PCCP

Accurate Three-body Noncovalent Interactions: the Insights from Energy Decomposition

Journal:	*Physical Chemistry Chemical Physics*
Manuscript ID	CP-ART-08-2023-003938.R1
Article Type:	Paper
Date Submitted by the Author:	05-Oct-2023
Complete List of Authors:	Ochieng, Sharon; Auburn University, Chemistry and Biochemistry Patkowski, Konrad; Auburn University, Chemistry and Biochemistry

PCCP

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Accurate Three-body Noncovalent Interactions: the Insights from Energy Decomposition†

Sharon A. Ochieng[a] and Konrad Patkowski[a*]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

An impressive collection of accurate two-body interaction energies for small complexes has been assembled into benchmark databases and used to improve the performance of multiple density functional, semiempirical, and machine learning methods. Similar benchmark data on nonadditive three-body energies in molecular trimers are comparatively scarce, and the existing ones are practically limited to homotrimers. In this work, we present a benchmark dataset of 20 equilibrium noncovalent interaction energies for a small but diverse selection of 10 heteromolecular trimers. The new 3BHET dataset presents complexes that combine different interactions including $\pi-\pi$, anion$-\pi$, cation$-\pi$, and various motifs of hydrogen and halogen bonding in each trimer. A detailed symmetry-adapted perturbation theory (SAPT)-based energy decomposition of the two- and three-body interaction energies shows that 3BHET consists of electrostatics- and dispersion-dominated complexes. The nonadditive three-body contribution is dominated by induction, but its influence on the overall bonding type in the complex (as exemplified by its position on the ternary diagram) is quite small. We also tested the extended SAPT (XSAPT) approach which is capable of including some nonadditive interactions in clusters of any size. The resulting three-body dispersion term (obtained from the many-body dispersion formalism) is mostly in good agreement with the supermolecular CCSD(T)$-$MP2 values and the nonadditive induction term is similar to the three-body SAPT(DFT) data, but the overall three-body XSAPT energies are not very accurate as they are missing the first-order exchange terms.

1 Introduction

Noncovalent intermolecular interactions, besides determining spectroscopic and scattering data for bimolecular complexes, are critical for the properties of clusters, liquids, and solids. In these cases, the structure and properties of the system at hand is determined by the cooperative effect of all interactions[1], not just the two-body forces between pairs of molecules. While the nonadditive three-body interactions are individually much smaller in comparison to their two-body counterparts, their importance increases rapidly with the cluster size, thus very subtly influencing the electronic structure of the interacting species[2–4].

The magnitude of interactions and the importance of three-body nonadditive effects influences the choice of methods used to describe them either theoretically or experimentally. In principle, the most stringent test of a theoretical model is its comparison with experimental data. However, the interaction potentials between molecules are usually measured indirectly through various kinds of gas-phase spectroscopy[5,6]. This, however, comes with a limitation as the experimental data samples a range of intermolecular configurations simultaneously, and one also has to subtract the zero-point vibrational energy (ΔZPVE), making it hard to obtain interaction energies at the desired level of accuracy[7]. These limitations have made it a common practice to test the interaction energies of more approximate theoretical methods against accurate benchmark calculations. Therefore, benchmark databases such as S22[8], S66[9], X40[10], or S12L[11,12] have been vital in quantifying the accuracy of computational methods as well as optimizing new density functionals, semiempirical methods, or machine learning approaches[7,13].

Assessing the accuracy, reliability, and transferability of an electronic structure method requires broad testing on systems that are chemically diverse[13,14]. A well balanced benchmark set should aim to provide a statistically meaningful overview of the performance and allow for the comparison of various methods. For example, the X40 dataset[10] provides a focused description of halogenated complexes including hydrogen and halogen bonding, while the S66 dataset[9] includes a broader description of biorelevant dimers, including varied motifs of hydrogen bonding and $\pi-\pi$ interactions. While both of these databases focus on the van der Waals minimum geometries, they also have their X40x10 and S66x8 counterparts that include points across the entire dissoci-

[a] Department of Chemistry and Biochemistry, Auburn University, Auburn, Alabama 36849, United States

* E-mail: patkowsk@auburn.edu

† Electronic Supplementary Information (ESI) available: Figures with SAPT(DFT) potential energy curves and XYZ coordinates of all complexes in this study. See DOI: 10.1039/cXCP00000x/

ation curve[9,10]. Such databases involving both equilibrium and non-equilibrium points provide data for the parameterization of empirical and semiempirical methods towards the description of attractive and repulsive noncovalent interactions[7].

In recent years, the number and breadth of benchmark databases of bimolecular complexes has greatly expanded. Burns et al.[15] sampled the entire Protein Data Bank of crystal structures identifying 3380 different conformations of interacting amino acid side chains. Řezač et al.[16] are in the process of developing an atlas of noncovalent interactions aimed at mapping the entire chemical space, one specific interaction type at a time. The first datasets of the project describe hydrogen bonding in neutral and ionic organic dimers[16], also those including sulfur, phosphorus or halogens[17]. The next dataset[18] focuses on repulsive contacts in the same kinds of organic molecules. Additional datasets have been constructed to populate the atlas, focusing on sigma-hole interactions (halogen, chalcogen, and pnictogen bonds) and London dispersion interactions, sampling an extensive range of the chemical space[19,20]. Also recently, DiStasio and coworkers developed an extensive benchmark dataset of ∼8000 non-equilibrium geometries of bimolecular complexes[21], aiming to adequately describe noncovalent interactions across the entire potential energy surfaces. The use of such broad databases ensures that the observed performance of approximate methods is transferable between systems and between different regions of the potential energy surface.

Compared to this extensive coverage of two-body complexes, there is limited benchmark data available to describe diverse types of nonadditive three-body interactions relevant to molecular clusters and condensed-phase systems. The most extensive dataset available is 3B-69[22] which comprises accurate interaction energies in 69 trimers extracted from 23 molecular crystal structures (thus, every trimer has three identical molecules). Accurate data is also available for numerous structures of specific systems, such as the water trimer[23] and the benzene trimer[24,25]. Gordon and coworkers[26] constructed a number of trimers from the dimers in the S22 set[8] to create the S22(3) dataset. Their study focuses on investigating and quantitatively estimating the many-body dispersion effects for different bonding motifs in molecular clusters, and the trimers have 1–2 types of nonequivalent monomers (thus, they are of either the AAA or AAB type). More recently, Low et al.[27] combined the 3B-69 and S22(3) datasets with a larger, 509-trimer collection extracted from an enzyme-inhibitor complex to develop a machine learning approach to predict three-body interaction energies. This new collection involves trimers that are heteromolecular and also larger (up to 77 atoms), but the reference interaction energies were only computed with a tailored version of the spin-component-scaled MP2 approach[28] and it is not entirely clear if they all are of benchmark CCSD(T) quality.

While the 3B-69 and S22(3) datasets provide valuable reference energies and establish the performance of various electronic structure methods, primarily for complexes of identical molecules, they provide limited data for modelling different kinds of interactions simultaneously. Thus, one often needs a broader set of reference values, including also heteromolecular trimers. In the long run, such a broad dataset can establish a "dictionary" of three-body interaction types, translating between intuitive (but "fuzzy") classifications of the constituent dimers (hydrogen bonding, $\pi-\pi$ stacking, halogen bonding, $\cdots$) and quantitative numerical data obtained from benchmark supermolecular calculations and symmetry-adapted perturbation theory (SAPT) energy decomposition. This allows one to provide a chemically based classification of different interactions present in trimers. As a step towards this goal, we construct a new dataset of accurate three-body intermolecular interactions in ten heteromolecular trimers (20 structures) computed at the CCSD(T) level of theory at the complete basis set (CBS) limit. The complexes include up to 3 distinct monomers and represent a range of interactions including hydrogen and halogen bonding, π-π stacking, cation/anion-π, CH-π, and OH-π bonds. The dataset is a mix of global and local minima.

A wavefunction-based, CCSD(T)-level benchmark reference is essential for quantifying nonadditive interactions and assessing the performance of more approximate methods. Density functional theory (DFT), using multiple exchange-correlation functionals, has failed to predict accurately three-body interaction energies, especially dispersion[29]. According to Jankiewicz et al.[29], the main reason for the failure is the inability to remove the deficiencies of the functional using various dispersion corrections. This issue persists even with the many-body dispersion (MBD) extension by Tkatchenko and coworkers[30], which formally incorporates correct many-body physics but does not necessarily predict more accurate three-body energies than the bare functionals. It should be noted that, in the MBD formalism, a "body" is an atom rather than an entire molecule, so this approach captures, to a reasonable approximation, all-order long-range correlation effects of both intra- and intermolecular character. The three-body dispersion (where each "body" is a different molecule) can be extracted from the MBD calculations in a standard supermolecular manner, by subtracting the pairwise terms from the interaction energy of the trimer.

To quantitatively analyze the diversity of interactions in our dataset, we employ the energy decomposition analysis provided by SAPT at both the two-body and three-body level. We compare the decompositions predicted by different SAPT variants including SAPT(DFT)[31,32], SAPT2+(3)[33], and XSAPT[34], allowing us to assess the strengths and weaknesses of each approach (SAPT2+(3) is restricted to pairwise additive interactions while both SAPT(DFT) and XSAPT account for some nonadditive effects). To visualize and interpret these interactions, we expand upon the ternary diagram representation commonly used for bimolecular complexes[14,35]. In this extended representation, we incorporate the nonadditive three-body effects for both SAPT(DFT) and XSAPT. This approach enables us to analyze and interpret the complex interplay between different interaction components in our dataset, providing valuable insights into the nature and strength of these interactions.

2 Computational Methods

2.1 Composition of the Data Set

The goal of this work is to design a small but diverse database of heteromolecular trimers with an equally diverse range of non-covalent interaction types. The 3BHET dataset constructed here consists of 10 trimers, with the starting geometries obtained from dimers in the BEGDB database[8] by adding the third molecule and reoptimizing the geometry. The species included (see Figures 1–2 for pictures and the Electronic Supplementary Information† for the XYZ geometries) were specifically chosen to represent a variety of interactions relevant to bioorganic chemistry that include hydrogen and halogen bonding, π-π stacking, cation/anion-π, CH-π, and OH-π interactions. For each system, two geometries are studied, with one being the global minimum of the trimer, and the other one a local minimum of a different binding character than the global one (whenever possible).

2.2 Geometries

Geometry optimizations were carried out using MP2/aug-cc-pVDZ with counterpoise correction and tight convergence criteria. Frequency calculations were also carried out to ascertain that the geometries obtained were true global or local minima. Out of the 20 structures considered, 19 turned out to correspond to actual minima (no imaginary frequencies) while the last one (complex 5.2) was a high-symmetry saddle point. Both global and local minima obtained were used for the benchmark calculations and SAPT analysis, and several of them were later selected as starting points for radial potential energy curves. In such a case, the geometries and mutual orientations of all molecules are conserved as the intermolecular center-of-mass distance is scaled. This scaling is done by shifting one molecule away from the other two molecules in the complex, thereby creating a 1D potential curve. The minimum-energy intermolecular distance is scaled along this direction by factors ranging from 0.7 to 3.0. Note that three curves of this kind are obtained for each trimer as any one molecule can be translated relative to the other two.

We performed the benchmark and two-body SAPT(DFT) calculations using the MOLPRO2012.1 program[36]. For SAPT2+(3) calculations, the Psi4 code was used[37]. Nonadditive SAPT(DFT) results were obtained using the three-body module of the SAPT package[38], importing the DFT quantities from DALTON 2.0[39]. Finally, the XSAPT computations utilized the Q-Chem code[40].

2.3 Benchmark Two-Body and Three-Body Interaction Energies

The nonadditive three-body interaction energy E^3_{int} is defined as the difference between the trimer interaction energy and the sum of the pair interaction energies[41]. The total interaction energy in a trimer

$$E_{int} = E_{ABC} - E_A - E_B - E_C \quad (1)$$

can be decomposed as

$$E_{int} = E^2_{int} + E^3_{int} \quad (2)$$

where the two-body part is the sum of pairwise interactions:

$$E^2_{int} = (E_{AB} - E_A - E_B) + (E_{BC} - E_B - E_C) + (E_{AC} - E_A - E_C) \quad (3)$$

The monomer and dimer calculations were performed in the full trimer basis set to eliminate basis set superposition error (BSSE). In addition, the interaction energies were computed using fixed monomer geometries ignoring all monomer deformation effects.

