DALHOUSIE UNIVERSITY

Exchange-correlation effects for non-covalent interactions in density-functional theory

A. Otero-de-la-Roza,*,†,‡ Gino A. DiLabio,*,†,‡ and Erin R. Johnson*,¶

† *National Institute for Nanotechnology, National Research Council of Canada, 11421 Saskatchewan Drive, Edmonton, Alberta, Canada T6G 2M9.*
‡*Department of Chemistry, University of British Columbia, Okanagan, 3247 University Way, Kelowna, British Columbia, Canada V1V 1V7.*
¶ *Department of Chemistry, Dalhousie University, 6274 Coburg Road, Halifax, Nova Scotia, Canada B3H 4R2.*

E-mail: aoterodelaroza@gmail.com; gino.dilabio@ubc.ca; erin.johnson@dal.ca

Abstract

In this article, we develop an understanding of how errors from exchange-correlation functionals affect the modeling of non-covalent interactions in dispersion-corrected density-functional theory. Computed CCSD(T) reference binding energies for a collection of small-molecule clusters are decomposed via a molecular many-body expansion, and used to benchmark density-functional approximations, including the effect of semilocal approximation, exact-exchange admixture, and range separation. Three sources of error are identified. Repulsion error arises from the choice of semilocal functional approximation. This error affects intermolecular repulsions, and is present in all n-body exchange-repulsion energies with a sign that alternates with the order (n) of the interaction. Delocalization error is independent of the choice of semilocal functional but does depend on the admixture of exact exchange. Delocalization error misrepresents the induction energies, leading to overbinding in all induction n-body terms, and underbinds the electrostatic contribution to the 2-body energies. Deformation error affects only monomer relaxation (deformation) energies, and behaves similarly to bond dissociation energy errors. Delocalization and deformation errors affect systems with significant intermolecular orbital interactions (hydrogen- and halogen-bonded systems, for instance), whereas repulsion error is ubiquitous. Many-body errors from the underlying exchange-correlation functional greatly exceed, in general, the magnitude of the many-body dispersion energy term. A functional built to accurately model non-covalent interactions must contain a dispersion correction, semilocal exchange and correlation components that minimize repulsion error independently, and must also incorporate exact exchange in such a way that delocalization error is absent.

1 Introduction

Modeling intermolecular interactions in density-functional theory (DFT) requires functionals that account for dispersion effects.[1–22] Common density functionals do not account for dispersion, and so the electronic energies obtained using these methods must be augmented by one of several available dispersion-correction approaches.[1] Regardless of the method used to obtain it, the dispersion energy is combined with a pre-existing exchange-correlation functional (the base functional), which accounts for the non-dispersive component of the non-covalent interaction energy. Hence, by using dispersion-corrected DFT, a tacit assumption

is made that the base functional gives an accurate representation of electrostatics, induction, and exchange-repulsion, as well as any other non-dispersive intermolecular effects.[23] In this article, we show that this presumption is overly optimistic and that most base functionals fail at describing non-dispersive intermolecular interactions, sometimes spectacularly so. Dispersion corrections are, in general, used to correct for these failings via error cancellation.

Dispersion and base functional combinations have been extensively tested against benchmark sets composed of high-level reference data. Virtually all of these sets are limited to gas-phase molecular dimers, for which coupled-cluster calculations combined with complete-basis-set extrapolation techniques are used to obtain very accurate reference data.[24–34] For larger systems, where coupled-cluster calculations are not possible (except for fragment-based approaches[35–38]), back-corrected experimental binding energies are used, although this approach has limitations regarding the accuracy limit of the benchmark set.[33,39–42] Comparatively, there are only a few benchmark data sets for small-molecule clusters that can bridge the gap between isolated dimers and condensed phases, and the majority of these comprise water clusters.[35,43–54]

