limit. To do this, we make a linear fit on a subset of time steps below (and including) 0.02 au and calculate the difference of the extrapolated estimates from the linear fit against the original cubic fit. The final error estimate is taken to be the larger of the stochastic errors of the cubic fit or the difference between the linear fit and cubic fit, as discussed in Sec. S2.4 of the supplementary material.

The parameters chosen within the present work follow from previous DMC calculations for large molecules in Ref. 49 as well as molecular crystals in Refs. 40 and 41. Within these studies of noncovalent interactions, the LDA trial wavefunction was shown to be valid, either by comparison to experiments or when using trial wavefunctions with other DFAs. For the case of the AcOH dimer system (ID 20), we have performed our validation tests on the choice of trial wavefunction as well as localization approximation in Sec. S7 of the supplementary material.

B. Coupled cluster theory

Several CCSD(T) estimates^{64,75,77,79,80,88–92} of the S66 interaction energies are available in the literature. Here, we compare DMC to the average of three recent CCSD(T) calculations:^{75,89,92} the revised calculations from Řezáč *et al.*⁷⁵ the "SILVER" estimates from Kesharwani *et al.*⁸⁹ and the "14k-GOLD" estimates from Nagy *et al.*⁹² A brief description of the three different CCSD(T) calculations is reported in Sec. S3 of the supplementary material.⁹³

Schäfer *et al.*⁵⁰ have recently demonstrated that there exists an empirical relationship for dispersion-dominated complexes between the (cT) and the (T) correlation contributions to the total energy using the CCSD and MP2 correlation energies. The resulting (cT)-fit is of the form,

$$\frac{(\mathrm{T})}{(\mathrm{cT}) - \mathrm{fit}} = a + b \cdot \frac{\mathrm{MP2 \ corr.}}{\mathrm{CCSD \ corr.}},$$
(2)

where *a* and *b* were parameters fitted from comparing CCSD(cT) to CCSD(T) calculations for a set of dispersion-bound complexes. We have recomputed the CCSD(cT)-fit values from the original CCSD(T) "SILVER" estimates from Kesharwani *et al.*, adding the difference between (cT) and (T) to the final (averaged) CCSD(T) estimates, as given in Sec. S4 of the supplementary material.

III. RESULTS AND DISCUSSION

We report the final DMC estimates for the entire S66 dataset in Figs. 1–3. This dataset comprises a diverse range of interactions, and we have separated the systems according to the original S66 categories of hydrogen-bonded, dispersion-bonded, and "mixed"character systems in Figs. 1–3, respectively. The corresponding dimer complexes are visualized in Fig. S1 of the supplementary material. We report the DMC estimate of the interaction energy $\Delta E_{\rm int.}$ above the label of each S66 complex. In all cases, the errors on $\Delta E_{\rm int.}$ estimates are below 0.12 kcal/mol, with the majority below 0.10 kcal/mol, facilitating reliable comparisons to CCSD(T).

The strength of $\Delta E_{int.}$ varies significantly across the systems, from as large as -20.17 ± 0.07 kcal/mol for complex 20 (acetic acid dimer) to as weak as -1.11 ± 0.06 kcal/mol for complex 30 (benzene-ethene dimer), being stronger in the H-bonded systems. With gray crosses, we plot the difference between DMC and CCSD(T) estimates (as described in the Methods) for the three classes of interactions. We use DMC as the reference (i.e., zero), and plot blue vertical bars along the horizontal zero axis representing the errors in the DMC estimates. There is overall excellent agreement, with a mean absolute deviation (MAD) of 0.21 kcal/mol across the entire S66 dataset. We find systematic trends in the differences between CCSD(T) and DMC, with CCSD(T) predicting weaker binding compared to DMC for hydrogen-bonded systems in Fig. 1, with an MAD of ~0.24 kcal/mol, while predicting a stronger binding for dispersion-dominated systems in Fig. 2, with an MAD of ~0.24 kcal/mol. For the "mixed" character systems in Fig. 3, the MAD is lower at 0.14 kcal/mol.

The stronger binding of DMC over CCSD(T) has not been (systematically) reported before, with the acetic acid dimer (ID 20)



FIG. 1. Comparison between DMC interaction energies $\Delta E_{int.}$ calculated in the present work against CCSD(T) for a subset of systems in the S66 dataset with hydrogen bonds. The CCSD(T) estimate is taken as an average from three previous calculations,^{75,89,92} with the corresponding standard deviation as an error. The deviation of CCSD(T) from the DMC is plotted with gray crosses, with the statistical errors (corresponding to one standard deviation σ). The complex ID and label are provided below the *x*-axis, while the number above each *x*-axis tick represents the DMC $\Delta E_{int.}$ estimate, with the error on the reported digit given in parentheses. The uracil dimer (ID 17) is in its base-pair (BP) configuration.