The benchmark interaction energies at both the two-body and three-body level were calculated using the composite CCSD(T)/CBS approach where correlation consistent basis sets fully augmented with diffuse functions were used; in this text, aug-cc-pVXZ (X = D,T,Q,5,...) are abbreviated as aXZ[42]. The composite interaction energy is constructed from the contributions of the Hartree-Fock method, the MP2 correlation energy, and a small-basis CCSD(T) correction:

$$E^{(\mathrm{CCSD(T)/CBS})} = E^{\mathrm{HF}} + E^{\mathrm{MP2/CBS}}_{\mathrm{corr}} + \delta E^{\mathrm{CCSD(T)}} \quad (4)$$

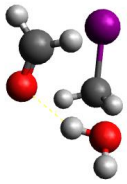
In this way, the MP2 term covers a large section of interaction energy at the CBS limit. However, MP2 is known to overestimate two-body dispersion effects in many systems[7] and it entirely misses three-body dispersion; therefore, a CCSD(T) correction $\delta E^{\mathrm{CCSD(T)}} = E^{\mathrm{CCSD(T)}} - E^{\mathrm{MP2}}$ in a moderate basis set is needed for benchmark accuracy. With the latter correction computed in a double-zeta basis set, this level of theory can be classified as the *silver standard* of accuracy, suitable for nearly all benchmark purposes[43,44]. As each trimer contains three pairwise interactions plus a comparatively small nonadditive effect, it is reasonable to expect a silver-standard approach to yield absolute interaction energy errors about three times larger than in the dimer case, which is still adequate for our benchmark purposes.

The Hartree-Fock energy converges fast with the basis set size. Therefore, a single HF calculation in a large enough basis set is sufficient, and we have used the a5Z basis set for this purpose. The MP2 correlation energy is extrapolated to the CBS limit from the aQZ and a5Z basis sets using Helgaker's X^{-3} formula[45]. The coupled cluster correction $\delta E^{\mathrm{CCSD(T)}}$ is calculated in aDZ for larger complexes and aTZ for a few smaller ones. Density fitting of two-electron integrals was employed in all MP2 computations in order to make calculations for larger complexes feasible.

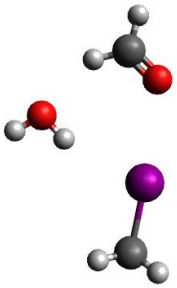
2.4 SAPT Energy Decomposition

To adequately describe and interpret the supermolecular benchmark data obtained in this work, we need to know the nature of the three-body binding in each trimer in relation to its two-body counterpart. Therefore, the two-body and three-body interaction energies are decomposed into physically meaningful quantities using several SAPT variants. In the two-body and three-body SAPT based on the HF wavefunctions, the Hamiltonian is partitioned into the Fock operator F, the intermolecular interaction operator V, and the intramolecular correlation operator W encompassing all monomers:

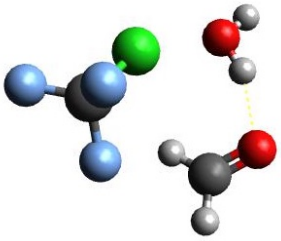
$$H = F + V + W \quad (5)$$



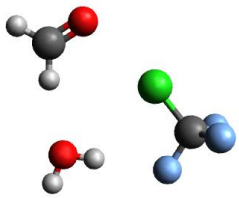
1.1 Iodomethane-Formaldehyde-Water (hydrogen bonds)



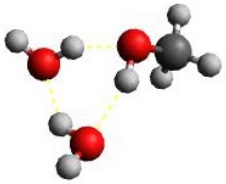
1.2 Iodomethane-Formaldehyde-Water (hydrogen and halogen bonds)



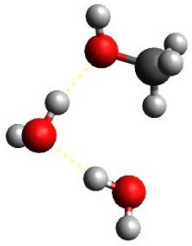
2.1 Trifluorochloromethane-Formaldehyde-Water (hydrogen bonds)



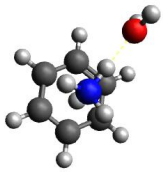
2.2 Trifluorochloromethane-Formaldehyde-Water (halogen and hydrogen bonds)



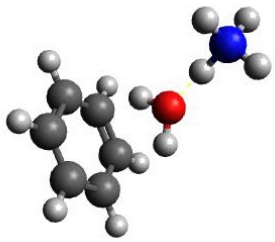
3.1 Methanol-Water-Water (hydrogen bonds)



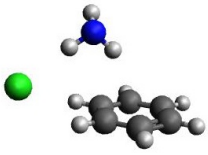
3.2 Methanol-Water-Water (hydrogen bonds)



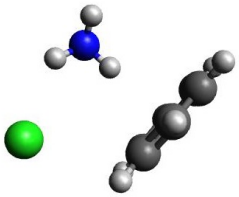
4.1 Ammonium-Benzene-Water (cation-π and hydrogen bonds)



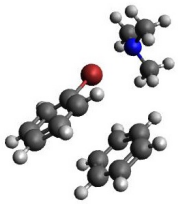
4.2 Ammonium-Benzene-Water (OH-π and hydrogen bonds)



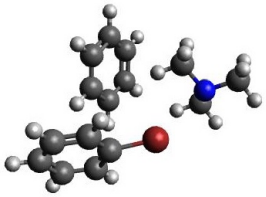
5.1 Chloride-Ammonia-Benzene (NH-π and hydrogen bonds)



5.2 Chloride-Ammonia-Benzene (NH-π, anion-π, and hydrogen bonds)



6.1 Bromobenzene-Trimethylamine-Benzene (halogen and $\pi-\pi$ bonds)



6.2 Bromobenzene-Trimethylamine-Benzene (CH-π bonds)

Fig. 1 Global and local-minimum geometric arrangements of trimers present in the dataset.

where

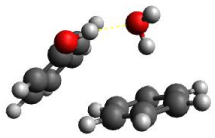
$$F = F_A + F_B + F_C \quad V = V_{AB} + V_{BC} + V_{AC} \quad W = W_A + W_B + W_C \quad (6)$$

The simplest approximation that provides qualitatively reasonable total interaction energies is the SAPT0 one, which totally neglects the intramonomer electron correlation[41]. The two-body SAPT0 interaction energy is calculated as

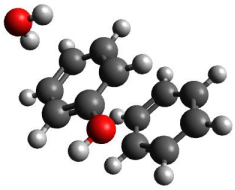
$$E_{int}^{\mathrm{SAPT0}} = E_{elst}^{(10)} + E_{exch}^{(10)} + E_{ind,resp}^{(20)} + E_{exch-ind,resp}^{(20)} + E_{disp}^{(20)} + E_{exch-disp}^{(20)} + \delta E_{HF}^{(2)} \quad (7)$$

In a SAPT correction $E^{(kl)}$ in Eq. (7) and throughout the text, the subscripts k and l denote orders of perturbation theory with respect to V and W, respectively. SAPT0 provides the physical components that represent the electrostatic, first-order exchange, induction, exchange-induction (both with the inclusion of monomer relaxation/response effects as denoted by the additional subscript "resp"), dispersion, exchange-dispersion, and $\delta E_{HF}^{(2)}$ contributions. The $\delta E_{HF}^{(2)}$ term is meant to estimate the higher-order induction and exchange-induction effects from a supermolecular HF calculation[33,41]. This energy decomposition allows one to classify complexes according to the relative importance of electrostatics, induction, and dispersion for their binding, for example, using a ternary diagram[14].

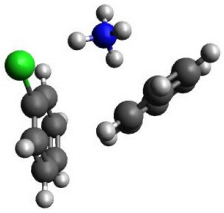
To attain quantitative accuracy of total energies and their contributions, a higher-order SAPT level that includes both intra- and intermonomer electron correlation effects is required. In this



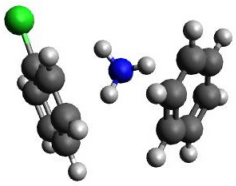
7.1 Phenol-Benzene-Water ($\pi-\pi$, OH-π, and hydrogen bonds)



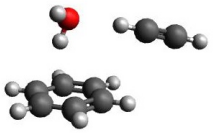
7.2 Phenol-Benzene-Water ($\pi-\pi$, OH-π, and hydrogen bonds)



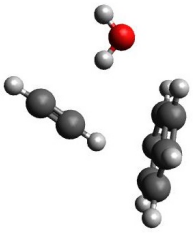
8.1 Chlorobenzene-Benzene-Ammonium (cation-π, CH-π, and hydrogen bonds)



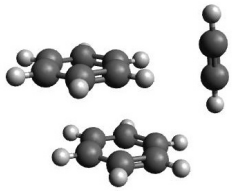
8.2 Chlorobenzene-Benzene-Ammonium (cation-π and hydrogen bonds)



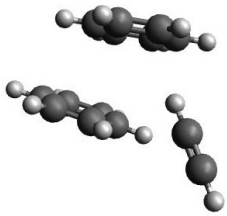
9.1 Acetylene-Benzene-Water (OH-π and CH-π bonds)



9.2 Acetylene-Benzene-Water (OH-π and CH-π bonds)



10.1 Acetylene-Benzene-Benzene (CH-π and π-π bonds)



10.2 Acetylene-Benzene-Benzene (CH-π and π-π bonds)

Fig. 2 Global and local-minimum geometric arrangements of trimers present in the dataset.

study, the two-body interaction energy decomposition was performed by wavefunction-based SAPT at the SAPT2+(3) level of theory, simultaneously including the effects of the W operator on most SAPT0-level corrections and some terms that are third-order in V:[33]

$$\begin{aligned} E_{int}^{\mathrm{SAPT2+(3)}} =& E_{elst}^{(10)} + E_{exch}^{(10)} + E_{ind,resp}^{(20)} + E_{exch-ind,resp}^{(20)} \\ &+ E_{disp}^{(20)} + E_{exch-disp}^{(20)} + \delta E_{HF}^{(2)} \\ &+ E_{elst,resp}^{(12)} + E_{exch}^{(11)} + E_{exch}^{(12)} + E_{ind}^{(22)} + E_{exch-ind}^{(22)} \\ &+ E_{disp}^{(21)} + E_{disp}^{(22)} + E_{disp}^{(30)} + E_{elst,resp}^{(13)} \end{aligned} \quad (8)$$

2.4.1 SAPT(DFT)

An alternative, Kohn-Sham based variant of SAPT (SAPT(DFT)) was also used to decompose the two-body and three-body interaction energies[32,46,47]. In SAPT(DFT), the Fock operator F in the partitioned Hamiltonian is replaced by the Kohn-Sham operator $K = K_A + K_B + K_C$. Because the Kohn-Sham operator includes effective electron correlation, the intramonomer electron correlation operator W can be neglected altogether, making SAPT(DFT) scale much better with system size as compared to wavefunction-based SAPT[46]. In the early studies of the two-body SAPT(DFT) method, the observed inaccurate interaction energies were attributed to the common DFT functionals being unable to provide exchange-correlation potentials with the correct asymptotic behavior[32,46]. This problem can be remedied by performing the SAPT(DFT) calculations (using the PBE0 functional as recommended in the literature[48]) with an asymptotically corrected exchange-correlation potential. To do this, the difference between the highest occupied molecular orbital (HOMO) energies of the individual monomers and their true ionization potentials was used. However, the asymptotic correction alone is still insufficient to obtain an accurate description of the complete interaction energy. The induction and dispersion SAPT(DFT) terms, both two-body and three-body, need to be computed from monomer frequency-dependent density susceptibilities (density response functions) within the coupled KS (CKS) formalism.

