

Fragment-Based Local Coupled Cluster Embedding Approach for the Quantification and Analysis of Noncovalent Interactions: Exploring the Many-Body Expansion of the Local Coupled Cluster Energy

Soumen Ghosh, Frank Neese, Róbert Izsák, and Giovanni Bistoni*

Cite This: *J. Chem. Theory Comput.* 2021, 17, 3348–3359

Read Online

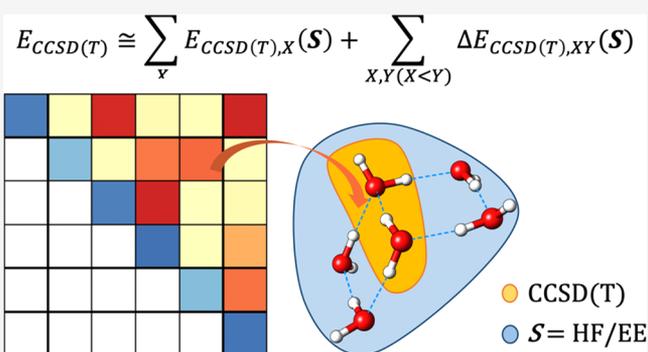
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Herein, we introduce a fragment-based local coupled cluster embedding approach for the accurate quantification and analysis of noncovalent interactions in molecular aggregates. Our scheme combines two different expansions of the domain-based local pair natural orbital coupled cluster (DLPNO-CCSD(T)) energy: the many-body expansion (MBE) and the local energy decomposition (LED). The low-order terms in the MBE are initially computed in the presence of an environment that is treated at a low level of theory. Then, LED is used to decompose the energy of each term in the embedded MBE into additive fragment and fragment-pairwise contributions. This information is used to quantify the total energy of the system while providing at the same time in-depth insights into the nature and cooperativity of noncovalent interactions. Two different approaches are introduced and tested, in which the environment is treated at different levels of theory: the local coupled cluster in the Hartree–Fock (LCC-in-HF) method, in which the environment is treated at the HF level; and the electrostatically embedded local coupled cluster method (LCC-in-EE), in which the environment is replaced by point charges. Both schemes are designed to preserve as much as possible the accuracy of the parent local coupled cluster method for total energies, while being embarrassingly parallel and less memory intensive. These schemes appear to be particularly promising for the study of large and complex molecular aggregates at the coupled cluster level, such as condensed phase systems and protein–ligand interactions.



1. INTRODUCTION

Noncovalent interactions (NCIs) play an important role in all fields of chemical research, determining the formation of intermolecular complexes, the structure and properties of large biomolecules, and the properties of condensed phase systems. Two types of computational approaches—perturbative and supermolecular—can be used to study NCIs in large and complex systems of many interacting molecules.

Among the perturbative approaches, symmetry-adapted perturbation theory (SAPT)^{1–5} has found many applications in the study of weak NCIs due to its great accuracy and the fact that the various perturbation terms can be easily interpreted chemically.^{6,7} Unfortunately, in most implementations, SAPT is only applicable to intermolecular interactions between a pair of monomers in a dimer. Going beyond dimers within the SAPT framework, although possible theoretically,^{8–10} would necessarily increase the computational cost of the method, limiting its applicability to relatively small systems.^{11,12} Hence, to tackle larger systems while remaining within the SAPT framework, one needs to devise practical approximations to the induction and

dispersion terms, as it is done in the recently developed XSAPT approach.^{13–15} In the most recent implementations of this method, the zeroth-order wave function is defined as the product of the monomer wavefunctions at the semiempirical XPol level,¹⁶ while the dispersion energy is replaced with an atom-pairwise empirical dispersion correction.

Within a supermolecular approach, NCIs can be computed as the energy difference between a molecular aggregate (also called “supersystem”) and its constituting monomers. Electronic structure methods of different accuracy and computational cost can be used to compute the individual energies. For example, the “gold standard” coupled cluster method with singles, doubles, and perturbatively included triple excitations,

Received: January 4, 2021

Published: May 26, 2021



CCSD(T),¹⁷ is considered to be the most accurate computational method for studying NCIs among those applicable to systems of medium size. Unfortunately, CCSD(T) in its canonical implementation scales as $O(N^7)$ with the system size, which significantly limits its applicability to relatively small systems with a few hundred basis functions.

A possible strategy to avoid the inherent steep scaling of accurate electronic structure calculations relies on the many-body expansion (MBE) of the total energy of the supersystem.^{18–22} Within the MBE, the energy of the supersystem is expressed exactly as a sum of contributions from all possible monomers (one-body), dimers (two-body), trimers (three-body), etc. Hence, the calculation on the supersystem is reduced to a series of smaller, independent calculations on the subsystems, making the method embarrassingly parallel. Of course, the MBE needs to be truncated in practical applications, and the accuracy of truncated MBEs depends on the nature of the NCIs in the system.^{23,24}

In particular, a defining feature of NCIs is that their strength and nature may change drastically in the presence of other NCIs.^{25,26} This property is sometimes termed “cooperativity” and plays an important role in a large variety of systems, including biomolecules,^{27–29} van der Waals complexes of π -systems,³⁰ and hydrogen³¹ and halogen-bonded³² adducts. In the MBE framework, a “strong” cooperativity manifests itself in large contributions to the energy from the terms of order higher than 2.²² Clearly, truncated MBEs are not accurate for systems where such “many-body” effects play an important role.

To overcome this limitation, embedding can be used to accelerate the convergence of the MBE by effectively incorporating many-body effects into the one- and two-body terms.^{18,19,33–48} Hence, the low-order terms are computed in the presence of the remaining molecules (denoted hereafter as the “environment”), which are treated at a lower level of theory. Kitaura and co-workers first proposed the fragment molecular orbital (FMO) method, in which terms are added to the Hamiltonians of the fragments and dimers to account for the electrostatic potential of the environment. Then, the densities of the fragments and dimers are converged self-consistently.^{33,35} Later, the effective fragment molecular orbital (EFMO) approach was introduced.⁴⁹ This method combines the FMO method with an effective fragment potential (EFP) force field. In addition, a series of electrostatically embedded MBE (EE-MBE) approaches have been proposed by Truhlar and co-workers, in which the environment is represented by point charges computed at a low level of theory.^{36,41} A more sophisticated approach was introduced by Manby and co-workers, who combined the use of atom-centered spherical gaussian charges to represent the electrostatic potential of the environment with an empirical “exchange-repulsion” potential to account for the repulsive interactions between the environment and the subsystems.³⁷ Quantum embedding approaches have also been proposed, in which different levels of electronic structure theory are used to model the environment,^{19,40,45–48,50} such as the popular wavefunction-in-DFT embedding method developed by Manby, Miller, and co-workers.^{48,50} Moreover, the recently proposed embedded mean-field theory⁵¹ (EMFT) of Manby, Miller, and co-workers, which permits the use of different mean-field approaches for the fragments and the environment, was recently used by Head Gordon and co-workers as a basis for the development of a “polarized MBE” quantum embedding scheme, which was applied to the calculation of binding energies in small molecular clusters.¹⁹

Finally, the cluster-in-molecule (CIM) method is a quantum embedding scheme that is based on localized molecular orbitals, which can be used to accelerate the calculation of correlation energies obtained with post Hartree–Fock (HF) methods.^{52,53} To reduce the cost of the underlying HF calculations, the FMO approach has also been combined with CIM.⁵⁴

In this work, we present an approach to quantify and analyze NCIs accurately and in great detail, which is based on an embedded MBE of the CCSD(T) energy of the supersystem. In particular, our approach relies on the domain-based local pair natural orbital variant of the CCSD(T) method, that is, DLPNO-CCSD(T).^{55–61} A large number of benchmark studies demonstrated that the DLPNO-CCSD(T) method can be used to compute relative energies for systems with hundreds of atoms and thousands of basis functions with great accuracy.^{60,62,63}

In our method, the well-established local energy decomposition (LED) analysis^{63–66} is used to decompose exactly the low-order terms in the embedded MBE of the DLPNO-CCSD(T) energy of the supersystem into a series of additive contributions corresponding to: (i) the energy of the embedded monomers; (ii) the interaction energy between pairs of embedded monomers (for the terms of order higher than 1); (iii) the energy of the environment; and (iv) the interaction energy between the environment and each of the monomers. This information is used to quantify the total energy of the supersystem, leading to an embarrassingly parallel method. Moreover, an in-depth, term-by-term comparison between the decomposed contributions of different orders provides unique insights into the nature of cooperativity in noncovalent interactions.

Two different approaches are tested: the local coupled cluster in the Hartree–Fock (LCC-in-HF) embedding scheme, in which the one- and two-body terms in the MBE are treated at the DLPNO-CCSD(T) level in the presence of an Hartree–Fock (HF) environment; and the local coupled cluster in an electrostatic embedding (LCC-in-EE) method, in which the environment is replaced by a series of point charges.

