Approaching the basis set limit of CCSD(T) energies for large molecules with local natural orbital coupled-cluster methods

Péter R. Nagy* and Mihály Kállay

Department of Physical Chemistry and Materials Science, Budapest University of Technology and Economics, H-1521 Budapest, P.O.Box 91, Hungary

E-mail: nagyrpeter@mail.bme.hu

Abstract

Recent optimization efforts and extensive benchmark applications are presented illustrating the accuracy and efficiency of the linear-scaling local natural orbital (LNO) coupled-cluster with single-, double,- and perturbative triple excitations [CCSD(T)] method. A composite threshold combination hierarchy (Loose, Normal, Tight, etc.) is introduced, which enables black box convergence tests and is useful to estimate the accuracy of the LNO-CCSD(T) energies with respect to CCSD(T). We also demonstrate that the complete basis set limit (CBS) of LNO-CCSD(T) energies can be reliably approached via basis set extrapolation using large basis sets including diffuse functions. Where reference CCSD(T) results are available, the mean (maximum) absolute errors of LNO-CCSD(T) reaction and intermolecular interaction energies with the default Normal threshold combination are below 0.2-0.3 (0.6-1.0) kcal/mol, while the same measures with the Tight setting are 0.1 (0.2-0.5) kcal/mol for all the tested systems including highly-complicated cases. The performance of LNO-CCSD(T) is also compared

*To whom correspondence should be addressed

with that of other popular local CCSD(T) schemes. The exceptionally low hardware requirements of the present scheme enables the routine calculation of benchmark-quality energy differences within chemical accuracy of CCSD(T)/CBS for systems including a few hundred atoms. LNO-CCSD(T)/CBS calculations can also be performed for more than 1000 atoms with 45000 atomic orbitals using a single, 6-core CPU, about 100 GB memory, and comparable disk space.

1 Introduction

The coupled-cluster (CC) method is one of the most reliable and accurate of our current tools applicable to theoretical simulation of matter at the atomic scale. CC energies and other properties are systematically improvable along the hierarchies of the CC wave function expansions and the single-particle orbital basis sets,[1,2] which can also be exploited to estimate the accuracy of the CC results. Assuming that a single-reference treatment is sufficient, the CC model with single and double excitations (CCSD) augmented with perturbative triples correction [CCSD(T)],[3] is considered the "gold standard" of quantum chemistry since CCSD(T) often provides an accuracy comparable to experiment within chemical accuracy. Unfortunately, the steep, sixth- and seventh-power scaling operation count and the fourth-power scaling data storage requirement are highly limiting even for the most advanced conventional CCSD and CCSD(T) implementations.[4–16]

Due to the substantial progress in the field of reduced-cost approximate CC methods local correlation based approaches are becoming competitive alternatives for accurate computations of molecules also well above the 20-30 atom limit of conventional CC codes. The vast knowledge accumulated in the field over the decades is greatly summarized in recent reviews,[17–19] and a somewhat more theory oriented introduction is also available in our related reports.[20–27] The common starting point of almost all the local methods is to exploit the relatively rapid decay of electron correlation with the distance via switching to a localized molecular orbital (LMO) basis. Many other commonly applied ideas date back to the pio-

neering work of Pulay and Saebø.[28–31] Pair approximations utilize that the pair correlation contribution of distant LMO pairs can be cost-efficiently approximated or even neglected. It is also beneficial to obtain such pair energies using a restricted, spatially close list of correlating orbitals (domain approximation). Recent local methods also take advantage of the sparsity of CC wave functions not only in real space but also in their orbital expansion by employing some sort of natural orbitals (NOs) constructed at the level of second-order Møller–Plesset (MP2) perturbation theory. One group of methods employ LMO pair specific pair natural orbitals (PNOs) popularized recently by Neese, Valeev, Riplinger, Guo, Pinski, and co-workers,[32–37] and taken over also by Werner, Ma, and Schwilk[38–42] as well as by Hättig and Tew.[43,44] Our methods make use of an alternative, LMO-specific NO set, the local natural orbitals (LNOs)[20–22,24,27] for the compression of both the occupied and virtual space of the domains. The virtual LNOs can also be interpreted as weighted average of PNOs corresponding to a specific LMO and all of its strongly interacting LMO pairs.[27]

The most challenging part of local CCSD(T) computations is the handling of the enormous operation count and data storage requirement emerging already for medium-sized systems, in which aspect current solutions deviate the most. The latest PNO-based approaches determine all the CCSD(T) amplitudes in a compressed LMO/PNO representation for the entire molecule at once but have to bear the appearance of an extremely large amount of intermediates originating from the redundancy of the PNOs among different LMO pairs.[37,41,44] An alternative group of local methods chooses to decouple the computation of the amplitudes for different LMOs via fragmentation approximations[20,21,45–56] allowing for an asymptotically constant data storage requirement. A variety of such approaches have been developed up to the CC level, including the cluster-in-molecule (CIM) method of Li, Li, Piecuch, Guo, Gordon, and their co-workers,[55–59] the divide-expand-consolidate scheme of Jørgensen et al.,[54] the divide-and-conquer method of Li and Li[60] and Kobayashi and Nakai.,[49,50] the many-body expansion scheme of Herbert et al.,[52] and the incremental schemes of Friedrich, Dolg, and their co-workers.[48,61] The source of high computational demands in fragmentation-

based methods stems from the overlap of the fragments in real space and the consequent high prefactor for the operation count of the independent CC calculations performed in the fragments.

Our recent developments attempt to find a balanced combination of the above approaches.[20–24,27] We introduced efficient, Laplace-transform based local MP2 (LMP2) and (T) algorithms which are completely free from redundant amplitude evaluations.[23,24] We chose to employ LNOs and obtain the CCSD amplitudes for each LMO separately in order to avoid the large intermediates appearing with the highly-redundant PNO basis. In turn, extensive, overlapping domains emerge in our scheme for both the integral evaluation part in the molecular orbital (MO) basis of LMP2 and in the CCSD amplitude determination. The negative effects of these drawbacks are mitigated by our extensively optimized, integral-direct, in-core, and parallel integral transformation and CCSD(T) algorithms, which optimally utilize the available computer power.[23,27] The resulting LNO-CCSD(T) algorithm is fully ab initio, i.e., free from real-space cutoffs, manual fragment definitions via input atom lists, bond cutting, scaling parameters, etc., the domain construction is completely automatic and adopts to the complexity of the wave function of the systems. The LNO-CC algorithm was also extended to general order CC schemes, such as CCSDT(Q)[62] [CCSD with iterative triple and perturbative quadruple excitations].[20,27]

The current LNO-CCSD(T) implementation runs effectively on both moderate and high-performance processors, it is OpenMP parallelized, benefits from point group symmetry (even non-Abelian), restartable in the case of, e.g., power failure, especially memory economic, and requires negligible hard disk space.[27] For extremely large systems, LNO-CC calculations can be embedded into the electrostatic potential of effective point charges via the well established quantum mechanics/molecular mechanics (QM/MM) technique,[63] and/or a quantum mechanical embedding is also feasible where density functional theory (DFT) is employed for the environment.[25,26] On top of those layers, the correlation energy calculation can also be accelerated by combining higher-level methods for the most important

regions [e.g., LNO-CCSD(T) or even LNO-CCSDT(Q)] with more cost-efficient choices for the environment layer [e.g., LNO-CC with less accurate local approximations or LMP2].[25,26] The previous, much less optimized version of LNO-CCSD(T) has already been successfully applied to chemical problems by us[64,65] and also by other researchers.[66–68] For instance, in a study aiming at accurate formation enthalpy, Paulechka and Kazakov benchmarked three reduced-cost CCSD(T) implementations against canonical CCSD(T) references and found even the previous version of our LNO-CCSD(T) approach to be the most accurate and efficient.[68]

Besides the accuracy of the local approximations the relatively slow convergence of CCSD(T) energies with the atomic orbital (AO) basis set size also has to be addressed in realistic applications. Extrapolation towards the complete basis set (CBS) limit using systematically improving AO basis set hierarchies was proven reliable to minimize the basis set incompleteness error (BSIE)[69] if the employed basis sets are sufficiently large (at least triple- and quadruple-ζ quality). Alternatively, the BSIE of CCSD can also be decreased using explicitly correlated or F12 methods,[70] which solution has also become available for some of the recent local CC implementations.[36,40,44,61] Besides the higher complexity of F12 methods from the perspective of the evaluation of local CC intermediates[36,40,44] or gradients,[71] the explicitly correlated treatment of the (T) contribution is also a not yet completely solved issue.[40,44,72] Nevertheless, the agreement of CBS extrapolated and explicitly correlated CC results, as also found, e.g., in Sects. 5 and 6 of the present report, verifies the applicability of both approaches.

The paper is organized as follows. Sect. 2 provides a summary of the LNO-CCSD(T) method and its implementation, including the most recent algorithmic developments, focusing on the aspects relevant from the perspective of applications. The behavior of the local approximations with the individual truncations and the presently introduced composite threshold sets is analyzed in Sect. 3. The accuracy of LNO-CCSD(T) is extensively benchmarked against canonical CCSD(T) reference energies (Sect. 5) and comparisons are

also provided with some of the most popular and efficient, alternative local CCSD(T) implementations.[35–37,40] The systematic convergence of LNO-CCSD(T) energies with respect to both the local approximations and the AO basis set completeness is in the focus of Sect. 6. Sect. 7 provides a practical perspective collecting wall-time and other hardware requirement measurements on the examples of a realistic application for an organocatalysis reaction involving up to 90 atoms and for an entire protein system containing 1023 atoms and almost 45000 AOs.

2 Theoretical background

This section briefly introduces the theoretical background of our local correlation schemes focusing on the aspects relevant from the perspective of practical applications. For more details regarding the theory, derivations, algorithms, etc., we refer to previous publications.[22–24,27]

2.1 The LNO-CCSD(T) approach

The correlation energy expressions considered in the context of our local methods are inspired by the formulation of the CIM approach.[56–59] In general, E^{corr} [where corr = MP2, CCSD, (T), CCSD(T), CCSDT, ...] are expressed as sums over occupied (and virtual) orbital indices, hence the invariance of the sums to unitary rotations among the occupied (and virtual) orbitals can be exploited. In the following the conventional canonical occupied orbital indices $(i, j, k, \dots)$ will be replaced by localized MOs $(i', j', k', \dots)$ in order to facilitate the introduction of local approximations. For instance, the canonical CCSD(T) energy can be exactly rearranged as

$$E^{\mathrm{CCSD(T)}} = \sum_i \delta E_i^{\mathrm{CCSD(T)}} = \sum_{i'} \delta E_{i'}^{\mathrm{CCSD(T)}} \,, \tag{1}$$

where $\delta E_i^{\mathrm{CCSD(T)}}$ and $\delta E_{i'}^{\mathrm{CCSD(T)}}$ are the correlation energy contributions of the i^{th} canonical and localized MO, respectively.

The CCSD and (T) terms of $E^{\mathrm{CCSD(T)}}$, E^{CCSD} and $E^{(\mathrm{T})}$, respectively, are also written as sums of contributions of individual occupied orbitals. Using CCSD as an example, the explicit form of the correlation energy contribution reads as

$$E^{\mathrm{CCSD}} = \sum_{i'} \delta E_{i'}^{\mathrm{CCSD}} = \sum_{i'} \left[2 \sum_{a} f_{ai'} t_{i'}^{a} + \sum_{abj} L_{i'j}^{ab} (t_{i'j}^{ab} + t_{i'}^{a} t_{j}^{b}) \right], \tag{2}$$

where f_{ai} is an element of the Fock-matrix, t_i^a and t_{ij}^{ab} are cluster amplitudes for single and double excitations, respectively, and $a, b, \ldots$ are virtual orbital indices. The quantity

$$L_{i'j}^{ab} = 2(ai'|bj) - (aj|bi') \tag{3}$$

is defined using $(ai|bj)$, which is an electron repulsion integral (ERI) in the Mulliken notation. Note that the $\delta E_{i'}^{\mathrm{corr}}$ correlation energy contributions of individual LMOs can be analogously expressed for MP2,[22,23] (T),[21,24] and higher-order CC methods as well.[20,27]

To reduce the scaling of the number of operations in the E^{corr} expressions to asymptotically linear the operation count for each $\delta E_{i'}^{\mathrm{corr}}$ has to be asymptotically constant. Here we assume that contribution $\delta E_{i'}^{\mathrm{corr}}$ is computed for each LMO correlated in the calculation. However, let us add that the scaling of the correlation energy evaluation can be further improved to sublinear, e.g., by employing embedding and multi-level correlation techniques,[25,26] where we can also limit the number of occupied orbitals for which $\delta E_{i'}^{\mathrm{corr}}$ is evaluated at a higher, LNO-CC level.

2.2 Local approximations and their implementation

In order to achieve asymptotically constant scaling for the individual $\delta E_{i'}^{\mathrm{corr}}$ terms the summations over all occupied and virtual indices have to be restricted (in an i' orbital specific

manner) to an asymptotically constant number of terms. For instance, the summations over indices j, a, and b in Eq. (2) are performed for less than the complete number of occupied and virtual orbitals of the entire system. This is achieved in our LNO-CC method, and in many other local correlation schemes,[28,32,39] by introducing pair and domain approximations. Our approximations are discussed in detail in Refs. 23, 24, and 27, while only a compact summary is given in this section.

The occupied orbitals contributing to $\delta E_{i'}^{\mathrm{LNO-CCSD(T)}}$, the LNO-CCSD(T) analogue of $\delta E_{i'}^{\mathrm{CCSD(T)}}$, at the CCSD(T) level are selected on the basis of pair correlation energy estimates, $\delta E_{i'j'}$, obtained for all $i'-j'$ LMO pairs. For that purpose sufficiently large, $i'-j'$ specific pair domains ($\mathcal{P}_{i'j'}$) are formed consisting of orbitals i', j', correlating orbitals in the form of projected atomic orbitals (PAOs) surrounding i' and j', and AOs necessary to span the MOs of $\mathcal{P}_{i'j'}$ up to a certain threshold (99.9% by default).[23,27] In those pair domains MP2 pair correlation energy estimates [$\delta E_{i'j'}(\mathcal{P}_{i'j'})$] are evaluated very efficiently and accurately using multipole expansions including contributions up to octupole moments.[23,27] After that only those strongly correlated pairs, or, in brief, strong pairs are kept for the following CC calculations which have a pair correlation energy higher than a threshold (ε_{w}), i.e., $\delta E_{i'j'}(\mathcal{P}_{i'j'}) \geq \varepsilon_{\mathrm{w}}$. The $\delta E_{i'j'}(\mathcal{P}_{i'j'})$ correlation energy contributions of the remaining pairs, the so-called distant pairs [with $\delta E_{i'j'}(\mathcal{P}_{i'j'}) < \varepsilon_{\mathrm{w}}$] are added to the total correlation energy [see Eq. (5) below].

The length of the strong pair list of each LMO is independent of the size of the molecule (for sufficiently large, non-metallic systems) due to the fast decay of the pair correlation energies. To restrict the summations for the virtual indices in the $\delta E_{i'}^{\mathrm{corr}}$-type expressions the domain approximation is employed. To that end a so-called extended domain (ED, $\mathcal{E}_{i'}$) is formed around each LMO i', and the LMO i' is called the central MO of its ED.[23,27] The MO space of each ED consists of the central LMO, its strong pair LMOs, and PAOs centered at the atoms of a domain (called PAO center domain, PCD) surrounding the centers of the LMOs of the ED. The PCD is defined so that, even with the shortened PAO list, the

ED's MP2 correlation energy contribution $[\delta E_{i'}^{\mathrm{MP2}}(\mathcal{E}_{i'})]$ will be accurate allowing only a few hundredths of a percent error. We have also showed that the necessary LMO-PAO three-center ERIs of the ED can be accurately fitted using auxiliary functions (AF) centered on the atoms of the PCD.[23,27] Finally, the AO list of the ED is composed as the union of LMO specific atom lists, governed by the T_{EDo} threshold. The default $T_{\mathrm{EDo}} = 0.9999$ setting ensures that the projection of each LMO onto the ED's AO list is accurate up to 99.99%.

We observed for large, 3D systems, such as proteins, that the average number of atoms in the EDs saturates around 100-150 even if there are an order of magnitude more atoms in the entire molecule.[23,27] Naturally, the radius of a closely packed 3D system of about 100-150 atoms is in the range of 5-10 Ångströms, and the correlation of the electron pairs closer than that distance has to be treated at a higher level. Nevertheless, several further efficiency improving techniques have to be combined in order to obtain $\delta E_{i'}^{\mathrm{MP2}}(\mathcal{E}_{i'})$ for all LMOs in a reasonable time and with affordable memory requirement in the range of tens of a gigabyte (GB). Relying on those results in highly efficient calculations even if some of the EDs contain all atoms of the entire molecule for 3D systems of below about 100 atoms. In brief, three-center ERIs are evaluated in an integral direct manner in a pseudo-canonical LMO-PAO basis, and only the absolutely necessary, non-redundant set of MP2 amplitudes is evaluated using a Laplace-transform/Cholesky-decomposition based approach.[22,23,27] The same MP2 amplitude list is also useful to construct MP2 natural orbitals, the LNOs of the ED. For the rest of the domain calculation occupied and virtual LNOs having occupation numbers smaller than ε_{o} and ε_{v} are kept frozen and not correlated at the CC level. The full LMO and PAO bases of the ED are then compressed by transforming them to the truncated LNO list, yielding the local interacting subspace (LIS, $P_{i'}$) of the central LMO. Since the LNO list is highly-compressed compared to the LMO/PAO basis of the ED, the accurate density fitting (DF) of the ERIs in the LNO basis requires much fewer AFs than those that are used for the local DF in the EDs, that is, the AFs located on the atoms of the PCD. For that purpose the so-called natural auxiliary functions (NAFs)[73] are employed, which are, in a system specific

manner, optimal for the fitting of the ERIs in the LNO basis.[22,23,27] The NAF selection step, in analogy with the LNO selection, is controlled by the $\varepsilon_{\rm NAF}$ threshold. On top of the above, LNO- and NAF-based cost-saving approaches an operation count efficient scheme has also been introduced recently for the construction of the still numerous two-external three-center integrals of the compressed LNO and NAF basis.[27]

The final correlation energy contribution of each LMO includes the CCSD(T) and MP2 correlation energies computed in the LNO basis of the LIS, $\delta E_{i'}^{\rm CCSD(T)}(P_{i'})$ and $\delta E_{i'}^{\rm MP2}(P_{i'})$, respectively. The LIS MP2 and CCSD correlation energy is evaluated in the complete LNO basis via conventional, DF-based formulations, however, the same approach would be demanding for the (T) part. For the latter, only a non-redundant list of triples amplitudes is computed using our highly-optimized Laplace (T) code,[24] which brings down the scaling of the (T) correlation energy calculation from seventh to sixth power with the size of the LIS. The accuracy of the Laplace-quadrature is governed by $T_{\rm LT}$ as discussed in Ref. 24.

