

Understanding discrepancies in noncovalent interaction energies from wavefunction theories for large molecules

Received: 25 November 2024

Accepted: 4 September 2025

Published online: 14 October 2025

 Check for updatesTobias Schäfer^{1,2}✉, Andreas Irmmler^{1,2}✉, Alejandro Gallo¹ & Andreas Grüneis¹✉

Are the currently used reference methods to approximately solve the many-electron Schrödinger equation accurate enough? Here, we investigate recently reported discrepancies of noncovalent interaction energies for large molecules predicted by two of the most widely-trusted many-electron theories: diffusion quantum Monte Carlo and coupled-cluster theory. We are able to unequivocally pin down the main source of the puzzling discrepancies and present modifications to widely-used coupled-cluster methods needed for more accurate noncovalent interaction energies of large molecules on the hundred-atom scale. This is of critical impact for a wide range of applications which rely on highly-accurate interaction energies between large and polarizable molecules.

The calculation of electronic transition energies for a single hydrogen atom in 1926 marks the beginning of an incredible successful era of quantum mechanics^{1,2}. Shortly after this breakthrough, Paul Dirac famously noted that the underlying physical laws necessary for much of physics and all of chemistry are now completely known. However, he also pointed out that the exact application of these laws leads to equations that are much too complicated to solve³. This paradigm of theoretical chemistry and physics is prevalent until today. In particular, the exponential growth of the computational complexity of the many-electron problem with system size still makes an exact solution of the electronic Schrödinger equation for more than a few atoms impossible. As a consequence, a hierarchy of increasingly accurate methods that is capable of producing reference results at the expense of tractable yet high computational cost has emerged. These reference results are pivotal in order to develop, assess, and further improve computationally more efficient but in general less accurate approximations. In this context, a prime example was the numerical prediction of highly accurate ground state energies for the uniform electron gas using diffusion Monte Carlo methods, leveraging the development of approximate exchange-correlation functionals that ultimately led to the breakthrough of density functional theory in computational materials science during the last decades^{4–6}.

At present, quantum mechanical many-electron calculations of systems containing more than 100 atoms have become possible thanks to methodological developments and considerable growth in

computing power. These methodological improvements are often based on taking advantage of the relative nearsightedness of many-electron correlation effects^{7–9}. In this manner, the scaling of the computational complexity with respect to system size can be lowered. However, recently several works^{10–13} showed that there exist alarming discrepancies between predicted interaction energies for large molecules when using two of the most widely-trusted highly accurate many-electron theories: DMC and CCSD(T), which stand for diffusion Monte Carlo and coupled-cluster theory using single, double, and perturbative triple particle-hole excitation operators, respectively. These observations are a source of great concern in the electronic structure theory community because, in the case of noncovalent interactions between molecular complexes, both CCSD(T) and DMC are considered highly reliable benchmark methods^{14–16}. Furthermore, the observed discrepancies are large enough to cause qualitative differences in calculated properties of materials, which can have scientific and technological implications. For example, accurate crystal structure predictions are crucial in drug design^{17–19}. Similarly, reliable reference methods are essential for discovering and designing new functional materials for applications such as renewable energy storage and conversion, including catalysis, or solar cells^{20–22}. Finally, as machine learning increasingly pervades all areas of computational first-principles physics, the accuracy of these reference methods, which provide the training data, becomes even more critical^{23–25}.

¹Institute for Theoretical Physics, TU Wien, Wiedner Hauptstraße 8–10/136, Vienna, Austria. ²These authors contributed equally: Tobias Schäfer, Andreas Irmmler. ✉ e-mail: tobias.schaefer@tuwien.ac.at; andreas.irmmler@tuwien.ac.at; andreas.grueneis@tuwien.ac.at