The article is organized as follows. In Sections 2.1 and 2.2, we recall the basic principles of the MBE and of the LED approaches, respectively. In Section 2.3, it is shown how the LED and MBE schemes can be combined to study cooperativity of NCIs. Besides being a valuable interpretation tool for the analysis of NCIs in chemical applications, the combination of MBE and LED allows us to thoroughly test the accuracy of the embedded MBE approaches proposed here. In Sections 2.4 and 2.5, the LCC-in-HF and the LCC-in-EE methods are introduced, respectively, while computational details are given in Section 2.6. In Section 3.1, the LCC-in-HF and the LCC-in-EE methods are tested on a series of water clusters, which are known to be challenging for local coupled clusters techniques^{60,62,63} and to feature large many-body effects.^{41,67,68} Initially, the optimal DLPNO-CCSD(T) settings for H-bonded systems are determined (Section 3.1.1) and used to obtain accurate reference data. These results are then used to discuss the accuracy of the LCC-in-HF and the LCC-in-EE approaches (Section 3.1.2). In Section 3.2, the accuracy and efficiency of these methods are further discussed on a larger example, namely, the interaction of the imidacloprid ligand with a nicotinic acetylcholine receptor (nAChR) model.⁶⁹

2. COMPUTATIONAL DETAILS AND THEORETICAL ASPECTS

2.1. Many-Body Expansion (MBE). For a system of N interacting fragments, the total energy can be exactly expressed using the MBE as a sum of one-body (E^1), two-body (ΔE^2), three-body (ΔE^3), ..., N -body terms (ΔE^N)

$$E_{\text{tot}} = E^1 + \Delta E^2 + \Delta E^3 + \dots + \Delta E^N \quad (1)$$

where

$$E^1 = \sum_X^N E_X \quad (2)$$

$$\Delta E^2 = \sum_{X,Y (X<Y)}^N E_{XY} - E_X - E_Y = \sum_{X,Y (X<Y)}^N E_{XY} \quad (3)$$

and so on, where E_X is the energy of the X th monomer and E_{XY} is the energy of the XY dimer. Hence, ΔE_{XY} represents the interaction energy of the XY dimer.

For example, the total binding energy (also called interaction energy or association energy) of the system, ΔE

$$\Delta E = E_{\text{tot}} - \sum_X^N E_X \quad (4)$$

can be written using the MBE as

$$\Delta E = \Delta E^2 + \Delta E^3 + \Delta E^4 + \dots + \Delta E^N \quad (5)$$

As it was previously suggested in ref22, it is useful to define the overall “cooperativity effect” ΔE_{coop} , as

$$\Delta E_{\text{coop}} = \Delta E - \Delta E^2 = \Delta E^3 + \Delta E^4 + \dots + \Delta E^N \quad (6)$$

Hence, ΔE_{coop} denotes the contribution to the total binding energy ΔE originating from the many-body terms.

2.2. Local Energy Decomposition of the DLPNO-CCSD(T) Energy. An alternative decomposition of the total energy is provided by the LED scheme in the DLPNO-CCSD(T) framework.^{63–66} By assigning the localized occupied orbitals onto the fragments in which they are dominantly localized, E_{tot} in eq 1 can be decomposed exactly into fragment-pairwise contributions

$$E_{\text{tot}} = \sum_X^N E_X^{\text{intra}} + \sum_{X,Y (X<Y)}^N E_{XY}^{\text{int}} \quad (7)$$

where E_X^{intra} represents the energy of monomer X when its electronic structure is perturbed by the presence of all other monomers. It is obtained by summing the contributions to the DLPNO-CCSD(T) energy associated with particles (nuclei and electrons) assigned to fragment X .^{63–66} E_{XY}^{int} represents the interaction between monomers X and Y in the presence of all other monomers. Notably, using LED, the electrostatics, exchange, and dispersion contributions to E_{XY}^{int} can also be quantified.^{63–66}

Using eq 4, we can now rewrite ΔE as

$$\begin{aligned} \Delta E &= \sum_X^N E_X^{\text{intra}} - \sum_X^N E_X + \sum_{X,Y (X<Y)}^N E_{XY}^{\text{int}} \\ &= \sum_X^N \Delta E_X^{\text{el-prep}} + \sum_{X,Y (X<Y)}^N E_{XY}^{\text{int}} \end{aligned} \quad (8)$$

where $\Delta E_X^{\text{el-prep}}$ is called “electronic preparation energy” and determines the energy required to change the electronic structure of monomer X from its ground state to the one that is optimal for the interaction with all of the other monomers. This is by definition a positive (repulsive) contribution to the binding energy.^{63–66}

2.3. Local Energy Decomposition of the MBE of the DLPNO-CCSD(T) Energy. In this section, a theoretical framework is presented for the accurate quantification and analysis of cooperativity effects, which combines the MBE with the LED approach. The LED scheme can be used to decompose all of the terms of the MBE of eq 1. In particular, the two-body term ΔE^2 can be decomposed using eq 8 as

$$\Delta E^2 = \sum_X^N \Delta E_X^{\text{el-prep},2} + \sum_{X,Y (X<Y)}^N E_{XY}^{\text{int},2} \quad (9)$$

where $\Delta E_X^{\text{el-prep},2}$ is the total two-body electronic preparation of monomer X , defined as

$$\Delta E_X^{\text{el-prep},2} = \sum_{Y \neq X}^{N-1} E_{X(XY)}^{\text{intra},2} - (N-1)E_X \quad (10)$$

in which $E_{X(XY)}^{\text{intra},2}$ is the energy of monomer X when its electronic structure is perturbed by the presence of monomer Y , while $E_{XY}^{\text{int},2}$ is the interaction between X and Y . By subtracting eqs 8 and 9, it is possible to write

$$\begin{aligned} \Delta E_{\text{coop}} &= \sum_X^N (\Delta E_X^{\text{el-prep}} - \Delta E_X^{\text{el-prep},2}) \\ &+ \sum_{X,Y (X<Y)}^N (E_{XY}^{\text{int}} - E_{XY}^{\text{int},2}) \\ &= \sum_X^N \Delta E_{\text{coop},X}^{\text{el-prep}} + \sum_{X,Y (X<Y)}^N E_{\text{coop},XY}^{\text{int}} \end{aligned} \quad (11)$$

where $\Delta E_{\text{coop},X}^{\text{el-prep}}$ represents the change in the energy of monomer X due to cooperativity effects. Similarly, $E_{\text{coop},XY}^{\text{int}}$ represents the contribution of cooperativity to the interaction between monomers X and Y . To summarize, by combining the LED and the MBE schemes, one can decompose the total binding energy ΔE , its two-body contribution ΔE^2 , and the total cooperativity effect ΔE_{coop} into additive fragment and fragment-pairwise contributions, providing unique insights into the interplay of various noncovalent interactions.

In the following, all of these components are presented as heat plot matrices where the diagonal elements represent electronic preparation energies while off-diagonal elements represent interactions between pairs of fragments. These plots are denoted as “LED interaction maps”.⁶⁹ As it will be demonstrated below, this type of analysis appears to be particularly useful for testing the performance of novel embedding schemes aimed at recovering many-body effects accurately in truncated MBEs.

2.4. Local Coupled Cluster-in-HF Approach. The theory behind the DLPNO-CCSD(T) methodology has already been discussed in detail in a series of publications.^{55–61} An important feature of this method is the so-called “multilevel” approach, which can be used to treat different parts of the system at different levels of theory.⁷⁰

Herein, the multilevel implementation of the DLPNO-CCSD(T) method is used to define a polarized MBE quantum embedding approach, in which the one-body and the two-body

components of the energy (eq 1) are computed at the DLPNO-CCSD(T) level of theory in the presence of an HF environment (which is treated as an additional fragment). The interaction between the environment and the dimer is also treated at the HF level. A smaller basis set can be used for the environment to increase the efficiency of the approach. Thus, for each pair of fragments, the coupled cluster equations are initially solved for the embedded system. Then, the LED scheme is used to extract the DLPNO-CCSD(T) components of the energy, while the energy of the environment and its interaction with the DLPNO-CCSD(T) subsystem is neglected. From here onwards, this approach is denoted as the local coupled cluster-in-HF (LCC-in-HF) method. If different basis sets, *e.g.*, basis1 and basis2, are used for the subsystem and the environment, respectively, the method is denoted as LCC(basis1)-in-HF(basis2).

For example, LCC-in-HF can be used to compute approximate binding energies (eq 4) using the following steps:

- **Step 1.** Compute the energy of all of the isolated monomers E_X at the DLPNO-CCSD(T) level;
- **Step 2.** Compute the energy of the monomers in the presence of the HF embedding $E_X(\mathbf{S})$ using the multilevel implementation, where \mathbf{S} denotes the environment. Using LED, we obtain approximate electronic preparation energies

$$\Delta E_X^{\text{el-prep},1}(\mathbf{S}) = E_X^{\text{intra},1}(\mathbf{S}) - E_X \quad (12)$$

where the superscript “1” is used to indicate quantities obtained using the embedded one-body approximation.

- **Step 3.** Compute the energy of the dimers in the HF embedding using the multilevel implementation: $E_{XY}(\mathbf{S})$. Using LED, we obtain a series of embedded interfragment contributions $E_{XY}^{\text{int},2}(\mathbf{S})$, where the superscript “2” indicates quantities computed within the embedded two-body approximation. Moreover, we also obtain a series of intrafragment energies $E_{X(XY)}^{\text{intra},2}(\mathbf{S})$, which correspond to the DLPNO-CCSD(T) energies of fragment X when the dimer XY is treated at the DLPNO-CCSD(T) level and all other fragments are described by the HF environment. This can be used to compute the final electronic preparation energies $\Delta E_X^{\text{el-prep},2}(\mathbf{S})$

$$\Delta E_X^{\text{el-prep},2}(\mathbf{S}) = \sum_{Y \neq X}^{N-1} (E_{X(XY)}^{\text{intra},2}(\mathbf{S}) - E_X^{\text{intra},1}(\mathbf{S})) + \Delta E_X^{\text{el-prep},1}(\mathbf{S}) \quad (13)$$

- **Step 4.** The total binding energy is then obtained as

$$\Delta E = \sum_X^N \Delta E_X^{\text{el-prep},2}(\mathbf{S}) + \sum_{X,Y (X<Y)}^N E_{XY}^{\text{int},2}(\mathbf{S}) \quad (14)$$