Once all the independent LIS CCSD(T) calculations are completed the total LNO-CCSD(T) correlation energy is evaluated as

$$E^{\rm LNO-CCSD(T)} = \sum_{i'} \left[\delta E_{i'}^{\rm CCSD(T)}(P_{i'}) + \Delta E_{i'}^{\rm MP2} \right], \tag{4}$$

where $\Delta E_{i'}^{\rm MP2}$ corrects for the above pair, LNO, NAF, and other approximations at the MP2 level:

$$\Delta E_{i'}^{\rm MP2} = \delta E_{i'}^{\rm MP2}(\mathcal{E}_{i'}) - \delta E_{i'}^{\rm MP2}(P_{i'}) + \frac{1}{2} \sum_{j'}^{\rm distant} \delta E_{i'j'}(\mathcal{P}_{i'j'}) \ . \tag{5}$$

We note that an accurate local MP2 correlation energy,

$$E^{\rm LMP2} = \sum_{i'} \left[\delta E_{i'}^{\rm MP2}(\mathcal{E}_{i'}) + \frac{1}{2} \sum_{j'}^{\rm distant} \delta E_{i'j'}(\mathcal{P}_{i'j'}) \right] , \tag{6}$$

also emerges from the above procedure, which can be useful, for instance, to obtain low-cost basis set incompleteness corrections to LNO-CC energies at the LMP2 level.[74]

2.3 Recent algorithmic improvements

On top of the algorithmic optimization summarized above and presented previously [23,24,27] a significant amount of additional advances has been made since our latest report, [27] which are already available in the most recent release of the MRCC program suite (release date February 9, 2019). [75] All of these developments are aimed to make the evaluation of the above LNO-CC energies more efficient with less computational resources. In other words, the LNO-CC Ansatz and the correlation energies remained the same as discussed in Ref. 27. The thorough description of these improvements would require an in depth introduction of many aspects of our approach and algorithms and would be highly technical. For that reason this is left to a theory oriented report, and here we only compile the features most relevant for the application of the new algorithm in practice.

From the perspective of approaching the CBS limit of LNO-CC correlation energies for large systems the most important task was to handle numerical challenges emerging from the near linear dependency of the AO basis set. The redundancy in triple- and quadruple-ζ (or even larger) basis sets can be severe for extended molecules, especially if diffuse functions are also present. Unfortunately, the most common approach of near linear dependency removal, the Löwdin canonical orthogonalization would require the mixing of the AO basis functions with each other, which negatively effects the locality of the LMOs expanded in the atom-centered AO basis. Our aim was to retain the full extent of the locality of the LMOs, hence our solution was to work with the complete, possibly linearly dependent AO basis and adjust those steps of the calculation that can be prone to numerical instabilities. The resulting algorithm is proven to be quite stable even with extremely large, diffuse basis sets. So far we were able to perform LNO-CCSD(T) calculations with diffuse, quadruple-ζ basis sets for systems of up to 1000 atoms and with diffuse, quintuple-ζ basis sets with up to 100 atoms. [74]

The wall time required for LNO-CCSD(T) in OpenMP-parallel calculations was decreased by enhancing the parallel efficiency of our DF-CCSD implementation. To that end, now

a fully integral-direct, in-core DF-CCSD(T) implementation[16] is employed for the CCSD and Laplace-transform (T) calculations of each LIS. Hence, four-center LNO integrals are not stored on disk any more, all the necessary integrals and intermediates are stored in memory or recomputed when required. This solution eliminates many steps, such as the input/output (I/O) of four-center integral lists, which cannot scale properly with the number of available cores, and consequently it improves the parallel efficiency. Compared to our previous CCSD implementation[21] the scaling with the number of cores and also the wall times have been improved by about a factor of 2 measured on a 6-core CPU. This new DF-CCSD(T) implementation[16] is superior over the previous one not only in wall times and OpenMP scaling, but its minimal memory requirement is almost by a factor of 2 smaller, and it uses negligible disk space. The latter is particularly useful to avoid extensive data traffic if only network file systems are available as in many current computer clusters.

In order to also minimize the potentially slow disk use in the integral transformation steps the order of operations has been rearranged. Previously, the three-center ERIs in the LNO/NAF basis were evaluated first for each LIS, these compressed integral lists were stored on disk, and finally the CCSD(T) calculations were performed in each LIS in a separate loop for all LISs. In contrast, the current algorithm computes a set of LNO/NAF three-center ERIs for a domain at a time, keeps it in memory, and immediately performs the CCSD(T) calculation for that LIS. In this way neither any three-center nor the assembled four-center LNO integrals are stored on disk, but they are kept in memory only as long as they are needed and discarded as soon as the LIS calculation is completed.

Let us note that the in-core storage of the LNO/NAF integrals does not increase the memory requirement of the integral transformation part since it is usually dominated by the three-center LMO/PAO ERI lists of the ED. As it was illustrated in Table 11 of Ref. 27, the memory use of our previous approach was already relatively small being in the range of a few tens of GBs for average cases. In order to facilitate LNO-CCSD(T) calculations with even smaller minimal memory requirement or for systems with very large, diffuse basis sets,

we have also implemented alternative algorithms for the memory intensive operations in the ED. If needed the minimal memory requirement can now be reduced significantly (depending on the system by up to about a factor of 4) in exchange for slightly increased operation count or for storing some of the large intermediates on disk.

3 Accuracy of local approximations

We have performed thorough investigations on the effect of all individual approximations on the accuracy of both correlation energies and energy differences in the context of our LMP2,[23] LNO-dRPA,[22] and LNO-CC[20,21,24,27] methods. Additionally, default combinations of these thresholds were identified which reproduce the results obtained with the canonical methods with high accuracy. For instance, LNO-CCSD(T) with its default settings recovered conventional CCSD(T) on the average within chemical accuracy in all the test cases studied so far, where exact CCSD(T) results were available. Here we introduce a few additional, pre-defined threshold combinations so that convergence tests utilizing the systematically improvable nature of our local approximations can be performed in a simple, black box manner. For that purpose, first, we review the conclusions of the convergence studies performed by changing one parameter at a time and devise parameter sets yielding balanced accuracy and computational performance.

3.1 Threshold dependence of individual approximations

The thresholds determining the quality of the pair-energy estimates, namely the order of the multipole expansion, the size of the pair-domains, etc. were analyzed by comparisons to exact MP2 pair energies.[23] The largest system in these tests contains 103 atoms and it is about 28 Å long ensuring that a wide range of pair-energy values are assessed. We have found that the cumulated errors of the above approximations lead to an error of lower than 0.02% in the total correlation energy if the pair-energies of distant pairs are added up to

the default threshold of $\varepsilon_{\rm w} = 10^{-5}$ $\rm E_h$, while the same error drops below 0.01% if a tighter parameter, $\varepsilon_{\rm w} = 3 \times 10^{-6}$ $\rm E_h$ is employed.[23] Varying the strong/distant pair threshold, $\varepsilon_{\rm w}$ effects the accuracy of the MP2 energy contributions of the EDs $[\delta E_{i'}^{\rm MP2}(\mathcal{E}_{i'})]$ in a larger extent. Benchmarks performed on molecules containing more than 100 atoms and triple-ζ basis sets revealed that even for such large systems correlation energy errors in the range of only 0.03% to 0.08% can be attributed to the pair approximation with the default $\varepsilon_{\rm w} = 10^{-5}$ $\rm E_h$ value, while a factor of 2-3 improvement comes with the tighter $\varepsilon_{\rm w} = 3 \times 10^{-6}$ $\rm E_h$ setting.[23,27] The good accuracy provided by the $\varepsilon_{\rm w} = 10^{-5}$ $\rm E_h$ choice is in accord with the findings of independent studies. For instance, the above default value matches the default close and weak pair separating threshold in the local PNO family of methods of Werner and co-workers[39] and is also equal to the strong and weak pair separating threshold employed in the TightPNO variant of the domain-based local PNO (DLPNO) methods of Neese and co-workers.[76] The final threshold needed for the definition of the EDs, $T_{\rm EDo}$, which controls the size of the atom and AO lists, has a smaller impact on the energetics. Correlation energy errors of about 0.01% to 0.02% are caused when using the default $T_{\rm EDo} = 0.9999$ value and the results are more than twice as good with the tighter $T_{\rm EDo} = 0.99995$ setting.[23,27] The combined effect of the above pair and domain errors is systematic, the absolute correlation energy errors in the MP2 energy of the ED are positive (higher negative number due to the exclusion of MOs from the ED energy expressions) and found to be below 0.1% for all systems of a test set containing large molecules of 90-260 atoms and a wide range of basis sets.[23,27] The overall contributions of the pair and domain truncations to the error of LNO-CCSD(T) energy differences are also sufficiently small, for the tested systems only 0.2-0.3 kcal/mol ($< 1\%$) absolute (relative) deviations were found even for quite challenging, large systems of up to 100 atoms and triple-ζ bases extended with diffuse functions.[27]

The remaining thresholds determine the quality of the correlation energies obtained in the LIS. The cost of tightening the thresholds for the NAF truncation and the Laplace-transform as far as achieving negligible errors turned out to be affordable when using the default or

tighter LNO and pair truncation parameters.[24,27] The additional benefit of choosing such tight $\varepsilon_{\mathrm{NAF}}$ and T_{LT} as default is that exploratory convergence studies can be started by focusing on the effect of the dominant pair and LNO truncations. From the same perspective, it was beneficial to tie together the occupation number thresholds for the occupied and virtual LNOs since their behavior was found to correlate irrespective of the type of the investigated molecules.[21,27] We have also invested sizable resources to test the accuracy of the combined LNO threshold on large systems and we have found rapid convergence for both correlation and reaction energies. Considering average to relatively challenging cases of medium-sized systems, compared to canonical references we measured relative errors of about 0.01-0.06% with triple-ζ basis sets and somewhat larger, up to 0.07-0.09% errors with quadruple-ζ bases. These correlation energy errors decrease below 0.02-0.03% for both basis set types when the tighter, $\varepsilon_{\mathrm{v}} = 3\times10^{-7}$ value is set.[27] The correlation energy errors caused by discarding LNOs were found to be almost as systematic as the pair and domain errors, but of the opposite sign, hence LNO-CCSD(T) may also benefit from the partial compensation of those two major error sources. The contribution of the LNO truncation to reaction energy errors is usually acceptable but can get close to 1 kcal/mol for complicated systems with large reaction energies, such as the ISOL4 reaction (Fig. 1). Nevertheless, we have found for the tested systems that the tighter ε_{v} threshold of 3×10^{-7} can provide reaction energies which are accurate up to a few tenth of a kcal/mol, and the relative errors in energy differences were also excellent, below 1% already with the default choice.[27]

3.2 Composite truncation threshold combinations

Building on the numerical experience collected on the behavior of our local approximations we determined a default threshold combination, hereafter referred to as the Normal threshold set, leading to affordable and accurate LNO-CCSD(T) results.[27] The performance of the Normal set can be characterized by its satisfactory average correlation and reaction energies of 0.07% and 0.34 kcal/mol, respectively, which were measured using a test set containing

26 CCSD(T) correlation energies of medium-sized systems (CEMS26).[27] The resulting set collects molecules of 30-63 atoms and employs realistic, at least triple-ζ quality basis sets, hence it is suitable to predict the accuracy of local CCSD(T) methods for average applications. However, not all the local approximations operate at full extent even on 3D molecules of up to 63 atoms, therefore here we present additional benchmarks on larger molecules containing 58–92 atoms. The examples of the present report are also suitable to characterize the convergence of energies and energy differences towards the CBS limit and the limitations and maximum expected errors of our method by analyzing its behavior on some of the most complicated cases presented in the literature so far, the ISOL4 and AuAmin reactions and the dimerization of coronene (Fig. 1).

Such tests are performed by exploiting that both the local approximations and the quality of the AO basis set are systematically improvable. Since there are numerous different local approximations in our method, it is beneficial to package them together and perform convergence tests by varying only a single, composite parameter. For that purpose, besides the Normal set, Loose and Tight threshold sets were defined by balancing the performance and accuracy of the various forms of truncations (see Table 1). Mainly for the purpose of performing convergence tests when the exact reference is unavailable an even tighter combination is also introduced, which will be dubbed very Tight or vTight (see the corresponding entry of Table 1). The improvement of the main truncation thresholds in Table 1 are roughly exponential, i.e., there is about a factor of three improvement in the ε_{o}, ε_{v}, and ε_{w} thresholds when going one step further. Thus large steps are made towards very small thresholds ensuring the rapid and reliable approach of conventional CCSD(T).

4 Computational details and test systems

The above LNO-CCSD(T) and related approaches have been implemented in the MRCC suite of quantum chemical programs, and they are available in the latest release of the program

Table 1: Threshold combinations defined in this study.[a]

Symbol	Keyword[b]	Loose	Normal	Tight	vTight
ε_{o}	`lnoepso`	3×10^{-5}	10^{-5}	3×10^{-6}	10^{-6}
ε_{v}	`lnoepsv`	3×10^{-6}	10^{-6}	3×10^{-7}	10^{-7}
ε_{w} [$\mathrm{E_h}$]	`wpairtol`	3×10^{-5}	10^{-5}	3×10^{-6}	10^{-6}
T_{EDo}	`bpedo`	0.9999	0.9999	0.99995	0.99999
$\varepsilon_{\mathrm{NAF}}$ [$\mathrm{E_h}$]	`naf_cor`	10^{-2}	10^{-2}	10^{-2}	10^{-3}
T_{LT}	`laptol`	10^{-1}	10^{-2}	10^{-3}	10^{-3}

[a] The various threshold combinations can be set by the `lcorthr` keyword of the MRCC suite if keyword `localcc` is set to `2018`. [b] Name of the corresponding MRCC keywords.

suit.[75] For comparison PNO-CCSD(T)-F12 results were collected from the literature.[40,42] DLPNO-CCSD($\mathrm{T_0}$) and DLPNO-CCSD(T) results computed with and without relying on the semi-canonical triples approximation ($\mathrm{T_0}$)[35,37] were obtained using the 4.0 and 4.1 versions of the ORCA package.[77]

In the calculations presented here Dunning's (augmented) correlation-consistent polarized valence X-tuple-ζ basis sets [(aug-)cc-pVXZ, X = T, Q, and 5],[78–80] and the triple- and quadruple-ζ valence basis sets including polarization functions (def2-TZVP and def2-QZVP) optimized by Weigend and Ahlrichs[81] were used. Dunning's (aug-)cc-pVXZ bases were applied to the atoms of the reactants and products of the AuAmin reaction (see below in Sect. 4.1), except for the phosphorus atom, for which the revised (aug-)cc-pV(X+d)Z set was used,[82] and for the gold atom, for which the augmented correlation consistent valence triple-ζ pseudopotential (aug-cc-pVTZ-PP) basis set of Peterson[83] was employed together with the corresponding effective core potential (ECP) for the inner 60 electrons of the atom.[84] For the sake of brevity, the basis set combination used for the AuAmin reaction will simply be labeled by (aug-)cc-pVXZ throughout the text. If not indicated otherwise, for the above AO basis sets the corresponding auxiliary bases of Weigend and co-workers were applied.[85–87] To simplify the notation wherever clarity allows the (a)XZ shorthand is introduced referring the to corresponding Dunning-type basis set of (aug-)cc-pVXZ.

For most of our calculations the DF approximation was also invoked at the HF calculations without further approximations (DF-HF), while for the protein calculation (Fig. 3) the

exchange contribution in the DF-HF calculations was computed with local fitting domains as described in Ref. 23. The core electrons were kept frozen in all the correlation calculations. In accordance with Ref. 40, 8 electrons were not correlated (on top of the 60 electrons described via ECP) on the gold atom.

The CBS limit of the HF and correlation energies is approached via two-point extrapolation techniques. The compact notation of (a)(X,Y)Z will be employed to denote extrapolation using the (a)XZ and (a)YZ bases. For instance, extrapolated results based on the aug-cc-pVTZ and aug-cc-pVQZ sets will be labeled by a(T,Q)Z. For the extrapolation of HF energies the two-point formula suggested by Karton and Martin[88] is used:

$$E^{\mathrm{HF}}_{X(X-1)} = E^{\mathrm{HF}}_{X} + \frac{(E^{\mathrm{HF}}_{X} - E^{\mathrm{HF}}_{X-1})(X+1)}{X \exp\left(\gamma(\sqrt{X} - \sqrt{X-1})\right)}\,, \tag{7}$$

where E^{HF}_{X} is the HF energy obtained with the (a)XZ basis set and γ is 6.57 for $X = 4$.[88] Correlation energies are extrapolated using the formula introduced by Helgaker *et al.*[69] as

$$E^{\mathrm{corr}}_{X(X+1)} = \frac{X^3 E^{\mathrm{corr}}_{X} - (X+1)^3 E^{\mathrm{corr}}_{X+1}}{X^3 - (X+1)^3}\,, \tag{8}$$

with E^{corr}_{X} being the correlation energy with the (a)XZ basis set.

The reported computation times are wall-clock times determined on a machine with 128 GB of main memory, a standard 7200 rpm hard disk, and a single 6-core 3.5 GHz Intel Xeon E5-1650 v2 processor with a theoretical peak performance of 168 billion floating point operations per second.

4.1 Benchmark sets and test systems

The accuracy of the LNO-CCSD(T) correlation, reaction, and non-covalent interaction energies was assessed on tests sets composed of small to medium sized systems of up to 63 atoms. The common characteristic of these tests is that the conventional CCSD(T) results

were available as reference. First, the test set of Neese, Wennmohs, and Hansen (NWH)[89] was studied, which is assembled from 23 reactions including molecules of up to 36 atoms. In this case the relatively small system size allowed us to obtain the CCSD(T) reference with both triple- and quadruple-ζ quality basis sets so the accuracy of the basis set extrapolated results are also analyzed. Furthermore, weak interaction energies were tested for the dimers collected in the S66 benchmark set of Řezáč and co-workers.[90] Regarding the basis set limit of interaction energies, it is not possible to perform quadruple-ζ CCSD(T) calculations for all members of the S66 set, but recent benchmark data obtained with explicitly correlated methods are available for comparison.[42,91] In the recently complied CEMS26 set well-reproducible CCSD(T) reference energies were obtained with at least triple-ζ quality basis sets for as large molecules as possible (see Fig. 6 of Ref. 27 for a pictorial representation of the entries of CEMS26).

To illustrate the convergence with the local approximations and with the AO basis set towards the CBS limit of DF-CCSD(T) six additional, larger, and more challenging systems were selected, containing 58 to 92 atoms (see Figs. 1 and 2).