Dispersion-dominated systems



FIG. 2. Comparison between DMC interaction energies $\Delta E_{int.}$ calculated in the present work against CCSD(T) for a subset of systems in the S66 dataset dominated by dispersion interactions. The first six dimers are $\pi - \pi$ stacked. Refer to the caption of Fig. 1 for the plot details. The CCSD(T) estimate is taken as an average from three previous calculations, ^{75,89,92} with the corresponding standard deviation as an error. Additional CCSD(cT)-fit estimates are reported with golden crosses. These are calculated by scaling the CCSD(T) estimates based on their MP2 and CCSD contributions using the approach described in Ref. 50.

	1 2									Miz	ked s	syste	ms									
	1.2 -	× CCSD(T) [MAD: 0.14]]																
	0.6 -																					
Difference against DMC [kcal/mol]	0.0 -	I - ×	-¥	×	-¥-	<u> </u>	- 7-	-¥-	Ţ	- <u></u> - ×	÷	- <u>Ť</u> -	-*-	Ť	- Ť -	- <u>+</u> - ×	- -	- <u></u> -	×	- `I -	-¥-	_
	-0.6 -	-2.61(8)	-3.43(8)	-3.11(8)	-2.90(6)	-1.54(4)	-4.64(8)	-4.25(7)	-3.19(6)	-3.93(7)	-3.03(7)	-5.11(8)	-4.22(12)	-3.05(4)	-4.98(6)	-2.63(8)	-3.08(7)	-3.51(8)	-2.78(7)	-4.26(6)	-3.81(6)	
	-1.2 -	Benzene…Benzene (TS) - 47 -	Pyridine…Pyridine (TS) - 48 -	BenzenePyridine (TS) - 49 -	Benzene…Ethyne (CH- π) - 50 -	Ethyne…Ethyne (TS) - 51 -	Benzene…AcOH (OH-π) - 52 -	Benzene···AcNH $_2$ (NH- π) - 53 –	Benzene…Water (OH-π) - 54 -	BenzeneMeOH (OH-π) - 55 -	κ Benzene····MeNH ₂ (NH- π) - 56 -	a Benzene…Peptide (NH-π) - 57 -	PyridinePyridine (CH-N) - 58 -	Ethyne…Water (CH-O) - 59 -	Ethyne…AcOH (OH-π) - 60 -	Pentane…AcOH - 61 -	Pentane…AcNH ₂ - 62 -	BenzeneAcOH - 63 -	Peptide…Ethene - 64 -	PyridineEthyne - 65 -	MeNH ₂ Pyridine - 66 -	

FIG. 3. Comparison between DMC interaction energies $\Delta E_{int.}$ calculated in the present work and CCSD(T) for a subset of systems in the S66 dataset with mixed bonding character. The systems consist of T-shaped (TS) aromatic ring complexes as well as X–H··· π (X = C,O,N) interactions. Refer to the caption of Fig. 1 for the plot details. The CCSD(T) estimate is taken as an average from three previous calculations, ^{75,89,92} with the corresponding standard deviation as an error.

showing the maximum deviation of 0.8 kcal/mol across all S66 systems. Within Sec. S7 of the supplementary material, we have confirmed that the computed $\Delta E_{int.}$ estimate (-20.17 ± 0.07 kcal/mol) does not depend on the chosen pseudopotential (eCEPP), localization scheme (DLA), or trial wavefunction (LDA). For example, we have also performed all-electron calculations, giving an estimate of -20.32 ± 0.12 kcal/mol. We have also performed tests using PBE and PBE0 trial wavefunctions, showing that the DMC ΔE_{int} has a negligible (< 0.15 kcal/mol) dependence on the nodal surface for the DFAs considered. Furthermore, we also computed estimates for two other localization schemes: T-move and determinant localization T-move, both of which were within the statistical uncertainties of our original estimate. It should be noted that while the absolute value of the difference can be significant for some hydrogen-bonded systems, the relative difference (normalized against the DMC $\Delta E_{int.}$) is significantly smaller, with a mean relative difference of 2.45% compared to 8.21% for the dispersion-dominated systems (see Sec. S6 of the supplementary material).

The weaker binding of DMC over CCSD(T) for dispersiondominated systems is now relatively well-documented,⁴ and there is evidence that it can be improved by replacing the perturbative triples (T) contribution with the recent $(cT)^{32}$ contribution. We plot the difference between an empirically CCSD(cT)-fit formulation as yellow crosses for the dispersion-dominated systems. In all cases, CCSD(cT) has weaker binding than CCSD(T), leaning closer toward DMC, leading to an MAD of 0.09 kcal/mol and a mean relative difference of 3.44%. However, this does not fully resolve the discrepancies across all of the dispersion systems, with a significant discrepancy of ~ 0.5 kcal/mol remaining for the uracilcyclopentane dimer (ID 42). Such a significant discrepancy makes this a worthwhile system to investigate further and could give clues on remaining discrepancies between DMC and CCSD(T) observed in Ref. 50. The analysis reported above highlights an important outcome of this work: the identification of smaller, simpler systems that show notable discrepancies between DMC and coupled cluster methods. Specifically, we found a discrepancy of approximately ~0.8 kcal/mol for the hydrogen-bonded acetic acid dimer (ID 20) and ~0.5 kcal/mol for the dispersion-dominated uracil-cyclopentane dimer (ID 42), which contain 64 and 98 (total) electrons, respectively. These medium-sized systems represent an almost tenfold reduction in electron count compared to the larger C₆₀@[6]CPPA buckyball-ring system (672 electrons) studied in Ref. 49. Thus, they might offer practical, cost-effective models for further exploring the discrepancy between DMC and coupled-cluster.