More in general, the expression for the total LCC-in-HF energy of a system of many interacting molecules reads

$$E_{\text{tot}} = \sum_{X,Y (X<Y)}^N [E_{X(XY)}^{\text{intra},2}(\mathbf{S}) + E_{Y(XY)}^{\text{intra},2}(\mathbf{S}) + E_{XY}^{\text{int},2}(\mathbf{S})] - (N-2) \sum_X^N E_X^{\text{intra},1}(\mathbf{S}) \quad (15)$$

There are some important points that deserve to be discussed in more detail:

- LCC-in-HF avoids the computationally demanding (in terms of memory and computational time) DLPNO-CCSD(T) calculation on the full system by performing a series of less-expensive DLPNO-CCSD(T) calculations on the embedded monomers and dimers, which can be run in parallel.
- As it will be demonstrated numerically in Sections 3.1 and 3.2, the LCC-in-HF method retains to a large extent the accuracy of the parent DLPNO-CCSD(T) method.
- The LCC-in-HF method can be used to compute the interaction between specific pairs of fragments in an environment accurately, *e.g.*, the interaction between a ligand and a residue in the active site of a protein or that between two species in a prereactive complex in solution. An example of such applications is given in Section 3.2.
- If the same basis set is used for the subsystem and for the HF environment in LCC-in-HF, the same reference orbitals can be used for computing all of the one-body and two-body contributions, *i.e.*, the HF calculation on the whole system needs to be carried out only once. In this case, eq 15 can be rewritten as

$$E_{\text{tot}} = E_{\text{HF}} + \sum_{X>Y}^N E_{XY,\text{corr}}(\mathbf{S}) - (N-2) \sum_X^N E_{X,\text{corr}}(\mathbf{S}) \quad (16)$$

where E_{HF} is the HF energy of the system and $E_{XY,\text{corr}}(\mathbf{S})$ and $E_{X,\text{corr}}(\mathbf{S})$ are the correlation components of $E_{XY}(\mathbf{S})$ and $E_X(\mathbf{S})$, respectively. This allows us to skip the LED part of the calculation, thus providing a small but often noticeable saving of computational time ($\sim 10\%$). The downside is that by switching off LED, we lose the analysis of the NCIs in the system. It is also worth mentioning that the approach exemplified by eq 16 conceptually resembles the cluster-in-molecules DLPNO-CCSD(T) method (CIM-DLPNO-CCSD(T)⁷¹) but with some important differences. In CIM-DLPNO-CCSD(T), the total correlation energy E_{cor} is expressed as a sum of contributions from the localized MOs $E_{i,\text{cor}}$

$$E_{\text{cor}} = \sum_i E_{i,\text{cor}} = \sum_i \sum_{jab} [2(ialjb) - (iblja)] \tau_{ij}^{ab} \quad (17)$$

where indexes i and j denote occupied localized MOs; a, b virtual orbitals; and τ_{ij}^{ab} cluster amplitudes. The summation over j is truncated in such a way so as to include only localized MOs that are spatially close to i , and the amplitudes τ_{ij}^{ab} associated with different $E_{i,\text{cor}}$ contributions are not coupled.

2.5. Local Coupled Cluster-in-EE Method. The LCC-in-HF results will be compared with those obtained using electrostatically embedded approaches, in which the HF

environment is replaced by point charges computed at the HF level of theory. Hereafter, these methods are denoted as LCC-in-EEX(PC_type* α), where “PC_type” identifies the type of point charges used, α is a scaling factor, $X = 1$ indicates that point charges are obtained from single point calculations on the whole system, and $X = 2$ indicates that point charges are obtained from separate calculations on the individual fragments. The approach with $X = 2$ has the advantage that allows us to skip the HF/aDZ calculation on the whole system. As an example, LCC-in-EE2(NPA/aDZ*1.2) denotes the use of natural population analysis (NPA) charges⁷² computed for the individual fragments at the HF/aDZ level and scaled by a factor of 1.2. The optimal PC_type/ α combination must be determined for each specific application, e.g., by means of benchmark studies on small model systems.

In the LCC-in-EEX approaches defined here, the total energy of the system is computed using eq 15. Hence, these approaches are not equivalent to the previously published EE-MBE schemes.⁶⁸ One important difference is that, by construction, the $E_X^{\text{intra},1}(\text{S})$ and $E_X^{\text{intra},2}(\text{S})$ terms do not incorporate the interaction between the electrons and the environment. In fact, they represent the energy of the isolated monomers when their electronic structure is distorted in the presence of the embedding. The second important difference is that the total energy in this approach can be decomposed without any additional computational cost into a series of additive contributions (e.g., dispersion and electrostatics^{63–66}), thus providing new information into the nature of noncovalent interactions in the system. For the sake of simplicity, this further decomposition is not discussed in the article but illustrative examples are reported in the Supporting Information (SI) (Figure S4).

Consistent with other electrostatically embedded approaches, the accuracy of LCC-in-EEX schemes depends on the actual point charges used for the embedding. The optimal choice might change depending on the nature of the system considered, and defining a universal set of point charges for general applications goes beyond the scope of the present work. This aspect is currently under investigation in our group. However, the results reported below on water clusters and protein–ligand interactions show that LCC-in-EE1(NPA/aDZ*1.16) and LCC-in-EE2(NPA/aDZ*1.25) provide errors in NCI energies that are typically below 1 kcal/mol (see Section 3.1.2). Importantly, the LED interaction maps discussed in Section 2.3 provide in-depth information that can be used for testing the accuracy of the various embedding approaches, thus aiding to the development of accurate LCC-in-EEX schemes specifically tailored for any given application.

2.6. Computational Details. All calculations were performed using a development version of the ORCA suits of programs based on version 4.2.^{73,74} For LCC-in-HF and LCC-in-EE calculations, a Python program interfaced to ORCA was used to generate the many input files needed for the embedded MBE calculations as well as for data analysis.

For water clusters, all single point DLPNO-CCSD(T) and LED calculations were carried out using the aug-cc-pVTZ⁷⁵ (aTZ) and aug-cc-pVQZ⁷⁵ (aQZ) basis sets. Matching auxiliary/C basis sets were used in all cases.⁷⁶ HF and correlation energies computed using aTZ and aQZ basis sets were extrapolated to the estimated complete basis set (CBS) limit using a two-point extrapolation scheme,⁷⁷ as detailed in refs 78, 79. All valence electrons were included in the correlation treatment, as detailed in ref 80. “TightPNO” settings were used

in all cases. To reduce the PNO truncation error, energies obtained using $T_{\text{CutPNO}} = 10^{-6}$ and $T_{\text{CutPNO}} = 10^{-7}$ settings were extrapolated to the estimated $T_{\text{CutPNO}} = 0$ limit, as detailed in ref 63. This approach is denoted as “6/7 PNO extrapolation”.

In particular, the extrapolated energy was computed using the following equation

$$E = E^X + F \cdot (E^Y - E^X)$$

in which E is the target correlation energy (for a given basis set) at the complete PNO space, E^X and E^Y are the DLPNO-CCSD(T) correlation energies obtained with $T_{\text{CutPNO}} = 10^{-6}$ and $T_{\text{CutPNO}} = 10^{-7}$, respectively, and F is a parameter that is set to 1.5. As discussed extensively in ref 63, 6/7 extrapolation calculations are typically 2 times faster than the corresponding $T_{\text{CutPNO}} = 10^{-8}$ calculations and are also less memory intensive. This indicates that the PNO extrapolation scheme provides a cost-effective alternative to the tightening of the T_{CutPNO} threshold.

In LED, LCC-in-HF, and LCC-in-EE calculations, each water molecule was considered as a fragment, unless mentioned otherwise. In the LCC-in-HF methods, the environment was treated at the HF level using various basis sets, namely, aug-cc-pVDZ⁷⁵ (aDZ), cc-pVDZ⁷⁵ (DZ), def2-SVPD (dSVPD),⁸¹ and def2-TZVPP⁸¹ (dTZ) basis sets. Optimized structures for water hexamer isomers were obtained from ref 79. In all cases, the perturbative triples contribution (T) was calculated using the iterative algorithm.⁸²

For the interaction of imidacloprid with the nAChR model, DLPNO-CCSD(T)/LED calculations were carried out using essentially the same computational settings, as described in ref 69. Hence, single point DLPNO-CCSD(T)/dTZ calculations were carried out using matching def2/C auxiliary basis sets. NormalPNO settings were used except that the TCutPairs threshold was set to the TightPNO value of 10^{-5} . The RJCOSX approximation^{83,84} was used in the HF part (the def2/JK basis set was used) with an extremely fine integration grid. In the LCC-in-HF method, the environment was treated at the HF level using the same basis set. The structure of the imidacloprid-nAChR adduct (resistant insect model) was taken from ref 69. As fragments, we selected the imidacloprid ligand, the water molecules (H_2O), and the W53, W143, R55, Y185, Y192, C187-188, A108, and M14-L122 residues, as described in ref 69.

3. RESULTS AND DISCUSSION

In this section, the efficiency and accuracy of the LCC-in-HF and LCC-in-EE schemes are tested on water clusters (Section 3.1) and on protein–ligand interactions (Section 3.2).

3.1. Water Hexamers. In Section 3.1.1, we initially obtain accurate DLPNO-CCSD(T) reference values for the binding energies of water hexamer clusters. This information is used in Section 3.1.2 to test the accuracy of the LCC-in-HF and LCC-in-EE schemes.