The importance of many-body (many-molecule) contributions to intermolecular binding energies, such as cooperative polarization effects and 3-body dispersion,[53,55–59] increases for molecular aggregates compared to dimers. In this context, it is natural to employ a many-body expansion of the total energy as a sum of n-molecule terms of decreasing magnitude for increasing n. Many-body expansions have been applied in the past to analyze the energetic contributions to water clusters,[53,58–63] benzene trimers,[64] nucleic acid tetramers,[65] noble gases,[53] and others.[54] Wavefunction theory can be combined with polarizable force fields in a many-body interaction model to study molecular clusters[66] or crystals.[67,68] The many-body expansion has also been used successfully with electronic embedding methods in studies of water clusters and of ice and carbon dioxide crystals.[61,69] Many-body calculations have also allowed the highly-accurate determination of the lattice energy of benzene.[70,71]

The most popular approach to incorporate dispersion effects in density-functional theory is to use a damped asymptotic pairwise dispersion expression. One example of such an approach is the exchange-hole dipole moment (XDM) dispersion model,[2,72] in which the coefficients that enter the pairwise dispersion expression are calculated non-empirically. This approach has been shown to give excellent results for gas-phase dimers[4,73] as well as molecular crystals.[42,74,75] In the framework of a many-body expansion of the energy, a pairwise dispersion correction can affect almost exclusively the 2-body energies, except for the minor variations in the dispersion energy caused by changing dispersion coefficients. Therefore, errors in the treatment of 3-body and higher-order terms coming from the base functional remain uncorrected, regardless of how the dispersion correction is parametrized. Since $n > 2$ terms scale faster with the number of molecules than the pairwise terms, it is likely that these errors become important for systems composed of many interacting molecules, particularly if these are densely packed.[53]

We assemble a benchmark for small-molecule clusters in this article, consisting of 2 to 6 monomers and covering a range of interaction types. The cluster binding energies are decomposed into monomer, pairwise, 3-body, and higher-order contributions, and calculated using high-level wavefunction theory. The resulting reference data is then used to benchmark different density-functional approximations (DFA). The underlying reasons behind the performance of a particular base and dispersion functional combination are poorly understood, and it is difficult to predict *a priori* the adequacy of a particular combination for a given purpose.[76–78] Our choice of DFAs includes several popular generalized-gradient approximations (GGA), and we explore the effect of varying the fraction of exact exchange in hybrids derived from the chosen semilocal functionals. Range-separated hybrid functionals are also considered. Our work extends previous benchmark studies that were re-

stricted to water and rare-gas clusters or a reduced set of semilocal functionals.[53] Delocalization error and exchange-repulsion[53,59,61,77,78] have been suggested as likely candidates, and we analyze the contribution of each of these errors in this article.

Our coupled-cluster results show that pairwise interactions are dominant for dispersion-bound clusters, but 3-body and higher-order effects are important for hydrogen-bonded (H-bonded) complexes. In agreement with previous studies,[53] the errors in the many-body energies caused by the choice of DFA are much larger than the many-body dispersion effect, which explains the difficulties in their application when a damped asymptotic dispersion expression is used.[3,39] The DFA results on our benchmark set of dispersion-bound clusters are mostly explained by the errors in the intermolecular exchange-repulsion component (repulsion error), which is controlled by the underlying GGA,[53] and which alternates in sign with the order of the interaction (the sign for the 2-body error is the opposite as the 3-body error, but the same as the 4-body error, etc.). However, repulsion error alone does not explain the DFA performance in H-bonded clusters, where induction and electrostatic effects are important. In these systems, a second source of deviation from the reference data is delocalization error[79–82] coming from the base functional. The amount of delocalization error depends solely on the admixture of exact exchange and affects the induction energy, resulting in overbinding of all n-body induction terms, and the electrostatic energy, which results in an underbinding contribution to the 2-body term. The combined effect of delocalization error on the 2-body energies depends on the relative contribution of induction and electrostatics to overall binding, and hence on the type of non-covalent interaction. Monomer distortion (deformation) is also shown to be relatively important for H-bonded complexes, and the performance of different DFAs regarding deformation error is directly related to their ability to predict bond-dissociation energies.