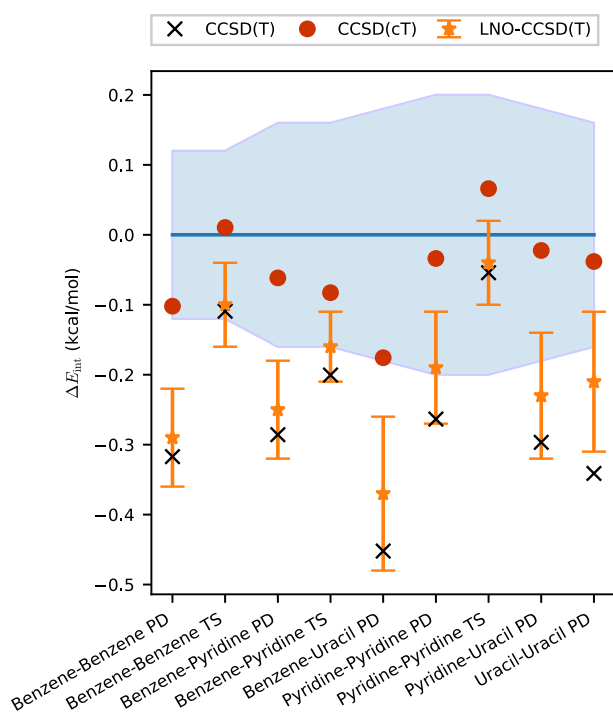


Fig. 1 | Deviations of coupled cluster results from diffusion quantum Monte Carlo (DMC) results for noncovalently bound dimers with up to 24 atoms. The dimers are in parallel displaced (PD) or T-shaped (TS) configurations. The CCSD(T) and CCSD(cT) values are calculated complete basis set (CBS) estimates obtained from basis set extrapolation using aug-cc-pVTZ and aug-cc-pVQZ basis sets^{46,47} (details in the supplementary information). The LNO-CCSD(T) and DMC results as well as their uncertainties are taken from ref. 11. The uncertainty of LNO-CCSD(T) is shown by the error bars. The uncertainty of DMC is shown by the blue area. Source data are provided as a Source Data file.

energies. The bra- and ket-states correspond to a triple excited and reference state, respectively. The (T) approximation disregards the term $[[\hat{V}, \hat{T}_2], \hat{T}_2]$, which is included in (cT) and also occurs in full CCSDT theory. This term effectively screens the bare Coulomb interaction of the $[\hat{V}, \hat{T}_2]$ term and has an opposite sign, making it crucial for systems with large polarizability. However, for small and weakly polarizable systems the $[[\hat{V}, \hat{T}_2], \hat{T}_2]$ contribution is small, making the (T) and (cT) approximation agree, as it was already shown for a set of small molecules²⁹.

We now demonstrate that using CCSD(cT) instead of CCSD(T) theory restores excellent agreement for noncovalent interaction energies with DMC findings. First, we consider again the coronene dimer. Table 1 shows that the binding energy for the coronene dimer calculated on the level of CCSD(cT) theory is by almost 2 kcal/mol closer to the DMC estimate compared to CCSD(T) theory, achieving chemical accuracy (1 kcal/mol) in comparison to DMC after subtracting error bars. Next, we investigate the accuracy of CCSD(T) and CCSD(cT) compared to DMC for noncovalent interactions in smaller molecules. To this end, we study a set of dimers containing up to 24 atoms that were also investigated in ref. 11. This gives us another opportunity to assess the effect of local approximations at the level of CCSD(T) theory. Figure 1 depicts the deviations of all computed interaction energies from DMC reference values taken from ref. 11. It should be noted that DMC references and differences to LNO-CCSD(T) interaction energies are shown with error bars¹¹. Using our massive computational parallelization approach, we are able to add canonical CCSD(T) interaction energies extrapolated to the CBS limit to the comparison to DMC. For these relatively small molecules, we can employ sufficiently

large basis sets, reducing the remaining uncertainty to -0.01 kcal/mol (see supplementary information). Importantly, our canonical CCSD(T) results are in good agreement with LNO-CCSD(T) findings to within its error bars. The only minor exception is observed for the parallel displaced uracil dimer, where canonical CCSD(T) predicts a slightly stronger interaction. A comparison to DMC reveals that CCSD(T) theory predicts on average about 0.3 kcal/mol stronger interaction energies. Based on LNO-CCSD(T) and DMC data alone such a statement cannot be made due to the relatively large and mostly overlapping error bars. However, our well converged canonical CCSD(T) findings allow drawing such conclusions. Only for the T-shaped pyridine and benzene dimers, DMC and CCSD(T) binding energies agree to within the DMC errors. Note that these systems have a smaller (T) contribution to the interaction energy, compared to the parallel displaced systems. All other systems exhibit small but significant discrepancies between CCSD(T) and DMC results, which is consistent with the even larger discrepancies reported for the larger molecules in ref. 11. Similar to our findings for the coronene dimer reported in Table 1, Fig. 1 shows that CCSD(cT) interaction energies agree significantly better with DMC values than their CCSD(T) counterparts.