3.1.1. Optimal DLPNO-CCSD(T) Settings for Water Clusters. Second order Møller–Plesset perturbation theory (MP2) and CCSD(T) have been extensively used to study small water clusters.⁷⁹ MP2 is known to slightly underestimate many-body interaction energies in these systems,^{79,85–87} and in fact the deviation between MP2 and CCSD(T) results increases with the size of the cluster.⁷⁹ In this section, we provide the optimal DLPNO-CCSD(T) settings for the calculation of binding energies (see eq 1) of water clusters using canonical CCSD(T) results as the reference. As a first prototype case study, we

Table 1. Interaction Energies (kcal/mol) Calculated with DLPNO-CCSD(T) Using Different T_{CutPNO} Thresholds for Six Isomers of the Water Hexamer Cluster^a

isomer	HF	DLPNO-CCSD(T)/CBS $T_{\text{CutPNO}} 10^{-6}$	DLPNO-CCSD(T)/CBS $T_{\text{CutPNO}} 10^{-7}$	DLPNO-CCSD(T)/CBS 6/7 PNO Extr.	canonical CCSD(T)/CBS ^b
prism	-32.32	-47.74	-48.26	-48.64	-49.04
cage	-32.25	-47.46	-47.92	-48.26	-48.73
book	-33.49	-47.32	-47.73	-48.04	-48.36
bag	-32.79	-46.81	-47.21	-47.50	-47.85
cyclic	-34.56	-46.40	-46.78	-47.06	-47.28
boat	-33.47	-45.42	-45.79	-46.06	-46.26
MAE ^c	14.77	1.06	0.64	0.33	

^aAll energies are extrapolated to the estimated CBS limit using aTZ and aQZ basis sets. ^bTaken from ref 85. ^cComputed using canonical CCSD(T)/CBS as the reference.

consider the six isomers of the hexamer water cluster, shown in Figure S1.

The canonical- and DLPNO-CCSD(T) binding energies obtained with different computational settings are shown in Table 1. As previously shown for other water clusters,^{62,63} the dependence of the DLPNO-CCSD(T)/CBS results on the value of the T_{CutPNO} threshold appears to be significant. For example, default TightPNO settings ($T_{\text{CutPNO}} = 10^{-7}$) give a mean absolute error (MAE) of 0.6 kcal/mol with respect to canonical CCSD(T)/CBS. The MAE becomes even higher if looser PNO thresholds ($T_{\text{CutPNO}} = 10^{-6}$) are used, reaching up to 1.1 kcal/mol. In contrast, 6/7 PNO extrapolation⁶³ reduces the MAE of the DLPNO-CCSD(T) results to 0.3 kcal/mol, providing results with essentially canonical CCSD(T) accuracy. Unless specified otherwise, 6/7 PNO extrapolation will be used for all of the DLPNO-CCSD(T), LCC-in-HF, and LCC-in-EE calculations discussed in the following.

3.1.2. Accuracy of LCC-in-HF and LCC-in-EE. Having established the optimal DLPNO-CCSD(T) settings for quantifying binding energies in water clusters, we now turn to the evaluation of the accuracy of the LCC-in-HF and LCC-in-EE approaches. Table 2 shows the HF and correlation components of the binding energy of the prism isomer obtained with different methods and computational settings. As it will be demonstrated numerically below, the main results obtained for this system remain valid for all of the water clusters examined here.

Table 2. Binding Energies (kcal/mol) for the Prism Isomer of the Hexamer Water Cluster Calculated Using Embedding MBE Approaches Described in Section 2.3^a

	ΔE (HF)	ΔE (Corr)	ΔE
DLPNO-CCSD(T)/CBS	-32.32	-16.32	-48.64
LCC(CBS)-in-HF(CBS)	-32.32	-16.38	-48.70
two-body DLPNO-CCSD(T)/CBS ^b	-22.73	-16.69	-39.42
LCC(CBS)-in-HF(aDZ)	-32.33	-16.70	-49.03
LCC(CBS)-in-HF(DZ)	-33.55	-16.60	-50.05
LCC(CBS)-in-HF(dSVPD)	-32.32	-16.54	-48.86
LCC(CBS)-in-EE1(CHELPG ^c /aDZ)	-27.43	-17.00	-44.43
LCC(CBS)-in-EE1(NPA/aDZ)	-29.77	-17.05	-46.82
LCC(CBS)-in-EE1(NPA/aDZ*1.16)	-32.21	-17.10	-49.31
LCC(CBS)-in-EE2(NPA/aDZ*1.16)	-30.60	-17.07	-47.66
LCC(CBS)-in-EE2(NPA/aDZ*1.25)	-31.85	-17.09	-48.94

^aCorrelation energies were extrapolated to the estimated PNO limit using 6/7 extrapolation. ^b ΔE^2 in eq 3. ^cCHELPG atomic charges were used, see ref 88 for details.

The overall DLPNO-CCSD(T)/CBS binding energy ΔE in this cluster amounts to -48.64 kcal/mol, while the contribution from the two-body component (ΔE^2 in eq 3) is only -39.42 kcal/mol. Hence, many-body effects increase the stability of this system by ~ 10 kcal/mol. In-depth information into the chemical origin of such effects can be obtained by analyzing the LED interaction maps (see Section 2.3). This analysis is reported in the SI.

Remarkably enough, the LCC(CBS)-in-HF(CBS) method essentially demonstrates the same accuracy as the parent DLPNO-CCSD(T)/CBS method. In fact, the HF contribution to the binding energy is not affected by the use of an HF embedding (by definition), while the error introduced in the correlation contribution to the binding energy amounts to only 0.06 kcal/mol. These results demonstrate that the use of an HF environment is a cost-effective strategy for computing NCIs between pairs of monomers at the DLPNO-CCSD(T) level in large and complex systems. In particular, the accuracy of the HF embedding proposed here appears to be similar to that obtained using projection-based embedding techniques on similar systems.⁵⁰

Interestingly, if the basis set used for describing the environment is reduced to aDZ, DZ, and dSVPD, the error in the HF component of the binding energy becomes 0.01, 1.22, and 0.00 kcal/mol, respectively. The corresponding errors in the correlation binding energies become 0.38, 0.28, and 0.22 kcal/mol. For comparison, the error obtained in the HF and correlation components of the binding energy by truncating the MBE to the two-body terms (without embedding) is 9.59 kcal/mol and 0.36, respectively. These results demonstrate (i) the importance of including diffuse basis functions in the HF environment to recover mean-field many-body effects accurately, (ii) and the inherent difficulty of describing many-body correlation effects accurately using embedding approaches. In fact, the error in the correlation binding energy in the presence of an HF embedding with a reduced basis set is similar to that obtained by neglecting many-body correlation effects entirely.

Importantly, the LED interaction maps for the HF and correlation components of the binding energy (Figures 1b,c and 2a-c, respectively) show that the LCC-in-HF method not only provides accurate binding energies but also accurate electronic preparation and fragment pairwise interaction energies. Hence, the LED interaction maps demonstrate that the accuracy of this approach is not due to error cancellation but to the accurate description of the entire system of NCIs that contribute to the stability of this complex. This type of analysis appears to be particularly important for the development of novel embedding schemes.

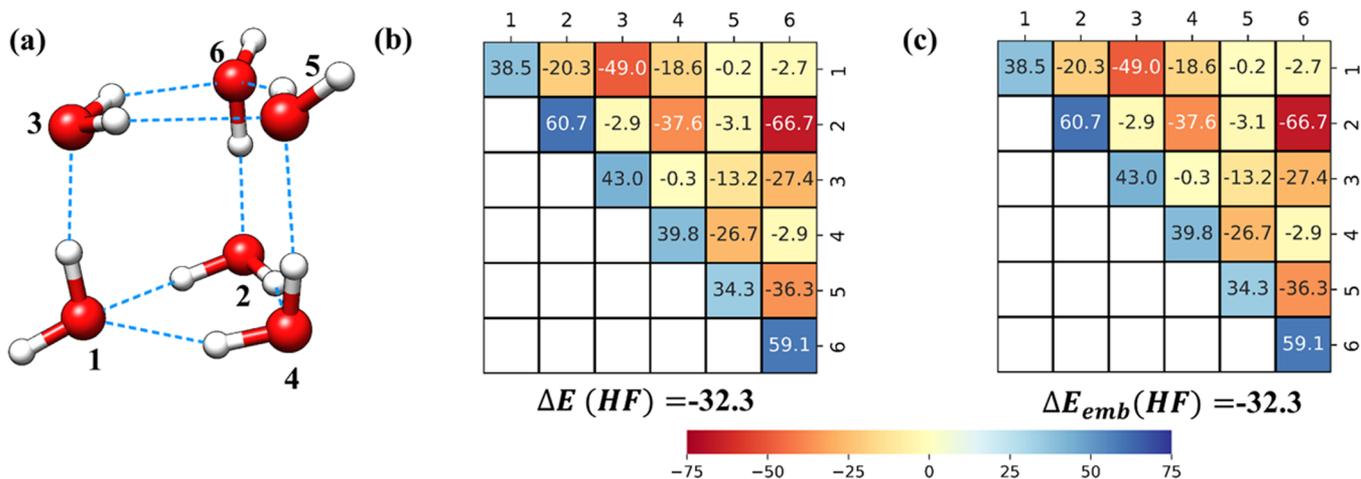


Figure 1. HF component of the LED interaction maps associated with the H-bonding interactions in the (a) prism isomer of the water hexamer; (b) decomposition of the HF-binding energy according to eq 8; and (c) decomposition of the HF-binding energy obtained at the LCC(CBS)-in-HF(aDZ) level, according to eq 14. The color code is shown at the bottom of the figure, with all values in kcal/mol. Bluish colors indicate repulsive interactions, while orange/red indicate attractive interactions.