We now focus on the difference between DMC and CCSD(T) as a function of the dispersion and electrostatic contributions to the interaction energy. In particular, we find that the relative differences between DMC and CCSD(T) for each system within the S66 dataset can be correlated to the relative strength of the dispersion and electrostatic interactions that make up its $\Delta E_{int.}$. We used the Symmetry Adapted Perturbation Theory (SAPT) calculations (at the sSAPT0 level with the jun-cc-pVDZ basis set⁹⁵) from Burns *et al.*,⁹⁴ which decomposes $\Delta E_{int.}$ into contributions from electrostatics (ELST), exchange, induction, and dispersion (DISP). Notably, we show in Fig. 4 that there is a strong linear trend $(R^2 = 0.78)$ between the natural logarithm of the ELST and DISP contributions, log $(\frac{ELST}{DISP})$, and the relative difference (in %) between



FIG. 4. Error decomposition analysis. We report the difference between DMC and CCSD(T) relative to the DMC magnitude, that is, $(E_{\rm DMC} - E_{\rm CCSD(T)})/|E_{\rm DMC}|$, as a function of the natural logarithm of the electrostatic (ELST) to dispersion (DISP) ratio contribution to the binding energy. The ELST to DISP ratio is determined from the SAPT analysis from Ref. 94. The color code is red for H-bonded systems (ID from 1 to 23), blue for dispersion-dominated systems (ID from 24 to 46), and green for mixed systems (ID from 47 to 66).

CCSD(T) and DMC. In Sec. S9 of the supplementary material, we show that this strong linear trend remains at the more sophisticated SAPT2+(3) (CCD)/aug-cc-pVTZ level.⁸³ This analysis confirms our prior observations on the trends between DMC and CCSD(T). For example, the more dominant the DISP contribution to $\Delta E_{int.}$ (i.e., a more negative log $\left(\frac{\text{ELST}}{\text{DISP}}\right)$), the more CCSD(T) is found to underbind with respect to DMC. Similarly, the stronger the ELST contribution to $\Delta E_{int.}$ (i.e., a more positive log $\left(\frac{\text{ELST}}{\text{DISP}}\right)$), the more CCSD(T) is found to overbind with respect to DMC. We expect that this cheap descriptor can be used in the future to identify more challenging systems with larger discrepancies between DMC and CCSD(T).

Finally, we discuss briefly the potential origins of the observed discrepancies between DMC and CCSD(T) based on the current literature. For H-bonded systems, CCSDT(Q) estimates are available for the A24 dataset^{36,62}—a set of $\Delta E_{int.}$ for small dimer complexes. Nakano et al.⁹⁶ have performed DMC calculations for the entire A24 dataset, where there is a notable discrepancy of 0.26 ± 0.07 and 0.34 ± 0.07 kcal/mol for the water—ammonia and HCN dimer complexes, in line with our observation of stronger binding in DMC. The CCSDT(Q) references, albeit at small basis sets, find negligible (< 0.01 kcal/mol) changes relative to CCSD(T). For dispersionbound systems, there exist CCSDT(Q) estimates for the paralleldisplaced (PD) benzene dimer by Semidalas et al.⁵⁶ and by Karton and Martin,⁹⁷ which report differences to CCSD(T) of -0.085 and -0.058 kcal/mol, respectively, using small truncated doublezeta quality basis sets. The reported difference between DMC and CCSD(T) is -0.37 ± 0.08 kcal/mol for this system, which indicates that the majority of this difference is not covered when going to CCSDT(Q). For both H-bonded and dispersion-bound systems (as well as those of mixed character), the observed differences could arise either from higher-order excitations and larger basis sets needed from coupled cluster theory or biases in the FN-DMC evaluations, likely coming from the fixed-node approximation.

IV. CONCLUSIONS

To summarize, we have computed highly accurate estimates for the S66 dataset-one of the most widely used databases for noncovalent interactions in biological and organic molecules-with fixed-node diffusion quantum Monte Carlo. These estimates have provided new insights into recent discussions on its discrepancies with another widely trusted method: coupled cluster theory with single, double, and perturbative triple excitations [CCSD(T)]. Our data show systematic trends, with DMC predicting stronger binding in hydrogen-bonded systems than CCSD(T) and weaker binding in dispersion-dominated systems. We show that there is a correlation between the relative strength of these discrepancies with the nature of the interaction, specifically the relative ratio of the electrostatic and dispersion contributions to the interaction energy as provided by previous Symmetry Adapted Perturbation Theory (SAPT) calculations.94 In addition, we show that the discrepancy between DMC and CCSD(T) on dispersion-dominated systems can be reduced using a recently proposed CCSD(cT) formulation, albeit with still significant remaining differences. While this work does not identify the origin of the disagreement between DMC and CCSD(T), it has identified the type of interactions where it is particularly prevalent, and importantly, we have identified model systems within the S66 dataset where these errors are prominent. These results have strong implications for the electronic structure theory community, addressing the knowledge gap on the trends of DMC interaction energies for noncovalent molecular complexes. Furthermore, the accurate reference data produced within this work is expected to benefit the wider materials modeling community, being instrumental for benchmarking applications ranging from the development of machine-learned interatomic potentials to crystal structure prediction, drug design, and renewable energy.