Given the good agreement between DMC and CCSD(cT) for the systems studied above, an important question to ask is if CCSD(cT) is really more accurate than CCSD(T) for noncovalent interaction energies? To answer this question we compare interaction energies of both approaches to higher-level CC methods for complexes from the S22 data set. As can be observed in Fig. 2, we find that while (T) is in good agreement with T for total energies, it overestimates interaction energies. Here T stands for the triples contribution to the correlation energy, $E^T = E^{\text{CCSDT}} - E^{\text{CCSD}}$. While (cT) total energies do not match the accuracy of (T) when compared to T, this is a secondary consideration in the present work. Our primary focus is on interaction energies, where we demonstrate that (cT) significantly improves the accuracy compared to (T). As shown in Fig. 2, (cT) closely matches the T interaction energies, indicating its superior accuracy for weakly bound complexes. This effect is particularly strong for interaction energies with large triples correlation contributions. Here, additional triples contributions, neglected in the (T) model, are required to resolve the systematic overestimation of (T). In the supplementary information we present a more detailed analysis, which indicates that the most important contributions indeed come from fifth-order ring diagram contributions depicted in Eq. (1).

Estimating the overcorrelation of (T) for weak interactions

In summary, we have demonstrated that CCSD(cT) theory achieves excellent agreement for noncovalent interaction energies between molecular complexes compared to DMC and CCSDT theory. However, we stress that the CC series of methods (CCSD, CCSDT and CCSDTQ) is observed to yield monotonic and rapidly converging interaction energies for small and weakly bound complexes³⁰. Based on this knowledge, we emphasize that the Q contribution to the interaction energies can be expected to be smaller than its T counterpart, but could possibly yield a significant contribution. Indeed, this is part of the reason for the success of the CCSD(T) approximation for very small molecules, where CCSD(T) is often fortuitously closer to CCSDTQ than CCSDT³⁰. Here, we argue that this error cancellation no longer functions in the case of large molecular complexes involving strongly polarizable systems such as C2C2PD, C3GC and C60@[6]CPPA. A similar problem is known to occur in Møller-Plesset theory, where the truncation of the perturbation series also leads to significantly too strong interaction energies for systems with large polarizability, although MP2 yields relatively accurate interaction energies for systems with an intermediate polarizability²⁷. To quantify and support the statements above, Fig. 2c illustrates that there exists a correlation between the ratio of (T) and (cT) with the ratio of the MP2 and CCSD correlation energy contributions to the interaction energies of all

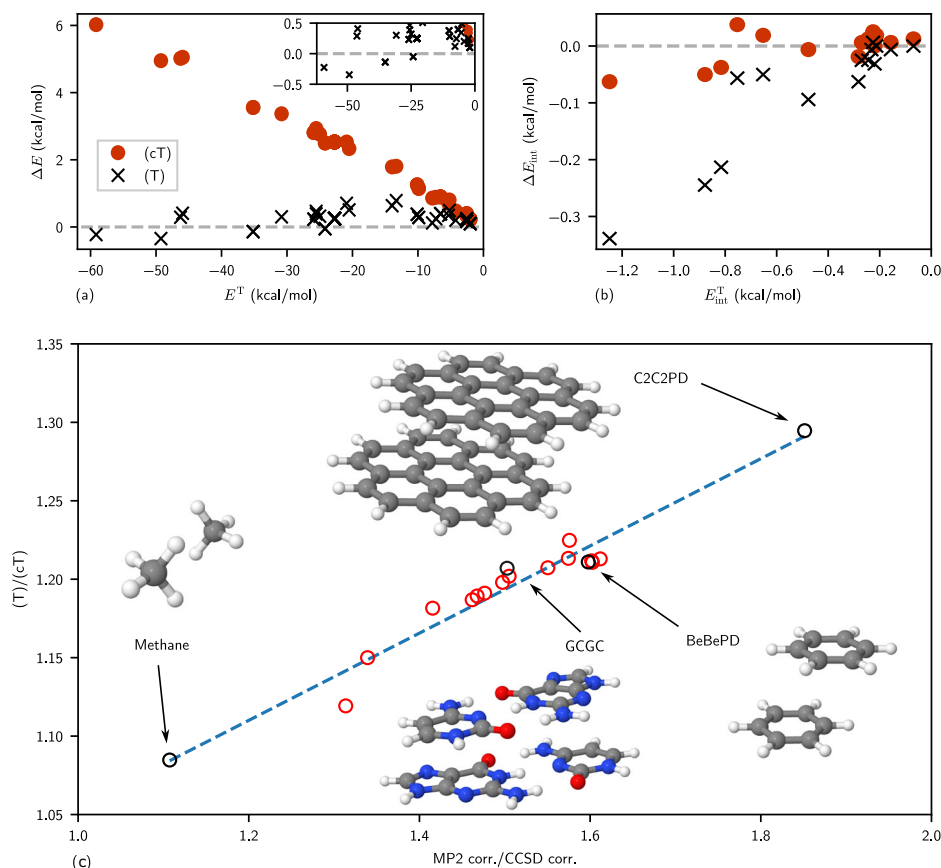


Fig. 2 | Comparison between the full triples and the perturbative triples approaches, (cT) and (T) for a set of molecules contained in the S22 data set⁴⁸. The total triples correlation energy contribution E^T on the x-axis is compared to both differences between the (T), (cT) correlation energy contributions and E^T for (a) total energies (ΔE) and (b) interaction energies (ΔE_{int}). The inset in (a) depicts a zoom of the figure. The abscissa is represented by a gray dashed line. Source data are provided as a Source Data file. c Correlation between the ratio of (T) and (cT)