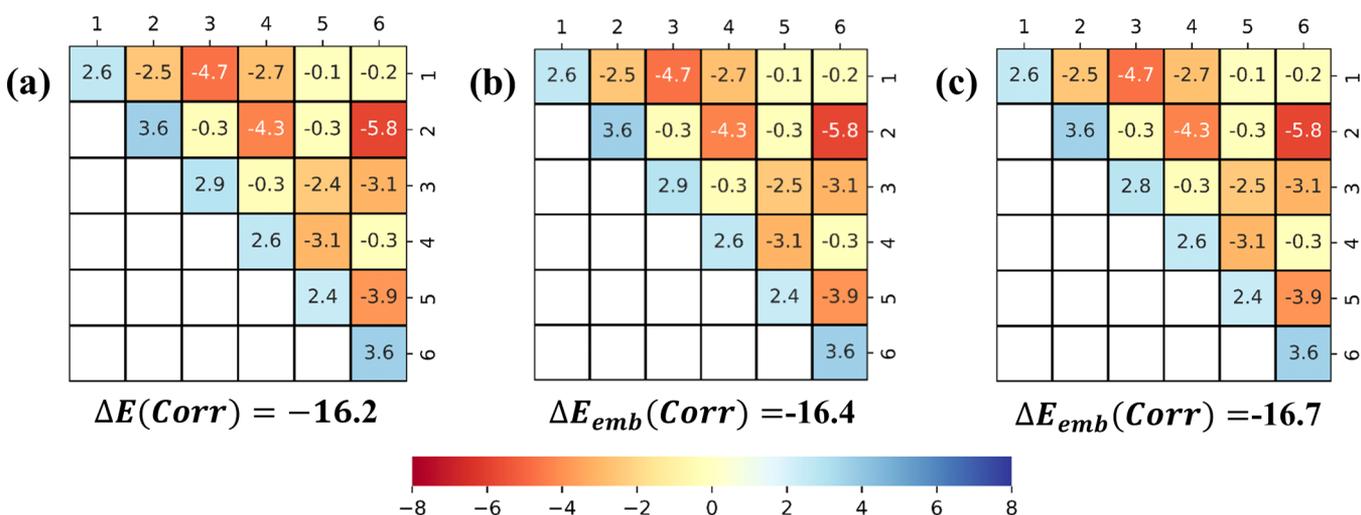


Figure 2. Correlation components of the LED interaction maps associated with the H-bonding interactions in the prism isomer of the water hexamer shown in Figure 1a. (a) Decomposition of the DLPNO-CCSD(T) correlation binding energy,⁸⁹ according to eq 8; (b) decomposition of the LCC(CBS)-in-HF(CBS) correlation binding energy, according to eq 14; and (c) decomposition of the LCC(CBS)-in-HF(aDZ) correlation binding energy, according to eq 14. The color code is shown on the right side, with all values in kcal/mol. Bluish colors indicate repulsive interactions, while orange/red indicate attractive interactions.

It is now instructive to compare the results just discussed with those obtained using the LCC-in-EE approach defined in Section 2.5. Consistent with previously published results using similar schemes, the accuracy of the mean-field component from such electrostatically embedded methods depends on the actual point charges used for the embedding. In contrast, the correlation binding energy is not affected by the specific point charges used. However, the errors in the correlation binding energy are consistently larger than those obtained with the HF embedding. These results suggest that a more sophisticated HF embedding should be preferred when many-body correlation effects are expected to play an important role.

In our example, accurate mean-field binding energies can be obtained using LCC(CBS)-in-EE1(NPA/aDZ*1.16), in which the charges on the oxygen and hydrogen atoms are on average ca. -1.2 and $+0.6$, respectively. The LCC-in-EE2(NPA/aDZ*1.16) scheme uses slightly smaller charges (on average, -1.1 and $+0.5$ for oxygen and hydrogen atoms, respectively),

which causes the error on the mean-field binding energy to increase. More accurate values can be obtained by increasing the scaling factor to 1.25. The resulting LCC(CBS)-in-EE2(NPA/aDZ*1.25) scheme provides binding energies of accuracy comparable to that of LCC(CBS)-in-EE1(NPA/aDZ*1.16).

As mentioned above, the results just discussed remain valid for all of the isomers of the water hexamers, as shown in Table 3 (note that the LCC-in-EE2 results were not computed for all systems because they are expected to be similar to those obtained at the LCC-in-EE1 level).

Being embarrassingly parallel, the computational cost of LCC-in-HF and LCC-in-EE calculations is limited by that of the largest dimer considered, provided that enough computational resources are available. In the present case, all dimers have a similar computational cost. For example, the energy calculation for an embedded water dimer at the LCC(aDZ)-in-EE2(NPA/aDZ*1.16) level ($T_{\text{CutPNO}} 10^{-7}$) took on average 0.13 h using four cores from a single cluster node equipped with Intel Xeon

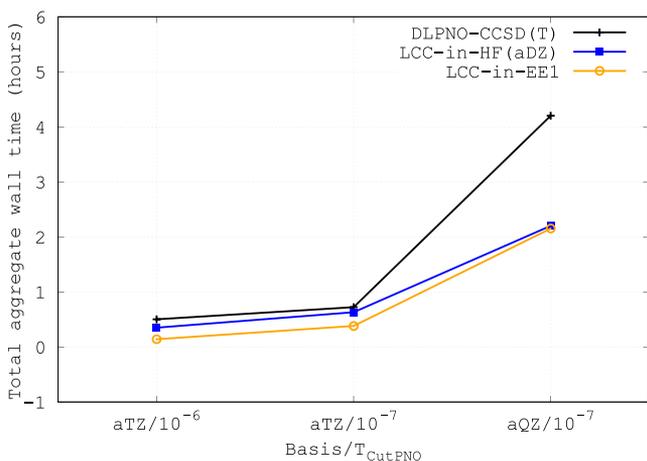
Table 3. Binding Energies (kcal/mol) Calculated with Different Approaches for Six Isomers of the Water Hexamer Cluster

isomers	LCC(CBS)-in-HF(CBS)	LCC(CBS)-in-HF(aDZ)	LCC(CBS)-in-EE1(NPA/aDZ*1.16)
prism	-48.70	-49.03	-49.31
cage	-48.36	-48.58	-48.51
book	-48.07	-48.27	-48.72
bag	-47.51	-47.80	-47.49
cyclic	-47.12	-47.38	-48.60
boat	-46.10	-46.41	-47.33
MAE ^a	0.05	0.32	0.74
MAE(HF) ^b	0.00	0.12	0.68
MAE(Corr.) ^c	0.05	0.19	0.39

^aMAE computed using as reference DLPNO-CCSD(T)/CBS results (see Table 1). ^bMAE obtained for the mean-field component of the energy. ^cMAE obtained for the correlation component of the energy using HF/CBS as the reference.

E5-2690v2 CPUs and 64 GB of RAM (the SCF part of the calculation took only a few seconds). In contrast, a DLPNO-CCSD(T)/aQZ calculation on the “Bag” isomer took 4 h using the same computational resources (only 4 min were required for the SCF part of the calculation). For comparison, the coupled cluster part of LCC(aQZ)-in-HF(aDZ) and LCC(aQZ)-in-HF(aQZ) calculations on the embedded dimers using the same DLPNO settings took 0.13 and 0.5 h, respectively.

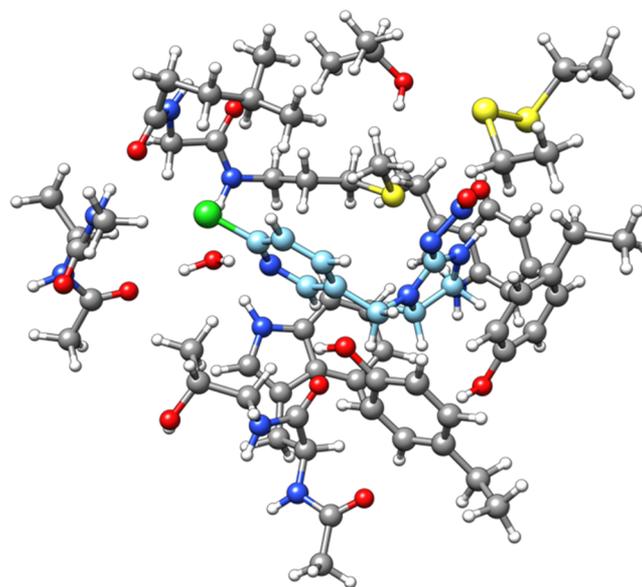
The total aggregate wall time required to compute the energy of the Bag isomer with various computational settings is shown in Figure 3. For the purpose of this comparison, all calculations

**Figure 3.** Total aggregate wall time required to compute the energy of the Bag isomer with various computational settings. All calculations were carried out using four cores from a single cluster node equipped with Intel Xeon E5-2690v2 CPUs and 64 GB of RAM.

were run sequentially on a single node with four cores, meaning that we did not exploit the embarrassingly parallel nature of the LCC-in-EE and LCC-in-HF methods. For example, timings could be trivially reduced by running the individual embedded dimer calculations on separate nodes. Nevertheless, LCC-in-EE and LCC-in-HF are still more efficient than their parent DLPNO-CCSD(T) method, and the efficiency difference increases with large basis sets. It is also worth mentioning that for this system, LCC-in-EE1 and LCC-in-HF show similar performances. This is due to the fact that the computational cost in this case is dominated by the coupled cluster part, which is

similar for both methods. However, for larger systems, LCC-in-EE1 will become more efficient than LCC-in-HF, as the former does not require to perform HF calculations on the supersystem. This makes LCC-in-EE1 and LCC-in-EE2 particularly promising for the study of condensed phase systems.

3.2. Protein–Ligand Interactions. In this section, the efficiency and the accuracy of the LCC-in-HF and LCC-in-EE schemes are discussed on a model system representing the interaction of the imidacloprid ligand with the active site of the nicotinic acetylcholine receptor (nAChR).⁹⁰ Imidacloprid is a neonicotinoid synthetic insecticide and it is important to study its interaction with insect nAChRs to understand the selective toxicity of these types of insecticides.⁹¹ The structure of the imidacloprid-nAChR adduct is shown in Figure 4.⁶⁹

**Figure 4.** Cluster model for the binding of the imidacloprid ligand to a nAChR model. The system has 223 atoms. The structure is taken from ref 69. Carbon atoms of the imidacloprid ligand are shown in light blue color.