For such systems, of course, the exact CCSD(T) reference cannot be computed with reasonable basis sets, and comparisons can only be made to results obtained with very tight thresholds or to recent calculations performed with alternative local CCSD(T) schemes.[36,40] Among these systems the formation of androstendione from its precursor is the simplest, serving as a model of a one-step organic reaction.[40] A much more challenging case is the fourth reaction of the isomerization test set (ISOL4) of Grimme and co-workers,[92] because the two intermediate steps in the biosynthesis of cholesterol, lanosterol [ISOL4 educt] and (S)-2,3-oxidosqualene [ISOL4 product] are markedly different, separated by many elementary steps of the net reaction, hence one cannot rely on any error compensation between the educt and product. The "AuAmin" ($AuC_{41}H_{45}N_4P$) organometallic reaction was also introduced by the Werner group as an especially hard case for local correlation methods,[38] because of the extensive correlation energy contribution of the numerous important but individually

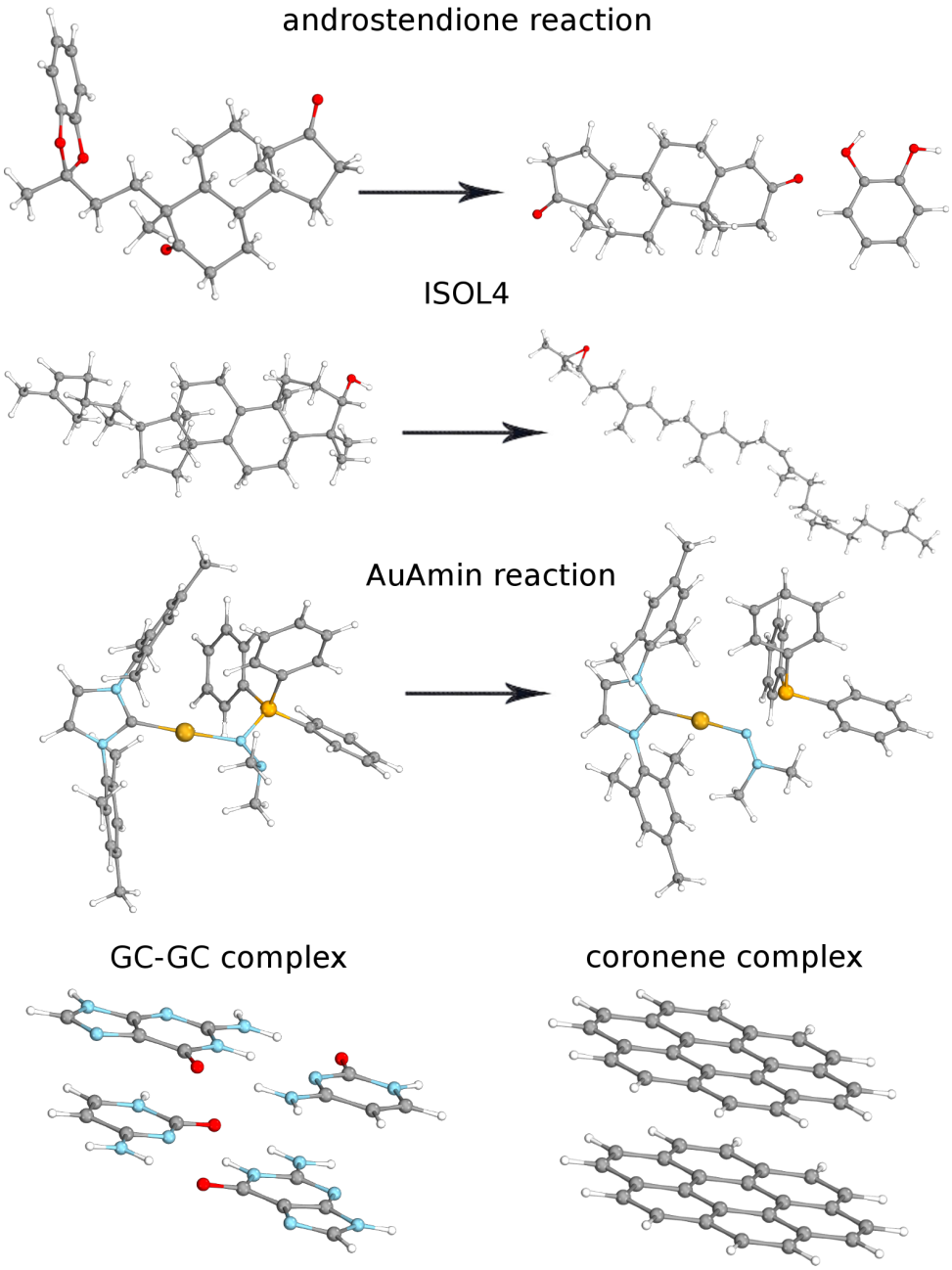

Figure 1: The reactions and dimers investigated in the convergence tests of Sect. 6: formation of androstendione,[38] fourth isomerization of the ISOL test set[92] (ISOL4), the AuAmin reaction,[38] and dimers of the guanine–cytosine complex and coronene taken from the L7 test set.[93]

small non-covalent interactions to the reaction energy. We also considered two notoriously complicated examples taken from the L7 test set of Hobza and co-workers,[93] namely the

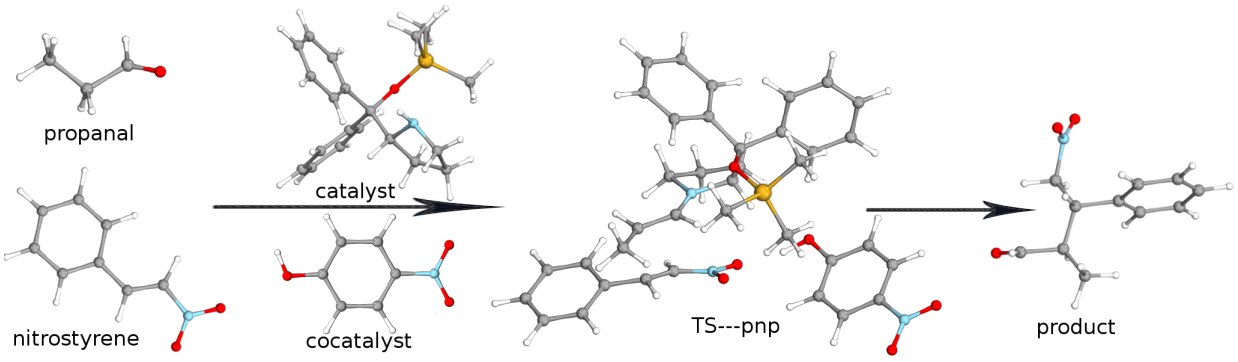

Figure 2: Reactants (propanal and β-nitrostyrene), catalyst and cocatalyst (p-nitrophenol), and the transition state of the C–C bond formation step ($\mathrm{TS}^{\mathrm{RS}}_{\mathrm{CC}} \cdots \mathrm{pnp}$) leading to the major stereogenic product. See Sect. 7.1 and Ref. 64 for further details.

dimerization energies of the guanine–cytosine complex and coronene (GCGC and C2C2PD, respectively), in order to assess the performance of the LNO-CCSD(T) method directly on the most challenging type of non-covalent interactions, that is, on the example of extended π-π stacking interactions.

A case study on a typical organocatalysis reaction will also be presented as an illustrative example of the application of the LNO-CCSD(T) approach to realistic chemical problems. The reactants, product, and transition state (TS) considered here for the Michael-addition reaction of Ref. 64 are depicted in Fig. 2. Finally, large-scale calculations illustrating the current capabilities of our LNO-CCSD(T) implementation will be presented for a lipid transfer protein[94] (LTP, PDB ID: 1N89) containing 1023 atoms and up to 44712 atomic orbitals (see Fig. 3).

The structures of all the investigated molecules are available in the above cited reports. The Cartesian coordinates of most of the systems and the employed (DF-)CCSD(T) reference energies are collected in the Supporting Information (SI) of Ref. 27. The structure used for the LTP calculations is provided in the SI.

For the statistical analysis the maximum absolute error (MAX), the mean absolute error (MAE), and the standard deviation of these errors (STD) were applied. Relative energy differences with respect to a reference energy (E^{ref}) are obtained as $(100\%)\cdot(E^{\mathrm{LNO-CCSD(T)}} -$

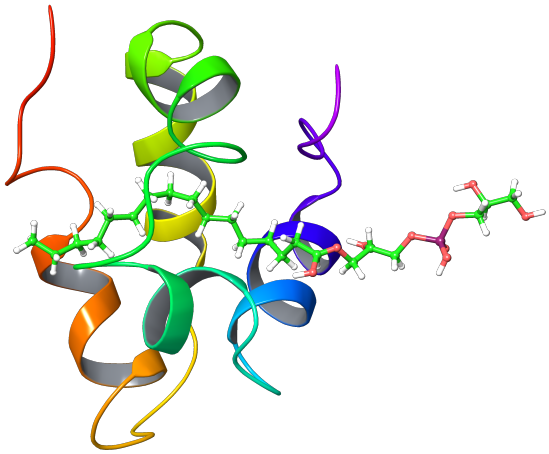

Figure 3: Lipid transfer protein[94] containing 1023 atoms, including its ligand of 79 atoms.

$E^{\text{ref}})/E^{\text{ref}}$. Exact CCSD(T) or LNO-CCSD(T) correlation energies obtained with very tight thresholds will be used as E^{ref}.

5 Benchmarks for medium-sized systems

The most appropriate and precise assessment of the quality of local correlation energies and energy differences is to compare them to exact (DF-)CCSD(T) references. The extremely high cost of conventional CCSD(T), especially if sufficiently large, at least triple-ζ quality AO basis sets are employed, does not allow such reference calculations to be performed on systems larger than those in the CEMS26 compilation. Nevertheless, at least in this range of up to 63 atoms one can benefit from the exact (DF-)CCSD(T) references in the given AO basis sets, and a relevant portion of the chemical space can be sampled.

5.1 Reference energies for the benchmark sets

Correlation and reaction energies are tested first on the test set employed in one of the local PNO studies of Ref. 89. The NWH set consists of 23 organic reactions of 47 different molecules. All systems are below 36 atoms, which allowed us to perform conventional CCSD(T) calculations using the cc-pVTZ, aug-cc-pVTZ, and also the cc-pVQZ basis set.[27]

We showed in a detailed analysis[27] that (at least for the NWH set) the DF-CCSD(T)/cc-pVTZ correlation energies are systematically overestimated compared to CCSD(T)/cc-pVTZ without DF by about 0.05%. Consequently, this deviation has to be taken into account when comparing local CCSD(T) energies (including DF errors) to conventional CCSD(T) references without DF, otherwise local errors in the usually underestimated correlation energies would seem about 0.05% smaller due to the DF error present only in the former.[27] To avoid this bias we have chosen to minimize the DF error in the presented local CCSD(T) calculations [for LNO-CCSD(T) and both DLPNO-CCSD(T) variants] by fitting the integrals of the (aug-)cc-pVXZ basis with auxiliary functions of the (aug-)cc-pV($X+2$)Z-RI basis. As expected, the reference reaction energies are affected by the DF approach to a negligible extent.[27]

Correlation and non-covalent interaction energies are benchmarked for the 66 organic molecular dimers of the S66 compilation.[90] We performed DF-CCSD(T) reference calculations with the aug'-cc-pVTZ basis set (aug-cc-pVTZ basis for non-hydrogen, cc-pVTZ for hydrogen atoms)[27] for all 3×66 species, containing at most 34 atoms. It was not possible to carry out DF-CCSD(T) for the entire set with larger basis sets, hence the CBS extrapolated energies are compared to the most recent "SILVER" reference values of Martin and co-workers, including half counterpoise (CP) corrected MP2-F12/aug-cc-pV(T,Q)Z-F12 improved by [CCSD(F12*) − MP2-F12]/aug-cc-pVTZ-F12 and [CCSD(T) − CCSD]/aug'-cc-pV(D,T)Z corrections.[91] This introduces, of course, a small uncertainty when CBS extrapolated local CCSD(T) energies obtained with a'(T,Q)Z are assessed against the SILVER values. Recently, Werner and Ma performed PNO-CCSD(T*)-F12/aug-cc-pVQZ-F12 calculations and found about 0.02 and 0.06 kcal/mol MAE and MAX deviations, respectively, compared to the same SILVER reference.[42] On the basis of that comparison the mean uncertainty of the SILVER reference values can be estimated to be about 0.02 kcal/mol, which is in good agreement with the uncertainty estimate of Martin and co-workers.[91]

For the species of the CEMS26 test set the available reference energies also include both

CCSD(T) and DF-CCSD(T) values obtained from the literature or computed by us.[27,95] Again, to avoid the potential bias of the DF error in the correlation energies, the "$(X+2)$Z-RI" type auxiliary basis sets were utilized in both the LNO and the DLPNO calculations when the reference values were calculated without the DF approximation (see Table 8 and the supplementary material of Ref. 27 for further details.)

5.2 Accuracy of correlation energies

Mean absolute and maximum correlation energy errors and the STD measures of LNO-CCSD(T) results are collected for all three compilations in Table 2. The systematic convergence of the LNO-CCSD(T) results can be observed when comparing the Loose, Normal, Tight, and vTight LNO-CCSD(T) values with the canonical (DF-)CCSD(T) references. Generally, the convergence of all three error measures are monotonic and fast, both the average absolute and maximum errors improve in almost all the cases by at least a factor of two when switching to a better threshold set. This exponential rate of improvement in the accuracy observed with the tighter and tighter settings is in accord with the exponential rate of decrease in the truncation thresholds (see Table 1). It is also highly satisfactory that, for a wide range of molecules, concerning both the type and the size (2-63 atoms), and for a good selection of medium to large basis sets the MAE for correlation energies is below 0.069% already with the default, Normal settings. This MAE value is significantly better, below 0.029% for the smaller species of the NWH and S66 compilations with the (aug)-cc-pVTZ basis sets. Since the orbitals are noticeably more spread when diffuse functions are added to the AO basis set, it is challenging to design local approximations with balanced accuracy for AO bases with and without diffuse functions. Considering the maximum errors, except for 6 species, namely the [2.2]paracyclophane in the NWH set with cc-pVQZ basis, and for the two porphyrin derivatives, the two $(H_2O)_{17}$ conformers, and octamethylsilsesquioxane of the CEMS26 set, all correlation energy errors are below the 0.1% mark already with the Normal settings. The Tight setting ensures that, for all studied systems and with any basis

set considered, the LNO-CCSD(T) correlation energies are accurate to at least 99.9%. Even the more challenging NWH/cc-pVQZ and the CEMS26 MAE values get below 0.028% with the Tight thresholds on the average. The Normal and Tight STD values are in the range of 0.022-0.044% and 0.009-0.019%, respectively, for all species and basis sets, hence some beneficial compensation of errors can be expected when energy differences are formed from chemically similar systems. Regarding the performance of the Loose settings, excluding for a moment the NWH/cc-pVQZ case, both the MAE and the STD values are below 0.11%. Thus the Loose setting serves its purpose of providing a very economic but still relatively reliable estimate of the better converged LNO-CCSD(T) energies, at least for triple-ζ basis sets. The higher errors of the NWH/cc-pVQZ energies with the Loose settings are explained by the fact that the same, relatively loose LNO occupation number threshold truncates the larger, quadruple-ζ basis more severely.

Finally, we also included the same error measures using the vTight settings for two purposes. First, it is important to see that the exponential improvement in the correlation energies continues when even better settings are employed. As we showed previously, following this approach, i.e., using tighter and tighter settings for the individual truncations, the conventional correlation energy can be recovered with arbitrary accuracy[23,27] due to the fact that all of our local approximations are systematically improvable. The results presented here also support this statement. Second, we are aiming to employ the vTight results as reference for larger systems, where it is impossible to carry out the canonical CCSD(T) calculations. This is justified by the rapid convergence of the energies and the excellent overall accuracy of the vTight setting with at most 0.02% MAE and 0.01% STD.

The analogous DLPNO-CCSD(T_0) and DLPNO-CCSD(T) correlation energy errors obtained using the same reference are included in Table S1 of the SI with both the NormalPNO and the TightPNO settings.[33] Concluding the detailed analysis of Sect. S1 of the SI in brief, the (T_0) approximation-free DLPNO-CCSD(T) is systematically, on the average by about 0.2% closer to the canonical reference than DLPNO-CCSD(T_0). Still, both

Table 2: Statistical measures for relative LNO-CCSD(T) correlation energy errors in percents for the species of the NWH,[89] S66,[90] and CEMS26[27] test sets with respect to canonical CCSD(T) reference.

Threshold	MAE	MAX	STD
NWH, cc-pVTZ			
Loose	0.051	0.126	0.034
Normal	0.024	0.092	0.019
Tight	0.013	0.052	0.010
vTight	0.006	0.036	0.006
NWH, aug-cc-pVTZ			
Loose	0.043	0.164	0.046
Normal	0.024	0.084	0.019
Tight	0.013	0.060	0.010
vTight	0.006	0.045	0.007
NWH, cc-pVQZ			
Loose	0.214	0.364	0.104
Normal	0.069	0.143	0.036
Tight	0.028	0.094	0.016
vTight	0.019	0.073	0.011
S66, aug'-cc-pVTZ			
Loose	0.046	0.134	0.031
Normal	0.029	0.094	0.022
Tight	0.017	0.046	0.009
vTight	0.007	0.018	0.005
CEMS26			
Loose	0.109	0.364	0.090
Normal	0.067	0.145	0.044
Tight	0.026	0.065	0.019
vTight	0.014	0.033	0.008

the MAEs of 0.12-0.24% and 0.04-0.10% obtained with the NormalPNO and TightPNO settings of DLPNO-CCSD(T), respectively, are above the MAE of 0.02-0.07% provided by LNO-CCSD(T) already with its Normal settings. This comparison of correlation energies is mostly useful to make correspondence between the different composite threshold nomenclature of the two schemes. The characterization of energy differences, such as reaction, interaction, and conformation energies (next section), is naturally of more interest from the perspective of chemical applications.

5.3 Reaction, interaction and conformation energies

We have collected the same statistical measures for the deviations of LNO-CCSD(T) energy differences with respect to the conventional CCSD(T) reference for the above three test sets in Table 3. Additionally, the average and standard deviation of the signed errors are depicted in the form of normal distributions in Fig. 4 and Fig. S1 of the SI. Looking at Table 3 and Fig. 4 we again observe a systematic and rapid convergence when switching to a better approximation set. The errors with the Normal settings are about 2-3 times better than those obtained with the Loose combination, whereas the Tight results are about 1.5-3 times better than the ones with the Normal settings. The MAE values for the reaction and interaction energies of the smaller species of the NWH and S66 sets are below 0.16 kcal/mol, which grows to 0.34 kcal/mol for the larger examples in the CEMS26 set. The Tight setting ensures highly-accurate energy differences with MAE values below 0.12 kcal/mol across the three studied benchmark compilations and with various, triple- and quadruple-ζ bases. Maximum errors are also promising being below 1 kcal/mol for all but one studied examples already with the default thresholds. The 1.01 kcal/mol error is obtained for the complexation energy of Li^+ with two 12-crown-4 molecules for which the CCSD(T) reference value is 125.2 kcal/mol. This corresponds to a relative error of about 0.8% in this case which is acceptable from the perspective of most applications. Nevertheless, the Tight settings provide an excellent accuracy of below 0.46 kcal/mol for all of our examples. We

again observe a slower convergence with the cc-pVQZ basis set, the largest errors of the Tight threshold set are obtained for this case. However, if comparing the correlation and reaction energy errors, especially for the Normal settings, one finds that the cc-pVQZ energy difference deviations are comparable to the (aug-)cc-pVTZ ones even if the LNO-CCSD(T) correlation energies are noticeably worse with the cc-pVQZ basis set. This can be at least partly explained by the relatively low, 0.036% and 0.016% STD of the cc-pVQZ correlation energy errors obtained with the Normal and Tight settings, respectively. In other words, the CCSD(T) wave function is truncated in a larger extent using the same LNO occupation number thresholds with the larger, cc-pVQZ basis, but the correlation energy appears to be more sensitive to that truncation than the reaction energy. The STD of the reaction energy errors with the Normal and Tight settings are found below 0.30 kcal/mol and 0.13 kcal/mol, respectively, indicating that fast convergence can be expected also when differences between reaction energies or barrier heights are of interest.