with the ratio of the MP2 and CCSD correlation energy contributions to interaction energies of a set of dispersion-dominated complexes from the S22, L7 and S66 benchmark datasets^{48–50}. The blue dashed line is a linear fit of the data points. Selected cases are labeled and visualized: Methane dimer, GCGC (guanincytosine tetramer), BeBePD (benzene dimer parallel displaced), C2C2PD (coronene dimer parallel displaced). All the data are available in the supplementary information. Source data are provided as a Source Data file.

studied molecules in this work with dispersion-dominated interactions. This demonstrates that (T) exhibits a tendency to overestimate the absolute binding energy in a similar manner as MP2 for more polarizable systems. Although an overestimation of the (T) binding energy contribution compared to its (cT) counterpart by about 10% might yield a fortuitously better agreement between CCSD(T) and CCSDTQ, we argue that 20–30% overestimation is expected to yield significantly too strong interaction energies. For example, the values of (T)/(cT) for the Benzene-Benzene PD and coronene dimer are 1.2 and 1.3, respectively.

Having demonstrated and explained the reasons for the overestimation of absolute interaction energies on the level of CCSD(T) theory for small molecules with up to 24 atoms and the C2C2PD system, we now want to turn to the discussion of the remaining large molecular complexes where significant absolute discrepancies between DMC and CCSD(T) have been observed. These systems include C3GC from the L7 data set and the C60@[6]CPPA buckyballing. Here, substantial differences in the binding energies of 2.2 kcal/mol and 7.6 kcal/mol were reported, respectively after subtracting error bars. Although CCSD(cT) calculations for systems of that size are currently not feasible using our approach, we now introduce a simplified model that allows us to estimate the change in interaction energies from CCSD(T) to CCSD(cT) in an approximate manner. Given the linear trend between the different correlation energy contributions to the interaction energy depicted in Fig. 2c, it is possible to estimate the (cT) contribution for systems where only MP2, CCSD and (T) are

known. These numbers can be calculated using a computationally efficient LNO-CCSD(T) implementation^{31–36}. Results computed in this manner are denoted as CCSD(cT)-fit. Details about this procedure and the corresponding error estimates are provided in the supplementary information. Furthermore, the supplementary information includes a detailed justification of the CCSD(cT)-fit approach, which is based on extrapolation methods used to estimate the interaction energy between molecules and substrates modeled by planar molecules.

Table 2 gives our estimated CCSD(cT) interaction energies in comparison to CCSD(T) and DMC findings for seven large molecular complexes. A comparison between CCSD(cT)-fit and the explicitly calculated CCSD(cT) results for GGG, GCGC and C2C2PD shows that the linear regression model is sufficiently reliable for the systems studied in this work. Moreover, we have computed basis set converged (cT)-(T) estimates for C3A and PHE, which have been added to the LNO-CCSD(T) estimates to approximate CCSD(cT). These values and the agreement between CCSD(cT) and CCSD(cT)-fit explicitly confirm the CCSD(cT)-fit approach.

For comparison Table 2 also summarizes the DMC interaction energies from refs. 10 and 11, which agree to within at least 1 kcal/mol for GGG, C2C2PD, PHE and C3GC. For the remaining systems the DMC estimates show a larger discrepancy and for C60@[6]CPPA only one DMC estimate is available. Although the DMC binding energies have overlapping error bars, the remaining uncertainties are relatively large, illustrating that obtaining highly accurate interaction energies for these large molecules is also challenging for DMC.

Table 2 | Comparison of the interaction energy for large molecular complexes in kcal/mol as calculated by different levels of theory

System	CCSD(T)	LNO-CCSD(T) ^{††}	CCSD(cT)	CCSD(cT)-fit	DMC ^{††}	DMC ^{†‡}
GGG	-1.5 ± 0.5	-2.1 ± 0.2	-1.2 ± 0.5	-1.8 ± 0.2	-1.5(6)	-2.0(8)
GCGC	-13.1 ± 0.5	-13.6 ± 0.4	-12.5 ± 0.5	-12.8 ± 0.5	-12.4(8)	-10.6(12)
C2C2PD	-21.1 ± 0.5	-20.6 ± 0.6	-19.3 ± 0.5	-18.9 ± 0.7	-18.1(8)	-17.5(14)
C3A		-16.5 ± 0.8	-16.0 ± 0.8*	-15.3 ± 0.9	-15.0(10)	-16.6(18)
PHE		-25.4 ± 0.2	-25.1 ± 0.2*	-25.0 ± 0.2	-26.5(13)	-24.9(12)
C3GC		-28.7 ± 1.0		-26.7 ± 1.1	-24.2(13)	-25.1(18)
C ₆₀ @[6]CPPA		-41.7 ± 1.7		-36.9 ± 2.3	-31.1(14)	