The total energy of the system computed at the DLPNO-CCSD(T)/dTZ, LCC(dTZ)-in-HF(dTZ), and LCC(dTZ)-in-EE2(NPA/aDZ*1.25) levels is shown in Table 4. In the LCC-in-HF and LCC-in-EE case, the total energy was computed using eq 15 (see Section 2.6 for the fragment definition).

For this system, all methods reproduce the total DLPNO-CCSD(T) energy extremely well, with an error of only -0.11

Table 4. Total Energy E_{tot} , HF energy E_{HF} , and Correlation Energy E_{cor} (in Hartree) for the Imidacloprid-nAChR Adduct Shown in Figure 4 Computed at the DLPNO-CCSD(T)/dTZ Level of Theory, as Described in Section 2.6^a

	DLPNO-CCSD(T)/dTZ energy (Hartree)	LCC(dTZ)-in-HF(dTZ) Δ (kcal/mol)	LCC(dTZ)-in-EE2(NPA/aDZ*1.25) Δ (kcal/mol)
E_{HF}	-6281.59409	0.00	1.40
E_{cor}	-20.4988543	-0.11	-0.95
E_{tot}	-6302.09294	-0.11	0.45

^a Δ (in kcal/mol) is the deviation obtained for the same energy contributions computed at various levels of theory with respect to the DLPNO-CCSD(T)/dTZ reference.

and 0.45 kcal/mol for LCC(dTZ)-in-HF(dTZ) and LCC(dTZ)-in-EE2(NPA/aDZ*1.25), respectively. In the LCC(dTZ)-in-EE2(NPA/aDZ*1.25) case, the error in the mean-field component of the energy amounts to 1.40 kcal/mol and it partially cancels out with the -0.95 kcal/mol error observed in the correlation energy.

To determine whether or not this remarkable accuracy is the result of error cancellation between different interaction terms, we examined the individual errors associated with the key interactions responsible for the binding of the ligand to the protein, *i.e.*, those between the ligand and the residues and in the active site.⁶⁹ In the DLPNO-CCSD(T) case, these are quantified by the E_{XY}^{int} terms appearing in eq 7, while in embedded approaches these are given by the $E_{XY}^{\text{int},2}(\mathcal{S})$ terms in eq 15.

The comparison between the E_{XY}^{int} and $E_{XY}^{\text{int},2}(\mathcal{S})$ terms associated with the same XY dimer (Table 5) revealed that the

Table 5. LED Interactions E_{XY}^{int} between the Imidacloprid Ligand and the Residues in the Active Site of the “Resistant” nAChR Model at the DLPNO-CCSD(T)/def2-TZVPP Level^a

residue	DLPNO-CCSD(T)/dTZ	LCC(dTZ)-in-HF(dTZ)		LCC(dTZ)-in-EE2(NPA/aDZ*1.25)	
	E_{XY}^{int} (kcal/mol)	Δ (kcal/mol)	wall time ^b (h)	Δ (kcal/mol)	wall time (h)
H20	-42.80	-0.09	4.2	-0.39	0.7
W53	-22.94	-0.05	5.1	-0.01	2.3
W143	-70.65	-0.14	12.0	-1.00	5.5
R55	-0.48	0.00	4.2	0.02	0.9
Y185	-48.43	-0.10	6.8	-0.39	3.4
Y192	-19.40	-0.08	4.5	-0.32	2.2
C187-188	-18.89	-0.03	4.8	-0.48	1.1
A103	-5.90	-0.05	8.2	-0.44	1.3
M14-L112	-65.61	-0.16	4.2	-0.51	3.9

^a Δ (in kcal/mol) is computed as $E_{XY}^{\text{int},2}(\mathcal{S}) - E_{XY}^{\text{int}}$, where $E_{XY}^{\text{int},2}(\mathcal{S})$ is obtained using different embedding schemes. For LCC-in-HF and LCC-in-EE2, the wall time for computing the energy of the corresponding (embedded) dimers is reported in hours.^bWall time for the HF/dTZ calculation on the full protein–ligand adduct is 4.5 h. In the LCC(dTZ)-in-HF(dTZ) scheme, this calculation needs to be carried out only once. The resulting orbitals are used to compute all of the terms in the MBE.

error observed in the individual interaction terms is of the same order of that observed for the total energy of the system. Thus, the difference between E_{XY}^{int} and $E_{XY}^{\text{int},2}(\mathcal{S})$ is relatively small, reaching a maximum of 0.14 kcal/mol with LCC(dTZ)-in-HF(dTZ) and of 1.00 kcal/mol with LCC(dTZ)-in-EE2(NPA/aDZ*1.25). For comparison, the deviation that one would obtain without any embedding ($E_{XY}^{\text{int},2} - E_{XY}^{\text{int}}$) is significantly larger, reaching up to 4.2 kcal/mol for the imidacloprid-W53 interaction. Hence, the LCC-in-HF and LCC-in-EE methods provide the right results for the right reasons, even though the calculation of total energies benefit from favorable error compensation between the terms in eq 15.

In terms of efficiency, a DLPNO-CCSD(T) single point energy calculation on the whole system required 64 h using 12 cores from a single cluster node equipped with Intel Xeon E5-2687Wv4 CPUs and 256 GB of RAM. In contrast, the

LCC(dTZ)-in-HF(dTZ) and LCC(dTZ)-in-EE2(NPA/aDZ*1.25) results were obtained from separate calculations, and the associated wall times are also reported in Table 5. In the present case, in which 12 cores were assigned to each subsystem, the most expensive dimer calculations took 12.0 and 5.5 h for LCC(dTZ)-in-HF(dTZ) and EE2(NPA/aDZ*1.25), respectively.

Finally, it is worth emphasizing that in some chemical applications, one is only interested in the quantification and analysis of selected interactions at the coupled cluster level. In this case, the approaches introduced here appear to be particularly promising, as they allow us to focus the computational resources on the subsystem of interest. For example, they can be used to monitor how the interaction between the ligand and a specific residue or solvent molecule changes in various snapshots of a molecular dynamic simulation, without the need to treat the whole system at the coupled cluster level.

4. CONCLUSIONS

We have introduced the LCC-in-HF and the LCC-in-EE methods for the quantification and analysis of NCIs in molecular aggregates at the local coupled cluster level. These methods were defined by combining two different expansions of the DLPNO-CCSD(T) energy, *i.e.*, the MBE and the LED. In the LCC-in-HF scheme, the one- and two-body terms in the MBE are computed in the presence of an HF embedding, while in the LCC-in-EE scheme, the environment is replaced by point charges. In both cases, the LED scheme was used to decompose the terms in the embedded MBE into contributions from:

- (1) the embedded monomers;
- (2) the interaction energy between pairs of embedded monomers (for the terms of order higher than 1)
- (3) the energy of the environment; and
- (4) the interaction energy between the environment and each of the monomers.

This information is used to determine the total energy of the supersystem, while providing at the same time quantitative information into the underlying pattern of NCIs as well as their cooperativity effects.

Both approaches were tested on a series of water clusters and on the adduct between the imidacloprid ligand and a nAChR model. It was found that the LCC-in-HF scheme essentially retains the same accuracy of the parent DLPNO-CCSD(T) method, provided that a sufficiently large basis set is used in the HF embedding. In contrast, the accuracy of the LCC-in-EE approach depends on the actual charges used for the embedding. For the systems considered in this work, chemical accuracy can be achieved using NPA charges computed at the HF level for the individual fragments and scaled by a suitable factor. Notably, the use of the more sophisticated HF embedding is recommended for those systems in which many-body correlation effects are expected to be especially important.

Finally, the LCC-in-HF and the LCC-in-EE schemes are embarrassingly parallel and require less memory than their parent DLPNO-CCSD(T) method. Hence, they appear to be particularly promising for the study of large and complex systems in the condensed phase.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.1c00005>.

Analysis of cooperative effects for the prism isomer of the water hexamer and XYZ structures (PDF)

AUTHOR INFORMATION

Corresponding Author

Giovanni Bistoni – Max-Planck-Institut für Kohlenforschung, D-45470 Mülheim an der Ruhr, Germany; orcid.org/0000-0003-4849-1323; Email: giovanni.bistoni@kofo.mpg.de

Authors

Soumen Ghosh – Max-Planck-Institut für Kohlenforschung, D-45470 Mülheim an der Ruhr, Germany; orcid.org/0000-0003-0850-4855

Frank Neese – Max-Planck-Institut für Kohlenforschung, D-45470 Mülheim an der Ruhr, Germany; orcid.org/0000-0003-4691-0547

Róbert Izsák – Max-Planck-Institut für Kohlenforschung, D-45470 Mülheim an der Ruhr, Germany; orcid.org/0000-0001-9236-9718

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jctc.1c00005>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We gratefully acknowledge the Max Planck Society for financial support. Soumen Ghosh also acknowledges generous support from the Alexander von Humboldt Foundation.