Moving onto the SAPT(DFT) description of nonadditive three-body effects[47], it is important to note that electrostatics and second-order dispersion are rigorously pairwise additive and do not constitute part of the three-body interaction energy; thus, one needs to go to the third order of SAPT to estimate nonadditive dispersion. The bulk of the nonadditive interaction energy is typically accounted for by induction; additionally, all SAPT exchange corrections are nonadditive starting from the first-order one. The exchange counterparts of the three-body dispersion and induction terms are scaled according to the ratios of the CKS and KS values for the corresponding induction and dispersion corrections (equations (9) and (10), with the additional subscript ",3" denoting the nonadditive three-body correction):[46,47]

$$\widetilde{E}_{exch-ind,3}^{(2)}(\mathrm{CKS}) = E_{exch-ind,3}^{(2)}(\mathrm{KS}) \frac{E_{ind,3}^{(2)}(\mathrm{CKS})}{E_{ind,3}^{(2)}(\mathrm{KS})} \quad (9)$$

$$\widetilde{E}_{exch-disp,3}^{(2)}(\mathrm{CKS}) = E_{exch-disp,3}^{(2)}(\mathrm{KS}) \frac{E_{disp,3}^{(3)}(\mathrm{CKS})}{E_{disp,3}^{(3)}(\mathrm{KS})} \quad (10)$$

For two-body SAPT2+(3) and SAPT(DFT), the calculations used the aTZ basis set (with the aTZ-PP effective core potential (ECP) for the iodine atom). Furthermore, density fitting with the aTZ-RI (SAPT(DFT)) and def2-QZVPP-RI (SAPT2+(3)) auxiliary bases was used to reduce the computational cost. The three-body SAPT(DFT) calculations, performed with the three-body code[47] from the SAPT2008 suite[38], are computationally demanding, making the aDZ basis the only practical option to accommodate larger complexes. The truncated jun-cc-pVDZ[49] basis set was used for one complex (**6**) as it was not feasible to run it in the fully augmented basis, and a fully decontracted aug-cc-pVTZ-DK3 set was used for the iodine atom in complex **1** as this calculation does not support ECPs. Both the two-body and three-body SAPT(DFT) methods require an asymptotic correction to overcome the qualitative inaccuracy of the exchange-correlation potential and kernel. In the three-body SAPT(DFT) method, we employed the Fermi-Amaldi asymptotic correction, which involves shifting the appropriate potential for the asymptotic region by $(I+\varepsilon_{HOMO})$ and smoothly splicing[50] it to match a short-range potential without derivative discontinuity[46,51]. The two-body SAPT(DFT) method employed the gradient-regulated asymptotic connection (GRAC) procedure. This scheme enables the construction of smooth asymptotically corrected potentials[52,53] by splicing the PBE0 functional with the asymptotically correct LB94 functional developed by van Leeuwen and Baerends[54].

The characterization of each of the complexes in terms of two-body and three-body interaction types was done based on the effects that dominated the binding (attractive) contribution. The numerical values in two-body SAPT2+(3) and SAPT(DFT) calculations are obtained by summing the interaction energy components from each dimer in the trimer. The nonadditive interaction energy is entirely obtained from the three-body SAPT(DFT) calculation, with second-order exchange terms scaled according to Eqs. (9) and (10):

$$E_{int,3}^{\mathrm{SAPT(DFT)}} = E_{exch,3}^{(1)}(\mathrm{KS}) + E_{ind,3}^{(2)}(\mathrm{CKS}) + \widetilde{E}_{exch-ind,3}^{(2)}(\mathrm{CKS}) + E_{disp,3}^{(3)}(\mathrm{CKS}) + \widetilde{E}_{exch-disp,3}^{(2)}(\mathrm{CKS}) + \delta E_{int,3}^{\mathrm{HF}} \quad (11)$$

In equation (11), higher-order terms derived for the three-body wavefunction-based SAPT, such as third-order induction and induction-dispersion corrections[55], are not included as they are likely mostly quenched by their exchange counterparts which have not been derived. Therefore, the higher-order nonadditive induction and exchange-induction effects are estimated with supermolecular HF[46,47]. This correction covers the difference between the nonadditive HF interaction energy term and the sum of its SAPT contributions, the first-order exchange and second-order induction and exchange-induction:

$$\delta E_{int,3}^{\mathrm{HF}} = E_{int,3}^{\mathrm{HF}} - E_{exch,3}^{(10)} - E_{ind,3}^{(20)} - E_{exch-ind,3}^{(20)} \quad (12)$$

Here, $E_{int,3}^{\mathrm{HF}}$ represents the nonadditive interaction energy from a supermolecular Hartree-Fock approach, and the other components are obtained from the three-body SAPT0 calculations.

2.4.2 Extended SAPT (XSAPT)

The computational cost of performing SAPT calculations quickly becomes prohibitive beyond simple molecular trimers. Employing a different flavor of SAPT (XSAPT) that incorporates many-body polarization into the monomer wavefunctions attempts to remedy this issue, also allowing the SAPT calculations to be extended to an arbitrary number of monomers[34,56]. This many-body polarization is included within the explicit polarization (XPol) scheme[57] by partitioning the system into noncovalently interacting fragments and performing DFT or HF calculations for each fragment in the embedding of point charges representing the electric field of its neighbors. This process is iterated until the atomic charges resulting from the DFT/HF electron density are the same as the atomic charges defining the embedding. At convergence, the densities and orbitals are used to compute standard pairwise-additive SAPT0 or SAPT(KS) energy corrections. Thanks to the XPol scheme, these corrections additionally contain both nonadditive (three- and higher-body) induction and some higher-order two-body induction effects, for example, the third-order term resulting from the density of molecule A polarized by unperturbed molecule B interacting with the density of molecule B polarized by unperturbed molecule A[58]. The dispersion and exchange-dispersion contributions in XSAPT are not calculated explicitly, but are instead taken from the many-body dispersion (MBD) model of Tkatchenko and co-workers[30]. This leads to significant computational savings (as dispersion and exchange-dispersion terms are by far the most time-consuming part of SAPT0 or SAPT(KS)), avoids the prevalent overestimation of dispersion by SAPT0 (note that the SAPT(KS) dispersion is no better unless it is computed in the CKS formalism like in SAPT(DFT)), and improves basis set convergence[59]. The resulting XSAPT method provides an estimate of the nonadditive three-body induction (from the XPol embedding) and dispersion (from MBD) but neglects the nonadditive first-order exchange.

The total interaction energy in XSAPT is computed as

$$E_{int}^{\mathrm{XSAPT}} = E_{elst}^{(1)} + E_{exch}^{(1)} + E_{Disp}^{(2)} + [E_{ind}^{(2)} + E_{exch-ind}^{(2)} + \sum_A \sum_{B>A} \delta E_{AB}^{\mathrm{HF}} + \sum_A \sum_{B>A} (E_{AB}^{\mathrm{XSAPT}} - E_{AB}^{\mathrm{SAPT(KS)}}) + E_{int}^{\mathrm{MB}}] \quad (13)$$

where

$$\delta E_{AB}^{\mathrm{HF}} = E_{int}^{HF} - (E_{elst}^{(10)} + E_{exch}^{(10)} + E_{ind,resp}^{(20)} + E_{exch-ind,resp}^{(20)}) \quad (14)$$

with all terms on the r.h.s. computed for the dimer AB. The $\delta E_{AB}^{\mathrm{HF}}$ term takes into account higher-order induction effects for the A-B pair. Thus, it is assumed that, for systems with multiple bodies (molecules), this correction can be treated as a sum of pairwise contributions ($\delta E_{int}^{\mathrm{HF}} = \sum_A \sum_{B>A} \delta E_{AB}^{\mathrm{HF}}$). The terms in Eq. (13) were obtained from various stages of the XSAPT calculation. The first-order electrostatics and exchange terms, and a part of the overall (induction plus exchange-induction) effect, were obtained from SAPT(KS) calculations performed without electrostatic embedding. The $\delta E_{int}^{\mathrm{HF}}$ effect, which is classified as part of the overall induction energy, was computed from pairwise HF and SAPT0 cal-

Table 1 Benchmark two-body and three-body interaction energy data (kcal/mol). The CCSD(T)/CBS estimate was obtained in a composite manner according to Eq. (4). The numbering of complexes is specified in Figs. 1 and 2.

Complex	MP2/CBS			CCSD(T)/CBS			Chemical Formula	Interaction Type[a]
	E^2_{int}	E^3_{int}	E_{int}	E^2_{int}	E^3_{int}	E_{int}		
1.1	−9.917	−0.563	−10.480	−9.772	−0.546	−10.317	$CH_3I-H_2CO-H_2O$	HB
1.2	−6.698	−0.577	−7.275	−6.466	−0.574	−7.041	$CH_3I-H_2CO-H_2O$	HB,XB
2.1	−8.306	−0.428	−8.734	−8.293	−0.405	−8.697	$CF_3Cl-H_2CO-H_2O$	HB
2.2	−8.079	−0.566	−8.645	−8.129	−0.559	−8.688	$CF_3Cl-H_2CO-H_2O$	HB,XB
3.1	−14.601	−2.491	−17.091	−14.573	−2.421	−16.994	$CH_3OH-H_2O-H_2O$	HB
3.2	−12.561	−1.619	−14.180	−12.523	−1.593	−14.115	$CH_3OH-H_2O-H_2O$	HB
4.1	−40.779	3.387	−37.391	−39.618	3.386	−36.232	$NH_4^+-C_6H_6-H_2O$	HB,cation-π
4.2	−28.656	−4.641	−33.297	−27.766	−4.623	−32.389	$NH_4^+-C_6H_6-H_2O$	HB,OH-π
5.1	−19.492	1.225	−18.267	−18.888	1.338	−17.550	$Cl^--NH_3-C_6H_6$	HB,NH-π
5.2[b]	−13.798	1.127	−12.671	−12.974	1.171	−11.803	$Cl^--NH_3-C_6H_6$	HB,NH-π,anion-π
6.1	−11.479	0.232	−11.248	−7.206	0.347	−6.859	$C_6H_5Br-(CH_3)_3N-C_6H_6$	π-π, XB
6.2	−10.195	−0.227	−10.422	−7.211	−0.116	−7.327	$C_6H_5Br-(CH_3)_3N-C_6H_6$	CH-π
7.1	−15.593	−1.677	−17.271	−12.968	−1.559	−14.527	$C_6H_5OH-C_6H_6-H_2O$	HB,π-π,OH-π
7.2	−9.905	0.013	−9.891	−6.240	−0.009	−6.249	$C_6H_5OH-C_6H_6-H_2O$	HB,π-π,OH-π
8.1	−40.830	3.746	−37.084	−37.576	3.789	−33.788	$C_6H_5Cl-C_6H_6-NH_4^+$	HB,CH-π,cation-π
8.2	−40.069	4.084	−35.986	−37.114	4.152	−32.962	$C_6H_5Cl-C_6H_6-NH_4^+$	HB,cation-π
9.1	−8.193	−0.726	−8.919	−7.261	−0.669	−7.930	$HCCH-C_6H_6-H_2O$	OH-π,CH-π
9.2	−8.090	−0.550	−8.640	−6.883	−0.510	−7.392	$HCCH-C_6H_6-H_2O$	OH-π,CH-π
10.1	−9.655	−0.300	−9.955	−6.746	−0.200	−6.947	$HCCH-C_6H_6-C_6H_6$	CH-π,π-π
10.2	−8.452	−0.042	−8.494	−4.975	0.056	−4.919	$HCCH-C_6H_6-C_6H_6$	CH-π,π-π

[a]HB and XB stand for hydrogen and halogen bonding, respectively.
[b]This structure, unlike all the other ones, is a saddle point rather than a minimum.

culations. The remaining parts of the XSAPT induction energy in Eq. (13) come from the effects of mutual polarization accounted for by the XPol embedding. Such effects can be partitioned into a pairwise additive and nonadditive part. The former is computed as a difference between the "total energies" of each pair computed using XSAPT (with the XPol embedding) and SAPT(KS) (without embedding) and thus takes into account some of the higher-order but two-body induction and exchange-induction effects. Note that the "total energies", the sums of monomer energies and the interaction term, are not commonly employed in the SAPT formalism, but are necessary in XSAPT as the XPol embedding effectively mixes terms of different orders of perturbation theory. The nonadditive part of the embedding effect, denoted E^{MB}_{int} in Eq. (13), contains the additional $E^{\mathrm{XSAPT}} - E^{\mathrm{SAPT(KS)}}$ "total energy" difference for the trimer beyond the differences resulting from the pairwise dimer calculations. Thus, this effect approximates the actual induction nonadditivity in the trimer, in all orders of perturbation theory. Finally, the dispersion term $E^{(2)}_{Disp}$ (including, effectively, the exchange-dispersion contribution) was exclusively treated with the many-body dispersion model[30] (note that, in the MBD context, the bodies are individual atoms; in this sense, MBD accounts for both intra- and intermolecular dispersion nonadditivity).