SUPPLEMENTARY MATERIAL

See the supplementary material for details on the DMC calculations, comprising the convergence of the calculations with respect to the time step, the influence of the choice of the monomer geometry on the dimer interaction energy, as well as tests on the pseudopotential localization error and the Jastrow optimization.

ACKNOWLEDGMENTS

We acknowledge the computational resources from Cambridge Service for Data Driven Discovery (CSD3), operated by the University of Cambridge Research Computing Service, provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council (capital Grant Nos. EP/T022159/1 and EP/P020259/1) and DiRAC funding from the Science and Technology Facilities Council (www.dirac.ac.uk). We are further grateful for computational support from the UK National High Performance Computing Service, ARCHER2, for which access was obtained via the UKCP consortium and funded by EPSRC Grant Ref. No. EP/X035891/1. This research also used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract No. DE-AC05-00OR22725. A.M. and B.X.S acknowledge support from the European Union under the "n-AQUA" European Research Council project (Grant No. 101071937). D.A. and A.Z. acknowledge support from Leverhulme Grant No. RPG-2020-038 and from the European Union under the Next generation EU (Project Nos. 20222FXZ33 and P2022MC742).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

B.X.S. and F.D.P. contributed equally to this work.

Benjamin X. Shi: Conceptualization (equal); Data curation (lead); Formal analysis (equal); Investigation (lead); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing - original draft (lead); Writing - review & editing (equal). Flaviano Della Pia: Conceptualization (equal); Data curation (lead); Formal analysis (equal); Investigation (lead); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing - original draft (lead); Writing - review & editing (equal). Yasmine S. Al-Hamdani: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Visualization (equal); Writing review & editing (equal). Angelos Michaelides: Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing review & editing (equal). Dario Alfè: Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Supervision (equal); Writing - review & editing (equal). Andrea Zen: Conceptualization (lead); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (lead); Project administration (equal); Resources (equal); Software (equal); Supervision (lead); Validation (equal); Visualization (equal); Writing - original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its supplementary material. All analysis can be found on GitHub at https://github.com/zenandrea/FNDMC-S66 and can be viewed interactively online through associated Jupyter notebooks (via Google Colab), with links provided in the corresponding GitHub repository.

14 April 2025 14:50:46

REFERENCES

¹G. J. O. Beran, "Modeling polymorphic molecular crystals with electronic structure theory," Chem. Rev. **116**, 5567–5613 (2016).

²K. E. Riley and P. Hobza, "Noncovalent interactions in biochemistry," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **1**, 3–17 (2011).

³M. Raynal, P. Ballester, A. Vidal-Ferran, and P. W. N. M. van Leeuwen, "Supramolecular catalysis. Part 1: Non-covalent interactions as a tool for building and modifying homogeneous catalysts," Chem. Soc. Rev. **43**, 1660–1733 (2014).

⁴N. Planas, A. L. Dzubak, R. Poloni, L.-C. Lin, A. McManus, T. M. McDonald, J. B. Neaton, J. R. Long, B. Smit, and L. Gagliardi, "The mechanism of carbon dioxide adsorption in an alkylamine-functionalized metal–organic framework," J. Am. Chem. Soc. 135, 7402–7405 (2013).

⁵R. Poloni, B. Smit, and J. B. Neaton, "Ligand-assisted enhancement of CO₂ capture in metal-organic frameworks," J. Am. Chem. Soc. 134, 6714-6719 (2012).
⁶X. Du, Y. Li, Y.-L. Xia, S.-M. Ai, J. Liang, P. Sang, X.-L. Ji, and S.-Q. Liu, "Insights into protein-ligand interactions: Mechanisms, models, and methods," Int. J. Mol. Sci. 17, 144 (2016).

⁷F. R. Rehak, G. Piccini, M. Alessio, and J. Sauer, "Including dispersion in density functional theory for adsorption on flat oxide surfaces, in metal–organic frameworks and in acidic zeolites," Phys. Chem. Chem. Phys. 22, 7577–7585 (2020).

⁸K. Müller-Dethlefs and P. Hobza, "Noncovalent interactions: A challenge for experiment and theory," Chem. Rev. 100, 143–168 (2000).

⁹W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, "Quantum Monte Carlo simulations of solids," Rev. Mod. Phys. **73**, 33–83 (2001).

¹⁰R. J. Bartlett and M. Musiał, "Coupled-cluster theory in quantum chemistry," Rev. Mod. Phys. **79**, 291–352 (2007).