REFERENCES

- (1) Szalewicz, K.; Jeziorski, B. Symmetry-Adapted Double-Perturbation Analysis of Intramolecular Correlation Effects in Weak Intermolecular Interactions: The He-He Interaction. *Mol. Phys.* **1979**, *38*, 191–208.
- (2) Rybak, S. A.; Jeziorski, B.; Szalewicz, K. Many-Body Symmetry-Adapted Perturbation Theory of Intermolecular Interactions. H₂O and Hf Dimers. *J. Chem. Phys.* **1991**, *95*, 6576–6601.
- (3) Korona, T. A Coupled Cluster Treatment of Intramonomer Electron Correlation within Symmetry-Adapted Perturbation Theory: Benchmark Calculations and a Comparison with a Density-Functional Theory Description. *Mol. Phys.* **2013**, *111*, 3705–3715.
- (4) Misquitta, A. J.; Szalewicz, K. Intermolecular Forces from Asymptotically Corrected Density Functional Description of Monomers. *Chem. Phys. Lett.* **2002**, *357*, 301–306.
- (5) Heßelmann, A.; Jansen, G. First-Order Intermolecular Interaction Energies from Kohn–Sham Orbitals. *Chem. Phys. Lett.* **2002**, *357*, 464–470.
- (6) Szalewicz, K.; Bukowski, R.; Jeziorski, B., *Theory and Applications of Computational Chemistry: The First Forty Years*; Elsevier: New York; 2005.
- (7) Patkowski, K. Recent Developments in Symmetry-Adapted Perturbation Theory. *WIREs Comput. Mol. Sci.* **2020**, *10*, No. e1452.
- (8) Lotrich, V. F.; Szalewicz, K. Symmetry-Adapted Perturbation Theory of Three-Body Nonadditivity of Intermolecular Interaction Energy. *J. Chem. Phys.* **1997**, *106*, 9668–9687.
- (9) Lotrich, V. F.; Szalewicz, K. Symmetry-Adapted Perturbation Theory of Three-Body Nonadditivity in Ar Trimer. *J. Chem. Phys.* **1997**, *106*, 9688–9702.
- (10) Podeszwa, R.; Szalewicz, K. Three-Body Symmetry-Adapted Perturbation Theory Based on Kohn–Sham Description of the Monomers. *J. Chem. Phys.* **2007**, *126*, No. 194101.
- (11) Groenenboom, G. C.; Mas, E. M.; Bukowski, R.; Szalewicz, K.; Wormer, P. E. S.; van der Avoird, A. Water Pair and Three-Body

Potential of Spectroscopic Quality from Ab Initio Calculations. *Phys. Rev. Lett.* **2000**, *84*, 4072–4075.

(12) Mas, E. M.; Bukowski, R.; Szalewicz, K. Ab Initio Three-Body Interactions for Water. I. Potential and Structure of Water Trimer. *J. Chem. Phys.* **2003**, *118*, 4386–4403.

(13) Jacobson, L. D.; Herbert, J. M. An Efficient, Fragment-Based Electronic Structure Method for Molecular Systems: Self-Consistent Polarization with Perturbative Two-Body Exchange and Dispersion. *J. Chem. Phys.* **2011**, *134*, No. 094118.

(14) Lao, K. U.; Herbert, J. M. Accurate Intermolecular Interactions at Dramatically Reduced Cost: Xpol+ Sapt with Empirical Dispersion. *J. Phys. Chem. Lett.* **2012**, *3*, 3241–3248.

(15) Lao, K. U.; Herbert, J. M. Accurate and Efficient Quantum Chemistry Calculations for Noncovalent Interactions in Many-Body Systems: The Xsapt Family of Methods. *J. Phys. Chem. A* **2015**, *119*, 235–252.

(16) Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. The Variational Explicit Polarization Potential and Analytical First Derivative of Energy: Towards a Next Generation Force Field. *J. Chem. Phys.* **2008**, *128*, No. 234108.

(17) Bartlett, R. J.; Stanton, J. F. Applications of Post-Hartree–Fock Methods: A Tutorial. *Rev. Comput. Chem.* **1994**, 65–169.

(18) Grabowski, S. J.; Leszczynski, J. The Enhancement of X–H···Π Hydrogen Bond by Cooperativity Effects—Ab Initio and Qctaim Calculations. *Chem. Phys.* **2009**, *355*, 169–176.

(19) Veccham, S. P.; Lee, J.; Head-Gordon, M. Making Many-Body Interactions Nearly Pairwise Additive: The Polarized Many-Body Expansion Approach. *J. Chem. Phys.* **2019**, *151*, No. 194101.

(20) Mandal, A.; Prakash, M.; Kumar, R. M.; Parthasarathi, R.; Subramanian, V. Ab Initio and Dft Studies on Methanol–Water Clusters. *J. Phys. Chem. A* **2010**, *114*, 2250–2258.

(21) Rincón, L.; Almeida, R.; García Aldea, D. Many-Body Energy Decomposition Analysis of Cooperativity in Hydrogen Fluoride Clusters. *Int. J. Quantum Chem.* **2005**, *102*, 443–453.

(22) Saha, S.; Sastry, G. N. Quantifying Cooperativity in Water Clusters: An Attempt Towards Obtaining a Generalised Equation. *Mol. Phys.* **2015**, *113*, 3031–3041.

(23) Elrod, M. J.; Saykally, R. J. Many-Body Effects in Intermolecular Forces. *Chem. Rev.* **1994**, *94*, 1975–1997.

(24) Xantheas, S. S. Ab Initio Studies of Cyclic Water Clusters (H₂O)_N, N = 1–6. II. Analysis of Many-Body Interactions. *J. Chem. Phys.* **1994**, *100*, 7523–7534.

(25) Mahadevi, A. S.; Sastry, G. N. Cooperativity in Noncovalent Interactions. *Chem. Rev.* **2016**, *116*, 2775–2825.

(26) Hunter, C. A.; Anderson, H. L. What Is Cooperativity? *Angew. Chem., Int. Ed.* **2009**, *48*, 7488–7499.

(27) Perutz, M. F.; Perutz, M. F. *Mechanisms of Cooperativity and Allosteric Regulation in Proteins*; CUP Archive: 1990.

(28) Sleat, D. E.; Wiseman, J. A.; El-Banna, M.; Price, S. M.; Verot, L.; Shen, M. M.; Tint, G. S.; Vanier, M. T.; Walkley, S. U.; Lobel, P. Genetic Evidence for Nonredundant Functional Cooperativity between Npc1 and Npc2 in Lipid Transport. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5886–5891.

(29) Zandany, N.; Ovadia, M.; Orr, I.; Yifrach, O. Direct Analysis of Cooperativity in Multisubunit Allosteric Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 11697–11702.

(30) Kim, K. S.; Tarakeshwar, P.; Lee, J. Y. Molecular Clusters of Π-Systems: Theoretical Studies of Structures, Spectra, and Origin of Interaction Energies. *Chem. Rev.* **2000**, *100*, 4145–4186.

(31) Estarellas, C.; Frontera, A.; Quiñonero, D.; Deyà, P. M. Interplay between Cation–Π and Hydrogen Bonding Interactions: Are Non-Additivity Effects Additive? *Chem. Phys. Lett.* **2009**, *479*, 316–320.

(32) Li, Q.; Li, R.; Liu, Z.; Li, W.; Cheng, J. Interplay between Halogen Bond and Lithium Bond in Mcn-Licn-Xcch (M = H, Li, and Na; X = Cl, Br, and I) Complex: The Enhancement of Halogen Bond by a Lithium Bond. *J. Comput. Chem.* **2011**, *32*, 3296–3303.

(33) Kitaura, K.; Sawai, T.; Asada, T.; Nakano, T.; Uebayasi, M. Pair Interaction Molecular Orbital Method: An Approximate Computa-

tional Method for Molecular Interactions. *Chem. Phys. Lett.* **1999**, *312*, 319–324.

(34) Gao, J.; Wang, Y. *Communication: Variational Many-Body Expansion: Accounting for Exchange Repulsion, Charge Delocalization, and Dispersion in the Fragment-Based Explicit Polarization Method*; American Institute of Physics, 2012.

(35) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment Molecular Orbital Method: An Approximate Computational Method for Large Molecules. *Chem. Phys. Lett.* **1999**, *313*, 701–706.

(36) Dahlke, E. E.; Truhlar, D. G. Electrostatically Embedded Many-Body Expansion for Large Systems, with Applications to Water Clusters. *J. Chem. Theory Comput.* **2007**, *3*, 46–53.

(37) Bygrave, P.; Allan, N.; Manby, F. The Embedded Many-Body Expansion for Energetics of Molecular Crystals. *J. Chem. Phys.* **2012**, *137*, No. 164102.

(38) Beran, G. J. Approximating Quantum Many-Body Intermolecular Interactions in Molecular Clusters Using Classical Polarizable Force Fields. *J. Chem. Phys.* **2009**, *130*, No. 164115.

(39) Beran, G. J.; Nanda, K. Predicting Organic Crystal Lattice Energies with Chemical Accuracy. *J. Phys. Chem. Lett.* **2010**, *1*, 3480–3487.

(40) Schmitt-Monreal, D.; Jacob, C. R. Frozen-Density Embedding-Based Many-Body Expansions. *Int. J. Quantum Chem.* **2020**, *120*, No. e26228.

(41) Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. Evaluation of the Electrostatically Embedded Many-Body Expansion and the Electrostatically Embedded Many-Body Expansion of the Correlation Energy by Application to Low-Lying Water Hexamers. *J. Chem. Theory Comput.* **2008**, *4*, 33–41.

(42) 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.

(43) Jiang, N.; Ma, J.; Jiang, Y. Electrostatic Field-Adapted Molecular Fractionation with Conjugated Caps for Energy Calculations of Charged Biomolecules. *J. Chem. Phys.* **2006**, *124*, No. 114112.

(44) Liu, J.; Herbert, J. M. Pair–Pair Approximation to the Generalized Many-Body Expansion: An Alternative to the Four-Body Expansion for Ab Initio Prediction of Protein Energetics Via Molecular Fragmentation. *J. Chem. Theory Comput.* **2016**, *12*, 572–584.