The XSAPT approach is also capable of computing additional three-body induction couplings, derived and implemented in Ref. 60. These couplings do not occur in dimer calculations and can be exclusively attributed to the nonadditive part of the interaction energy. Note that the calculation of the couplings is quite cumbersome as they are not available in the newer and more efficient, atomic orbital-based implementation of XSAPT[61], and one has to use the older molecular-orbital code. The XSAPT results were computed in the aTZ basis set for all atoms other than iodine and a fully decontracted aug-cc-pVTZ-DK3 set for iodine. In the XSAPT data below, the three-body induction couplings will be omitted unless explicitly stated otherwise. Overall, the XSAPT treatment includes the three-body nonadditivity in induction/exchange-induction (through the effects of the XPol embedding and possibly also the three-body induction couplings) and dispersion/exchange-dispersion (through the MBD formalism), but neglects the nonadditivity in the first-order exchange energy. At the two-body level, one might wonder if the addition of the $\delta E^{\mathrm{HF}}_{int}$ correction to XSAPT is rigorous, as this term accounts for some of the same higher-order induction effects as the XPol embedding. However, any possible double counting seems to be practically inconsequential and the addition of $\delta E^{\mathrm{HF}}_{int}$ substantially improves the XSAPT accuracy for hydrogen-bonded complexes[59]. At the nonadditive three-body level, the $\delta E^{\mathrm{HF}}_{int}$ term does not contribute so no double counting can occur.

3 Results and Discussion

3.1 Benchmark Interaction Energies

The construction of the 3BHET dataset was targeted at creating a dictionary of complexes that provides an example translation between the chemical, qualitative interaction types and the numerical, quantitative data. Such a dictionary tries to match the SAPT decompositions to the interaction types present in the trimolecular complex. For this purpose, first, reference energies were established by performing single-point interaction energy calculations on two minimum geometries for each complex (in one case, only one minimum was found so we picked a high-symmetry saddle

point geometry as the second one).

Table 1 provides a summary of the CCSD(T)/CBS benchmark total and nonadditive interaction energies. The complete interaction energies range from −36.2 kcal/mol down to −4.9 kcal/mol. Complexes with an ionic monomer exhibited the largest binding energies whereas $\pi-\pi$ stacked complexes showed interaction energies that were smaller in magnitude as compared to the rest of the dataset. The benchmark nonadditive contributions in Table 1 can be positive or negative and are up to 4–5 kcal/mol in magnitude, with the largest contributions observed for complexes where one of the monomers is ionic. The relative magnitude of the nonadditive terms ranges from nearly zero for complex **7.2** (in which two of the molecules are completely separated by the third one), to 1–5% for nonpolar systems **6** and **10**, to up to 14% for the hydrogen-bonded and/or ionic systems **3** and **4**.

The benchmark data in Table 1 allow us to estimate the accuracy of some approximate methods of calculating the total interaction energy and its nonadditive component. For this purpose, we use the mean unsigned errors (MUE) and mean unsigned relative errors (MURE) relative to the benchmark values. In this way, in Table 2 we quantify the accuracy of several approximate methods in reproducing the benchmark data. The MP2/CBS data include the correlation energy extrapolated from aQZ and a5Z, while the SAPT2+(3) and SAPT(DFT) calculations were performed in aTZ and aDZ, respectively. However, three-body SAPT(DFT) calculations of one of the larger complexes (number 6) was only feasible in the jun-cc-pVDZ basis set.

Table 2 shows the errors averaged over the van der Waals minimum geometries for the 20 trimers considered in this work (two structures for each system). The SAPT2+(3) approach is only available for two-body interaction energies, and we did not calculate three-body uncoupled dispersion needed for the SAPT0 E^3_{int}, so the three-body errors for these two methods are omitted.

Table 2 Error statistics of two-body, nonadditive three-body, and total interaction energies for different methods with CCSD(T)/CBS as the reference for 20 trimolecular complexes studied in this work. The MUE data are in kcal/mol and MURE in percent.

Method	E^2_{int}MUE	E^3_{int}MUE	E_{int}MUE	E_{int}MURE
HF/a5Z	10.154	0.153	10.030	104.318
SAPT0	2.818		2.957	27.361
XSAPT	2.008	0.205	1.873	19.240
XSAPT + $3b_{ind}$	2.008	0.206	2.059	19.954
MP2/CBS	1.613	0.055	1.665	18.600
SAPT(DFT)	1.720	0.085	1.710	12.814
SAPT2+(3)	0.280		1.391	7.823
CCSD(T)[a]	1.589	0.014	1.582	15.307

[a] Pure CCSD(T) in the same aDZ/aTZ basis as used for the $\delta E^{\mathrm{CCSD(T)}}$ term in the benchmark data.

Upon examining the error statistics in Table 2, it is evident that XSAPT outperforms SAPT0 (a 19% vs. 27% MURE for the total interaction energy). However, the performance of XSAPT is inferior compared to SAPT(DFT), especially for the nonadditive three-body component. Moreover, the inclusion of three-body induction couplings actually makes the XSAPT nonadditive contributions and total interaction energies slightly worse (of course, this inclusion does not change the two-body XSAPT values). Among the approximate methods in Table 2, the best performer for the two-body interaction energy is by far SAPT2+(3). This excellent accuracy has to be to some extent coincidental, but it makes SAPT2+(3) give the lowest MUE (1.4 kcal/mol) and MURE (7.8%) for the total energy even though none of the nonadditive effects are recovered. Concerning the latter effects, even the simple HF method provides a decent description of the three-body term, with a MUE of 0.15 kcal/mol (although with a large MURE of 67% indicating that small nonadditive effects are not well described). Among the SAPT flavors studied, only SAPT(DFT) leads to more accurate three-body energies than HF, but the two correlated parts of the composite benchmark data, MP2/CBS and CCSD(T)/aDZ (aTZ for a few systems), each lead to even smaller three-body errors. The good performance of MP2/CBS for E^3_{int} indicates that nonadditive dispersion, missing at the MP2 level, is not particularly important for our systems.

3.2 Potential Energy Curves

Potential energy curves for three selected complexes were obtained through displacing a single molecule from the equilibrium geometry by scaling the intermolecular center-of-mass distance, keeping the separation and mutual orientation of the other two molecules unchanged. In this way, three 1D potential curves are generated passing through a particular global/local minimum. These curves have been computed at the same CCSD(T)/CBS level as all the minimum structures. In Figures 3–5, we highlight the potential energy curves of three complexes that showcase cation-π, halogen, and hydrogen bonding interactions. Furthermore, the interaction energy along these curves has been decomposed using two-body SAPT(DFT) – see the Supplementary Information. The intermolecular distances R reported in these figures are scaled relative to the global-minimum separation: therefore, 1.0 on the horizontal scale always represents the minimum configuration which is the same for all three curves (which correspond to the displacements of a different monomer relative to the other two). The interactions tend to be relatively small at larger separations but do not tend to zero. This is because the pairwise interaction between two monomers remains as only the third monomer is shifted away. At short-range, as expected, all interaction energies are seen to be positive due to the orbital overlap among the monomers and the resulting exchange repulsion.

Figure 3 shows the 1D potential energy curves for the ammonium-benzene-water complex. The strongest repulsive interaction occurs when the water molecule is brought closer from the equilibrium distance. When the NH_4^+ ion is moved away from the benzene-water dimer, the interaction energy decays slowly, as $1/R^2$, as it is dominated by the charge-dipole interaction. The closeness of two curves in Fig. 3a is accidental and we verified that the results diverge more at larger distances: after all, the decaying term in the interaction is dominated by a charge-dipole contribution when the water molecule is moving away, and a charge-quadrupole one when the benzene molecule is shifted away.

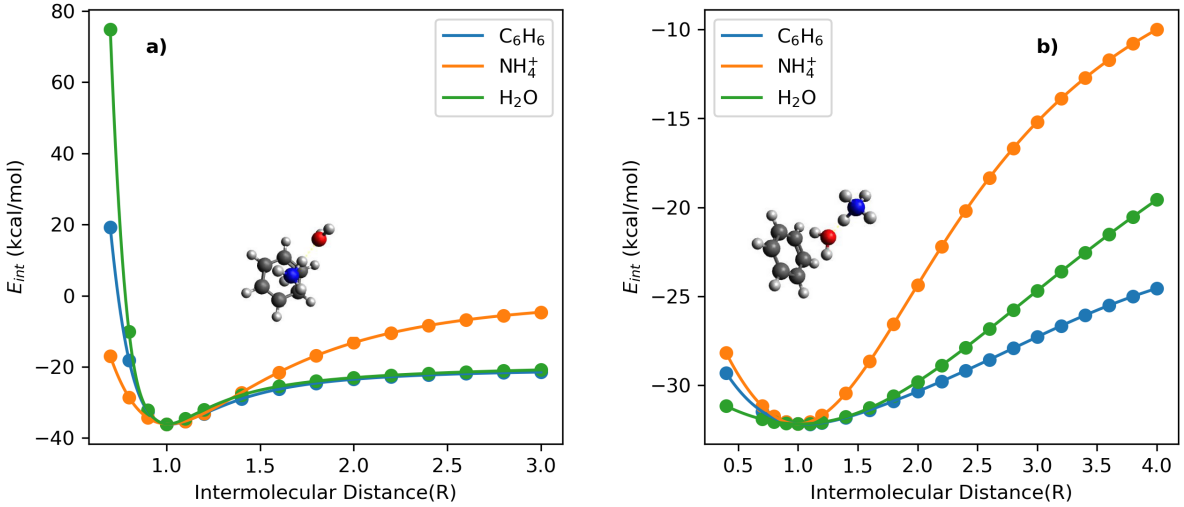


Fig. 3 Radial benchmark potential energy curves (in kcal/mol) for the ammonium-benzene-water complex 4.1 **(a)** and 4.2 **(b)**. The curves are obtained by shifting C_6H_6, NH_4^+, and H_2O relative to the center of mass of the entire complex. The relative geometry of the other two molecules stays unchanged for each panel.

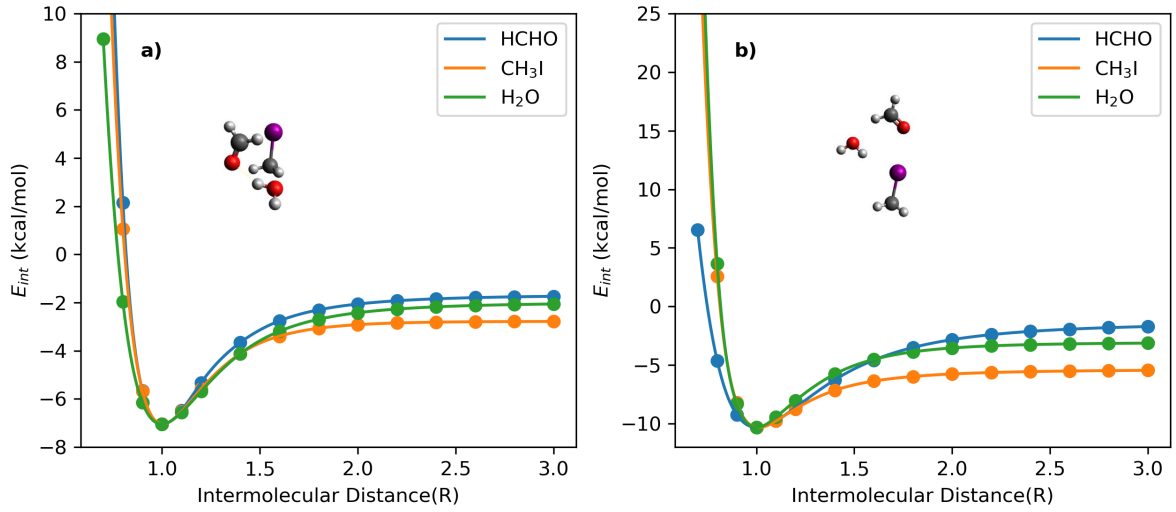


Fig. 4 Radial benchmark potential energy curves (in kcal/mol) for the iodomethane-formaldehyde-water complex 1.1 **(a)** and 1.2 **(b)**. The curves are obtained by shifting HCHO, H_2O, and CH_3I relative to the center of mass of the entire complex. The relative geometry of the other two molecules stays unchanged for each panel.

At equilibrium, the most dominant two-body interaction in the iodomethane-formaldehyde-water complex (Fig. 4) is electrostatics, followed by dispersion and induction. This ordering holds for both the halogen-bonded structure 1.2 and the non-halogen-bonded configuration 1.1. The largest contributions to both electrostatics and dispersion come from the hydrogen and halogen bonds.