2 Theory

Let us consider a cluster of n molecules. The binding energy, $\mathrm{BE}^{(n)}$, is defined as:

$$-\mathrm{BE}^{(n)} = -\mathrm{BE}^{(n)}_{\mathrm{relax}} = E^{(n)} - nE^{(1)}, \qquad (1)$$

where $E^{(n)}$ is the energy of the cluster and $E^{(1)}$ is the energy of a single monomer at its relaxed geometry. In this work, attractive interactions correspond to a positive binding energy.

It is convenient to consider the monomers at the geometries distorted by the cluster environment instead of isolated-monomer optimized geometries. We define the deformation energy $\Delta E^{(1)}$ as the energy difference between the monomers in their distorted geometries from the n-molecule cluster and their relaxed geometries:

$$-\Delta E^{(1)} = \sum_{i}^{1\subset n} E_i^{(1)} - nE^{(1)}. \qquad (2)$$

In this equation, the notation "$1 \subset n$" indicates that the sum runs over all possible monomers in the n-molecule cluster and $E_i^{(1)}$ is the energy of monomer i at the geometry of the monomer in that cluster. The total binding energy with respect to the rigid monomers is:

$$-\mathrm{BE}^{(n)}_{\mathrm{rigid}} = E^{(n)} - \sum_{i}^{1\subset n} E_i^{(1)}. \qquad (3)$$

In the rest of this section, unless otherwise stated, we will consider rigid-monomer binding energies only.

The leading-order contribution to the binding energy is the sum of the dimer binding energies, the pairwise or 2-body contribution:

$$\Delta E^{(2)} = \sum_{i<j}^{2\subset n} \mathrm{BE}^{(2)}_{ij}, \qquad (4)$$

$$-\Delta E^{(2)} = \sum_{i<j}^{2\subset n} E^{(2)}_{ij} - (n-1)\sum_{i}^{1\subset n} E_i^{(1)}. \qquad (5)$$

In equation 4, $\mathrm{BE}^{(2)}_{ij}$ represents the binding energy and $E^{(2)}_{ij}$ is the total energy of the dimer composed of monomers i and j. The encompassing sums run over all possible pairs in the

n-molecule cluster, with a total of $\binom{n}{2} = n(n-1)/2$ terms.

The 3-body energy is defined as the contribution to the binding energy of a given trimer that is not captured by the sum of the BEs of the three constituent dimers. The total 3-body energy for the n-molecule cluster is the sum of the 3-body energies, $\Delta E^{(3)}$, over all possible trimers within the cluster. Thus, the 3-body contribution to the total binding energy of the n-molecule cluster is the difference between the sum of trimer binding energies and 2-body contributions of their constituent dimers:

$$\Delta E^{(3)} = \sum_{i<j<k}^{3\subset n} \mathrm{BE}^{(3)}_{ijk} - (n-2)\sum_{i<j}^{2\subset n} \mathrm{BE}^{(2)}_{ij}, \tag{6}$$

$$-\Delta E^{(3)} = \sum_{i<j<k}^{3\subset n} E^{(3)}_{ijk} - (n-2)\sum_{i<j}^{2\subset n} E^{(2)}_{ij} + \frac{(n-1)(n-2)}{2}\sum_{i}^{1\subset n} E^{(1)}_{i}. \tag{7}$$

In equation 6, the first sum runs over all the trimers (indices i, j, k) in the n-molecule cluster, $\mathrm{BE}^{(3)}_{ijk}$ is the binding energy of a particular trimer, and $E^{(3)}_{ijk}$ is the total energy. There are $\binom{n}{3} = n(n-1)(n-2)/6$ terms in the trimer sum.

Using the same procedure, the 4-body and higher-order contributions, up to the n-body contributions, can be defined. The general m-body contribution to the total binding energy can be evaluated directly as:

$$-\Delta E^{(m)} = \sum_{k=1}^{m}(-1)^{m-k}\frac{(n-k)!}{(n-m)!(m-k)!}\sum_{i<j\ldots}^{k\subset n} E^{(k)}. \tag{8}$$