Showcasing partially large discrepancies between CCSD(T) and DMC on the one hand, and an excellent agreement between CCSD(cT) and DMC results for complexes up to the 100-atom scale on the other hand. CCSD(T) and CCSD(cT) results are obtained using our plane wave approach. Values for CCSD(cT)-fit are taken from the LNO-CCSD(T) results of^{††} and corrected by an estimate for $\Delta = (T) - (cT)$ obtained from a fit. CCSD(cT) estimates with a * were estimated by computing Δ separately and adding it to the LNO-CCSD(T) values. Details of the procedures can be found in the supplementary information. Abbreviations: CCSD, coupled cluster singles and doubles theory; (T) and (cT) refer to different perturbative triples methods described in the article; LNO refers to a local approximation referenced in the article; DMC, diffusion quantum Monte Carlo; GGG, Guanine trimer; GCGC, Guanine-Cytosine tetramer; C2C2PD, Coronene dimer parallel displaced; C3A, Circumcoronene adenine; PHE, Phenylalanine residues trimer; C3GC, Circumcoronene Guanine-Cytosine.

As already discussed in ref. 11, CCSD(T) interaction energies listed in Table 2 exhibit large discrepancies compared to DMC for C2C2PD, C3GC and C60@[6]CPPA. In contrast, CCSD(cT)-fit resolves these discrepancies for all systems on the hundred-atom scale, achieving excellent agreement with DMC estimates of Al-Hamdani and Nagy et al.¹¹ to within chemical accuracy (1 kcal/mol) after subtracting the error bars. Even for C60@[6]CPPA, which contains 132 atoms, a discrepancy of only 2.1 kcal/mol remains, although the error bar of CCSD(cT)-fit is relatively large in this case. We argue that the remaining discrepancies are potentially caused by uncertainties in DMC, CCSD(cT)-fit and the underlying LNO-CCSD(T) calculations. It should be noted that the error bars of LNO-CCSD(T) interaction energies are in some cases underestimated, as exemplified for the Uracil-Uracil PD dimer by the comparison between canonical CCSD(T) and LNO-CCSD(T) interaction energies shown in Fig. 1. Furthermore, the DMC interaction energy of C60@[6]CPPA has not yet been verified independently using a different DMC implementation as it was done for all other systems listed in Table 2. We also stress that in some cases the differences between the DMC estimates are larger than their respective error bars.

Discussion

Our work unequivocally demonstrates that, due to the employed truncation of the many-body perturbation series expansion, one of the most widely-used and accurate quantum chemistry approaches—CCSD(T) theory—in certain cases binds noncovalently interacting large molecular complexes too strongly. Our findings show that a simple yet efficient modification denoted as CCSD(cT) remedies these shortcomings, enabling highly reliable benchmark calculations of large molecular complexes on the hundred-atom scale that play a crucial role in scientific and technological problems, for example, drug design and surface science. We stress that the more accurate CCSD(cT) approximation can directly be transferred to computationally efficient low-scaling and local correlation approaches, which will substantially advance the applications of theoretical chemistry as well as physics in all areas of computational materials science where highly accurate benchmark results are urgently needed. We are witnessing an unremitting expansion of the frontiers of accurate electronic structure theories to ever larger systems which when combined with machine-learning techniques, has the potential to transform the paradigm of modern computational materials science.

Methods

Benchmark systems

We considered representative subsets of the L7, S22, and S66 benchmark sets of noncovalent interactions. These structures can be found in the begdb database³⁷. Nine systems from the S66 set were selected

to obtain extrapolated complete basis set (CBS) estimates, while 18 systems from the S22 set were studied at the CCSDT level of theory with modest basis sets. Furthermore, two additional groups of systems, parallel displaced polycyclic aromatic hydrocarbons, and single adenine on top of a polycyclic aromatic hydrocarbon have been studied. The benzene-benzene structures are taken from the S66 set, and the coronene dimer structure is taken from the L7 test database. All other structures not included in the abovementioned data sets, were obtained by geometry optimization using the NWChem³⁸ version 7.0.0 with the TPSS functional and def2-tzvp basis sets. Furthermore, we include the employed structures in the SI as xyz-files.