The water-water-methanol complex (Fig. 5) resembles the well known water trimer. Each monomer simultaneously acts as a H-bond donor and an H-bond acceptor in the cyclic trimer minimum geometry, and the fact that the non-hydrogen-bonded H atom of one of the water molecules is replaced by a methyl group brings a minimal perturbation to the bonding pattern of a water trimer. The interactions at long range are dominated by the dipole-dipole term, decaying as $1/R^3$, and this decay is very similar when each monomer, water or methanol, is shifted away from the other two, as indicated by the strong similarity of the three curves in Figure 5a. In the second, 3.2 structure of this complex (Fig. 5b), one of the hydrogen bonds is replaced by a weaker C-H$\cdots$O contact; as a result, the long-range interaction energy is more sensitive to the displacement of the water molecule taking part in both hydrogen bonds.

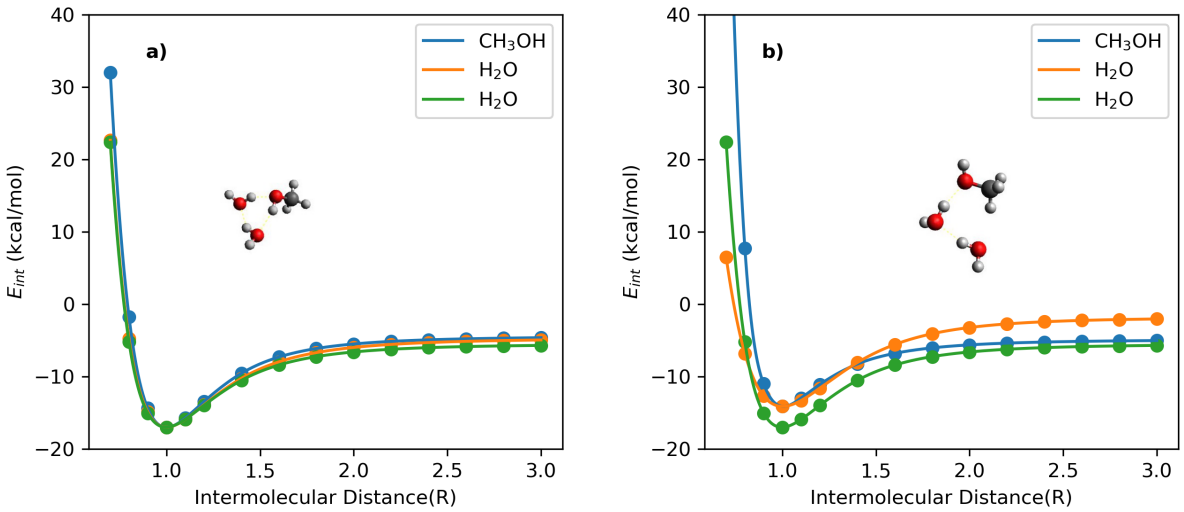


Fig. 5 Radial benchmark potential energy curves (in kcal/mol) for the water-water-methanol complex 3.1 **(a)** and 3.2 **(b)**. The curves are obtained by shifting CH_3OH, H_2O, and H_2O relative to the center of mass of the entire complex. The relative geometry of the other two molecules stays unchanged for each panel.

3.3 SAPT analysis of two-body interaction energy

The two-body interaction energy components for the twenty 3BHET structures are presented in Table 3 for SAPT(DFT), SAPT2+(3), and XSAPT. The physical meaning of these two-body components can be interpreted precisely. In the 3BHET dataset, the electrostatic interactions are all attractive because we picked favorable global- and local-minima geometries. First-order exchange interactions in dimers are always repulsive while the net induction and dispersion interactions are attractive (that is, the positive exchange-induction and exchange-dispersion terms attenuate, but not cancel, the negative induction and dispersion contributions). The magnitudes of individual terms vary depending on the geometry and complex size. The aromatic complexes are largely dispersion-bound while the complexes involving ions and the systems with strong hydrogen bonding are primarily electrostatics-bound. In addition, the systems 4, 5, and 8 exhibit large induction interaction energies due to the presence of charged species. The induction energy computed with all SAPT variants does include an appropriate δ^{HF} term, which accounts for higher-order induction and exchange induction interactions along with some small contributions from first-order effects[62].

As far as the three SAPT flavors in Table 3 are concerned, the best agreement between the first-order components occurs between SAPT2+(3) and XSAPT (with mean absolute deviations (MAD) of 0.3 kcal/mol for electrostatics and 0.7 kcal/mol for exchange), but the agreement with SAPT(DFT) is not much worse (with MAD values up to 0.5 kcal/mol for electrostatics and 1.6 kcal/mol for exchange). The second-order contributions deviate between the methods to a similar extent, with MAD values in the range 0.8–1.3 kcal/mol for induction (the largest discrepancy occurs between XSAPT and SAPT(DFT)) and 1.3–2.7 kcal/mol for dispersion (with the largest discrepancy between XSAPT and SAPT2+(3)).

The total two-body interaction energies from all three SAPT variants, shown in Table 4, exhibit varying magnitudes across the methods. The agreement between the total two-body energies is perhaps lower than expected, with an average absolute deviation between two "higher-level" methods, SAPT(DFT) and SAPT2+(3), amounting to 1.5 kcal/mol. The XSAPT two-body energies deviate on the average by 1.7 and 2.0 kcal/mol from SAPT(DFT) and SAPT2+(3), respectively.

3.4 SAPT analysis of three-body nonadditive interaction energy

For the decomposition of the three-body nonadditive interaction energy, we can no longer use the (purely two-body) SAPT2+(3) variant, so we now focus on SAPT(DFT) and XSAPT. The individual three-body components for both methods are given in Table 5. Note that while the former method includes the leading-order nonadditivity in all primary components except for the (entirely pairwise additive) electrostatics, the XSAPT approach neglects the nonadditivity of first-order exchange and includes the nonadditive induction (with or without three-body induction couplings) and dispersion (the part of the MBD value for the trimer that cannot be explained by pairwise MBD terms). For the 20 structures in 3BHET, the three-body exchange effects are always attractive, slightly lowering the two-body exchange repulsion. The nonadditive dispersion term is repulsive (again compensating a small part of the two-body dispersion attraction) while the largest one of the three, the nonadditive induction, can be both attractive and repulsive, and sometimes changes sign between two configurations of the same complex. While the nonadditive induction term ranges from -3.7 to 4.1 kcal/mol for the trimers in our dataset, the magnitude of the three-body exchange and dispersion does not exceed 0.9 and 0.4 kcal/mol, respectively.

Figure 6 displays a more extensive comparison of the total nonadditive induction energy, including all available minor variants

Table 3 Two-body SAPT(DFT), SAPT2+(3), and XSAPT interaction energy contributions (kcal/mol). The capitalized component names indicate the inclusion of the corresponding exchange effects, e.g., $E^{(2)}_{Disp} = E^{(2)}_{disp} + E^{(2)}_{exch-disp}$.

Complex	2-body SAPT(DFT)				SAPT2+(3)				XSAPT			
	$E^{(1)}_{elst}$	$E^{(1)}_{exch}$	$E^{(2)}_{Ind}$	$E^{(2)}_{Disp}$	$E^{(1)}_{elst}$	$E^{(1)}_{exch}$	$E^{(2)}_{Ind}$	$E^{(2)}_{Disp}$	$E^{(1)}_{elst}$	$E^{(1)}_{exch}$	$E^{(2)}_{Ind}$	$E^{(2)}_{Disp}$
1.1	−14.245	17.860	−4.143	−7.776	−14.223	18.468	−5.024	−8.823	−14.529	18.474	−4.803	−6.569
1.2	−9.408	11.778	−2.716	−5.679	−9.235	12.630	−3.361	−6.508	−9.671	13.651	−3.160	−5.130
2.1	−11.377	13.732	−3.086	−6.017	−11.572	13.865	−3.663	−6.811	−11.535	14.128	−3.325	−4.651
2.2	−11.708	13.867	−3.280	−5.559	−11.859	13.918	−3.879	−6.267	−11.907	14.323	−3.522	−4.167
3.1	−27.738	32.119	−7.842	−8.523	−27.965	32.758	−9.021	−9.672	−28.002	32.224	−8.817	−6.211
3.2	−22.049	25.231	−6.663	−6.996	−22.260	25.597	−7.632	−7.885	−22.311	25.509	−7.572	−4.904
4.1	−33.865	27.709	−21.270	−9.076	−34.310	28.422	−23.591	−9.915	−33.631	27.782	−26.704	−6.229
4.2	−35.256	36.690	−17.519	−9.568	−35.631	37.679	−19.441	−10.706	−35.237	37.004	−21.034	−6.723
5.1	−17.677	16.550	−11.656	−8.850	−18.967	21.906	−12.325	−10.125	−19.233	21.964	−11.350	−9.052
5.2	−9.315	11.925	−8.747	−8.278	−10.812	16.291	−9.424	−9.501	−11.158	16.295	−8.317	−7.821
6.1	−13.641	31.477	−3.521	−19.139	−13.752	32.472	−4.023	−21.895	−13.109	30.389	−3.320	−18.343
6.2	−11.752	27.635	−3.000	−17.704	−11.306	27.295	−3.367	−19.991	−11.161	26.238	−2.899	−16.683
7.1	−20.326	28.783	−6.928	−15.978	−21.207	33.784	−7.688	−17.921	−21.235	32.884	−7.632	−14.377
7.2	−8.657	18.712	−2.946	−14.438	−9.757	23.075	−3.143	−16.403	−9.350	21.333	−2.860	−13.356
8.1	−25.819	32.586	−24.645	−15.779	−26.335	34.119	−27.461	−17.623	−25.655	33.056	−29.759	−13.378
8.2	−23.020	27.384	−24.237	−13.816	−22.964	29.040	−26.997	−14.957	−22.587	27.557	−30.291	−10.975
9.1	−10.164	14.012	−3.090	−7.854	−10.158	14.860	−3.539	−8.645	−10.311	14.964	−3.464	−6.699
9.2	−9.059	12.928	−3.017	−8.175	−8.799	13.925	−3.308	−8.918	−8.826	13.976	−3.258	−7.489
10.1	−9.404	18.945	−2.956	−13.427	−9.150	20.161	−3.301	−14.804	−9.003	19.374	−3.226	−13.346
10.2	−8.083	19.684	−2.726	−13.799	−8.198	21.449	−2.983	−15.310	−7.896	19.989	−2.890	−13.488

Table 4 Total two-body and three-body SAPT interaction energies (kcal/mol). The benchmark CCSD(T)/CBS values from Table 1 are copied here to facilitate comparisons.