¹¹A. E. Mattsson, P. A. Schultz, M. P. Desjarlais, T. R. Mattsson, and K. Leung, "Designing meaningful density functional theory calculations in materials science—A primer," Modell. Simul. Mater. Sci. Eng. **13**, R1–R31 (2004).

¹²D. M. Ceperley and B. J. Alder, "Ground state of the electron gas by a stochastic method," Phys. Rev. Lett. 45, 566–569 (1980).

¹³S. Grimme, "Semiempirical GGA-type density functional constructed with a long-range dispersion correction," J. Comput. Chem. 27, 1787–1799 (2006).

¹⁴S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," J. Chem. Phys. **132**, 154104 (2010).

¹⁵S. Grimme, A. Hansen, J. G. Brandenburg, and C. Bannwarth, "Dispersion-Corrected mean-field electronic structure methods," Chem. Rev. **116**, 5105–5154 (2016).

¹⁶A. J. A. Price, A. Otero-de-la-Roza, and E. R. Johnson, "XDM-corrected hybrid DFT with numerical atomic orbitals predicts molecular crystal lattice energies with unprecedented accuracy," Chem. Sci. 14, 1252–1262 (2023).

¹⁷A. Zen, S. Sorella, M. J. Gillan, A. Michaelides, and D. Alfè, "Boosting the accuracy and speed of quantum Monte Carlo: Size consistency and time step," Phys. Rev. B **93**, 241118 (2016).

¹⁸J. T. Krogel, "Nexus: A modular workflow management system for quantum simulation codes," Comput. Phys. Commun. **198**, 154–168 (2016).

¹⁹M. C. Bennett, C. A. Melton, A. Annaberdiyev, G. Wang, L. Shulenburger, and L. Mitas, "A new generation of effective core potentials for correlated calculations," J. Chem. Phys. **147**, 224106 (2017).

²⁰ M. C. Bennett, G. Wang, A. Annaberdiyev, C. A. Melton, L. Shulenburger, and L. Mitas, "A new generation of effective core potentials from correlated calculations: 2nd row elements," J. Chem. Phys. **149**, 104108 (2018).

²¹ A. Annaberdiyev, G. Wang, C. A. Melton, M. C. Bennett, L. Shulenburger, and L. Mitas, "A new generation of effective core potentials from correlated calculations: 3d transition metal series," J. Chem. Phys. **149**, 134108 (2018).

²² A. Zen, J. G. Brandenburg, A. Michaelides, and D. Alfè, "A new scheme for fixed node diffusion quantum Monte Carlo with pseudopotentials: Improving reproducibility and reducing the trial-wave-function bias," J. Chem. Phys. **151**, 134105 (2019).

²³ R. J. Needs, M. D. Towler, N. D. Drummond, P. López Ríos, and J. R. Trail, "Variational and diffusion quantum Monte Carlo calculations with the CASINO code," J. Chem. Phys. **152**, 154106 (2020). ²⁴K. Nakano, C. Attaccalite, M. Barborini, L. Capriotti, M. Casula, E. Coccia, M. Dagrada, C. Genovese, Y. Luo, G. Mazzola, A. Zen, and S. Sorella, "TurboRVB: A many-body toolkit for *ab initio* electronic simulations by quantum Monte Carlo," J. Chem. Phys. **152**, 204121 (2020).

²⁵ P. R. C. Kent, A. Annaberdiyev, A. Benali, M. C. Bennett, E. J. Landinez Borda, P. Doak, H. Hao, K. D. Jordan, J. T. Krogel, I. Kylänpää, J. Lee, Y. Luo, F. D. Malone, C. A. Melton, L. Mitas, M. A. Morales, E. Neuscamman, F. A. Reboredo, B. Rubenstein, K. Saritas, S. Upadhyay, G. Wang, S. Zhang, and L. Zhao, "QMCPACK: Advances in the development, efficiency, and application of auxiliary field and real-space variational and diffusion quantum Monte Carlo," J. Chem. Phys. 152, 174105 (2020).

²⁶K. Nakano, O. Kohulák, A. Raghav, M. Casula, and S. Sorella, "TurboGenius: Python suite for high-throughput calculations of ab initio quantum Monte Carlo methods," J. Chem. Phys. **159**, 224801 (2023).

²⁷C. Riplinger and F. Neese, "An efficient and near linear scaling pair natural orbital based local coupled cluster method," J. Chem. Phys. **138**, 034106 (2013).

²⁸C. 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 (2016).

²⁹Q. Ma and H.-J. Werner, "Explicitly correlated local coupled-cluster methods using pair natural orbitals," WIREs Comput. Mol. Sci. 8, e1371 (2018).

³⁰P. 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).

³¹ 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).

³²N. Masios, A. Irmler, T. Schäfer, and A. Grüneis, "Averting the infrared catastrophe in the gold standard of quantum chemistry," Phys. Rev. Lett. **131**, 186401 (2023).

³³A. Jiang, Z. L. Glick, D. Poole, J. M. Turney, C. D. Sherrill, and H. F. Schaefer, III, "Accurate and efficient open-source implementation of domain-based local pair natural orbital (DLPNO) coupled-cluster theory using a t1-transformed Hamiltonian," J. Chem. Phys. **161**, 082502 (2024).