The MAE of Loose LNO-CCSD(T) results is acceptable if it is used to obtain a quick estimate. For the reaction and interaction energies of the smaller species of the S66 and NWH set a relatively good MAE of below 0.41 kcal/mol was found. The same Loose MAE jumps to 1.5 kcal/mol for the more realistic CEMS26 set. The maximum errors being in the range of 1.3-3.7 kcal/mol indicate that large deviations occur with the Loose settings, which are probably most useful for exploratory calculations and convergence tests. Considering the MAX errors of the NWH set for the Loose values, there is one outlier, the formation of [2, 2]paracyclophane from two *p*-xylenes, which has by far the largest errors in the range of 2.6-2.8 kcal/mol. For the second worst case the largest errors do not exceed 1.4 kcal/mol for any of the three basis sets and are thus more than twice better than for this complicated example.

Considering also the vTight setting, one finds that, in spite of the significant improvement in the correlation energies provided by the vTight setting over Tight, energy differences are not improved at the same extent. This can be explained by looking at the difference of

the correlation energy errors. For instance, the Tight and vTight correlation energy deviations are 0.053% and 0.017% for the Mg-porphyrin and 0.049% and 0.014% for porphyrin, respectively, thus correlation energies become better by about a factor of 3, but the reaction energy error of the complexation of Mg with porphyrin only changes from 0.13 kcal/mol to 0.11 kcal/mol when switching to the vTight threshold set. In the case of the NHW set, there is again an outlier, the challenging isomerization of 1,2,3,4,5,6-heptahexane to hepta-1,3,5-triyne. The MAX errors of 0.24-0.35 kcal/mol for this example in all three basis sets are about twice as large as the second highest absolute deviations of 0.12-0.15 kcal/mol. This is explained by the markedly different electronic structure of the two isomers: 1,2,3,4,5,6-heptahexane is a cumulene with six double bonds, while 1,3,5-heptatriyne has three triple bonds. Both of these motives occur relatively rarely and can be considered challenging for local methods due to the numerous unsaturated bonds. Nevertheless, the excellent overall performance of the vTight energy differences (MAX error is below 1 kJ/mol for all examples expect for the 1,2,3,4,5,6-heptahexane isomerization) validates our assumption that vTight LNO-CCSD(T) results can reliably substitute canonical CCSD(T) if such reference is not available.

The normal distribution plots of Fig. 4 (and Fig. S1 of the SI for NWH/aug-cc-pVTZ and NHW/cc-pVQZ) based on the mean signed deviations of energy differences and their STDs reveal some additional trends. For all three test sets and all AO basis choices, on the average, the Loose and Normal energy differences are slightly underestimated, while the Tight and vTight results are slightly overestimated. Looking at the individual cases over- and underestimations can both occur with almost the same probability with all threshold combinations and basis sets. Nevertheless, the mean signed errors are fairly close to 0 kcal/mol, and the convergence towards the exact CCSD(T) is again apparent from the narrower and narrower distributions. The figures also indicate that chemical accuracy can be expected in the majority of the cases already with the Normal settings, and the Tight results are mostly located within the highlighted 1 kJ/mol error region. Despite of the relatively good

Table 3: Statistical measures for LNO-CCSD(T) deviations in kcal/mol for the reaction energies of the NWH,[89] the interaction energies of the S66,[90] and the reaction and conformation energies in the CEMS26[27] test sets with respect to canonical CCSD(T) reference. See section 5.3 for more details.

Threshold	MAE	MAX	STD
NWH, cc-pVTZ			
Loose	0.41	2.78	0.68
Normal	0.13	0.64	0.23
Tight	0.09	0.32	0.09
vTight	0.04	0.24	0.06
NWH, aug-cc-pVTZ			
Loose	0.40	2.78	0.62
Normal	0.14	0.61	0.21
Tight	0.08	0.35	0.10
vTight	0.05	0.29	0.06
NWH, cc-pVQZ			
Loose	0.38	2.63	0.57
Normal	0.13	0.63	0.23
Tight	0.09	0.46	0.12
vTight	0.05	0.35	0.07
S66, aug'-cc-pVTZ			
Loose	0.32	1.34	0.36
Normal	0.16	0.58	0.15
Tight	0.05	0.20	0.05
vTight	0.03	0.07	0.02
CEMS26			
Loose	1.50	3.70	1.32
Normal	0.34	1.01	0.30
Tight	0.12	0.19	0.03
vTight	0.11	0.18	0.05

average performance of the Loose settings, the wide spread of the individual errors makes this choice less reliable. From the perspective of practical applications for medium to large systems both the Tight and the vTight LNO-CCSD(T) results can be basically considered benchmark quality. The trends are very similar for the NWH set with all three basis sets, thus the normal distribution plots for the aug-cc-pVTZ and cc-pVQZ bases sets are only provided in the supplementary material (see Fig. S1 of the SI).

Although the main focus of the present report is to benchmark the accuracy of the LNO-CCSD(T) method, for which the presented Normal, Tight, etc. threshold compositions were designed, it might also be interesting to consider the CCSD and (T) terms separately. It is well-documented, that CCSD in itself is often insufficient for chemical accuracy, and there is no practical reason to stop at the LNO-CCSD level, because an LNO-CCSD(T) calculation takes only about 30% more time than that of LNO-CCSD. Thus we recommend to always perform LNO-CCSD(T) instead of LNO-CCSD. Nevertheless, the same error measures are also evaluated and analyzed for LNO-CCSD in Sect. S6 of the SI. In brief, since the MP2 correction of Eq. (5) systematically overcorrects the CCSD energies,[21,27] we found that an equal distribution of $\Delta E_{i'}^{\mathrm{MP2}}$ to the CCSD and (T) terms drastically improves the accuracy of LNO-CCSD. While about one step tighter settings are needed for LNO-CCSD including the total MP2 correction to match the accuracy of LNO-CCSD(T), i.e., Tight LNO-CCSD is comparable to Normal LNO-CCSD(T), LNO-CCSD with 0.5 $\Delta E_{i'}^{\mathrm{MP2}}$ is as good as LNO-CCSD(T) even if the same thresholds are used.

Returning to the normal distribution plots of Fig. 4 and Fig. S1 of the SI, and to the statistical measures of Tables 3 and S2 of the SI, these data collections are also useful to compare LNO-CCSD(T) with the DLPNO-CCSD(T) variants against the same reference in a given AO basis set. The detailed discussion of the above data is provided in Sect. S3 of the SI, the conclusions are briefly summarized here. While the NormalPNO DLPNO results with the two (T) variants are found almost identical, at least for the studied reaction energies only DLPNO-CCSD(T) with iterative triples treatment was able to fully benefit from

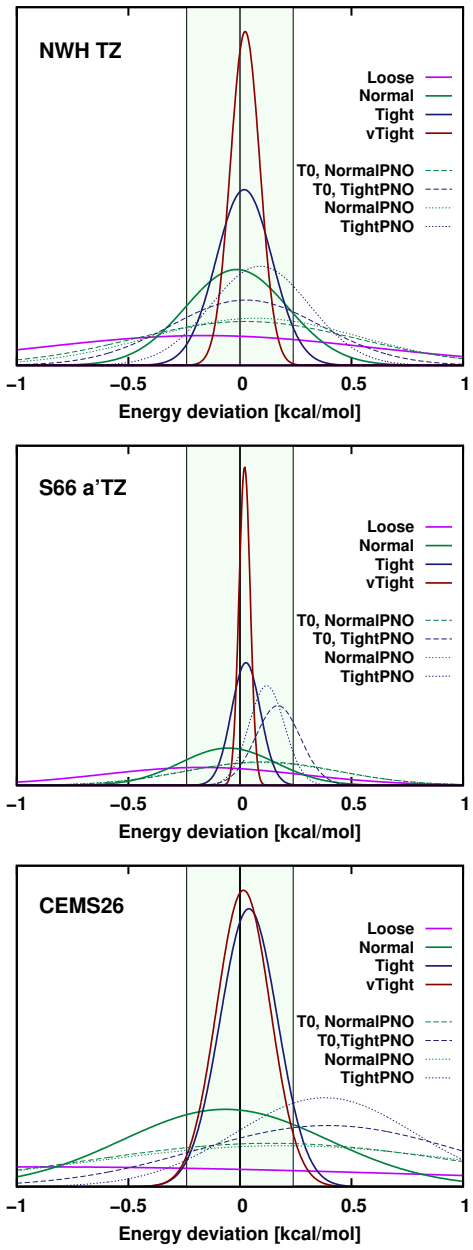

Figure 4: Normal distribution of the LNO-CCSD(T), DLPNO-CCSD(T_0), and DLPNO-CCSD(T) deviations from the reference CCSD(T) energy differences in kcal/mol for the NWH[89] set using the cc-pVTZ basis, for the S66[90] set using the aug'-cc-pVTZ basis, and for the CEMS26[27] set. The ± 1 kJ/mol error region is highlighted in the middle.

the accuracy of the TightPNO setting. Accordingly, for both the NWH and the CEMS26 sets the Normal LNO-CCSD(T) distributions of Fig. 4 are apparently narrower than both the "T0, NormalPNO" and the "T0, TightPNO" curves. The widths of the "TightPNO" curves are comparable to that of the Normal LNO-CCSD(T) for the aforementioned two test sets but exhibit up to almost 0.4 kcal/mol larger centroids. In the case of the S66 set, both "TightPNO" curves are situated about halfway between the Normal and Tight LNO-CCSD(T) ones. According to the analysis of Sect. S6 of the SI the performance of the two local methods are more comparable for the case of CCSD. NormalPNO DLPNO-CCSD is close to Normal LNO-CCSD with full MP2 correction, while TightPNO DLPNO-CCSD achieves excellent accuracy outperforming Tight DLPNO-CCSD(T) on the average and matching Tight LNO-CCSD with half MP2 correction.

5.4 Extrapolation to the basis set limit

In this section, we investigate the convergence of LNO-CCSD(T) towards the CBS limit of CCSD(T). For that purpose, as before, the local approximations will be improved systematically, and the AO basis set limit will be approached via conventional extrapolation techniques [see Eqs. (7) and (8) of Sect. 4]. As reference, for the NWH test set our canonical CCSD(T)/(T,Q)Z results are available to assess the effect of local approximations on the extrapolated reaction energies. Since the (a')QZ calculations are not feasible for the entire S66 test set, here we cannot employ the analogous CCSD(T)/a'(T,Q)Z reference. Thus the recent F12-based values of Martin and coworkers[91] are used for comparisons. This choice will introduce some uncertainty, because, in contrast to the case of the NWH set, for the interaction energies of S66 the local and basis set incompleteness errors cannot be completely separated. The benchmark calculations in Ref. 91 for the S66 and in Ref. 27 for the NWH set are very close to the limit of what is currently feasible with state of the art conventional CC implementations, thus obtaining such references for larger systems, e.g., for those in the CEMS26 compilation seems out of the question, at least in the near future.

Table 4: Statistical measures for CBS extrapolated LNO-CCSD(T) deviations in kcal/mol. References: canonical CCSD(T)/(T,Q)Z for the reaction energies of the NWH[89] set and the "SILVER" reference value of Ref. 91 for the interaction energies of the S66[90] set. See section 5.4 for more details.

Threshold	MAE	MAX	STD
NWH, cc-pVTZ			
Loose	0.59	2.79	0.68
Normal	0.48	2.50	0.55
Tight	0.49	2.48	0.57
vTight	0.47	2.46	0.55
NWH, cc-pVQZ			
Loose	0.24	0.96	0.25
Normal	0.20	0.84	0.20
Tight	0.20	0.83	0.21
vTight	0.19	0.82	0.20
NWH, CBS (T,Q)Z			
Loose	0.37	2.51	0.54
Normal	0.14	0.68	0.20
Tight	0.09	0.56	0.14
vTight	0.06	0.44	0.09
S66, aug'-cc-pVTZ			
Loose	0.66	2.97	0.72
Normal	0.51	2.38	0.53
Tight	0.41	1.90	0.41
vTight	0.41	1.84	0.38
S66, aug'-cc-pVQZ			
Loose	0.44	1.91	0.49
Normal	0.30	1.41	0.31
Tight	0.14	0.77	0.16
vTight	0.13	0.68	0.13
S66, CBS (a'T,a'Q)Z			
Loose	0.33	1.36	0.35
Normal	0.18	0.84	0.18
Tight	0.05	0.19	0.04
vTight	0.04	0.14	0.03

Statistical measures for the reaction and interaction energy deviations for the NWH and S66 sets are collected in Table 4. The corresponding normal distribution plots are provided in Fig. 5. Let us first compare the effect of local approximations on the basis set extrapolated case with the results obtained with the TZ and QZ bases (Table 3). This is an important indicator because the local and especially the NO approximation with the same thresholds might have unbalanced effect with different AO basis sets, as it is the case to some extent for the correlation energy errors calculated with the TZ and QZ bases for the NWH set (Table 2). However, the reaction energies formed from the extrapolated LNO-CCSD(T) energies (third block of Table 4) are almost exactly as accurate (compared to the corresponding conventional CCSD(T) results) as the reaction energies of Table 3 obtained in the TZ and QZ bases. In other words, the accuracy of reaction energies in the given basis set is preserved upon extrapolation. Additionally, in the extrapolated case the rate of convergence of the local approximations closely follows that observed for the TZ and QZ basis sets. For instance, the MAE values (in kcal/mol) of Table 4 in the series of Loose to vTight read 0.37, 0.14, 0.09, and 0.06, while the analogous error measures are within 0.04 and 0.02 kcal/mol of those in the TZ and QZ bases, respectively. The corresponding extrapolated MAX and STD values resemble similarly well the TZ and QZ counterparts, the agreement here is somewhat better with the QZ results.

The relatively slow convergence of the conventional CC energies is well-documented in the literature, at least for small molecules. The situation is not different in the case of local CC methods, and it is perhaps even more pronounced as larger and larger systems become accessible. In the first two blocks of Table 4, we compare LNO-CCSD(T) results obtained with TZ and QZ basis sets to the CCSD(T)/(T,Q)Z reference. Apparently, the BSIE dominates the deviation of both the TZ and the QZ results from the (T,Q)Z reference since the errors hardly change when better and better thresholds are employed for the local approximations. The sizable BSIE in CC energies, even with triple- or quadruple-ζ basis, is, of course, expected and well-known. However, as local correlation methods have become

more popular in recent years a growing number of reports appear in the literature where only triple-ζ or occasionally even smaller AO basis sets are employed. Thus we find it important to highlight here as well that the intrinsic accuracy of local correlation methods can only be utilized if a sufficiently large basis set is used. Conversely, as long as only at most triple-ζ basis sets are employed for any local CC method (computed without F12 contributions), the accuracy of the results will be determined by the BSIE and will hardly be affected by the threshold set governing the local approximations.

Looking at the analogous results presented for the S66 set the main trends are very similar. The LNO-CCSD(T)/(a'T,a'Q)Z results (last block of Table 4) are converging rapidly and monotonically towards the "SILVER" reference of Ref. 91. The accuracy of the extrapolated LNO-CCSD(T) interaction energies is again unaffected by the local approximations, i.e., the rate of convergence towards the local approximation-free reference is similar for the a'TZ (Table 3) and for the (a'T,a'Q)Z cases. This indicates that our local approximations are comparably suitable for both the a'TZ and a'QZ bases, at least also for interaction energies. There is one difference to observe compared to the case of the reaction energies above, which difference originates from the fact that it is not possible to use DF-CCSD(T)/(a'T,a'Q)Z values as reference. This difference is that the deviations of the vTight results from the reference are not as much better than those of the Tight ones as in the case of the a'TZ results, where the DF-CCSD(T)/a'TZ reference was available. This deviation is mostly caused by the remaining BSIE in the (a'T,a'Q)Z results compared to the "SILVER" reference. Indeed, we observe an even better agreement characterized by a MAE (MAX) of 0.04 (0.14) kcal/mol with respect to the "SILVER" reference when the interaction energies are CP corrected and/or calculated from LNO-CCSD(T)/(a'Q,a'5)Z results.[74] Additionally, we computed the DF-CCSD(T)/(a'T,a'Q)Z reference for the benzene-pyridine complex (1290 basis functions, see Fig. 6), for which the deviation of vTight LNO-CCSD(T)/(a'T,a'Q)Z is one of the largest with respect to the "SILVER" reference, that is, 0.094 kcal/mol. The same error is much smaller compared to the DF-CCSD(T)/(a'T,a'Q)Z reference, namely 0.037 kcal/mol. The

issue of the remaining BSIE will be addressed in detail in our forthcoming report,[74] using CP correction and/or (a'Q,a'5)Z extrapolation. Nevertheless, the MAE (MAX) errors of 0.18 (0.84) kcal/mol and 0.05 (0.19) kcal/mol obtained with the Normal and Tight settings, respectively, and with (a'T,a'Q)Z extrapolation are probably sufficient for most practical purposes.

Let us briefly look at the a'TZ and a'QZ deviations (fourth and fifth blocks of Table 4) compared to the "SILVER" CCSD(T) reference. The results again demonstrate that the BSIE dominates the a'TZ results, the accuracy with respect to the close to basis set limit reference cannot be improved by tightening the thresholds of the local approximations. In the case of a'QZ the local and BSIE errors become comparable, hence some improvement is observed when more accurate thresholds are employed. Nevertheless, the a'QZ accuracy is also limited by the BSIE with all four investigated threshold combinations, and the convergence pattern of the a'TZ errors of Table 3 is only approached with the (a'T,a'Q)Z extrapolated results.

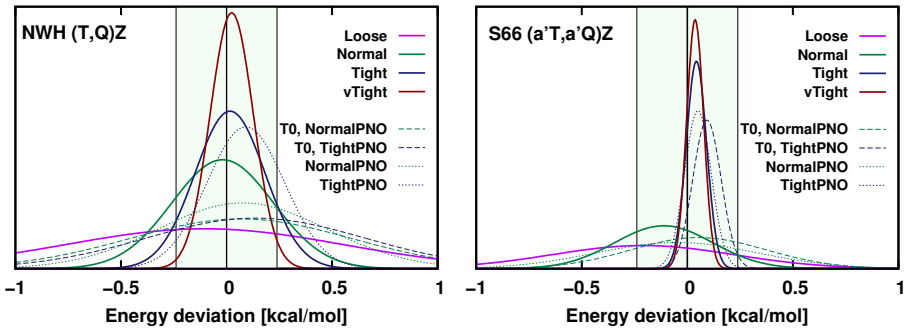

Figure 5: Normal distribution of the LNO-CCSD(T), DLPNO-CCSD(T_0), and DLPNO-CCSD(T) deviations from the reference CCSD(T) energy differences in kcal/mol. CCSD(T)/CBS(T,Q) was employed as reference for the NWH[89] set, while the "SILVER" reference of Ref. 91 was used for the interaction energies of the S66[90] set. The ± 1 kJ/mol error region is highlighted in the middle.