Complex	SAPT(DFT)			SAPT2+(3)			XSAPT			CCSD(T)/CBS
	E^2_{int}	E^3_{int}	E_{int}	E^2_{int}	E^2_{int}	E^3_{int}	E_{int}	E^2_{int}	E^3_{int}	E_{int}
1.1	−8.303	−0.582	−8.885	−9.603	−7.428	−0.505	−7.933	−9.772	−0.546	−10.317
1.2	−6.025	−0.538	−6.563	−6.474	−4.311	−0.396	−4.707	−6.466	−0.574	−7.041
2.1	−6.747	−0.408	−7.155	−8.182	−5.383	−0.381	−5.764	−8.293	−0.405	−8.697
2.2	−6.680	−0.533	−7.214	−8.086	−5.274	−0.473	−5.747	−8.129	−0.559	−8.688
3.1	−11.984	−2.547	−14.531	−13.899	−10.805	−2.264	−13.069	−14.573	−2.421	−16.994
3.2	−10.477	−1.708	−12.185	−12.180	−9.278	−1.649	−10.927	−12.523	−1.593	−14.115
4.1	−36.502	3.386	−33.116	−39.394	−38.782	2.682	−36.100	−39.618	3.386	−36.232
4.2	−25.653	−4.524	−30.176	−28.098	−25.989	−4.827	−30.815	−27.766	−4.623	−32.389
5.1	−21.632	1.694	−19.939	−19.511	−17.670	1.026	−16.645	−18.888	1.338	−17.550
5.2	−14.415	1.167	−13.248	−13.445	−11.001	0.795	−10.206	−12.974	1.171	−11.803
6.1	−4.824	0.436	−4.387	−7.198	−4.383	0.484	−3.900	−7.206	0.347	−6.859
6.2	−4.821	0.052	−4.768	−7.369	−4.506	−0.254	−4.760	−7.211	−0.116	−7.327
7.1	−14.448	−1.350	−15.798	−13.031	−10.360	−1.695	−12.055	−12.968	−1.559	−14.527
7.2	−7.329	0.004	−7.324	−6.228	−4.234	−0.033	−4.267	−6.240	−0.009	−6.249
8.1	−33.657	3.862	−29.795	−37.301	−35.737	3.234	−32.503	−37.576	3.789	−33.788
8.2	−33.689	4.254	−29.435	−35.878	−36.295	3.293	−33.001	−37.114	4.152	−32.962
9.1	−7.095	−0.618	−7.714	−7.482	−5.510	−0.637	−6.146	−7.261	−0.669	−7.930
9.2	−7.323	−0.491	−7.814	−7.099	−5.597	−0.536	−6.133	−6.883	−0.510	−7.392
10.1	−6.841	−0.133	−6.974	−7.093	−6.201	−0.193	−6.395	−6.746	−0.200	−6.947
10.2	−4.924	0.164	−4.761	−5.043	−4.285	0.009	−4.277	−4.975	0.056	−4.919

of XSAPT and SAPT(DFT) as well as SAPT0. For SAPT(DFT), one can include or omit the nonadditive $\delta E^{HF}_{int,3}$ term, Eq. (12). In the case of XSAPT, the HF delta term is pairwise additive and does not influence the three-body energy, but one can include or omit the three-body induction couplings[60]. Figure 6 shows that the SAPT0 approach and the SAPT(DFT) method without $\delta E^{HF}_{int,3}$ give results quite similar to each other, and smaller in magnitude than the other variants. The addition of the $\delta E^{HF}_{int,3}$ correction makes the SAPT(DFT) nonadditive induction term larger in magnitude and likely more accurate. Notably, the SAPT(DFT) values with $\delta E^{HF}_{int,3}$ are quite similar to the XSAPT data without three-body induction couplings, while the inclusion of the latter increases the magnitude of nonadditive induction even further. It should be stressed that while the description of nonadditive induction by SAPT(DFT) and XSAPT may appear similar, the total nonadditive interaction energies from SAPT(DFT) are much closer to the benchmark, cf. Table 2; the likely reason is the neglect of the nonadditive first-order exchange within XSAPT. We conclude that the iterative embedding scheme incorporated by XSAPT leads to a good description of three-body induction effects even without including the explicit three-body induction couplings; actually, the addition of the latter does not seem to improve the overall accuracy any further.

A scatter plot of the total nonadditive interaction energy

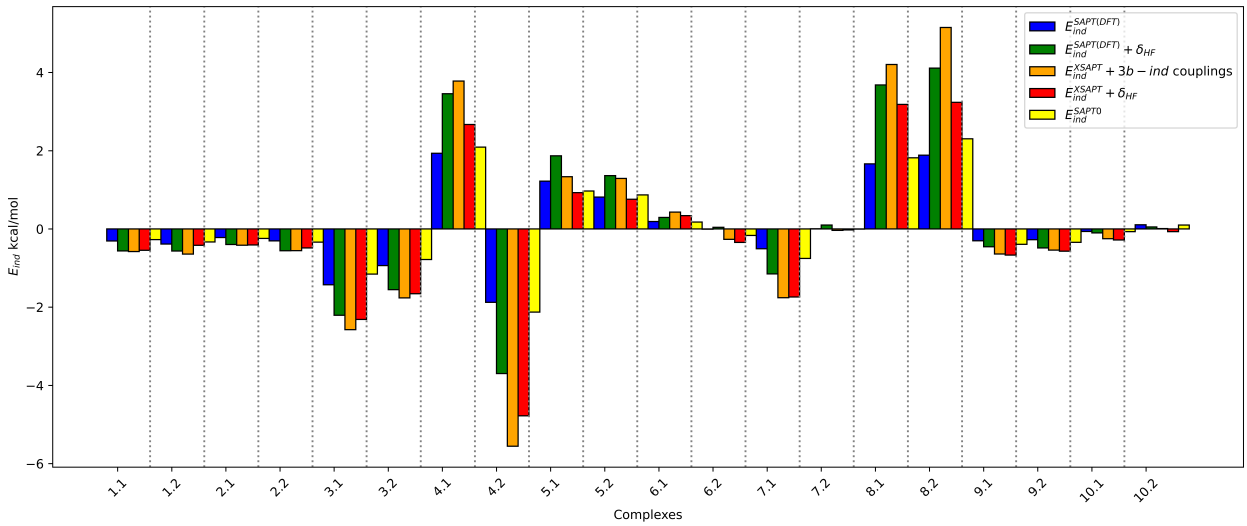


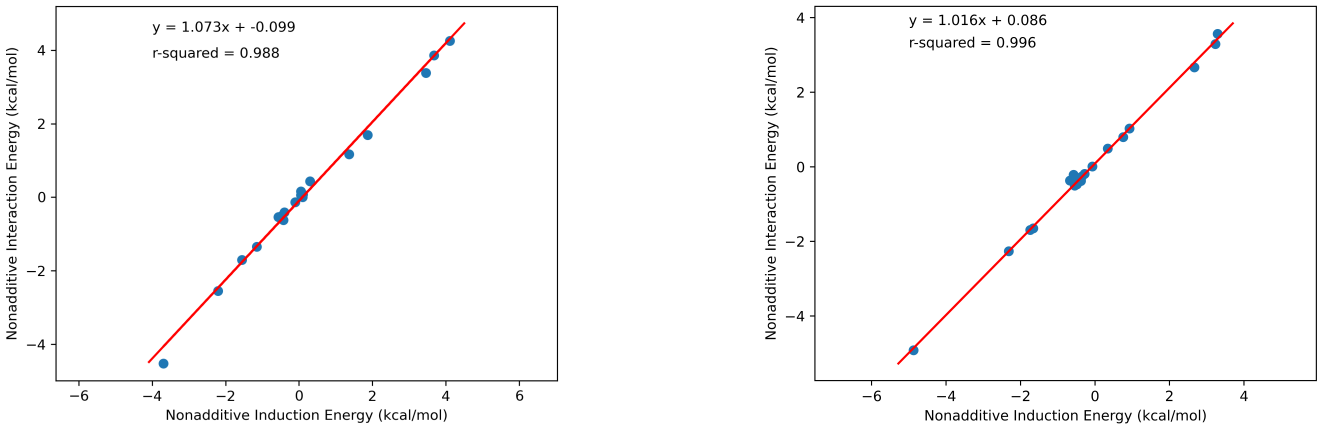
Fig. 6 Comparison of the nonadditive induction energies for the 3BHET dataset predicted by SAPT0, XSAPT, and SAPT(DFT).

Table 5 Three-body SAPT(DFT) and XSAPT interaction energy contributions (kcal/mol). The capitalized component names indicate the inclusion of the corresponding nonadditive exchange effects: specifically, $E^{(2)}_{Ind,3} = E^{(2)}_{ind,3} + E^{(2)}_{exch-ind,3}$ and $E^{(3)}_{Disp,3} = E^{(3)}_{disp,3} + E^{(2)}_{exch-disp,3}$. Three-body XSAPT induction couplings are not included.

Complex	SAPT(DFT)			XSAPT	
	$E^{(1)}_{exch,3}$	$E^{(2)}_{Ind,3}$	$E^{(3)}_{Disp,3}$	$E^{(2)}_{Ind,3}$	$E^{MBD}_{Disp,3}$
1.1	−0.172	−0.562	0.151	−0.545	0.040
1.2	−0.055	−0.566	0.083	−0.420	0.024
2.1	−0.129	−0.398	0.118	−0.409	0.029
2.2	−0.051	−0.557	0.075	−0.486	0.013
3.1	−0.573	−2.208	0.233	−2.315	0.051
3.2	−0.253	−1.553	0.098	−1.657	0.007
4.1	−0.205	3.457	0.134	2.671	0.010
4.2	−0.854	−3.697	0.027	−4.779	−0.048
5.1	−0.560	1.870	0.385	0.929	0.097
5.2	−0.434	1.362	0.239	0.760	0.035
6.1	−0.112	0.293	0.255	0.340	0.144
6.2	−0.306	0.041	0.317	−0.346	0.091
7.1	−0.501	−1.151	0.302	−1.739	0.045
7.2	−0.105	0.100	0.009	−0.026	−0.007
8.1	−0.034	3.681	0.215	3.185	0.049
8.2	−0.114	4.112	0.256	3.237	0.056
9.1	−0.326	−0.455	0.163	−0.670	0.033
9.2	−0.152	−0.485	0.147	−0.569	0.033
10.1	−0.328	−0.102	0.298	−0.283	0.090
10.2	−0.119	0.051	0.231	−0.071	0.079

against its induction component, displayed in Fig. 7, confirms that the latter accounts for the greatest percentage of the nonadditive three-body effect. Thus, accounting reliably for the nonadditive induction effects is the key to providing an accurate description of the entire E^3_{int} for the systems in our dataset, and those effects can be recovered, to a reasonable extent, already with a highly approximate method like HF. Moreover, Fig. 7 indicates that the total three-body effect is nearly proportional to its induction component, suggesting that a suitable scaling of the latter term might be a good way to approximate the entire nonadditive energy.

The different nonadditive dispersion estimates for the systems in our dataset are compared in Fig. 8. The calculations presented in this work provide three meaningful estimates of three-body dispersion: the sum $E^{(3)}_{disp,3}(\mathrm{CKS}) + \widetilde{E}^{(2)}_{exch-disp,3}(\mathrm{CKS})$ computed in three-body SAPT(DFT), the difference between the MBD dispersion energies for the trimer and dimers computed as part of XSAPT, and the difference between the supermolecular CCSD(T)/CBS and MP2/CBS nonadditive interaction energies. The latter difference is obviously not just due to three-body dispersion (for example, the improved multipole moments and polarizabilities in CCSD(T) will result in an altered value for the nonadditive induction), but since MP2 misses the pure three-body dispersion completely and CCSD(T) includes it, the difference has been used to estimate the nonadditive dispersion effects in the past[24]. Our results in Fig. 8 show that the nonadditive dispersion energy term, however small, is very hard to obtain accurately, as its three estimates computed in this work disagree with each other quite substantially. Most notably, the SAPT(DFT) estimate is consistently much larger than the other two, indicating that the CKS approach, very successful for computing two-body dispersion accurately and efficiently, might not be as reliable for three-body dispersion. This overestimation of nonadditive dispersion by SAPT(DFT) has been observed before[63] and might possibly be attributed to the lack of so-called "dispersion size-consistency" for many-molecule SAPT[64]. The fact that the SAPT(DFT) estimate requires scaling the exchange-dispersion term by the CKS/KS ratio[47] might lead to additional inaccuracies. The CCSD(T)−MP2 and XSAPT(MBD) estimates are more consistent with each other, but sometimes (e.g. for systems 4.2 and 7.1) also exhibit large relative differences. Note that the nonadditive (in the molecular sense) MBD three-body dispersion is obtained using a supermolecular approach, as the difference between the trimer energy and the sum of the pair energies, which might amplify residual errors. Overall, even though none of the methods presented in Fig. 8 can be considered an unambiguous benchmark measure



SAPT(DFT) nonadditive induction **XSAPT nonadditive induction**

Fig. 7 Total SAPT(DFT) and XSAPT nonadditive interaction energy plotted versus its respective induction contribution.

of pure nonadditive dispersion effects, the agreement between CCSD(T)−MP2 and XSAPT(MBD) for most (but not all) systems suggests that both approaches are quite trustworthy for reproducing this effect, which is small for systems in our dataset but important for many condensed-phase systems.