³⁴H.-Z. Ye and T. C. Berkelbach, "Periodic local coupled-cluster theory for insulators and metals," J. Chem. Theory Comput. **20**, 8948–8959 (2024).

³⁵M. Dubecký, P. Jurečka, R. Derian, P. Hobza, M. Otyepka, and L. Mitas, "Quantum Monte Carlo methods describe noncovalent interactions with subchemical accuracy," J. Chem. Theory Comput. **9**, 4287–4292 (2013).

³⁶J. Řezáč, M. Dubecký, P. Jurečka, and P. Hobza, "Extensions and applications of the A24 data set of accurate interaction energies," Phys. Chem. Chem. Phys. 17, 19268–19277 (2015).

³⁷A. Raghav, R. Maezono, K. Hongo, S. Sorella, and K. Nakano, "Toward chemical accuracy using the Jastrow correlated antisymmetrized geminal power *ansatz*," J. Chem. Theory Comput. **19**, 2222–2229 (2023).

³⁸E. Mostaani, N. D. Drummond, and V. I. Fal'ko, "Quantum Monte Carlo calculation of the binding energy of bilayer graphene," Phys. Rev. Lett. **115**, 115501 (2015).

³⁹A. Zen, J. G. Brandenburg, J. Klimeš, A. Tkatchenko, D. Alfè, and A. Michaelides, "Fast and accurate quantum Monte Carlo for molecular crystals," Proc. Natl. Acad. Sci. U. S. A. 115, 1724–1729 (2018).

⁴⁰F. Della Pia, A. Zen, D. Alfè, and A. Michaelides, "DMC-ICE13: Ambient and high pressure polymorphs of ice from diffusion Monte Carlo and density functional theory," J. Chem. Phys. **157**, 134701 (2022).

⁴¹ F. Della Pia, A. Zen, D. Alfè, and A. Michaelides, "How accurate are simulations and experiments for the lattice energies of molecular crystals?," Phys. Rev. Lett. 133, 046401 (2024).

⁴²O. Karalti, D. Alfè, M. J. Gillan, and K. D. Jordan, "Adsorption of a water molecule on the MgO(100) surface as described by cluster and slab models," Phys. Chem. Chem. Phys. 14, 7846–7853 (2012).

⁴³Y. S. Al-Hamdani, D. Alfè, and A. Michaelides, "How strongly do hydrogen and water molecules stick to carbon nanomaterials?," J. Chem. Phys. **146**, 094701 (2017).

⁴⁴T. Tsatsoulis, F. Hummel, D. Usvyat, M. Schütz, G. H. Booth, S. S. Binnie, M. J. Gillan, D. Alfè, A. Michaelides, and A. Grüneis, "A comparison between quantum

chemistry and quantum Monte Carlo techniques for the adsorption of water on the (001) LiH surface," J. Chem. Phys. 146, 204108 (2017).

⁴⁵Y. S. Al-Hamdani, M. Rossi, D. Alfè, T. Tsatsoulis, B. Ramberger, J. G. Brandenburg, A. Zen, G. Kresse, A. Grüneis, A. Tkatchenko, and A. Michaelides, "Properties of the water to boron nitride interaction: From zero to two dimensions with benchmark accuracy," J. Chem. Phys. **147**, 044710 (2017).

⁴⁶J. G. Brandenburg, A. Zen, M. Fitzner, B. Ramberger, G. Kresse, T. Tsatsoulis, A. Grüneis, A. Michaelides, and D. Alfè, "Physisorption of water on graphene: Subchemical accuracy from many-body electronic structure methods," J. Phys. Chem. Lett. **10**, 358–368 (2019).

⁴⁷B. X. Shi, A. Zen, V. Kapil, P. R. Nagy, A. Grüneis, and A. Michaelides, "Manybody methods for surface chemistry come of age: Achieving consensus with experiments," J. Am. Chem. Soc. **145**, 25372–25381 (2023).

⁴⁸B. X. Shi, V. Kapil, A. Zen, J. Chen, A. Alavi, and A. Michaelides, "General embedded cluster protocol for accurate modeling of oxygen vacancies in metal-oxides," J. Chem. Phys. **156**, 124704 (2022).

⁴⁹Y. 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 (2021).

⁵⁰T. Schäfer, A. Irmler, A. Gallo, and A. Grüneis, "Understanding discrepancies of wavefunction theories for large molecules," arXiv:2407.01442 (2024).

⁵¹F. Ballesteros, S. Dunivan, and K. U. Lao, "Coupled cluster benchmarks of large noncovalent complexes: The L7 dataset as well as DNA-ellipticine and buckycatcher-fullerene," J. Chem. Phys. **154**, 154104 (2021).

 52 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).

⁵³V. Fishman, M. Lesiuk, J. M. L. Martin, and A. D. Boese, "Another angle on benchmarking noncovalent interactions," J. Chem. Theory Comput. 21, 2311–2324 (2025).