(45) Tschumper, G. S. Multicentered Integrated Qm: Qm Methods for Weakly Bound Clusters: An Efficient and Accurate 2-Body: Many-Body Treatment of Hydrogen Bonding and Van Der Waals Interactions. *Chem. Phys. Lett.* **2006**, *427*, 185–191.

(46) Bates, D. M.; Smith, J. R.; Tschumper, G. S. Efficient and Accurate Methods for the Geometry Optimization of Water Clusters: Application of Analytic Gradients for the Two-Body: Many-Body Qm: Qm Fragmentation Method to (H₂O)_N, N = 3–10. *J. Chem. Theory Comput.* **2011**, *7*, 2753–2760.

(47) Mayhall, N. J.; Raghavachari, K. Molecules-in-Molecules: An Extrapolated Fragment-Based Approach for Accurate Calculations on Large Molecules and Materials. *J. Chem. Theory Comput.* **2011**, *7*, 1336–1343.

(48) Manby, F. R.; Stella, M.; Goodpaster, J. D.; Miller, T. F., III A Simple, Exact Density-Functional-Theory Embedding Scheme. *J. Chem. Theory Comput.* **2012**, *8*, 2564–2568.

(49) Steinmann, C.; Fedorov, D. G.; Jensen, J. H. Effective Fragment Molecular Orbital Method: A Merger of the Effective Fragment Potential and Fragment Molecular Orbital Methods. *J. Phys. Chem. A* **2010**, *114*, 8705–8712.

(50) Barnes, T. A.; Goodpaster, J. D.; Manby, F. R.; Miller, T. F., III Accurate Basis Set Truncation for Wavefunction Embedding. *J. Chem. Phys.* **2013**, *139*, No. 024103.

(51) Fornace, M. E.; Lee, J.; Miyamoto, K.; Manby, F. R.; Miller, T. F., III Embedded Mean-Field Theory. *J. Chem. Theory Comput.* **2015**, *11*, 568–580.

(52) Li, S.; Ma, J.; Jiang, Y. Linear Scaling Local Correlation Approach for Solving the Coupled Cluster Equations of Large Systems. *J. Comput. Chem.* **2002**, *23*, 237–244.

(53) Li, S.; Shen, J.; Li, W.; Jiang, Y. An Efficient Implementation of the “Cluster-in-Molecule” Approach for Local Electron Correlation Calculations. *J. Chem. Phys.* **2006**, *125*, No. 074109.

(54) 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*, 3587–3593.

(55) 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*, No. 114108.

(56) Riplinger, C.; Neese, F. An Efficient and near Linear Scaling Pair Natural Orbital Based Local Coupled Cluster Method. *J. Chem. Phys.* **2013**, *138*, No. 034106.

(57) 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*, No. 134101.

(58) Liakos, D. G.; Hansen, A.; Neese, F. Weak Molecular Interactions Studied with Parallel Implementations of the Local Pair Natural Orbital Coupled Pair and Coupled Cluster Methods. *J. Chem. Theory Comput.* **2011**, *7*, 76–87.

(59) 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*, No. 024109.

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

(61) 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*, 4054–4063.

(62) Liakos, D. G.; Guo, Y.; Neese, F. Comprehensive Benchmark Results for the Domain Based Local Pair Natural Orbital Coupled Cluster Method (Dlpno-Ccsd(T)) for Closed- and Open-Shell Systems. *J. Phys. Chem. A* **2020**, *124*, 90–100.

(63) Altun, A.; Neese, F.; Bistoni, G. Extrapolation to the Limit of a Complete Pair Natural Orbital Space in Local Coupled-Cluster Calculations. *J. Chem. Theory Comput.* **2020**, *16*, 6142–6149.

(64) Schneider, W. B.; Bistoni, G.; Sparta, M.; Saitow, M.; Riplinger, C.; Auer, A. A.; Neese, F. Decomposition of Intermolecular Interaction Energies within the Local Pair Natural Orbital Coupled Cluster Framework. *J. Chem. Theory Comput.* **2016**, *12*, 4778–4792.

(65) Altun, A.; Neese, F.; Bistoni, G. Effect of Electron Correlation on Intermolecular Interactions: A Pair Natural Orbitals Coupled Cluster Based Local Energy Decomposition Study. *J. Chem. Theory Comput.* **2019**, *15*, 215–228.

(66) Altun, A.; Saitow, M.; Neese, F.; Bistoni, G. Local Energy Decomposition of Open-Shell Molecular Systems in the Domain-Based Local Pair Natural Orbital Coupled Cluster Framework. *J. Chem. Theory Comput.* **2019**, *15*, 1616–1632.

(67) Cisneros, G. A.; Wikfeldt, K. T.; Ojamäe, L.; Lu, J.; Xu, Y.; Torabifard, H.; Bartók, A. P.; Csányi, G.; Molinero, V.; Paesani, F. Modeling Molecular Interactions in Water: From Pairwise to Many-Body Potential Energy Functions. *Chem. Rev.* **2016**, *116*, 7501–7528.

(68) Dahlke, E. E.; Truhlar, D. G. Electrostatically Embedded Many-Body Correlation Energy, with Applications to the Calculation of Accurate Second-Order Møller–Plesset Perturbation Theory Energies for Large Water Clusters. *J. Chem. Theory Comput.* **2007**, *3*, 1342–1348.

(69) Beck, M. E.; Riplinger, C.; Neese, F.; Bistoni, G. Unraveling Individual Host–Guest Interactions in Molecular Recognition from First Principles Quantum Mechanics: Insights into the Nature of Nicotinic Acetylcholine Receptor Agonist Binding. *J. Comput. Chem.* **2021**, *42*, 293–302.

(70) Sparta, M.; Retegan, M.; Pinski, P.; Riplinger, C.; Becker, U.; Neese, F. Multilevel Approaches within the Local Pair Natural Orbital Framework. *J. Chem. Theory Comput.* **2017**, *13*, 3198–3207.

(71) 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*, No. 124117.

(72) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural Population Analysis. *J. Chem. Phys.* **1985**, *83*, 735–746.

(73) Neese, F. The Orca Program System. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73–78.

(74) Neese, F. Software Update: The Orca Program System, Version 4.0. *WIREs Comput. Mol. Sci.* **2018**, *8*, No. e1327.

(75) Dunning, T. H., Jr. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(76) 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*, 3175–3183.

(77) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-Set Convergence in Correlated Calculations on Ne, N₂, and H₂O. *Chem. Phys. Lett.* **1998**, *286*, 243–252.

(78) Altun, A.; Neese, F.; Bistoni, G. HfId: A Nonempirical London Dispersion-Corrected Hartree–Fock Method for the Quantification and Analysis of Noncovalent Interaction Energies of Large Molecular Systems. *J. Chem. Theory Comput.* **2019**, *15*, 5894–5907.

(79) Alkan, M.; Xu, P.; Gordon, M. S. Many-Body Dispersion in Molecular Clusters. *J. Phys. Chem. A* **2019**, *123*, 8406–8416.

(80) Bistoni, G.; Riplinger, C.; Minenkov, Y.; Cavallo, L.; Auer, A. A.; Neese, F. Treating Subvalence Correlation Effects in Domain Based Pair Natural Orbital Coupled Cluster Calculations: An out-of-the-Box Approach. *J. Chem. Theory Comput.* **2017**, *13*, 3220–3227.

(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. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

(82) 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*, No. 011101.

(83) Neese, F.; Wennmohs, F.; Hansen, A.; Becker, U. Efficient, Approximate and Parallel Hartree–Fock and Hybrid Dft Calculations. A ‘Chain-of-Spheres’ Algorithm for the Hartree–Fock Exchange. *Chem. Phys.* **2009**, *356*, 98–109.

(84) Izsák, R.; Neese, F. An Overlap Fitted Chain of Spheres Exchange Method. *J. Chem. Phys.* **2011**, *135*, No. 144105.

(85) Gonthier, J. F.; Head-Gordon, M. Assessing Electronic Structure Methods for Long-Range Three-Body Dispersion Interactions: Analysis and Calculations on Well-Separated Metal Atom Trimers. *J. Chem. Theory Comput.* **2019**, *15*, 4351–4361.

(86) Piecuch, P. Spherical Tensor Theory of Long-Range Interactions in a System of N Arbitrary Molecules Including Quantum-Mechanical Many-Body Effects: I. Anisotropic Induction Interactions in the First Three Orders of Perturbation Theory. *Mol. Phys.* **1986**, *59*, 1067–1083.

(87) Piecuch, P. Towards Classification and Analytical Description of Molecular Interactions Including Quantum-Mechanical Many-Body Effects. In *Molecules in Physics*. In *Chemistry, and Biology*; Springer, 1988; pp 417–505.

(88) Breneman, C. M.; Wiberg, K. B. Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis. *J. Comput. Chem.* **1990**, *11*, 361–373.

(89) Note That the Value for the Correlation Component of the Binding Energy Differs from Those Reported in Table 2 by 0.13 Kcal/Mol. This Deviation Is Due to the “Delocalized” Component of the Perturbative Triples Correction, Which Is Associated with Triples of Electrons Located on Different Fragments.

(90) Jeschke, P.; Nauen, R.; Beck, M. E. Nicotinic Acetylcholine Receptor Agonists: A Milestone for Modern Crop Protection. *Angew. Chem., Int. Ed.* **2013**, *52*, 9464–9485.

(91) Selvam, B.; Graton, J.; Laurent, A. D.; Alamiddine, Z.; Mathé-Allainmat, M.; Lebreton, J.; Coqueret, O.; Olivier, C.; Thany, S. H.; Le Questel, J.-Y. Imidacloprid and Thiacloprid Neonicotinoids Bind More Favourably to Cockroach Than to Honeybee A6 Nicotinic Acetylcholine Receptor: Insights from Computational Studies. *J. Mol. Graphics Modell.* **2015**, *55*, 1–12.