3.5 Total SAPT interaction energies

The total interaction energies from three SAPT variants (SAPT2+(3), SAPT(DFT), and XSAPT) for the trimers in our dataset are presented in Table 4. Overall, the discrepancies between variants can be quite substantial, up to over 8 kcal/mol (or roughly double the nonadditive three-body term) for the strongly bound ionic complex 8. When these total interaction energies are compared to the reference CCSD(T)-level data from Table 1, the two-body SAPT2+(3) method is observed to overestimate and underestimate subsets of data to a similar extent — the mean signed error (MSE) of SAPT2+(3) is only 0.01 kcal/mol. More specifically, SAPT2+(3) exhibits too strong binding (by ≥ 1.5 kcal/mol) for systems 4.1, 5.1, 5.2, 8.1, and 8.2, and too weak binding (also by ≥ 1.5 kcal/mol) for 3.1, 3.2, and 4.2. Notably, for each of the structures listed in the previous sentence, the nonadditive three-body term is sizeable and of opposite sign to the two-body error, so an inclusion of even an approximate three-body term would substantially improve the accuracy of SAPT2+(3) total interaction energies. In contrast, the SAPT(DFT) and XSAPT variants underestimate the binding in the 3BHET complexes quite systematically, with an MSE of 1.05 kcal/mol for SAPT(DFT) and 1.87 kcal/mol for XSAPT. However, the main source of this underestimation is quite different in each case, with SAPT(DFT) leading to largest errors for systems involving the ammonium cation (4 and 8), while XSAPT gives the largest errors for the strongly hydrogen-bonded system 3, followed by halogen-containing (and halogen-bonded in some configurations) systems 2 and 6. On the positive side, the errors from SAPT(DFT), as well as those from XSAPT, are generally quite consistent between the two configurations of the same trimer, implying that both approaches might be better suited to examining relative energies between such configurations.

SAPT provides physical insights to examine the nature of the two- and three-body interaction energies, and a ternary diagram [14,35] provides a simple way to graphically assign a two-body interaction type. Such a SAPT(DFT) ternary diagram for the 20-point 3BHET dataset is presented in Figure 9 (blue dots), showing that the majority of the 3BHET trimers are mainly electrostatics and dispersion-bound. Using the SAPT(DFT) decomposition of the nonadditive three-body energy into induction, dispersion, and first-order exchange, we can enhance the ternary diagram by combining the respective SAPT contributions of a given type. The resulting relative importance of the electrostatic, induction, and dispersion two- plus three-body contributions is shown by red dots in Figure 9. We see that the blue and red dots clearly come in pairs and often overlap with each other, indicating that the three-body contribution provides only a minor modification of the two-body interaction type. However, the binding decomposition in Fig. 9 can be viewed as slightly incomplete, as the nonadditive first-order exchange is not displayed while it is attractive for the complexes studied here and provides an additional minor contribution to the binding. To account for all attractive two- and three-body interactions, Figure 10 contains a modified ternary diagram with the pure electrostatic data (which are pairwise additive) replaced by the entire first-order energies (electrostatics plus exchange, where the latter is not entirely additive). Compared to Fig. 9, the addition of exchange causes an obvious shift to more repulsive values, while the two-body (blue dots) and two- plus three-body (red dots data) still form closely lying pairs. Moreover, within each pair, the two points are displaced from each other in a direction roughly towards or away from the pure induction vertex of the diagram, confirming that the largest difference between the two-body and two- plus three-body binding character stems from nonadditive induction.

Finally, the information from two- and three-body SAPT de-

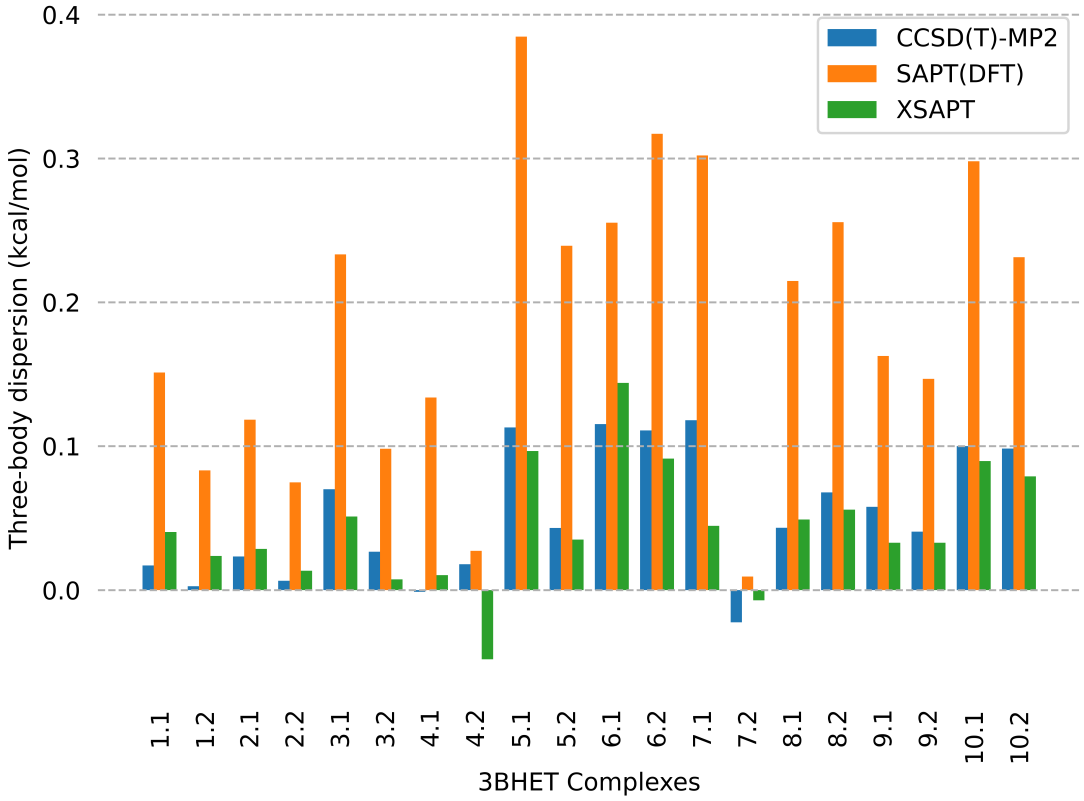


Fig. 8 Three-body nonadditive dispersion energy for the systems in the 3BHET dataset, estimated as the supermolecular CCSD(T)/CBS−MP2/CBS difference as well as computed with SAPT(DFT) and XSAPT (the latter uses the MBD formalism for the dispersion part).

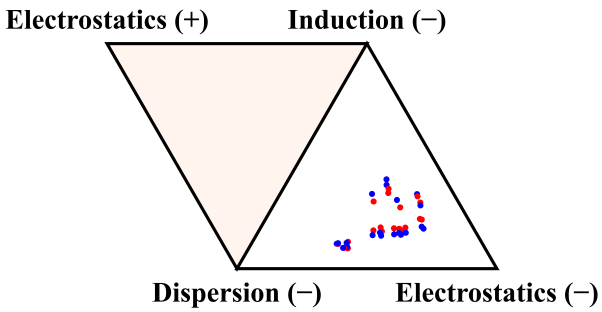


Fig. 9 Two-body (blue dots) and two- plus three-body (red dots) SAPT(DFT) interaction energy distribution for trimers studied here, displayed on a ternary diagram.

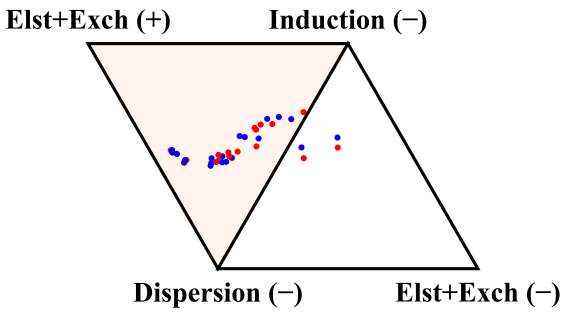


Fig. 10 Two-body (blue dots) and two- plus three-body (red dots) SAPT(DFT) interaction energy distribution, including first-order exchange effects, for trimers studied here, displayed on a ternary diagram.

compositions allows us to make connections between the chemical interaction type and the interaction energy contributions. The complexes involving ions exhibit the strongest binding overall, dominated by the two-body electrostatic and induction components. The same complexes have the largest magnitude of the three-body energy thanks to a sizable nonadditive induction. The later term can be attractive or repulsive, depending on whether the polarization of some monomer X by monomer Y makes the X–Z interaction more or less favorable. For example, the two structures of complex 4 (ammonium-benzene-water) exhibit opposite signs of nonadditive induction, which can be rationalized by looking at the most polarizable monomer (benzene) responding to the monomer producing the strongest electric field (the ammonium ion). When the π-electron density is pulled towards the positive ion, the change in the benzene-water interaction is unfavorable for 4.1 (where the negatively charged side of the water molecule points towards benzene) and favorable for 4.2 (where the positive side of water is closer to benzene), in accordance with the sign

of the nonadditive induction effect. The next strongest binding occurs in neutral complexes that are strongly hydrogen-bonded, such as 3 (methanol-water-water) and 7 (phenol-benzene-water). The induction energy no longer plays the most dominant role, and the binding in complex 3 is primarily due to electrostatics. Since complex 7 also involves two aromatic rings, dispersion energy is just as important for its stability as electrostatic energy. Looking now at complexes 1, 2, and 6 that can form halogen bonds (with a halogen atom X connected to C), we picked one halogen-bonded geometry (with the C-X bond pointing toward an electronegative atom in another molecule) and one non-halogen-bonded geometry for each system. For our complexes, no clear signature of halogen bonding is observed in the two-body energy contributions. However, the halogen bonded structures 1.2, 2.2, and 6.1 exhibit a smaller magnitude of SAPT(DFT) nonadditive exchange and dispersion corrections than their non-halogen-bonded counterparts. This might be coincidental and requires calculations on a broader class of systems to confirm any general trend. System 6 and the remaining two complexes (9 and 10) are the least polar and do not have hydrogen bonds. These complexes exhibit the weakest binding in general, and this binding is driven by dispersion followed by electrostatics. At the three-body level, induction is no longer a clearly dominant term and all three SAPT(DFT) contributions are roughly of similar importance. While nonadditive dispersion never plays a very large role in our dataset, its effect is notable in complexes 6 and 10 which involve two aromatic monomers.

4 Conclusions

Effective treatment of noncovalent interactions in many-body systems is vital as computational studies are moving towards bigger clusters and condensed phase systems. Exploring the binding mechanisms in molecular aggregates, solids, and liquids enables one to design new materials with tailored properties. However, it is essential to gather data on noncovalent interactions in vast regions of chemical space, not limited to dimers and homotrimers present in organic crystals. The new heteromolecular trimer benchmark dataset 3BHET developed here aims to aid in parameterizing and refining approximate methods that could be universally applicable throughout the chemical space. This dataset presents a diverse mixture of small and medium sized heteromolecular trimers that features various interactions involving both neutral and charged molecules. The benchmark two-body and three-body interaction energies were calculated at the CCSD(T)/CBS level.

Symmetry-adapted perturbation theory aids in the understanding and interpretation of noncovalent interaction energies at both two-body and three-body levels. Here, an extensive set of SAPT variants was tested on our new dataset, including SAPT2+(3), SAPT(DFT), and XSAPT. Overall, at the two-body level, the complexes are bound together by a varying mixture of electrostatics and dispersion, with an sizable induction effect present in complexes involving ions. The nonadditive three-body term can be favorable or unfavorable and is dominated by three-body induction. The other nonadditive effects, attractive first-order exchange and repulsive dispersion, are small but not negligible. To further elucidate the binding patterns and the degree of diversity present in our dataset, we extend the commonly used ternary diagrams from interactions in dimers to the complete two- and three-body interactions in trimers, illustrating also the importance of different nonadditive effects including induction, dispersion, and (optionally) first-order exchange. For our dataset, the balance of different pairwise and nonadditive interactions present in each trimer, despite their large chemical variety, resulted in a ternary diagram that features relatively similar points. This shows that the diversity of a dataset should be assessed by several complementary measures. A detailed and nuanced analysis of interactions provided by SAPT and ternary diagrams, in conjunction with the chemical intuition and a suitable selection of minimum and off-minimum separations and orientations, will be highly desired for modelling more diverse interactions in future benchmark studies. While the available mixed-trimer benchmark data are not yet sufficient for modelling a broad range of complexes, this study takes the first step forward by examining the relationship between the physical interaction types and quantitative two-body and three-body energy contributions on a reasonably broad range of small systems.