⁵⁴S. Lambie, D. Kats, D. Usyvat, and A. Alavi, "On the applicability of CCSD(T) for dispersion interactions in large conjugated systems," J. Chem. Phys. 162, 114112 (2025).

⁵⁵K. U. Lao, "Canonical coupled cluster binding benchmark for nanoscale noncovalent complexes at the hundred-atom scale," J. Chem. Phys. **161**, 234103 (2024).

⁵⁶E. Semidalas, A. D. Boese, and J. M. L. Martin, "Post-CCSD(T) corrections in the S66 noncovalent interactions benchmark," Chem. Phys. Lett. 863, 141874 (2025).
 ⁵⁷R. Pariser and R. G. Parr, "A semi-empirical theory of the electronic spectra and electronic structure of complex unsaturated molecules. I," J. Chem. Phys. 21,

466-471 (1953).
⁵⁸R. Pariser and R. G. Parr, "A semi-empirical theory of the electronic spectra and electronic structure of complex unsaturated molecules. II," J. Chem. Phys. 21, 767-776 (1953).

⁵⁹J. A. Pople, "Electron interaction in unsaturated hydrocarbons," Trans. Faraday Soc. 49, 1375–1385 (1953).

⁶⁰ R. Sure and S. Grimme, "Comprehensive benchmark of association (free) energies of realistic host–guest complexes," J. Chem. Theory Comput. **11**, 3785–3801 (2015).

⁶¹S. Grimme, "Supramolecular binding thermodynamics by dispersion-corrected density functional theory," Chem.– Eur. J. **18**, 9955–9964 (2012).

⁶²J. 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).

⁶³ 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).

⁶⁴J. Řezáč, 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).

⁶⁵M. O. Sinnokrot and C. D. Sherrill, "Highly accurate coupled cluster potential energy curves for the benzene dimer: Sandwich, T-shaped, and parallel-displaced configurations," J. Phys. Chem. A **108**, 10200–10207 (2004). ⁶⁶L. Goerigk, H. Kruse, and S. Grimme, "Benchmarking density functional methods against the S66 and S66x8 datasets for non-covalent interactions," ChemPhysChem **12**, 3421–3433 (2011).

⁶⁷T. Gao, H. Li, W. Li, L. Li, C. Fang, H. Li, L. Hu, Y. Lu, and Z.-M. Su, "A machine learning correction for DFT non-covalent interactions based on the S22, S66 and X40 benchmark databases," J. Cheminform. **8**, 24 (2016).

⁶⁸H. Peng, Z.-H. Yang, J. P. Perdew, and J. Sun, "Versatile van der Waals density functional based on a meta-generalized gradient approximation," Phys. Rev. X 6, 041005 (2016).

⁶⁹F. Yu and Y. Wang, "Dual-hybrid direct random phase approximation and second-order screened exchange with nonlocal van der Waals correlations for noncovalent interactions," J. Comput. Chem. **41**, 1018–1025 (2020).

⁷⁰S. Grimme, A. Hansen, S. Ehlert, and J.-M. Mewes, "r²SCAN-3c: A 'Swiss army knife' composite electronic-structure method," J. Chem. Phys. **154**, 064103 (2021).

⁷¹S. Ehlert, U. Huniar, J. Ning, J. W. Furness, J. Sun, A. D. Kaplan, J. P. Perdew, and J. G. Brandenburg, "r²SCAN-D4: Dispersion corrected meta-generalized gradient approximation for general chemical applications," J. Chem. Phys. 154, 061101 (2021).

⁷²M. Müller, A. Hansen, and S. Grimme, "ωB97X-3c: A composite rangeseparated hybrid DFT method with a molecule-optimized polarized valence double-ζ basis set," J. Chem. Phys. **158**, 014103 (2023).

⁷³T. Lu and Q. Chen, "Simple, efficient, and universal energy decomposition analysis method based on dispersion-corrected density functional theory," J. Phys. Chem. A **127**, 7023–7035 (2023).

⁷⁴M. Lee, B. Kim, M. Sim, M. Sogal, Y. Kim, H. Yu, K. Burke, and E. Sim, "Correcting dispersion corrections with density-corrected DFT," J. Chem. Theory Comput. **20**, 7155–7167 (2024).

⁷⁵ J. Řezáč, K. E. Riley, and P. Hobza, "Extensions of the S66 data set: More accurate interaction energies and angular-displaced nonequilibrium geometries," J. Chem. Theory Comput. 7, 3466–3470 (2011).

⁷⁶K. E. Riley, J. Řezáč, and P. Hobza, "The performance of MP2.5 and MP2.X methods for nonequilibrium geometries of molecular complexes," Phys. Chem. Chem. Phys. 14, 13187–13193 (2012).

⁷⁷A. 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).

⁷⁸J. Shee, M. Loipersberger, A. Rettig, J. Lee, and M. Head-Gordon, "Regularized second-order Møller–Plesset theory: A more accurate alternative to conventional MP2 for noncovalent interactions and transition metal thermochemistry for the same computational cost," J. Phys. Chem. Lett. **12**, 12084–12097 (2021).