Gaussian basis set calculations

For the S66 systems, we employed Dunning's augmented correlation-consistent basis sets (aug-cc-pVXZ, X = T, Q, 5). CBS estimates were obtained using standard two-point extrapolation, denoted as [34] (triple- ζ and quadruple- ζ) and [45] (quadruple- ζ and quintuple- ζ). Interaction energies were defined with and without counterpoise (CP) correction. Canonical MP2, CCSD, and CCSD(T) correlation energies were obtained with the MRCC package, interfaced to our Cc4s code. In all MRCC calculations, the program's default resolution-of-identity (RI) auxiliary basis sets were employed for both Hartree-Fock and post-HF correlation steps. We use the MRCC version released on March 18, 2022³¹⁻³³. Based on basis set convergence tests, the aug-cc-pVQZ basis set provides essentially converged Hartree-Fock energies, while the [34] extrapolation of MP2 correlation energies closely approximates the [45] CBS estimate. All of these observations hold for the counterpoise-corrected results. Given that CCSD(T) correlation energies typically converge faster with respect to basis set size than MP2, these findings justify using the [34] approach as a reliable estimate of the CBS limit for the S66 systems shown in Fig. 1.

For the S22 set, Hartree-Fock energies were computed with NWChem³⁸ version 7.0.0 using the cc-pVDZ basis, while post-Hartree-Fock methods (MP2, CCSD, CCSDT) were carried out via Cc4s. In this case, counterpoise corrections were not applied, and the reported interaction energies are based directly on the raw supermolecular results.

For the parallel displaced polycyclic aromatic hydrocarbons, and single adenine on top of a polycyclic aromatic hydrocarbon systems we employed aug-cc-pVTZ basis sets using MRCC. Results are given using the counterpoise correction. For the coupled-cluster calculations we used frozen natural orbitals (FNOs)³⁹ with $X = 8 - 12$ virtual orbitals per occupied orbital. CCSD energies are corrected using the Delta-MP2 approach, $E_{\text{CCSD}} = E_{\text{CCSD}}(X) + E_{\text{MP2}}(\text{aug-cc-pVTZ}) - E_{\text{MP2}}(X)$, and the (T) and (cT) energies are corrected using the (T*) method, $E_{(T^*)} = E_{(T)}(X) \cdot E_{\text{MP2}}(\text{aug-cc-pVTZ})/E_{\text{MP2}}(X)$ ^{40,41}.

Finally, we have calculated CBS estimates for C3A and PHE using Gaussian type orbitals with MRCC. Therefore, we took counterpoise corrected results using aug-cc-pVQZ basis set. It was shown that this choice is already sufficient for very accurate CBS estimates¹¹. In these calculations we employ approximative FNOs: 6 and 10 for C3A and 8 and 10 for PHE, respectively. For the final CCSD energies, we once more use Delta-MP2 corrected values, defined above. For these numbers we use MP2 CBS estimates using canonical MP2 calculations, employing [45] extrapolation. Similarly (T*) correction is used, again using [45] extrapolated values for the MP2 CBS estimate.

Plane-wave workflow

For larger molecular complexes, we employed a plane-wave workflow implemented in a development version of VASP 6.3 and Cc4s. Simulation cells were enlarged systematically to remove spurious interactions between periodic images, and plane-wave cutoffs were converged at 700 eV based on MP2 tests. Canonical Hartree-Fock orbitals were computed and subsequently transformed to approximate MP2 natural orbitals³⁹. Truncated sets of natural orbitals, characterized by the ratio of virtual to occupied states ($X = N_v/N_o$), were used both for efficient MP2 CBS corrections and as input for coupled-cluster calculations. Coulomb integrals were factorized using a singular-value decomposition of the auxiliary basis⁴², allowing for substantial reduction of the auxiliary dimension without loss of accuracy.

Final CCSD, CCSD(T), and CCSD(cT) calculations were performed with Cc4s on up to 50 compute nodes (128 cores each), exploiting the code's massively parallel implementation.

Data availability

Additional data supporting the findings of this study are available within the supplementary information. The supplementary data file contains molecular structures in xyz-format used for calculations presented in S6. Source data are provided with this paper.

Code availability

The Cc4s code⁴³ developed and used for the presented post-HF calculations can be downloaded from (<https://doi.org/10.5281/zenodo.16762158>⁴⁴). A documentation and more information about the code can be found in ref. 43. For the underlying HF calculations we employed different codes that are referenced in the Methods section.