The above trends are depicted in the form of normal distribution plots on Fig. 5. The extrapolated results for the NWH set are similar to the analogous plots of Fig. 4 and Fig. S1, the distributions indicate fast convergence with the local approximations, and the curves are centered very close to 0 kcal/mol. Chemical accuracy can already be expected

with the Normal setting, while the Tight and vTight accuracy levels provide errors within the highlighted 1 kJ/mol error range for most of the cases. The S66 plots do not follow the trends of the a'TZ curves of Fig. 4 as closely as in the case of the NWH set, again, because the DF-CCSD(T)/(a'T,a'Q)Z reference is unavailable. The accuracy of the Loose and Normal results is comparable to the a'TZ case because the local errors are still larger than the remaining BSIE for these settings. The Normal errors are still mostly within the 0.5 kcal/mol vicinity of the reference, and there is again a significant improvement with the Tight settings. However, the Tight and vTight distributions are slightly shifted from the center towards positive deviations indicating that the (a'T,a'Q)Z results are generally overestimated compared to the "SILVER" reference, which is assumed to be closer to the CBS limit.

Fig. 5 and Table S3 of the SI display the error measures for the corresponding CBS(T,Q)-extrapolated DLPNO reaction and interaction energies. Our results are in parallel with the study of Liakos *et al.*, where the local errors of the DLPNO-CCSD(T_0)/(D,T)Z results were not affected by the extrapolation.[33] The trends regarding the relative accuracy of the DLPNO energies with different threshold settings, (T) terms, and the comparison with LNO-CCSD(T) are thus almost identical to the ones discussed for the NWH/TZ, NWH/QZ, and S66/a'TZ cases. For that reason we do not repeat the conclusions of Sect. 5.3 and Sect. S3 of the SI here.

The above S66 results can also be compared to the recent PNO-CCSD(T)-F12 interaction energies of Ma and Werner.[42] The detailed discussion presented in Sect. S5 of the SI concludes that, for the S66 set, against the respective canonical reference the default of PNO-CCSD(T)-F12 is somewhat more accurate than the default of LNO-CCSD(T), while tighter PNO-CCSD(T)-F12 interaction energies are better than the Tight and comparable to the vTight LNO-CCSD(T) results. On the example of one of the largest species of S66 (uracil dimer), using almost identical processors and 20 cores wall times measured for Normal LNO-CCSD(T)/(a'T,a'Q)Z and default PNO-CCSD(T)-F12/aTZ-F12 or for Tight

LNO-CCSD(T)/(a'T,a'Q)Z and PNO-CCSD(T)-F12/aTZ-F12 were found basically identical within the uncertainty level of these timing measurements. Since these implementations and algorithms have different benefits and drawbacks which are also highly system specific, one cannot assume that the compared methods will yield similar accuracy and computational efficiency in all circumstances, but it is out of the scope of the present report to provide such a broad level comparison.

6 Convergence tests for larger systems

This section is devoted to larger examples where it is impossible to carry out conventional CCSD(T) calculations. Since this is the situation for most practical applications of any local CCSD(T) method, we find it important to show how the accuracy of the LNO-CCSD(T) energies can be improved or estimated via convergence tests along both the local approximation and the basis set completeness axes. We also define energy and error bar estimates for the LNO-CCSD(T) energies by utilizing the systematically improvable property of our individual approximations and the trend that the monotonic convergence of the energies is preserved in most of the cases even with the composite threshold combinations.

Figs. 6–10 show the convergence of LNO-CCSD(T) energy differences (reaction energies and barrier heights) as a function of the truncation thresholds and the AO basis set. Let us note that the AO basis set effect includes also the basis set dependence of the HF reference, which is, however, much smaller than that of the correlation energy contribution. All data are made available in Sect. S7 of SI which are necessary to reproduce the figures or decouple the basis set dependence of the HF reference and the correlation energy.

6.1 Introduction and verification of the error estimates

The potential of such convergence studies is illustrated first on two examples from the NWH and S66 sets, where the CBS estimated reference is available. The isomerization energy of

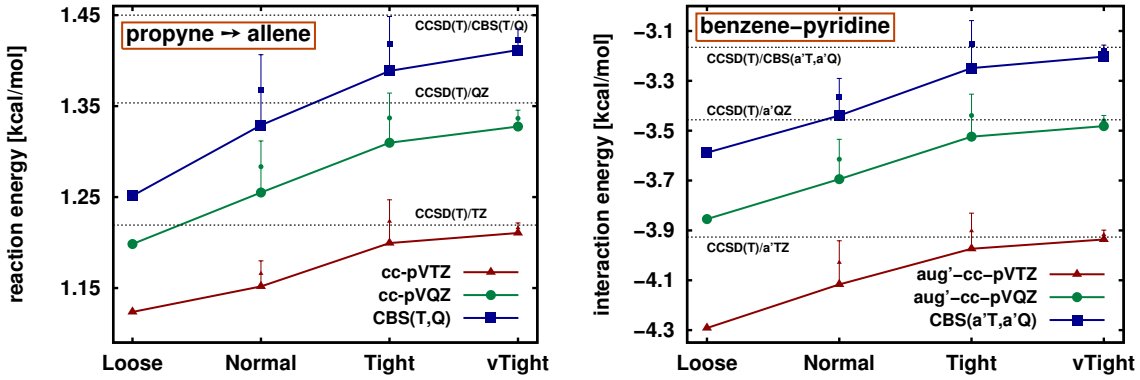

Figure 6: Convergence of LNO-CCSD(T) energies with systematically improving both the local approximations and the AO basis sets. The propyne → allene isomerization reaction energy of the NWH set is plotted on the left, while the benzene-pyridine interaction energy (No. 49 in the S66 set) is studied on the right side of the figure. Horizontal lines represent the exact (DF-)CCSD(T) reference in the AO basis set indicated by the nearby labels. The values depicted by smaller-sized symbols and the corresponding error bars represent an estimation for the next step in the local approximation hierarchy assuming systematic convergence. See Sect. 6 for further explanation.

propyne to allene (Fig. 6, left panel) and the interaction energy of the benzene–pyridine complex (Fig. 6, right panel) both exhibit rapid convergence towards the corresponding conventional reference (dashed horizontal lines) with the local approximations for both the (a')TZ and the (a')QZ basis sets, and consequently also for the extrapolated energies. Let us point out that the vertical range of the plots is relatively spread, but actually the difference between the Normal and the vTight (or the exact) values is only about 0.1 kcal/mol for the propyne–allene and about 0.2-0.3 kcal/mol for the benzene–pyridine cases. Let us also observe the general trend that the convergence curves from Loose to vTight of the (a')TZ and (a')QZ type basis sets are reliably parallel, especially starting from the Normal settings. This parallelity can be exploited to design cost-efficient BSIE corrections, e.g., the BSIE of Tight LNO-CCSD(T)/TZ energies can be effectively improved by adding a Normal LNO-CCSD(T)/(T,Q)Z based correction term. Our forthcoming study will construct and benchmark a wide selection of such accurate and inexpensive LNO-CCSD(T)-based CCSD(T)/CBS approximating schemes.[74] The previously mentioned trend that the local

errors are somewhat smaller with the TZ basis than with the QZ basis are also apparent in Fig. 6, hence the remaining local errors of the extrapolated energies are reminiscent of the ones with the QZ basis sets. The fact that the local errors with even the Loose settings are smaller than the BSIE of any of the (a')TZ results is also obvious from the plots, and it is in agreement with the general trend analyzed in Sect. 5.4.

Let us finally look at the energy and error estimates of Fig. 6 denoted by the small-sized symbols and attached error bars. These values, for a given threshold set, are obtained from the energy differences evaluated with the actual and the one level lower composite thresholds and estimate the value and uncertainty of the one level higher energy difference assuming monotonic convergence with the local approximations. For example, the small symbol and error bar depicted above the Normal values indicate that the Tight result is estimated to be in the region surrounded by the error bar. The small symbol in the middle of the error bar is placed to the Tight_estimate = Normal + (Normal - Loose)/2 position, and the endpoints of the error bar are at Tight_estimate ± (Normal - Loose)/2. Fig. 6 reveals that the vTight and also the exact results are indeed within the estimated region obtained from the Normal and Tight results. Also the Tight results are mostly within the range estimated by the Loose and Normal results, except for the TZ curve of the propyne–allene case and the CBS curve of the benzene–pyridine example, but these estimates miss their target by the very small margin of 0.02 kcal/mol and 0.04 kcal/mol, respectively. It is also worth noting that the convergence of the LNO-CCSD(T) energies is not necessarily monotonic towards the exact CCSD(T) value, and consequently in such cases the above simple, two-point extrapolation based energy estimates should be used with caution. If non-monotonic convergence is observed in a relatively small interval, then the convergence is often almost reached. In the rare case of pronounced non-monotonic convergence the best choice is to rely on the result obtained with the tightest settings or extrapolate from the two best converged results.

The relatively small size of the above two examples is probably not representative of the

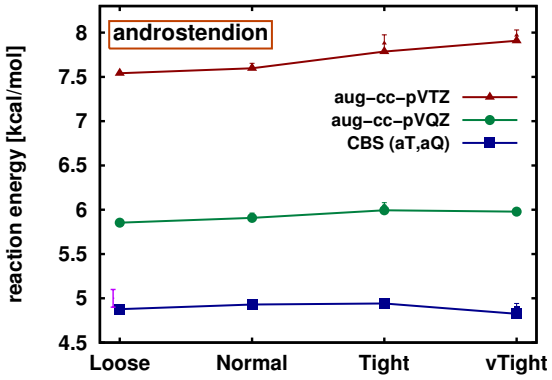

Figure 7: Convergence of LNO-CCSD(T) reaction energies with systematically improving both the local approximations and the AO basis sets for the formation of androstendione from its precursor (see Fig. 1). Available PNO-CCSD(T)-F12/cc-pVTZ-F12 results of various threshold settings are indicated by the purple bar in the bottom left corner. See Sect. 6 and the caption of Fig. 6 for further explanation.

common applications of local CC methods, hence we also looked at more extended systems, first the formation of androstendione from its precursor (Fig. 1). This reaction including medium-sized molecules is a good example of such organic reactions where the reactants and products are chemically similar, i.e., where rapid convergence of local errors can be expected. Indeed, there is only about 0.1-0.3 kcal/mol difference between the Loose and the vTight results with all three levels of basis sets (Fig. 7), indicating the convergence of LNO-CCSD(T) energy differences up to this accuracy. It is not possible to test the accuracy against the exact reference any more, but the deviation of the LMP2 and the exact DF-MP2 energies can give some indication at least about the accuracy of the domain and pair approximations. The Tight and vTight LMP2/aTZ reaction energies differ from DF-MP2 by 0.32 and 0.12 kcal/mol, respectively, while the same deviations with the aQZ basis is only 0.12 and 0.09 kcal/mol (see Table S12 of the SI). Both the LMP2 and the LNO-CCSD(T) correlation energies still change noticeably, by about 0.02-0.05% and 0.01-0.02%, respectively, when going from Normal to Tight and Tight to vTight settings, while the vTight LMP2 correlation energies are accurate to at least 99.993%.

Lacking the exact CCSD(T) reference comparisons to results obtained with alterna-

tive local CCSD(T) methods can yield additional information about the accuracy of LNO-CCSD(T). Ma and Werner presented PNO-CCSD(T)-F12/TZ-F12 reaction energies for the same reaction using multiple strategies for estimating the basis set limit of the triples contributions.[40] Their default settings yield 4.90–4.94 kcal/mol, while the use of tighter domain thresholds resulted in 5.07–5.11 kcal/mol. This is in good agreement with our LNO-CCSD(T)/(aT,aQ)Z results since 4.93, 4.94, and 4.83 kcal/mol were obtained with our Normal, Tight, and vTight settings, respectively. Additionally, about the third of the difference between the two local methods comes from the different reference energies: our HF/(aT,aQ)Z value is about 0.06 kcal/mol smaller than that of Ref. 40.

6.2 Convergence tests for challenging systems

This section collects four cases from the literature which have been established as especially challenging from the perspective of local correlation calculations. Thus these examples will indicate the limits of the accuracy that can be expected from our method in the worst-case scenario.

The first of such hard problems is the ISOL4 isomerization (Fig. 1), where the reactant and products are markedly different. Hence the source of the complication is that the local errors appearing for the quasi-linear product and for the much more compact educt are not canceled. The system size of 81 atoms also indicates that larger deviations from CCSD(T) can be expected solely on the basis of the extensivity of the local errors. Looking at the left panel of Fig. 8, it is promising to find the usual, satisfactory convergence pattern with the improving local approximations, although the difference of the Normal and vTight results is now about 0.8-0.9 kcal/mol, or about 1% in relative terms for all three basis sets. Here we find the Loose results outside of chemical accuracy, but both the Loose–Normal and Normal–Tight based error estimates operate perfectly. The Loose–Normal based interval completely covers both the Tight and the Normal–Tight based error bar, and similarly the Tight–vTight based intervals are the subsets of the Normal–Tight based ones. The small deviation

of the LMP2 reaction energies with respect to exact DF-MP2 changes as 0.41–0.72 kcal/mol, 0.13–0.20 kcal/mol, and 0.08–0.10 kcal/mol when the Normal, Tight, and vTight settings are employed, respectively, which indicates the excellent precision of the domain and pair approximations (see Table S12 of the SI). We again find our 67.56 kcal/mol (Tight) and 67.35 kcal/mol (vTight) LNO-CCSD(T)/(aT,aQ)Z results in agreement with the PNO-CCSD(T)-F12/TZ-F12 isomerization energies, which are 67.53–67.73 kcal/mol with the default and 67.80–67.99 kcal/mol with the tighter settings.[40] In this case the HF contributions of the reaction energies agree more closely [18.65 kcal/mol with HF/(aT,aQ)Z and 18.69 kcal/mol in Ref. 40]. Considering the relatively narrow, 0.1 kcal/mol error bars with our vTight settings and the fact that the LNO-CCSD(T)/(aT,aQ)Z and the PNO-CCSD(T)-F12/TZ-F12 results are getting further from each other when the tightest thresholds are used, a large part of the difference between the two methods can probably be attributed to remaining basis set errors. The LNO-CCSD(T)/aug-cc-pV5Z calculations would be certainly feasible, but those are not yet available to us and will be presented elsewhere. More importantly, this example illustrates well the capabilities of the very economical Normal energies and Loose–Normal based energy estimates, and the high accuracy of the Tight settings, at least if (aT,aQ)Z is used. Otherwise, without the basis set extrapolation the BSIE of the aTZ values being almost 10 kcal/mol or the BSIE of aQZ being still 3.5 kcal/mol would ruin the intrinsic accuracy of CCSD(T).

The next reaction, the dissociation of AuAmin (Fig. 1), can be characterized as one of the most notorious reactions in the local correlation literature due to the numerous, individually very small but collectively significant intermolecular interactions among the ligands.[41] Additionally, the less favorable locality of the wave function around metal atoms also makes transition-metal systems harder for local methods.[41] Moreover, the large size of the AuAmin complex (92 atoms) and its dissociation into two smaller species of comparable size (34 and 58 atoms) result in the most pronounced local correlation and basis set superposition error among the test cases of the present report. Nevertheless, in the right panel of Fig. 8, we

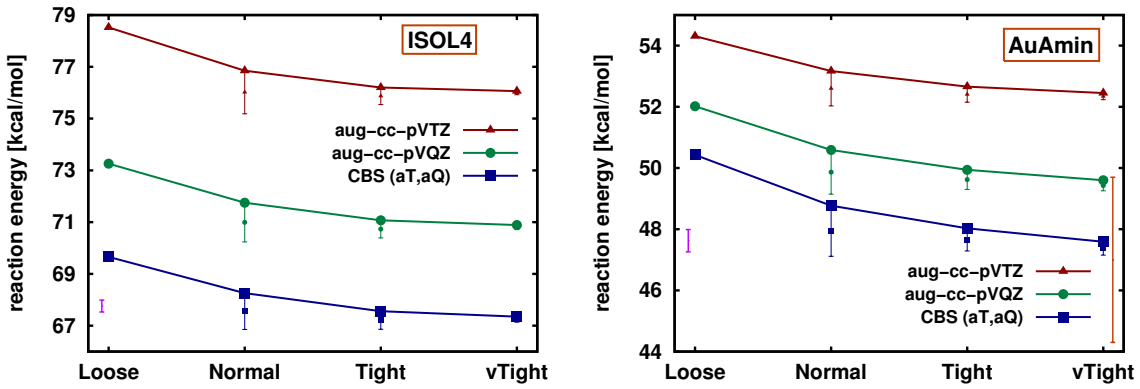

Figure 8: Convergence of LNO-CCSD(T) reaction energies with systematically improving both the local approximations and the AO basis sets for the ISOL4 (left) and AuAmin (right) reactions (see Fig. 1). Available PNO-CCSD(T)-F12/cc-pVTZ-F12 results of various threshold settings are indicated by the purple bars in the bottom left corners, while the experimental reaction energy of AuAmin is indicated by the orange bar in the bottom right corner of the plot. See Sect. 6 and the caption of Fig. 6 for further explanation.

again find the convergence pattern of the LNO-CCSD(T) reaction energies as satisfactory as before. The Normal results are within 1 kcal/mol of the vTight ones for both aTZ and aQZ, whereas the Tight reaction energies are only 0.3 kcal/mol apart from our most accurate values. Again, the Loose–Normal estimates closely envelope both the Normal–Tight and the Tight–vTight error bars. Both the Loose–Normal (47.94 ± 0.83 kcal/mol) and Normal–Tight (47.66 ± 0.37 kcal/mol) estimates and both the Tight (48.03 kcal/mol) and vTight (47.59 kcal/mol) LNO-CCSD(T)/(aT,aQ)Z reaction energies match closely the PNO-CCSD(T)-F12/cc-pVTZ-F12 alternatives, which are 47.26–47.55 kcal/mol and 47.70–47.99 kcal/mol with the default and the larger domain settings, respectively.[40] The HF/(aT,aQ)Z reaction energy of 21.88 kcal/mol is also very close to the 22.03 kcal/mol value computed by Ma and Werner.[40] Aside from this difference of 0.15 kcal/mol in the HF contribution the correlation energy contributions obtained by the two local CCSD(T) schemes are almost exactly match with only a few tenth of a kcal/mol difference. Note also that both methods are in excellent agreement with the reaction energies of 47.0 ± 2.7 kcal/mol obtained from gas-phase measurements from which zero-point corrections were subtracted.[41]

It has been shown in Sect. 5.3 that interaction energies are described generally very well by the LNO-CCSD(T) method, the average absolute (MAX) errors are below 0.2 kcal/mol already with the Normal (Tight) settings for the S66 set. The next two examples are the dimers of guanine–cytosine base pairs (GCGC, Fig. 9, left panel) and of coronene (C2C2PD, Fig. 9, right panel), which include π–π interactions of extended aromatic systems. They represent two of the most difficult intermolecular interaction cases that have been known so far for local methods. The complication in these cases lies in the fact that a crucial amount of the interaction energy comes from the correlation energy contribution. The numerous individually small contributions of such LMO pairs are especially important where the two LMOs are localized on different fragments. The key to accurate π–π interaction energies, since MP2 tends to overestimate them, is the proper selection of strong and distant pairs which are treated at the CCSD(T) and MP2 levels, respectively.