The total relaxed binding energy of the cluster is the sum of the deformation energy plus the n-body binding-energy terms with $n \geq 2$:

$$\mathrm{BE}_{\mathrm{relax}} = \sum_{k=1}^{n} \Delta E^{(k)}. \tag{9}$$

In this work we consider only the 2-body, 3-body, and a composite ≥4-body term. As we shall see in the next section, the higher-order interactions represent small contributions to the total binding energies. Thus, the ≥4-body contribution to the binding energy is simply evaluated as the difference between the total, rigid-monomer binding energy of the cluster and the sum of the 2-body and 3-body terms:

$$\Delta E^{(\geq 4)} = \mathrm{BE}^{(n)}_{\mathrm{rigid}} - \Delta E^{(2)} - \Delta E^{(3)}. \tag{10}$$

For all the terms defined above, we define the error as the difference between the approximate energy coming from a DFA and the reference energy from CCSD(T). Since positive values of the binding energy and the n-body terms indicate attractive interactions, positive values of the error correspond to overbinding, and negative values to underbinding behavior. In addition, in order to estimate the magnitude of the 3-body dispersion term to compare it to the 3-body errors originating from different DFA, we use the difference in the CCSD(T) and MP2 3-body energies, as done in previous works.[53,77,78]

3 Computational Methods

We consider homogeneous n-molecule (n = 2–6) clusters of noble gases (Ne, Ar), as well as molecules with interactions that are predominantly dispersion (N_2, CH_4, and CO), and hydrogen bonding (NH_3, H_2O, HF). The geometries for each of these clusters are provided in the supporting information (SI). Larger HF clusters with n = 7–9 molecules were also considered. The binding energies of the molecular clusters were calculated and decomposed into a series of contributions, as discussed in the previous section.

For the Ne and Ar clusters, geometries were constructed to have the maximum symmetry (tetrahedral, trigonal bipyramidal, octahedral) with fixed nearest-neighbor distances of 3.1 Å for Ne and 3.8 Å for Ar. Single-point energy calculations were performed using CCSD(T)/aug-cc-pV5Z with the Gaussian09 program,[83] with counterpoise corrections for basis set superposition error.[84] Because the counterpoise correction generally over-corrects for basis-set superposition error effects,[85–89] the counterpoise-

corrected and non-counterpoise results were averaged to obtain the final binding energies.[88,90,91]

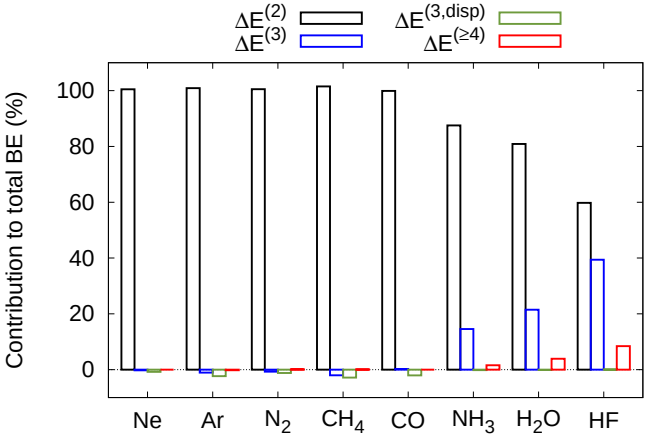
For the N_2 and CO clusters, initial geometries were obtained from the corresponding molecular crystals. For the NH_3 and HF clusters, initial geometries were chosen to involve a cyclic planar network of hydrogen bonds. In the case of HF, cycles are the most stable cluster structures, although they are in general non-planar.[92] The geometries for these four sets of clusters were then optimized using LC-ωPBE-XDM[4,93,94] with the aug-cc-pVDZ basis set. For the CH_4 clusters, geometries were optimized using a preliminary version of methane-specific dispersion-correcting potentials.[95] For the H_2O clusters, geometries were taken from the work of Shields *et al.*[48] In all of these cases, single-point energy calculations were performed using CCSD(T)/aug-cc-pVQZ with the Turbomole program.[96] As in the case of noble gases, the counterpoise- and non-counterpoise-corrected results were averaged to obtain the final binding energies. The only exception is the methane hexamer, for which we were unable to run the CCSD(T) calculation on the full cluster. For this case only, the $\geq$4-body effects were evaluated at the CCSD level of theory. It will be shown that, for dispersion dominated systems, >4 body effects are sufficiently small that our inability to evaluate this term via CCSD(T) will not impact our conclusions.