References

- Schrödinger, E. Quantisierung als eigenwertproblem. *Ann. der Phys.* **384**, 361–376 (1926).
- Pauli, W. Über das wasserstoffspektrum vom standpunkt der neuen quantenmechanik. *Z. fur Phys.* **36**, 336–363 (1926).
- Dirac, P. A. M. Quantum mechanics of many-electron systems. *Proc. R. Soc. Lond. Ser. A* **123**, 714–733 (1929).
- Kohn, W. & Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133–A1138 (1965).
- Ceperley, D. M. & Alder, B. J. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.* **45**, 566–569 (1980).
- Jones, R. O. Density functional theory: its origins, rise to prominence, and future. *Rev. Mod. Phys.* **87**, 897–923 (2015).
- Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *J. Chem. Phys.* **139**, 134101(2013).
- Ma, Q. & Werner, H.-J. Accurate Intermolecular interaction energies using explicitly correlated local coupled cluster methods [PNO-LCCSD(T)-F12]. *J. Chem. Theory Comput* **15**, 1044–1052 (2019).
- Szabó, P. B., Csóka, J., Kállay, M. & Nagy, P. R. Linear-scaling local natural orbital CCSD(T) approach for open-shell systems: algorithms, benchmarks, and large-scale applications. *J. Chem. Theory Comput.* **19**, 8166–8188 (2023).
- Benali, A., Shin, H. & Heinonen, O. Quantum Monte Carlo benchmarking of large noncovalent complexes in the L7 benchmark set. *J. Chem. Phys.* **153**, 194113 (2020).
- Al-Hamdani, Y. S. et al. Interactions between large molecules pose a puzzle for reference quantum mechanical methods. *Nat. Commun.* **12**, 3927 (2021).
- Ballesteros, F., Dunivan, S. & Lao, K. U. Coupled cluster benchmarks of large noncovalent complexes: the L7 dataset as well as DNA-ellipticine and buckycatcher-fullerene. *J. Chem. Phys.* **154**, 154104 (2021).
- Villot, C., Ballesteros, F., Wang, D. & Lao, K. U. Coupled cluster benchmarking of large noncovalent complexes in L7 and S12L as well as the C60 dimer, DNA-ellipticine, and HIV-indinavir. *J. Phys. Chem. A* **126**, 4326–4341 (2022).
- Řezáč, J. & Hobza, P. Benchmark calculations of interaction energies in noncovalent complexes and their applications. *Chem. Rev.* **116**, 5038–5071 (2016).
- Dubecký, M., Mitas, L. & Jurečka, P. Noncovalent interactions by quantum Monte Carlo. *Chem. Rev.* **116**, 5188–5215 (2016).
- Al-Hamdani, Y. S. & Tkatchenko, A. Understanding non-covalent interactions in larger molecular complexes from first principles. *J. Chem. Phys.* **150**, 010901 (2019).
- Firaha, D. et al. Predicting crystal form stability under real-world conditions. *Nature* **623**, 324–328 (2023).
- Hoja, J. et al. Reliable and practical computational description of molecular crystal polymorphs. *Sci. Adv.* **5**, eaau3338 (2019).
- Reilly, A. M. et al. Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallogr. Sect. B: Struct. Sci., Cryst. Eng. Mater.* **72**, 439–459 (2016).
- Chanussot, L. et al. Open Catalyst 2020 (OC20) dataset and community challenges. *ACS Catal.* **11**, 6059–6072 (2021).
- Sauer, J. The future of computational catalysis. *J. Catal.* **433**, 115482 (2024).
- Bokdam, M., Lahnsteiner, J., Ramberger, B., Schäfer, T. & Kresse, G. Assessing density functionals using many body theory for hybrid perovskites. *Phys. Rev. Lett.* **119**, 145501 (2017).
- Donchev, A. G. et al. Quantum chemical benchmark databases of gold-standard dimer interaction energies. *Sci. Data* **8**, 1–9 (2021).
- Eastman, P. et al. SPICE, a dataset of drug-like molecules and peptides for training machine learning potentials. *Sci. Data* **2022** *10:1* **10**, 1–11 (2023).
- Liu, P. et al. Combining machine learning and many-body calculations: coverage-dependent adsorption of CO on Rh(111). *Phys. Rev. Lett.* **130**, 078001 (2023).
- Raghavachari, K., Trucks, G. W., Pople, J. A. & Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **157**, 479–483 (1989).
- Nguyen, B. D. et al. Divergence of many-body perturbation theory for noncovalent interactions of large molecules. *J. Chem. Theory Comput.* **16**, 2258–2273 (2020).
- Shepherd, J. J. & Grüneis, A. Many-body quantum chemistry for the electron gas: convergent perturbative theories. *Phys. Rev. Lett.* **110**, 226401 (2013).
- Masios, N., Irmiler, A., Schäfer, T. & Grüneis, A. Averting the infrared catastrophe in the gold standard of quantum chemistry. *Phys. Rev. Lett.* **131**, 186401 (2023).
- Smith, D. G. A., Jankowski, P., Slawik, M., Witek, H. A. & Patkowski, K. Basis set convergence of the post-CCSD(T) contribution to non-covalent interaction energies. *J. Chem. Theory Comput.* **10**, 3140–3150 (2014).