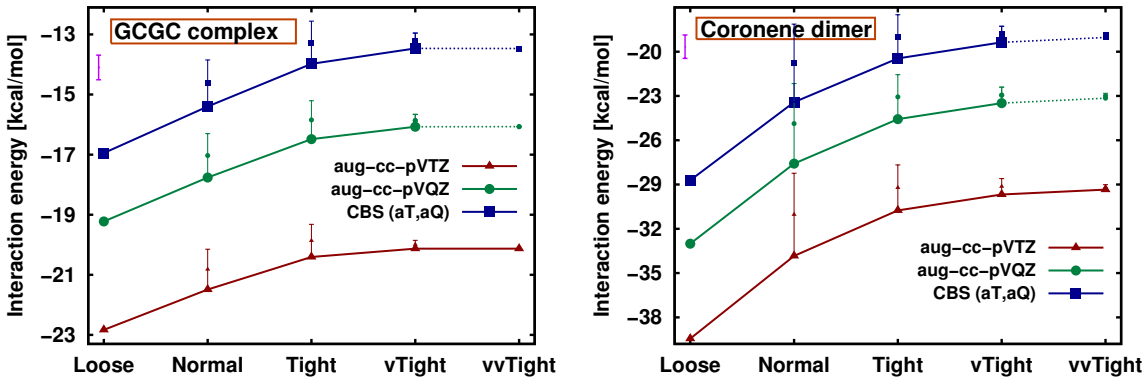

Figure 9: Convergence of LNO-CCSD(T) interaction energies with systematically improving both the local approximations and the AO basis sets for the guanine–cytosine (left) and coronene (right) dimers. Available DLPNO-CCSD(T)-F12/cc-pVDZ-F12[36] and PNO-CCSD(T)-F12/aug'-cc-pVTZ[41] results obtained with both default and tighter threshold settings are indicated by the purple bars in the top left corners of the plots. See Sect. 6 and the caption of Fig. 6 for further explanation.

Analogously to the case of the S66 benchmarks, we find for the GCGC and C2C2PD dimers that the Loose setting leads to large errors, even the Normal setting is about 1.3–1.9 and 4.0 kcal/mol away from the vTight ones, for the GC and coronene dimers, respectively.

Nevertheless, the Tight values and the Normal–Tight based estimates are acceptable even for such extreme cases: those deviate from the vTight results by about 0.3–0.5 and 0.2 kcal/mol for the GC dimer. The deviations of the Tight and vTight results are about 1.1 kcal/mol for the coronene dimer, but the Normal–Tight estimates are again within about 0.4 kcal/mol of the vTight results. Considering the slower convergence of these interaction energies we invested extensive resources to verify the quality of the vTight reference by performing even more accurate calculations. The "vvTight" label on the horizontal axes of the two plots of Fig. 9 refers to the configuration one step better than the vTight one following the exponential rule for the improvement of the individual thresholds ($\varepsilon_{\mathrm{o}} = 3\times10^{-7}$, $\varepsilon_{\mathrm{v}} = 3\times10^{-8}$, $\varepsilon_{\mathrm{w}} = 3\times10^{-7}$ E_{h}, and the vTight setting for the remaining thresholds). Such vvTight calculations were only performed with the aTZ basis set, the dashed lines and smaller symbol sizes for the aQZ and (aT,aQ)Z curves represent that the vTight basis set corrections were added to the vvTight aTZ results. For GCGC the vvTight LNO-CCSD(T) result is almost identical to the vTight one. This does not mean exact convergence, individual contributions, such as $\Delta E_{i'}^{\mathrm{MP2}}$ of Eq. (5), still change on the tenth of a kcal/mol scale. For the coronene dimer the difference of vvTight and vTight is not negligible but is convincingly small, 0.3 kcal/mol or about 1%.

Testing LMP2 against DF-MP2 reveals again the accuracy of the pair and domain approximations. For both dimers Normal LMP2 is accurate to about 1 kcal/mol, while the deviation drops below 0.5 kcal/mol and 0.2 kcal/mol with the Tight and vTight settings, respectively, as shown in Table S12 of the SI. The vvTight LMP2 energies are basically exact, they deviate from DF-MP2 by 0.003 and 0.06 kcal/mol for GCGC and C2C2PD, respectively.

As these two dimers are elements of the popular L7 set,[93] a larger number of high-level correlated calculations are available for comparison. For the GCGC dimer the recent TightPNO (-14.51 kcal/mol) and VeryTightPNO (-13.69 kcal/mol) DLPNO-CCSD(T)-F12/DZ-F12 results of Pavošević *et al.*[36] are shown in Fig. 9, the latter of which agrees

very closely with our Tight (13.98 kcal/mol) and vTight (13.47 kcal/mol) results. The most recent local CCSD(T) coronene dimer interaction energy estimates are collected in Table 5. LNO-CCSD(T)/(aT,aQ)Z and PNO-CCSD(T*)-F12/a'TZ results are available both with and without the CP correction. The complexity of this example is immediately apparent from the fact that there is a difference of about 1.1 (1.4) kcal/mol between both our Tight and vTight as well as the default and tight PNO-CCSD(T*)-F12 results obtained without (with) CP correction. The increment between the TightPNO and VeryTightPNO DLPNO-CCSD(T)-F12/DZ-F12 results is also sizable, 2.2 kcal/mol. Besides this the noticeable difference between the uncorrected and the CP-corrected results for both LNO-CCSD(T)/(aT,aQ)Z and PNO-CCSD(T*)-F12/a'TZ indicates that neither the F12 methods with DZ-F12 and a'TZ bases nor the plain (aT,aQ)Z extrapolation yield fully converged results regarding the basis set completeness. For the same reason the original QCISD(T) reference of the L7 set is probably too overbound. Indeed, our preliminary calculations indicate much better basis set convergence at the LNO-CCSD(T)/(aQ,a5)Z level, the difference of about 1.6 kcal/mol between the CP-corrected and uncorrected (aT,aQ)Z interaction energies decreases to about 0.4 kcal/mol with (aQ,a5)Z extrapolation. Further analysis of this issue will be in the scope of a forthcoming study.[74] Let us also add that the local errors, at least with the vTight setting, seem to be smaller than the basis set error at the (aT,aQ)Z level as indicated by the vvTight results. Since only vvTight LNO-CCSD(T)/aTZ is available, the vvTight (aT,aQ)Z can only be estimated by adding a BSIE correction at the vTight level, leading to the -19.04 (-20.73) kcal/mol estimate for the uncorrected (CP-corrected) vvTight LNO-CCSD(T)/(aT,aQ)Z.

7 Illustrative examples and computational requirements

The last two sections assessed the accuracy of LNO-CCSD(T) for a significant number of small- and medium-sized molecules (Sect. 5) and for a handful of especially tough larger

Table 5: CCSD(T) interaction energy estimates for the coronene dimer in kcal/mol. See Sect. 6.2 for explanation.

Method	Basis		
LNO-CCSD(T)	CBS (aT,aQ)Z	Tight	vTight
		-20.46	-19.37
LNO-CCSD(T)	CBS (aT,aQ)Z (CP)	Tight	vTight
		-22.47	-21.01
PNO-CCSD(T*)-F12[41]	a'TZ	default	domopt=tight
		-19.09	-20.46
PNO-CCSD(T*)-F12[41]	a'TZ (CP)	default	domopt=tight
		-18.87	-19.97
DLPNO-CCSD(T)-F12[36]	DZ-F12	TightPNO	VeryTightPNO
		-21.31	-19.14
QCISD(T)/a"DZ+ΔMP2/CBS (D,T)Z[93]		-24.36	

molecules (Sect. 6). In this section the performance of LNO-CCSD(T) is characterized on the realistic example of an organocatalysis reaction from the perspective of the practical applications of the method. Additionally, the results of wall-clock time measurements and the hardware requirements are also analyzed.

7.1 Organocatalysis reaction

We selected an organocatalytic Michael-addition in which the reaction of propanal and β-nitrostyrene is facilitated by a diphenylprolinol silyl ether catalyst and a *p*-nitrophenol (pnp) cocatalyst (Fig. 2). A similar model reaction for the Michael-addition was already investigated with a less optimized, two-year-old version of our LNO-CCSD(T) program.[64] With the current implementation LNO-CCSD(T)/a(T,Q)Z can be routinely obtained for the largest species along the reaction path, that is, the transition state of the carbon–carbon bond formation reaction step (TS_{CC}) leading to the RS stereoisomer ($TS_{CC}^{RS} \cdots$ pnp, 90 atoms). With the current version vTight LNO-CCSD(T)/a(T,Q)Z calculations are also feasible, thus the convergence with the composite local threshold can be analyzed as above. A Gibbs free energy profile of the overall reaction mechanism was obtained by our coworkers by combining kinetic measurements, kinetic simulations, and Gibbs free energy calculations, which is in

accord with the experimental observations available in the literature.[64] Consequently, comparison can be made with these reaction Gibbs free energies (ΔG) labeled by "experiment & kinetics & DFT" in Fig. 10. The ΔG values of Fig. 10 were evaluated by replacing the ωB97X-D/6-311++G(3df,3pd) electronic energies of Ref. 64 with LNO-CCSD(T) and keeping the original thermal, entropic, and solvent corrections of Ref. 64.

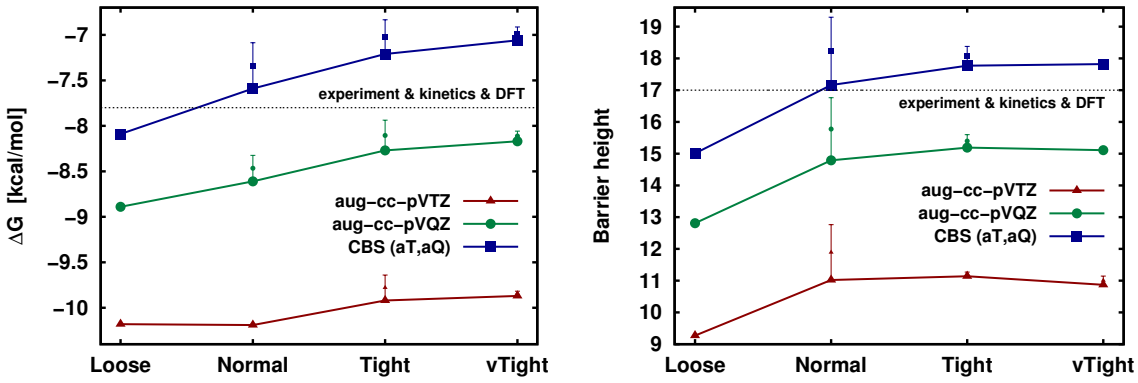

Figure 10: Convergence of LNO-CCSD(T) results for the ΔG and the barrier height of the Michael-reaction (see Fig. 2). See Sect. 7.1 and the caption of Fig. 6 for explanation.

The reactants and the product are of smaller size including at most 28 atoms. Correspondingly, the convergence of the Gibbs free energy of reaction with the local approximation is rapid (left panel of Fig. 10), the Normal (Tight) results are already within 0.5 (0.15) kcal/mol of the vTight values. The error estimates turn out to be useful again, for instance, vTight-quality results can be estimated utilizing the Normal and Tight energies. The computation of the barrier height is proven to be much more challenging. Since $\mathrm{TS}_{\mathrm{CC}}^{\mathrm{RS}} \cdots$pnp is the complex of three species (β-nitrostyrene, p-nitrophenol, and the enamine product of the condensation reaction between propanal and the main catalyst) the barrier height measured from the infinitely separate reactants and the complex of the two catalysts consists of the condensation reaction energy and non-covalent interaction energies of significant size. The latter are an important driving force of the stereochemistry of the overall reaction mechanism, thus their accurate description is crucial.[64] To illustrate the importance of accurate correlation energies the total LNO-CCSD(T)/a(T,Q)Z electronic energy difference of -4.1 kcal/mol can

be decomposed to the HF and correlation contributions of 39.2 kcal/mol and -43.3 kcal/mol, respectively. Moreover, the (T) contribution alone is also surprisingly significant being about -7.8 kcal/mol (18 %) out of the -43.3 kcal/mol LNO-CCSD(T)/a(T,Q)Z correlation energy contribution. In other words, not only MP2 [being about 12.7 kcal/mol below CCSD(T)] but CCSD in itself is also insufficient for the realistic description of this particular barrier height.

Returning to the analysis of Fig. 10, the convergence of the barrier height with the local approximations is again very satisfactory: the Normal (Tight) energies are less than 0.7 (0.3) kcal/mol apart from the vTight ones with all three basis choices. It is also apparent that for both the reaction energy and the barrier height the CCSD(T) energies have to be extrapolated to the CBS limit using large basis sets in order to get realistic results within chemical accuracy. For instance, the ΔG based on the ωB97X-D/6-311++G(3df,3pd) electronic energy is 15.9 kcal/mol for the barrier height, which is at least as good as or perhaps even better than the vTight LNO-CCSD(T)/aQZ result. In other words, a balanced choice of sufficiently large basis sets and local approximations is needed for LNO-CCSD(T) to provide more useful information than a well-selected density functional approximation.

We also computed the same reaction energy and barrier height using the DLPNO-CCSD(T)/aTZ scheme to obtain additional information from a separate source. Reaction Gibbs free energies computed with the default and tighter settings of the two schemes agree convincingly both with each other and with the -9.87 kcal/mol vTight LNO-CCSD(T)/aTZ result. In more detail, using the respective loose, normal, and tight settings the DLPNO-CCSD(T)/aTZ results are -10.35, -9.94, and -9.68 kcal/mol, while LNO-CCSD(T)/aTZ yields -10.18, -10.19, and -9.92 kcal/mol. The deviation is more pronounced for the barrier height: compared to the vTight LNO-CCSD(T)/aTZ result of 10.9 kcal/mol NormalPNO (LoosePNO) DLPNO-CCSD(T)/aTZ yields 12.3 (14.6) kcal/mol, while the Loose, Normal, and Tight LNO-CCSD(T)/aTZ barrier heights are found to be 9.3, 11.0, and 11.1 kcal/mol, respectively. Unfortunately, it was not possible to perform TightPNO DLPNO-

CCSD(T)/aTZ or any DLPNO-CCSD(T)/aQZ calculation with our hardware, which could probably provide valuable information about the source of this larger difference between the two barrier heights.

7.2 Scaling and performance for large molecules

Wall-clock times of this section were measured using a single, five-year-old, 6-core processor, thus these timings can be illustrative also for those users who do not necessarily wish to employ the latest hardware for local CC calculations. Besides the examples presented in this section, additional wall-time measurements are presented for all entries of the CEMS26 set in Table 11 of Ref. 27, which were obtained with a less optimized, previous LNO-CCSD(T) version.

Detailed timings are given in Table 6 separately for the HF, LMP2, two-external integral transformation, and the LNO-CCSD(T) calculation performed in the LIS. Using the DF approximation makes the HF calculation affordable for systems below a few hundred atoms even if a quartic-scaling algorithm is employed. Looking at the example of the $TS_{CC}^{RS} \cdots pnp$ system first, DF-HF took only 0.27 and 0.86 days with the aug-cc-pVTZ and aug-cc-pVQZ bases, containing 3155 and 5742 AOs, respectively. Our current Boys localization[96] algorithm is cubic-scaling, but the computation of the LMOs only takes seconds for this system. Besides the very fast PAO construction and pair-energy computation steps all the remaining operations in our LMP2 and LNO-CCSD(T) codes are asymptotically linear-scaling. Looking at the dimensions of the EDs in Table 6, it is obvious that we are far from that asymptotic limit for the $TS_{CC}^{RS} \cdots pnp$ system. Most of, and in particular cases all the atoms are included in the EDs, because of the relatively extended nature of the diffuse functions in the selected basis sets. Consequently, in average more than a third (half) of the LMOs (PAOs) are put into the EDs already with the Normal settings, while more than half of the LMOs and almost all PAOs are needed with the Tight threshold.

The LMP2 timings of Table 6 are very competitive being only about 1-2 (2-4) times longer

Table 6: Average (maximum) domain sizes, orbital space dimensions, DF-HF and correlation energies, wall-clock times in days (lower panel, obtained with a 6-core CPU), and memory requirements for LNO-CCSD(T) computations for large molecules.

Molecule	$TS^{RS}_{CC}\cdots$pnp			$TS^{RS}_{CC}\cdots$pnp		LTP	
No. of atoms	90			90		1023	
Basis	aug-cc-pVTZ			aug-cc-pVQZ		def2-TZVP	def2-QZVP
Total AOs	3155			5742		19067	44712
Total AFs	7001			11362		47093	101297
Total LMOs	123			123		1340	
Thresholds	Normal	Tight	vTight	Normal	Tight	Normal	Normal
Strong pairs [%]	39	53	67	39	53	3.6	3.7
Atoms in ED	78 (90)	84 (90)	89 (90)	76 (90)	83 (90)	123 (268)	121 (267)
LMOs in ED	49 (88)	66 (105)	83 (112)	49 (88)	66 (105)	53 (89)	54 (91)
PAOs in ED	1678 (2645)	2046 (2832)	2396 (2897)	2071 (4107)	3592 (5157)	889 (1594)	1984 (3793)
Occupied LNOs	29 (62)	42 (81)	56 (92)	29 (62)	42 (81)	28 (51)	29 (50)
Virtual LNOs	134 (273)	203 (413)	284 (563)	154 (300)	238 (465)	108 (202)	137 (241)
E^{DF-HF} [E_h]		-2328.083841		-2328.207262		-28235.5178	-28236.9430
$E^{LNO-CCSD(T)}$ [E_h]	-8.9503	-8.9515	-8.9524	-9.3412	-9.3358	-101.4099	-108.8656
HF[a]+localization	0.27			0.86		10.4	38.4
LMP2	0.38	0.62	1.4	1.6	3.4	1.1	6.9
Integral trf.	0.55	1.3	3.3	2.2	6.1	0.91	7.4
LNO-CCSD(T)	0.62	5.6	35.1	0.86	10.4	2.2	3.9
Total LNO-CCSD(T)	1.6	7.5	39.8	4.6	19.9	4.2	18.1
Memory [GB][b]	7 (26)	19 (36)	22 (43)	11 (67)	25 (74)	20 (20)	98 (98)

[a] Canonical DF-HF for $TS^{RS}_{CC}\cdots$pnp, DF-HF with local fitting domains employed for the exchange part [23] for LTP. The initial guess for the HF calculation was the density obtained with the corresponding basis set with a cardinal number of lower by 1, e.g., the aug-cc-pVQZ calculation was started from the aug-cc-pVTZ density, etc. [b] Minimal (optimal) memory requirement

than the DF-HF ones when the Normal (Tight) settings are used. Compared to LMP2 the integral transformation step takes similar, at most twice as much effort. The LNO-CCSD(T) part is just as manageable, at least with the Normal settings. This is explained by the fact that the LNO approximation compresses the MO space of the ED very efficiently, e.g., the number of virtual orbitals is cut on the average by about an order of magnitude for such extended systems. Switching to the Tight settings is not much more demanding concerning the operations in the ED, because the EDs formed with the Normal thresholds were already almost as large as the entire system. However, the joint effects of the Tight thresholds, the use of diffuse functions, and the numerous long-range intermolecular interactions lead to large LIS dimensions containing in average (up to) 245–280 (494–546) LNOs. Due to the sixth-power scaling of the CCSD(T) calculations in the LISs with the number of LNOs, the CCSD(T) calculations performed in the LISs of $TS^{RS}_{CC}\cdots$pnp were about 9–12 times more expensive with the Tight setting than with the Normal one. Consequently, the runtimes for "Total LNO-CCSD(T)", i.e., the entire correlation energy calculation are about 4.5 times

longer with the Tight settings.