The different SAPT variants tested in this work exhibit quite different performance. The benchmark two-body interaction energies are very well recovered by SAPT2+(3) (with a MUE below 0.3 kcal/mol), however, no treatment of nonadditive three-body effects commensurate with SAPT2+(3) is available. The XSAPT approach is computationally efficient and does include nonadditive induction and dispersion effects (the latter through the MBD formalism), but its three-body accuracy is hampered by the lack of nonadditive first-order exchange effects. The three-body induction couplings may be included in XSAPT and are required for formal completeness of the method, however, the effect of these couplings on the accuracy of XSAPT interaction energies is minor and actually slightly detrimental. The SAPT(DFT) variant is the only one that provides estimates of all important two-body and three-body effects. For the latter, the dominant induction term agrees quite well between SAPT(DFT) and XSAPT as long as the HF delta term is included in the former. The nonadditive dispersion energy is consistently overestimated by SAPT(DFT), while the MBD three-body dispersion included in XSAPT is of the same order as the difference between nonadditive CCSD(T) and MP2 energies (this difference is primarily but not exclusively dispersion) but the agreement is far from perfect. Overall, the total nonadditive energies from SAPT(DFT) are significantly more accurate than those from XSAPT, but still less accurate than supermolecular MP2 at the CBS limit. When it comes to total interaction energies, SAPT2+(3) emerges as the most accurate variant for our dataset despite the complete lack of three-body terms. In contrast, the accuracy of SAPT(DFT) and XSAPT interaction energies is far from satisfactory due to their inferior performance on two-body terms. The simplest SAPT0 variant is also the least accurate one, and XSAPT provides a substantial improvement over SAPT0.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by the U.S. National Science Foundation award CHE-1955328.

References

1 A. S. Mahadevi and G. N. Sastry, Chemical Reviews, 2016, **116**, 2775–2825.
2 U. Góra, R. Podeszwa, W. Cencek and K. Szalewicz, The Journal of Chemical Physics, 2011, **135**, 224102.
3 G. J. Beran, Chemical Reviews, 2016, **116**, 5567–5613.
4 M. Hapka, Ł. Rajchel, M. Modrzejewski, R. Schäffer, G. Chałasiński and M. M. Szczęśniak, Journal of Chemical Physics, 2017, **147**, 084106–084117.
5 J. M. Merritt, V. E. Bondybey and M. C. Heaven, Science, 2009, **324**, 1548–1551.
6 P. Jankowski, A. R. McKellar and K. Szalewicz, Science, 2012, **336**, 1147–1150.
7 J. Řezáč and P. Hobza, Chemical Reviews, 2016, **116**, 5038–5071.
8 P. Jurečka, J. Šponer, J. Černý and P. Hobza, Physical Chemistry Chemical Physics, 2006, **8**, 1985–1993.
9 J. Řezáč, K. E. Riley and P. Hobza, Journal of Chemical Theory and Computation, 2011, **7**, 2427–2438.
10 J. Řezáč, K. E. Riley and P. Hobza, Journal of Chemical Theory and Computation, 2012, **8**, 4285–4292.
11 S. Grimme, Chemistry – A European Journal, 2012, **18**, 9955–9964.
12 Y. S. Al-Hamdani and A. Tkatchenko, The Journal of Chemical Physics, 2019, **150**, 010901–010923.
13 K. Patkowski, Annual Reports in Computational Chemistry, 2017, **13**, 3–91.
14 D. G. A. Smith, L. A. Burns, K. Patkowski and C. D. Sherrill, Journal of Physical Chemistry Letters, 2016, **7**, 2197–2203.
15 L. A. Burns, J. C. Faver, Z. Zheng, M. S. Marshall, D. G. A. Smith, K. Vanommeslaeghe, A. D. MacKerell, K. M. Merz and C. D. Sherrill, Journal of Chemical Physics, 2017, **147**, 161727.
16 J. Řezáč, Journal of Chemical Theory and Computation, 2020, **16**, 2355–2368.
17 J. Řezáč, Journal of Chemical Theory and Computation, 2020, **16**, 6305–6316.
18 K. Kříž, M. Nováček and J. Řezáč, Journal of Chemical Theory and Computation, 2021, **17**, 1548–1561.
19 K. Kříž and J. Řezáč, Physical Chemistry Chemical Physics, 2022, **24**, 14794–14804.
20 J. Řezáč, Physical Chemistry Chemical Physics, 2022, **24**, 14780–14793.
21 Z. M. Sparrow, B. G. Ernst, P. T. Joo, K. U. Lao and R. A. DiStasio, Jr., The Journal of Chemical Physics, 2021, **155**, 184303.
22 J. Řezáč, Y. Huang, P. Hobza and G. J. Beran, Journal of Chemical Theory and Computation, 2015, **11**, 3065–3079.
23 V. Babin, G. R. Medders and F. Paesani, Journal of Chemical Theory and Computation, 2014, **10**, 1599–1607.
24 M. R. Kennedy, A. R. McDonald, A. E. DePrince, M. S. Marshall, R. Podeszwa and C. D. Sherrill, Journal of Chemical Physics, 2014, **140**, 121104.
25 J. Yang, W. Hu, D. Usvyat, D. Matthews, M. Schütz and G. K. L. Chan, Science, 2014, **345**, 640–643.
26 M. Alkan, P. Xu and M. S. Gordon, Journal of Physical Chemistry A, 2019, **123**, 8406–8416.
27 K. Low, M. L. Coote and E. I. Izgorodina, Journal of Chemical Theory and Computation, 2023, **19**, 1466–1475.
28 S. Tan, S. Barrera Acevedo and E. I. Izgorodina, Journal of Chemical Physics, 2017, **146**, 064108.
29 W. Jankiewicz, R. Podeszwa and H. A. Witek, Journal of Chemical Theory and Computation, 2018, **14**, 5079–5089.
30 A. Tkatchenko, R. A. DiStasio, R. Car and M. Scheffler, Physical Review Letters, 2012, **108**, 236402.
31 A. J. Misquitta, R. Podeszwa, B. Jeziorski and K. Szalewicz, Journal of Chemical Physics, 2005, **123**, 214103.
32 A. Heßelmann, G. Jansen and M. Schütz, The Journal of Chemical Physics, 2004, **122**, 014103.
33 T. M. Parker, L. A. Burns, R. M. Parrish, A. G. Ryno and C. D. Sherrill, Journal of Chemical Physics, 2014, **140**, 094106.
34 K. Carter-Fenk, K. U. Lao and J. M. Herbert, Accounts of Chemical Research, 2021, **54**, 3679–3690.
35 N. J. Singh, S. K. Min, D. Y. Kim and K. S. Kim, Journal of Chemical Theory and Computation, 2009, **5**, 515–529.
36 H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby and M. Schütz, Wiley Interdisciplinary Reviews: Computational Molecular Science, 2012, **2**, 242–253.
37 D. G. A. Smith, L. A. Burns, A. C. Simmonett, R. M. Parrish, M. C. Schieber, R. Galvelis, P. Kraus, H. Kruse, R. Di Remigio, A. Alenaizan, A. M. James, S. Lehtola, J. P. Misiewicz, M. Scheurer, R. A. Shaw, J. B. Schriber, Y. Xie, Z. L. Glick, D. A. Sirianni, J. S. O'Brien, J. M. Waldrop, A. Kumar, E. G. Hohenstein, B. P. Pritchard, B. R. Brooks, H. F. Schaefer, A. Y. Sokolov, K. Patkowski, A. E. DePrince, U. Bozkaya, R. A. King, F. A. Evangelista, J. M. Turney, T. D. Crawford and C. D. Sherrill, Journal of Chemical Physics, 2020, **152**, 84108–84129.
38 SAPT2012: An Ab Initio Program for Many-Body Symmetry-Adapted Perturbation Theory Calculations of Intermolecular Interaction Energies, by R. Bukowski, W. Cencek, P. Jankowski, M. Jeziorska, B. Jeziorski, S. A. Kucharski, V. F. Lotrich, A. J. Misquitta, R. Moszyński, K. Patkowski, R. Podeszwa, F. Rob, S. Rybak, K. Szalewicz, H. L. Williams, R. J. Wheatley, P. E. S. Wormer, and P. S. Żuchowski, University of Delaware and University of Warsaw (http://www.physics.udel.edu/∼szalewic/SAPT/SAPT.html).
39 K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekström, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. Hjertenæs, S. Høst, I. M. Høyvik, M. F. Iozzi, B. Jansík, H. J. A. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjærgaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen,

A. Ligabue, O. B. Lutnæs, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawlowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. V. Rybkin, P. Sałek, C. C. Samson, A. S. de Merás, T. Saue, S. P. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. Wilson, M. Ziolkowski and H. Ågren, Wiley Interdisciplinary Reviews: Computational Molecular Science, 2014, **4**, 269–284.

40 Y. Shao, Z. Gan, E. Epifanovsky, A. T. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey, P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kus, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C. M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. Distasio, H. Do, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. Hanson-Heine, P. H. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T. C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. D. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S. P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, E. Neuscamman, C. M. Oana, R. Olivares-Amaya, D. P. Oneill, J. A. Parkhill, T. M. Perrine, R. Peverati, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, S. M. Sharada, S. Sharma, D. W. Small, A. Sodt, T. Stein, D. Stück, Y. C. Su, A. J. Thom, T. Tsuchimochi, V. Vanovschi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, J. Yang, S. Yeganeh, S. R. Yost, Z. Q. You, I. Y. Zhang, X. Zhang, Y. Zhao, B. R. Brooks, G. K. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xu, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J. D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C. P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. Gill and M. Head-Gordon, Molecular Physics, 2015, **113**, 184–215.

41 K. Patkowski, Wiley Interdisciplinary Reviews: Computational Molecular Science, 2020, **10**, e1452.

42 R. A. Kendall, T. H. Dunning and R. J. Harrison, The Journal of Chemical Physics, 1992, **96**, 6796–6806.

43 L. A. Burns, M. S. Marshall and C. D. Sherrill, Journal of Chemical Physics, 2014, **141**, 234111.

44 M. Kodrycka and K. Patkowski, Journal of Chemical Physics, 2019, **151**, 070901.

45 A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen and A. K. Wilson, Chemical Physics Letters, 1998, **286**, 243–252.

46 A. J. Misquitta and K. Szalewicz, The Journal of Chemical Physics, 2005, **122**, 214109–214128.

47 R. Podeszwa and K. Szalewicz, Journal of Chemical Physics, 2007, **126**, 194101–194114.

48 T. Korona, Molecular Physics, 2013, **111**, 3705–3715.

49 K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li and T. L. Windus, Journal of Chemical Information and Modeling, 2007, **47**, 1045–1052.

50 D. J. Tozer and N. C. Handy, The Journal of Chemical Physics, 1998, **109**, 10180–10189.

51 A. J. Misquitta and K. Szalewicz, Chemical Physics Letters, 2002, **357**, 301–306.

52 M. Grüning, O. V. Gritsenko, S. J. A. Van Gisbergen and E. J. Baerends, The Journal of Chemical Physics, 2002, **116**, 9591–9601.

53 Y. Xie, D. G. A. Smith and C. D. Sherrill, Journal of Chemical Physics, 2022, **157**, 24801.

54 R. Van Leeuwen and E. J. Baerends, Physical Review A, 1994, **49**, 2421.

55 V. F. Lotrich and K. Szalewicz, The Journal of Chemical Physics, 1997, **106**, 9668–9687.

56 K. U. Lao and J. M. Herbert, Journal of Physical Chemistry A, 2015, **119**, 235–252.

57 W. Xie, L. Song, D. G. Truhlar and J. Gao, The Journal of Chemical Physics, 2008, **128**, 234108.

58 K. Patkowski, K. Szalewicz and B. Jeziorski, Journal of Chemical Physics, 2006, **125**, 154107.

59 M. Gray and J. M. Herbert, Journal of Chemical Theory and Computation, 2022, **18**, 2308–2330.

60 J. M. Herbert, L. D. Jacobson, K. U. Lao and M. A. Rohrdanz, Physical Chemistry Chemical Physics, 2012, **14**, 7679–7699.

61 K. U. Lao and J. M. Herbert, Journal of Chemical Theory and Computation, 2018, **14**, 2955–2978.

62 J. Garcia, R. Podeszwa and K. Szalewicz, Journal of Chemical Physics, 2020, **152**, 184109.

63 Y. Huang and G. J. Beran, Journal of Chemical Physics, 2015, **143**, 044113.

64 B. D. Nguyen, D. J. Hernandez, E. V. Flores and F. Furche, Electronic Structure, 2022, **4**, 014003.