31. Kállay, M. et al. The MRCC program system: accurate quantum chemistry from water to proteins. *J. Chem. Phys.* **152**, 074107 (2020).
32. Mester, D. et al. Overview of developments in the MRCC program system. *J. Phys. Chem. A* **129**, 2086–2107 (2025).
33. Kállay, M. et al. Mrcc, a quantum chemical program suite. www.mrcc.hu (2022).
34. Nagy, P. R. & Kállay, M. Optimization of the linear-scaling local natural orbital CCSD(T) method: redundancy-free triples correction using Laplace transform. *J. Chem. Phys.* **146**, 214106 (2017).
35. Nagy, P. R., Samu, G. & Kállay, M. Optimization of the linear-scaling local natural orbital CCSD(T) method: Improved algorithm and benchmark applications. *J. Chem. Theory Comput.* **14**, 4193 (2018).
36. Nagy, P. R. & Kállay, M. 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 (2019).
37. The Benchmark Energy & Geometry Database (BEGDB). <http://www.begdb.org/> (2024).
38. Valiev, M. et al. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* **181**, 1477–1489 (2010).
39. Grüneis, A. et al. Natural orbitals for wave function based correlated calculations using a plane wave basis set. *J. Chem. Theory Comput.* **7**, 2780–2785 (2011).
40. Knizia, G.; Adler, T. B.; Werner, H. J. Simplified CCSD(T)-F12 methods: theory and benchmarks. *J. Chem. Phys.* **130**, 054104 (2009).
41. Irmmler, A., Gallo, A. & Grüneis, A. Focal-point approach with pair-specific cusp correction for coupled-cluster theory. *J. Chem. Phys.* **154**, 234103 (2021).
42. Hummel, F., Tsatsoulis, T. & Grüneis, A. Low rank factorization of the Coulomb integrals for periodic coupled cluster theory. *J. Chem. Phys.* **146**, 124105 (2017).
43. CC4S developer team CC4S user manual. <https://manuals.cc4s.org/user-manual/>, (2024).
44. Grüneis, A.; Irmmler, A.; Schäfer, T.; Gallo, A. Coupled Cluster for Solids (CC4S). <https://doi.org/10.5281/zenodo.16762159> (2025).
45. Ma, Q. & Werner, H.-J. Explicitly correlated local coupled-cluster methods using pair natural orbitals. *WIREs Computational Mol. Sci.* **8**, e1371 (2018).
46. Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. i. the atoms boron through neon and hydrogen. *J. Chem. Phys.* **90**, 1007–1023 (1989).
47. Kendall, R. A., Dunning, T. H. & Harrison, R. J. Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions. *J. Chem. Phys.* **96**, 6796–6806 (1992).
48. Jurečka, P., Šponer, J., Černý, J. & Hobza, P. 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).
49. Sedlak, R. et al. Accuracy of quantum chemical methods for large noncovalent complexes. *J. Chem. Theory Comput.* **9**, 3364–3374 (2013).
50. Řezáč, J., Riley, K. E. & Hobza, P. S66: a well-balanced database of benchmark interaction energies relevant to biomolecular structures. *J. Chem. Theory Comput.* **7**, 2427–2438 (2011).

Acknowledgements

We gratefully acknowledge fruitful discussions on the post-CCSD(T) contributions to the interaction energies of large molecules with Adrian

Daniel Boese. We also thank Péter Nagy for estimating the CCSD(cT)-fit result of C60@[6]CPPA based on raw data of ref. 11. Tobias Schäfer acknowledges support from the Austrian Science Fund (FWF) [DOI: 10.55776/ESP335]. The computational results have been achieved using the Austrian Scientific Computing (ASC) infrastructure. Andreas Irmmler and Alejandro Gallo acknowledge support from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 951786 (The NOMAD CoE). Support and funding from the European Research Council (ERC) (Grant Agreement No. 101087184) is gratefully acknowledged.

Author contributions

T.S. and A.I. contributed equally to this work. Major numerical investigations were conducted by T.S. and A.I. alike. The work has been conceptualized by A.Gr. with additional contribution from A.I., T.S., and A.Ga. Software development has been carried out by A.I. and A.Ga. The original draft of the paper was written by A. Gr. Additional review and editing of the paper were undertaken by all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-64104-8>.

Correspondence and requests for materials should be addressed to Tobias Schäfer, Andreas Irmmler or Andreas Grüneis.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025