⁷⁹J. Lupi, S. Alessandrini, C. Puzzarini, and V. Barone, "junChS and junChS-F12 models: Parameter-free efficient yet accurate composite schemes for energies and structures of noncovalent complexes," J. Chem. Theory Comput. 17, 6974–6992 (2021).

⁸⁰E. Semidalas, G. Santra, N. Mehta, and J. M. L. Martin, "S66 noncovalent interactions benchmark re-examined: Composite localized coupled cluster approaches," AIP Conf. Proc. 2611, 020016 (2022).
 ⁸¹G. J. O. Beran, C. Greenwell, C. Cook, and J. Řezáč, "Improved description of

⁸¹G. J. O. Beran, C. Greenwell, C. Cook, and J. Řezáč, "Improved description of intra- and intermolecular interactions through dispersion-corrected second-order Møller–Plesset perturbation theory," Acc. Chem. Res. 56, 3525–3534 (2023).

⁸²A. S. Christensen, S. K. Sirumalla, Z. Qiao, M. B. O'Connor, D. G. A. Smith, F. Ding, P. J. Bygrave, A. Anandkumar, M. Welborn, F. R. Manby, and T. F. Miller, III, "OrbNet Denali: A machine learning potential for biological and organic chemistry with semi-empirical cost and DFT accuracy," J. Chem. Phys. 155, 204103 (2021).

⁸³C. Villot and K. U. Lao, "Ab initio dispersion potentials based on physics-based functional forms with machine learning," J. Chem. Phys. **160**, 184103 (2024).

⁸⁴ J. R. Trail and R. J. Needs, "Shape and energy consistent pseudopotentials for correlated electron systems," J. Chem. Phys. **146**, 204107 (2017).

⁸⁵P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Giron-coli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, "QUANTUM ESPRESSO: A

modular and open-source software project for quantum simulations of materials," J. Phys.: Condens. Matter **21**, 395502 (2009).

⁸⁶P. Giannozzi, O. Baseggio, P. Bonfà, D. Brunato, R. Car, I. Carnimeo, C. Cavazzoni, S. de Gironcoli, P. Delugas, F. Ferrari Ruffino, A. Ferretti, N. Marzari, I. Timrov, A. Urru, and S. Baroni, "Quantum ESPRESSO toward the exascale," J. Chem. Phys. **152**, 154105 (2020).

⁸⁷D. Alfe and M. J. Gillan, "Efficient localized basis set for quantum Monte Carlo calculations on condensed matter," Phys. Rev. B **70**, 161101 (2004).

⁸⁸B. Brauer, M. K. Kesharwani, S. Kozuch, and J. M. L. Martin, "The S66x8 benchmark for noncovalent interactions revisited: Explicitly correlated ab initio methods and density functional theory," Phys. Chem. Chem. Phys. 18, 20905–20925 (2016).

⁸⁹M. K. Kesharwani, A. Karton, N. Sylvetsky, and J. M. L. Martin, "The S66 noncovalent interactions benchmark reconsidered using explicitly correlated methods near the basis set limit," Aust. J. Chem. **71**, 238–248 (2018).

⁹⁰Q. Ma and H.-J. Werner, "Accurate intermolecular interaction energies using explicitly correlated local coupled cluster methods [PNO-LCCSD(T)-F12]," J. Chem. Theory Comput. 15, 1044–1052 (2019).

⁹¹G. Santra, E. Semidalas, N. Mehta, A. Karton, and J. M. L. Martin, "S66x8 noncovalent interactions revisited: New benchmark and performance of composite localized coupled-cluster methods," Phys. Chem. Chem. Phys. **24**, 25555–25570 (2022). ⁹²P. R. Nagy, L. Gyevi-Nagy, B. D. Lőrincz, 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 (2023).

⁹³Note that the difficulty of reaching the complete basis set (CBS) limit is the principle reason for several CCSD(T) evaluations of S66. As reaching the CBS limit remains a challenge, particularly in dispersion bound complexes.

⁹⁴L. A. Burns, J. C. Faver, Z. Zheng, M. S. Marshall, D. G. A. Smith, K. Vanommeslaeghe, A. D. MacKerell, Jr., K. M. Merz, Jr., and C. D. Sherrill, "The BioFragment database (BFDb): An open-data platform for computational chemistry analysis of noncovalent interactions," J. Chem. Phys. **147**, 161727 (2017).

⁹⁵T. M. Parker, L. A. Burns, R. M. Parrish, A. G. Ryno, and C. D. Sherrill, "Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies," J. Chem. Phys. **140**, 094106 (2014).

⁹⁶K. Nakano, B. X. Shi, D. Alfè, and A. Zen, "Basis set incompleteness errors in fixed-node diffusion Monte Carlo calculations on non-covalent interactions," arXiv:2412.00368 [physics] (2024).

⁹⁷A. Karton and J. M. L. Martin, "Prototypical π - π dimers re-examined by means of high-level CCSDT(Q) composite ab initio methods," J. Chem. Phys. **154**, 124117 (2021).