For the sake of completeness Table 6 also collects vTight data. However, we would like to highlight that, if performed at all, vTight calculations should almost always serve the purpose of assessing the local error in a small number of convergence tests, because the Tight or Normal-Tight calculations yield highly satisfactory accuracy except for the most extreme cases (e.g., coronene dimer). In brief, vTight settings in combination with diffuse basis sets result in the inclusion of a very high percentage of atoms and MOs into the EDs in this example. This would not be the case for larger systems or without diffuse functions. Nevertheless, the LNO approximations still compress the MO spaces effectively, the average LNO dimensions are about only twice as large as the ones obtained with the Normal settings. Consequently, the cost of the LNO-CCSD(T) calculation in the LISs increases by about a factor of 56 (almost exactly 2^6) compared to the Normal settings, which is costly but manageable even with this moderate processor. The accuracy of the Normal LNO-CCSD(T) is again apparent since its difference with respect to the vTight result is only 0.02%.

To better illustrate the trends in the computational costs with varying basis and threshold sets the "Total LNO-CCSD(T)" wall times are also depicted for $TS^{RS}_{CC}\cdots$pnp in Fig. 11. It is apparent from the plot, that, even if the HF is performed with an efficient DF implementation, the Loose LNO-CCSD(T) calculations are always feasible, taking less than twice as much time as DF-HF. Compared to Loose (DF-HF) the Normal LNO-CCSD(T) requires about 3 (6) times more effort for both basis sets. The Tight option is also useful because, at about 3–5 times the cost of Normal, the convergence of local approximations can be tested and the local errors can be estimated for a group of chemically similar systems. The most expensive, larger basis set Tight calculations can also be circumvented by adding cost-effective BSIE corrections to the lower-basis Tight LNO-CCSD(T) results.[74]

It is also interesting to look at the corresponding timings obtained by other local CCSD(T) methods. Compared to the 37.5 h (11.9 h) runtime of Normal (Loose) LNO-CCSD(T), using the same hardware and 6 parallel threads, the LoosePNO DLPNO-CCSD(T)/aTZ

calculation took 21 h for $TS^{RS}_{CC}\cdots$pnp, while the NormalPNO DLPNO-CCSD(T)/aTZ run required 19.6 days, which is however comparable to the wall time (19.9 days) of the Tight LNO-CCSD(T)/aQZ calculation.

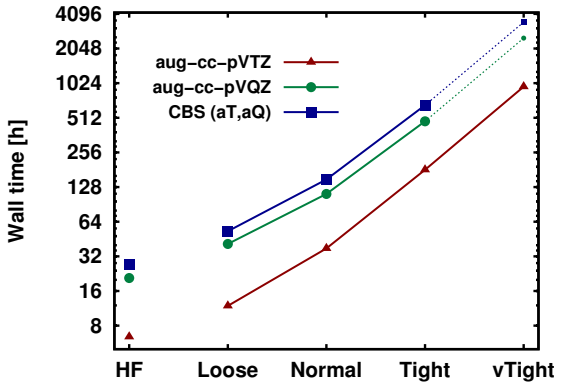

Figure 11: Wall-clock times (using a 6-core CPU) required for the HF and LNO-CCSD(T) calculations for $TS^{RS}_{CC}\cdots$pnp containing 90 atoms and 3155/5742 AOs with the aTZ/aQZ basis sets. The vTight LNO-CCSD(T)/aug-cc-pVQZ calculations were performed on a different hardware, the values with smaller symbol size and dashed line display the estimated wall-times for the same 6-core CPU.

Let us note that the wall times in Table 6 were determined for HF and LNO-CCSD(T) calculations exploiting the efficient DF approach. Thus the conditions for which our conclusions hold differ from the conditions under which the calculations of Ref. 34 were carried out, where HF and DFT wall times were compared to the times measured for DLPNO-CCSD(T_0) calculations. To the best of our knowledge the HF and DFT wall times were measured without the DF approach in Ref. 34, where NormalPNO DLPNO-CCSD(T_0) was found to be more accurate than the investigated density functional approximations (DFAs) at about twice as much cost for test sets containing small systems of at most 30–40 atoms. Obviously, HF (or DFT methods including HF exchange) without the DF approach, especially with aTZ or aQZ basis sets, would be rather costly for systems of increasing size, and it would not be possible to perform HF/aQZ for $TS^{RS}_{CC}\cdots$pnp, at least on our modest hardware. Thus we also find it important to compare our wall times with those for HF (or DFAs) accelerated by DF, in which case the Normal (Loose) LNO-CCSD(T) turned out to be less

than 6 (2) times more expensive than DF-HF even with the aTZ and aQZ basis sets. In other words, the computational expenses of LNO-CCSD(T) or other local CCSD(T) schemes are generally not yet as low as those of efficient DF-based HF/DFA implementations, especially if the faster convergence of DFAs with the AO basis set is taken into account. However, our comparison shows that, even for systems of about 100 atoms, the CCSD(T)/CBS limit can be closely approached at about an order of magnitude higher costs than those required for DFT calculations.

In turn, considering the long-range nature of the Coulomb interaction it is more complicated to break down the scaling of HF than that of the electron correlation methods. Because of that it is challenging to accelerate DF-HF in the region of 100–1000 atoms if the system is three-dimensional. On the other hand, the linear-scaling region of LNO-CCSD(T) is reached for systems of around 500 atoms, at least this is suggested by our experience gained for 3D protein systems.[27] This trend is also followed by LTP: using both triple- and quadruple-ζ basis sets the average (maximum) domain sizes are below 123 (268) atoms (see Table 6). Thus, the size of the EDs becomes independent of the size of the entire molecule, and consequently the computational cost becomes proportional to the number of occupied MOs above a certain system size. Even if local approximations are employed also for the construction of the domains of fitting functions at the evaluation of the HF exchange, the crossover of the compute times required by our third-power scaling DF-HF implementation[23] and the linear-scaling LNO-CCSD(T) is inevitable. This crossover point should be around a few hundred atoms, because DF-HF (with local DF) already takes more than twice as much time as Normal LNO-CCSD(T) for the 1023-atom LTP with both the def2-TZVP and the def2-QZVP bases. It is also worth noting that the acceleration of the HF exchange calculations with large, diffuse basis sets received relatively small attention in the literature for systems in the thousand atom region. Perhaps alternative basis sets other than Gaussian bases, such as multiresolution[97,98] or projected augmented wave[99] based unconventional approaches could offer more operation count economic solutions to approach the basis set limit

of HF in the near future.

Continuing with the discussion of Table 6, in spite of the larger ED sizes of LTP compared to $TS^{RS}_{CC}\cdots pnp$ and the additional diffuse functions present for $TS^{RS}_{CC}\cdots pnp$ the number of strong pair LMOs and occupied LNOs are almost identical for the two systems with the Normal setting because of their similar, organic chemical composition and 3D structure. The PAO/virtual LNO dimensions are, of course, somewhat, about 10–30 % larger with the larger and more diffuse Dunning-type basis sets. The case of LTP is far past the crossover point, where the LMP2 calculation becomes extremely efficient taking less than the time of two iterations of the HF wave function optimization. An additional computation time of about 2–3 times as long as the LMP2 calculation is required for Normal LNO-CCSD(T). As shown in Table 6, for the 1023 atoms of LTP the HF and LNO-CCSD(T) calculations took, respectively, 10.4 and 4.2 days with the def2-TZVP basis and 38.4 and 18.1 days with the def2-QZVP basis. The latter 4.2 and 18.1 days contain all steps besides the HF and Boys localization, including the respective LMP2 parts taking only 1.1 and 6.9 days with the two basis sets.

One of the most advantageous features of our algorithms is their especially low memory and disk space requirements, at least in comparison to other CC implementations. Currently, both our DF-HF implementation and the local correlation algorithm up to the pair energy evaluation require the storage of 6 n_{AO}^2-sized arrays, where n_{AO} is the number of AOs. In this respect the formal scaling of the memory requirement of our LMP2 and LNO-CCSD(T) implementation is quadratic, but because of its small prefactor the 6 arrays take up only about 89 GB of memory even for our largest example with $n_{AO} = 44712$. Following the pair energy evaluation, only two n_{AO}^2-sized arrays remain in memory, but their sparse storage would also be manageable if needed. On top of those two arrays, the size of all other ED- or LIS-specific quantities is asymptotically linear scaling. As a result of our recent optimization efforts, if there is sufficient memory available, almost all of these ED- and LIS-specific arrays are sorted in memory. Currently, hard disk use is limited to the storage of the DIIS vectors

appearing in the LNO-CCSD iteration of the LIS. The DIIS vectors are usually of negligible size with average LNO dimensions; e.g., the size of the 12 vectors is still below 64 GB for the largest LIS of the $TS_{CC}^{RS} \cdots$ pnp Tight LNO-CCSD(T)/aQZ calculation containing 81 (465) occupied (virtual) LNOs.

Considerable efforts were devoted to minimize the memory use of the ED and LIS calculations, and we also implemented alternative algorithms which can utilize disk if memory bottlenecks emerged for the largest domains of extended systems computed with tighter settings and large AO basis sets. The last row of Table 6 shows the minimum memory requirement of LNO-CCSD(T) if the disk-based algorithms are executed, whereas the values in parenthesis denote the memory usage of the in-core algorithms. Clearly, even the largest calculations require easily manageable storage resources, e.g., the $TS_{CC}^{RS} \cdots$ pnp Tight LNO-CCSD(T)/aQZ calculation can be performed using about 25 GB memory, and the optimal algorithms requiring the lowest I/O can be invoked with about 74 GB memory allocation. We note that the out-of-core algorithms require the I/O of about twice as much data as the difference of the memory requirement presented in Table 6 with and without the parenthesis, which is in the worst case still in the 100 GB range. The additional benefit of our current OpenMP parallelization strategy is that the largest arrays are shared among the threads, thus the minimal memory requirement only marginally grows with the number of OpenMP threads. The LTP/def2-QZVP example is somewhat an outlier, because the 6 n_{AO}^2-sized array needed for DF-HF and the pair energy calculation overgrow the memory need of the ED/LIS calculations, but their size still remains below 100 GB.

We also found these memory demands quite representative for systems of similar size. For instance, the memory need for the noticeably more complicated AuAmin molecule (92 atoms) is only about 10 (20) % larger than that of $TS_{CC}^{RS} \cdots$ pnp with the aTZ (aQZ) basis set for both the Normal and the Tight settings. In comparison, the default PNO-CCSD(T)/cc-pVTZ-F12 calculation of Ref. 41 performed for the same AuAmin system with 120 cores of 6 nodes took about 5 GB/core node-specific memory, an additional shared memory of about 300 GB

distributed among the nodes, and about 280 GB of disk space distributed over the local disks of the 6 nodes. The total storage use of the default PNO-CCSD(T)/cc-pVTZ-F12 calculation being roughly 1 TB was thus over an order of magnitude larger than that of the Tight LNO-CCSD(T)/aQZ calculation. In turn, the runtime of the default PNO-CCSD(T)/cc-pVTZ-F12 calculation in Ref. 41 is about 0.27 days on 120 cores. The corresponding wall time measurement for the Normal (Tight) LNO-CCSD(T)/aQZ calculation performed on the 6 cores of our two-generation-older processor is about 6 (15) days. Since the conditions of these computations could not be more different, more comparable measurements would be desirable to make more realistic comparison of the performance of the various local CCSD(T) implementations.

8 Summary and conclusions

We have presented a wide range of chemically relevant benchmark examples to illustrate the accuracy and efficiency of the LNO-CCSD(T) method[23,24,27] and introduced composite threshold sets to facilitate black box like convergence studies. The Loose, Normal, Tight, and vTight hierarchy largely retains the systematically convergent property of the individual local approximations. We have performed about 3400 benchmark calculations to assess the accuracy in comparison to exact CCSD(T) references. The default (Normal) settings of LNO-CCSD(T) provide excellent correlation energies: the mean absolute error (MAE) for the NWH and S66 test sets with the (aug-)cc-pVTZ bases is below 0.03%, while with cc-pVQZ or for the CEMS26 set containing larger and more complicated examples the MAE values are still below 0.07%. The Tight settings ensure that correlation energies are more accurate than 99.9% for every tested case. The energy differences are also highly reliable: the MAEs of the default LNO-CCSD(T) reaction energies and interaction energies for the moderate-sized test systems are below 0.18 kcal/mol even if the results are extrapolated to the basis set limit, while the MAE is still only 0.34 kcal/mol for the more challenging CEMS26

set. The Tight settings yield maximum errors below 0.46 kcal/mol for all the tested cases. Using the same test systems and basis sets Normal LNO-CCSD(T) is found more accurate than the DLPNO-CCSD(T_0)[35] method even with its TightPNO settings. The NWH and S66 energy differences of TightPNO DLPNO-CCSD(T) (including also an iterative triples treatment[37]) and Normal LNO-CCSD(T) are almost identical on the average, the MAE of Normal LNO-CCSD(T) for the CEMS26 set is still better by about 44%.

The sizable BSIE of CCSD(T) even with medium-sized basis sets is well-known at least for small molecules, and we have shown that the issue is even more pronounced for larger systems containing dozens of atoms. The deviation of local CCSD(T) compared to CCSD(T)/CBS is often dominated by the BSIE even with (a)TZ and (a)QZ bases. Compared to this BSIE the local errors of LNO-CCSD(T) are negligible already with the Normal settings. The extrapolation using the aTZ and aQZ bases reliably approaches the CBS limit of CCSD(T). Since the improved algorithm handles large and diffuse basis sets efficiently and can overcome the numerical complications appearing in severely near-linear dependent AO basis sets, CBS(T,Q) extrapolations of LNO-CCSD(T) energies become feasible for systems of up to about 1000 atoms, which is currently a unique capability of our implementation.

In order to establish the accuracy limits of our method we have shown that benchmark quality LNO-CCSD(T)/CBS(aTZ,aQZ) energies can be obtained with extremely tight settings for up to 92 atoms even for some of the most challenging systems studied in the context of local correlation to date. When canonical CCSD(T) is not available as reference, the Normal, Tight, etc. hierarchy of truncation thresholds provides a systematically improvable way both to estimate the remaining local error and to decrease it well below chemical accuracy. Relying on both the AO basis set and threshold hierarchies the CBS limit of CCSD(T) can be systematically approached, and a well estimated error bar can be assigned to the LNO-CCSD(T) results. We recommend to perform such convergence tests and error estimation at least for a small number of representative examples before large scale production runs. Our Tight LNO-CCSD(T)/CBS(aTZ,aQZ) energies also agree with most recent, tight PNO-

CCSD(T)-F12[41] results within 1-2 kJ/mol, which difference is comparable to the cumulative uncertainties originating from the remaining BSIE and local errors of the methods. The parallel decrease of the local errors with improving threshold sets in various AO bases indicates the generally balanced nature of our approximations and can be exploited to design cost-effective composite methods, such as Tight LNO-CCSD(T)/aTZ augmented with Normal LNO-CCSD(T)/CBS(aTZ,aQZ) BSIE correction.[74]

On the example of an organocatalytic Michael-reaction (including 90 atoms) we have demonstrated that it is fairly simple to perform well converged LNO-CCSD(T)/CBS(aT,aQ)Z calculations for realistic, three-dimensional systems at only about an order of magnitude higher cost than that of an efficient DF-HF calculation. Consequently, the performance of LNO-CCSD(T) is expected to be superior to DFAs above rung three implemented without DF, but more importantly the LNO-CCSD(T) electronic energy calculations would not be rate limiting in the frequently employed thermochemistry protocols including geometry optimizations and harmonic frequency evaluations performed using DFAs with HF exchange content. The presented LNO-CCSD(T) calculations were performed with a few tens of GBs of memory and a modest, 6-core CPU which has a performance comparable to a current mid-range laptop. Due to reaching the linear scaling regime at around a few hundred atoms even for three dimensional systems, such as proteins, LNO-CCSD(T)/CBS(T,Q) can even be more efficient than the preceding DF-HF calculations and can be performed for systems including more than 1000 atoms with an easily accessible hardware requirement.

Acknowledgement

Inspiring discussions with Imre Pápai, Tamás Földes, and Andrea Hamza (Hungarian Academy of Sciences) regarding the Michael-addition reaction are thankfully acknowledged. The authors are grateful for the financial support from the National Research, Development, and Innovation Office (NKFIH, Grant No. KKP126451). This work was also supported by the BME-Biotechnology FIKP grant of EMMI (BME FIKP-BIO). The work of PRN is supported

by the ÚNKP-18-4-BME-257 New National Excellence Program of the Ministry of Human Capacities and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. The computing time granted on the Hungarian HPC Infrastructure at NIIF Institute, Hungary and on the BME HPC Cluster (project number: TÁMOP - 4.2.2.B-10/1-2010-0009) is gratefully acknowledged.

Supporting Information Available

See supporting information for the structure of LTP (xyz), statistical measures for the DLPNO-CCSD(T_0) and DLPNO-CCSD(T) correlation, reaction, and interaction energy deviations, further discussion of LNO-CCSD, DLPNO-CCSD(T_0), DLPNO-CCSD(T), and PNO-CCSD(T)-F12 results available for the NWH, S66, and CEMS26 sets, and reaction and interaction energies of the convergence plots of Figs. 7-10 (pdf).

This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Bartlett, R. J.; Musiał, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291.

(2) Kállay, M.; Surján, P. R. Higher excitations in coupled-cluster theory. *J. Chem. Phys.* **2001**, *115*, 2945.

(3) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479.

(4) Deegan, M. J. O.; Knowles, P. J. Perturbative corrections to account for triple excitations in closed and open shell coupled cluster theories. *Chem. Phys. Lett.* **1994**, *227*, 321.

(5) Anisimov, V. M.; Bauer, G. H.; Chadalavada, K.; Olson, R. M.; Glenski, J. W.; Kramer, W. T. C.; Aprà, E.; Kowalski, K. Optimization of the Coupled Cluster Implementation in NWChem on Petascale Parallel Architectures. *J. Chem. Theory Comput.* **2014**, *10*, 4307–4316.

(6) DePrince, A. E.; Sherrill, C. D. Accuracy and Efficiency of Coupled-Cluster Theory Using Density Fitting/Cholesky Decomposition, Frozen Natural Orbitals, and a t1-Transformed Hamiltonian. *J. Chem. Theory Comput.* **2013**, *9*, 2687.

(7) Epifanovsky, E.; Zuev, D.; Feng, X.; Khistyaev, K.; Shao, Y.; Krylov, A. I. General implementation of the resolution-of-the-identity and Cholesky representations of electron repulsion integrals within coupled-cluster and equation-of-motion methods: Theory and benchmarks. *J. Chem. Phys.* **2013**, *139*, 134105.

(8) Janowski, T.; Pulay, P. Efficient Parallel Implementation of the CCSD External Exchange Operator and the Perturbative Triples (T) Energy Calculation. *J. Chem. Theory Comput.* **2008**, *4*, 1585–1592.

(9) Pitoňák, M.; Aquilante, F.; Hobza, P.; Neogrády, P.; Noga, J.; Urban, M. Parallelized implementation of the CCSD(T) method in MOLCAS using optimized virtual orbitals space and Cholesky decomposed two-electron integrals. *Collect. Czech. Chem. Commun.* **2011**, *76*, 713–742.

(10) Deumens, E.; Lotrich, V. F.; Perera, A.; Ponton, M. J.; Sanders, B. A.; Bartlett, R. J. Software design of ACES III with the super instruction architecture. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 895–901.

(11) Peng, C.; Calvin, J. A.; Valeev, E. F. Coupled-cluster singles, doubles and perturbative triples with density fitting approximation for massively parallel heterogeneous platforms. *Int. J. Quantum Chem.* **2019**, *119*, e25894.

(12) Kaliman, I. A.; Krylov, A. I. New algorithm for tensor contractions on multi-core CPUs, GPUs, and accelerators enables CCSD and EOM-CCSD calculations with over 1000 basis functions on a single compute node. *J. Comput. Chem.* **2017**, *38*, 842–853.

(13) Asadchev, A.; Gordon, M. S. Fast and Flexible Coupled Cluster Implementation. *J. Chem. Theory Comput.* **2013**, *9*, 3385–3392.

(14) Eriksen, J. J. Efficient and portable acceleration of quantum chemical many-body methods in mixed floating point precision using OpenACC compiler directives. *Mol. Phys.* **2017**, *115*, 2086.

(15) Dutta, A. K.; Neese, F.; Izsák, R. Accelerating the coupled-cluster singles and doubles method using the chain-of-sphere approximation. *Mol. Phys.* **2018**, *116*, 1428–1434.

(16) Gyevi-Nagy, L.; Kállay, M.; Nagy, P. R. **2019**, in preparation.

(17) Zaleśny, R.; Papadopoulos, M.; Mezey, P.; Leszczynski, J. *Linear-Scaling Techniques in Computational Chemistry and Physics: Methods and Applications*; Challenges and Advances in Computational Chemistry and Physics; Springer: Netherlands, 2011.

(18) Collins, M. A.; Bettens, R. P. A. Energy-Based Molecular Fragmentation Methods. *Chem. Rev.* **2015**, *115*, 5607.

(19) Raghavachari, K.; Saha, A. Accurate Composite and Fragment-Based Quantum Chemical Models for Large Molecules. *Chem. Rev.* **2015**, *115*, 5643–5677.

(20) Rolik, Z.; Kállay, M. A general-order local coupled-cluster method based on the cluster-in-molecule approach. *J. Chem. Phys.* **2011**, *135*, 104111.

(21) Rolik, Z.; Szegedy, L.; Ladjánszki, I.; Ladóczki, B.; Kállay, M. An efficient linear-scaling CCSD(T) method based on local natural orbitals. *J. Chem. Phys.* **2013**, *139*, 094105.

(22) Kállay, M. Linear-scaling implementation of the direct random-phase approximation. *J. Chem. Phys.* **2015**, *142*, 204105.

(23) Nagy, P. R.; Samu, G.; Kállay, M. An integral-direct linear-scaling second-order Møller–Plesset approach. *J. Chem. Theory Comput.* **2016**, *12*, 4897.

(24) 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.* **2017**, *146*, 214106.

(25) Hégely, B.; Nagy, P. R.; Ferenczy, G. G.; Kállay, M. Exact density functional and wave function embedding schemes based on orbital localization. *J. Chem. Phys.* **2016**, *145*, 064107.

(26) Hégely, B.; Nagy, P. R.; Kállay, M. Dual basis set approach for density functional and wave function embedding schemes. *J. Chem. Theory Comput.* **2018**, *14*, 4600.

(27) 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.* **2018**, *14*, 4193.

(28) Pulay, P. Localizability of dynamic electron correlation. *Chem. Phys. Lett.* **1983**, *100*, 151.

(29) Pulay, P.; Saebø, S. Orbital-invariant formulation and second-order gradient evaluation in Møller–Plesset perturbation theory. *Theor. Chem. Acc.* **1986**, *69*, 357.

(30) Saebø, S.; Pulay, P. Fourth-order Møller–Plesset perturbation theory in the local correlation treatment. I. Method. *J. Chem. Phys.* **1987**, *86*, 914.

(31) Saebø, S.; Pulay, P. Local configuration interaction: An efficient approach for larger molecules. *Chem. Phys. Lett.* **1985**, *113*, 13.

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

(33) Liakos, D. G.; Sparta, M.; Kesharwani, M. K.; Martin, J. M. L.; Neese, F. Exploring the Accuracy Limits of Local Pair Natural Orbital Coupled-Cluster Theory. *J. Chem. Theory Comput.* **2015**, *11*, 1525–1539.

(34) Liakos, D. G.; Neese, F. Is It Possible To Obtain Coupled Cluster Quality Energies at near Density Functional Theory Cost? Domain-Based Local Pair Natural Orbital Coupled Cluster vs Modern Density Functional Theory. *J. Chem. Theory Comput.* **2015**, *11*, 4054.

(35) Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. 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.* **2016**, *144*, 024109.

(36) Pavošević, F.; Peng, C.; Pinski, P.; Riplinger, C.; Neese, F.; Valeev, E. F. SparseMaps—A systematic infrastructure for reduced scaling electronic structure methods. V. Linear scaling explicitly correlated coupled-cluster method with pair natural orbitals. *J. Chem. Phys.* **2017**, *146*, 174108.

(37) Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D. G.; Minenkov, Y.; Cavallo, L.; Neese, F. 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.* **2018**, *148*, 011101.

(38) Schwilk, M.; Usvyat, D.; Werner, H.-J. Communication: Improved pair approximations in local coupled-cluster methods. *J. Chem. Phys.* **2015**, *142*, 121102.

(39) Schwilk, M.; Ma, Q.; Köppl, C.; Werner, H.-J. Scalable Electron Correlation Methods. 3. Efficient and Accurate Parallel Local Coupled Cluster with Pair Natural Orbitals (PNO-LCCSD). *J. Chem. Theory Comput.* **2017**, *13*, 3650–3675.

(40) Ma, Q.; Werner, H.-J. Scalable Electron Correlation Methods. 5. Parallel Perturbative Triples Correction for Explicitly Correlated Local Coupled Cluster with Pair Natural Orbitals. *J. Chem. Theory Comput.* **2018**, *14*, 198–215.

(41) Ma, Q.; Werner, H.-J. Explicitly correlated local coupled-cluster methods using pair natural orbitals. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, e1371.

(42) Ma, Q.; Werner, H.-J. Accurate Intermolecular Interaction Energies Using Explicitly Correlated Local Coupled Cluster Methods [PNO-LCCSD(T)-F12]. *J. Chem. Theory Comput.* **2019**, *15*, 1044.

(43) Schmitz, G.; Hattig, C.; Tew, D. P. Explicitly correlated PNO-MP2 and PNO-CCSD and their application to the S66 set and large molecular systems. *Phys. Chem. Chem. Phys.* **2014**, *16*, 22167–22178.

(44) Schmitz, G.; Hättig, C. Perturbative triples correction for local pair natural orbital based explicitly correlated CCSD(F12*) using Laplace transformation techniques. *J. Chem. Phys.* **2016**, *145*, 234107.

(45) Förner, W.; Ladik, J.; Otto, P.; Čížek, J. Coupled-cluster studies. II. The role of localization in correlation calculations on extended systems. *Chem. Phys.* **1985**, *97*, 251.

(46) Flocke, N.; Bartlett, R. J. A natural linear scaling coupled-cluster method. *J. Chem. Phys.* **2004**, *121*, 10935.

(47) Friedrich, J.; Dolg, M. Fully Automated Incremental Evaluation of MP2 and CCSD(T) Energies: Application to Water Clusters. *J. Chem. Theory Comput.* **2009**, *5*, 287.

(48) Fiedler, B.; Schmitz, G.; Hattig, C.; Friedrich, J. Combining Accuracy and Efficiency: An Incremental Focal-Point Method Based on Pair Natural Orbitals. *J. Chem. Theory Comput.* **2017**, *13*, 6023–6042.

(49) Kobayashi, M.; Nakai, H. Divide-and-conquer-based linear-scaling approach for traditional and renormalized coupled cluster methods with single, double, and noniterative triple excitations. *J. Chem. Phys.* **2009**, *131*, 114108.

(50) Nakano, M.; Yoshikawa, T.; Hirata, S.; Seino, J.; Nakai, H. Computerized implementation of higher-order electron-correlation methods and their linear-scaling divide-and-conquer extensions. *J. Comput. Chem.* **2017**, *38*, 2520–2527.

(51) Mochizuki, Y.; Yamashita, K.; Nakano, T.; Okiyama, Y.; Fukuzawa, K.; Taguchi, N.; Tanaka, S. Higher-order correlated calculations based on fragment molecular orbital scheme. *Theor. Chem. Acc.* **2011**, *130*, 515–530.

(52) Richard, R. M.; Lao, K. U.; Herbert, J. M. Aiming for Benchmark Accuracy with the Many-Body Expansion. *Acc. Chem. Res.* **2014**, *47*, 2828–2836.

(53) Yuan, D.; Li, Y.; Ni, Z.; Pulay, P.; Li, W.; Li, S. Benchmark Relative Energies for Large Water Clusters with the Generalized Energy-Based Fragmentation Method. *J. Chem. Theory Comput.* **2017**, *13*, 2696–2704.

(54) Eriksen, J. J.; Baudin, P.; Ettenhuber, P.; Kristensen, K.; Kjærgaard, T.; Jørgensen, P. Linear-Scaling Coupled Cluster with Perturbative Triple Excitations: The Divide–Expand–Consolidate CCSD(T) Model. *J. Chem. Theory Comput.* **2015**, *11*, 2984.

(55) Findlater, A. D.; Zahariev, F.; Gordon, M. S. Combined Fragment Molecular Orbital Cluster in Molecule Approach to Massively Parallel Electron Correlation Calculations for Large Systems. *J. Phys. Chem. A* **2015**, *119*, 3587.

(56) Li, W.; Piecuch, P.; Gour, J. R.; Li, S. Local correlation calculations using standard and renormalized coupled-cluster approaches. *J. Chem. Phys.* **2009**, *131*, 114109.

(57) Li, W.; Ni, Z.; Li, S. Cluster-in-molecule local correlation method for post-Hartree–Fock calculations of large systems. *Mol. Phys.* **2016**, *114*, 1447.

(58) Guo, Y.; Becker, U.; Neese, F. 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.* **2018**, *148*, 124117.

(59) Ni, Z.; Li, W.; Li, S. Fully optimized implementation of the cluster-in-molecule local correlation approach for electron correlation calculations of large systems. *J. Comput. Chem.* **2019**, *40*, 1130.

(60) Li, W.; Li, S. Divide-and-conquer local correlation approach to the correlation energy of large molecules. *J. Chem. Phys.* **2004**, *121*, 6649.

(61) Friedrich, J.; Walczak, K. Incremental CCSD(T)(F12)|MP2-F12—A Method to Obtain Highly Accurate CCSD(T) Energies for Large Molecules. *J. Chem. Theory Comput.* **2013**, *9*, 408.

(62) Kállay, M.; Gauss, J. Approximate treatment of higher excitations in coupled-cluster theory. *J. Chem. Phys.* **2005**, *123*, 214105.

(63) Hégely, B.; Bogár, F.; Ferenczy, G. G.; Kállay, M. A QM/MM program for calculations with frozen localized orbitals based on the Huzinaga equation. *Theor. Chem. Acc.* **2015**, *134*, 132.

(64) Földes, T.; Madarász, Á.; Révész, Á.; Dobi, Z.; Varga, S.; Hamza, A.; Nagy, P. R.; Pihko, P. M.; Pápai, I. Stereocontrol in Diphenylprolinol Silyl Ether Catalyzed Michael Additions: Steric Shielding or Curtin–Hammett Scenario? *J. Am. Chem. Soc.* **2017**, *139*, 17052.

(65) Bojtár, M.; Janzsó-Berend, P. Z.; Mester, D.; Hessz, D.; Kállay, M.; Kubinyi, M.; Bitter, I. An uracil-linked hydroxyflavone probe for the recognition of ATP. *Beilstein J. Org. Chem* **2018**, *14*, 747.

(66) Štejfa, V.; Bazyleva, A.; Fulem, M.; Rohlíček, J.; Skořepová, E.; Růžička, K.; Blokhin, A. V. Polymorphism and thermophysical properties of L- and DL-menthol. *J. Chem. Thermodyn.* **2019**, *131*, 524.

(67) Bakó, I.; Mayer, I.; Hamza, A.; Pusztai, L. Two- and three-body, and relaxation energy terms in water clusters: Application of the hierarchical BSSE corrected decomposition scheme. *J. Mol. Liq.* **2019**, *285*, 171.

(68) Paulechka, E.; Kazakov, A. Efficient Estimation of Formation Enthalpies for Closed-Shell Organic Compounds with Local Coupled-Cluster Methods. *J. Chem. Theory Comput.* **2018**, *14*, 5920.

(69) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set convergence of correlated calculations on water. *J. Chem. Phys.* **1997**, *106*, 9639.

(70) Hättig, C.; Klopper, W.; Köhn, A.; Tew, D. P. Explicitly Correlated Electrons in Molecules. *Chem. Rev.* **2012**, *112*, 4.

(71) Győrffy, W.; Werner, H.-J. Analytical energy gradients for explicitly correlated wave functions. II. Explicitly correlated coupled cluster singles and doubles with perturbative triples corrections: CCSD(T)-F12. *J. Chem. Phys.* **2018**, *148*, 114104.

(72) Köhn, A. Explicitly correlated coupled-cluster theory using cusp conditions. II. Treatment of connected triple excitations. *J. Chem. Phys.* **2010**, *133*, 174118.

(73) Kállay, M. A systematic way for the cost reduction of density fitting methods. *J. Chem. Phys.* **2014**, *141*, 244113.

(74) Nagy, P. R.; Kállay, M. *J. Chem. Theory Comput.* **2019**, in preparation.

(75) MRCC, a quantum chemical program suite written by M. Kállay, P. R. Nagy, Z. Rolik, D. Mester, G. Samu, J. Csontos, J. Csóka, B. P. Szabó, L. Gyevi-Nagy, I. Ladján-

szki, L. Szegedy, B. Ladóczki, K. Petrov, M. Farkas, P. D. Mezei, and B. Hégely. See http://www.mrcc.hu/ (Accessed May 15, 2019).

(76) Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *J. Chem. Phys.* **2013**, *139*, 134101.

(77) Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2018**, *8*, 1327.

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

(79) Kendall, R. A.; Dunning Jr., T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796.

(80) Woon, D. E.; Dunning Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358.

(81) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy integrals over Gaussian functions. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.

(82) Dunning Jr., T. H.; Peterson, K. A.; Wilson, A. K. Gaussian basis sets for use in correlated molecular calculations. X. The atoms aluminum through argon revisited. *J. Chem. Phys.* **2001**, *114*, 9244.

(83) Peterson, K. A.; Puzzarini, C. Systematically convergent basis sets for transition metals. II. Pseudopotential-based correlation consistent basis sets for the group 11 (Cu, Ag, Au) and 12 (Zn, Cd, Hg) elements. *Theor. Chem. Acc.* **2005**, *114*, 283.

(84) Figgen, D.; Rauhut, G.; Dolg, M.; Stoll, H. Energy-consistent pseudopotentials for group 11 and 12 atoms: adjustment to multi-configuration Dirac–Hartree–Fock data. *Chem. Phys.* **2005**, *311*, 227.

(85) Weigend, F. Hartree–Fock Exchange Fitting Basis Sets for H to Rn. *J. Comput. Chem.* **2008**, *29*, 167.

(86) Weigend, F.; Häser, M.; Patzelt, H.; Ahlrichs, R. RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chem. Phys. Lett.* **1998**, *294*, 143.

(87) Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002**, *116*, 3175.

(88) Karton, A.; Martin, J. M. L. Comment on: "Estimating the Hartree–Fock limit from finite basis set calculations". *Theor. Chem. Acc.* **2006**, *115*, 330.

(89) Neese, F.; Wennmohs, F.; Hansen, A. Efficient and accurate local approximations to coupled-electron pair approaches: An attempt to revive the pair natural orbital method. *J. Chem. Phys.* **2009**, *130*, 114108.

(90) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427.

(91) Kesharwani, M. K.; Karton, A.; Nitai, S.; Martin, J. M. L. The S66 Non-Covalent Interactions Benchmark Reconsidered Using Explicitly Correlated Methods Near the Basis Set Limit. *Aust. J. Chem* **2019**, *71*, 238.

(92) Huenerbein, R.; Schirmer, B.; Moellmann, J.; Grimme, S. Effects of London dispersion on the isomerization reactions of large organic molecules: a density functional benchmark study. *Phys. Chem. Chem. Phys.* **2010**, *12*, 6940.

(93) Sedlak, R.; Janowski, T.; Pitoňák, M.; Řezáč, J.; Pulay, P.; Hobza, P. Accuracy of Quantum Chemical Methods for Large Noncovalent Complexes. *J. Chem. Theory Comput.* **2013**, *9*, 3364.

(94) Pons, J.-L.; de Lamotte, F.; Gautier, M.-F.; Delsuc, M.-A. Refined Solution Structure of a Liganded Type 2 Wheat Nonspecific Lipid Transfer Protein. *J. Biol. Chem.* **2003**, *278*, 14249.

(95) Some typos made in Table 10 of Ref. 27 collecting reference energy differences are corrected in Sect. S8 of the SI.

(96) Foster, J. M.; Boys, S. F. Canonical Configurational Interaction Procedure. *Rev. Mod. Phys.* **1960**, *32*, 300.

(97) Yanai, T.; Fann, G. I.; Beylkin, G.; Gan, Z.; Harrison, R. J. Multiresolution quantum chemistry in multiwavelet bases. *J. Chem. Phys.* **2004**, *121*, 6680.

(98) Jensen, S. R.; Saha, S.; Flores-Livas, J. A.; Huhn, W.; Blum, V.; Goedecker, S.; Frediani, L. The Elephant in the Room of Density Functional Theory Calculations. *J. Phys. Chem. Lett.* **2017**, *8*, 1449.

(99) Xiong, X.-G.; Yanai, T. Projector Augmented Wave Method Incorporated into Gauss-Type Atomic Orbital Based Density Functional Theory. *J. Chem. Theory Comput.* **2017**, *13*, 3236.

Graphical TOC Entry