Subsequent calculations were performed using a range of dispersion-corrected density-functional methods. In addition, a series of six hybrid functionals with varying fractions of exact exchange were used. These hybrid functionals are based on semilocal functionals built from all six combination of B88,[97] PW86,[98] and PBE[99] exchange with LYP[100] and PBE[99] correlation. The mixing fraction of exact exchange was varied from 0-100% in 10% increments. Additional calculations were also performed using the B3LYP,[100,101] B971,[102,103] and PBE0 hybrid functionals,[104] and the LC-ωPBE[105] range-separated hybrid functional. All DFT calculations used the aug-cc-pVTZ basis set and ultrafine integration grids with the Gaussian 09 program. The dispersion corrections were obtained using the exchange-hole dipole moment (XDM) method. The XDM dispersion energies were calculated using the postg program[106] and added to the base DFT energies. Damping parameter values are the same as those in previous work.[107]

To investigate the perturbation theory energy components in the dimers, we have carried out calculations using symmetry-adapted perturbation theory (SAPT) using the psi4 program[108] and the aug-cc-pVQZ basis set at the SAPT2+3 level.[109] To illustrate the difference between hydrogen bonding and halogen bonding, we use the FBr$\cdots$NCH dimer from the XB18 set,[110] which we calculated using the aug-cc-pVTZ basis set.

4 Coupled-cluster results

Figure 1: Average percent contributions of the 2-body, 3-body, and $\geq$4-body terms to the total rigid-monomer binding energies of the molecular clusters. The 3-body dispersion contribution ($\Delta E^{(3,\mathrm{disp})}$) is estimated as the difference between the CCSD(T) and MP2 3-body terms.

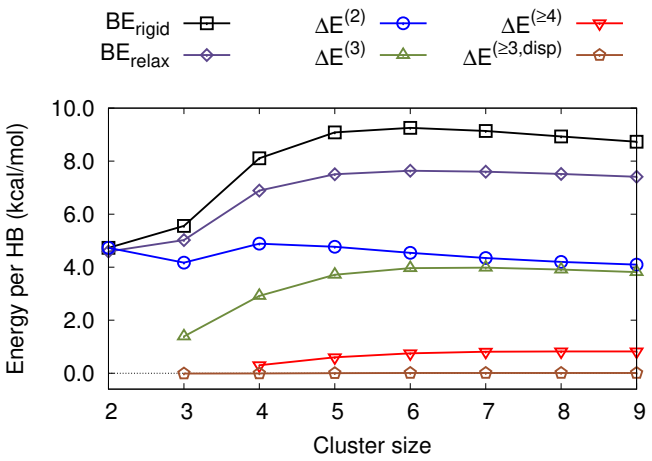


The CCSD(T) rigid and relaxed binding energies, n-body energies, and deformation energies for the molecular clusters are shown in Table 1 of the SI. In addition, we also show the 3-body dispersion contribution, estimated by the difference between CCSD(T) and MP2 3-body energies.[53,77,78] The average percent contributions to the total binding energies are represented graphically in Figure 1, in a style similar to

the figure presented by Góra *et al.* for water clusters.[63]

There are clearly two behavior regimes, depending on whether hydrogen bonds are present in the cluster or not. For non-H-bonded clusters (Ne, Ar, N_2, CO, CH_4) the pairwise interactions are almost the only contribution to the binding energy, and the higher-order effects are very small. The 3-body interaction is very slightly repulsive (except for CO), and slightly smaller in magnitude than the Axilrod-Teller-Muto dispersion contribution.[3,21,111,112] Higher-order contributions to the binding energies and deformation energies are negligible.

Figure 2: Total binding energies for the HF clusters and their component contributions, expressed as the absolute value of the energy, per hydrogen bond. The 3-body dispersion contribution ($\Delta E^{(3,\mathrm{disp})}$) is estimated as the difference between the CCSD(T) and MP2 3-body terms.



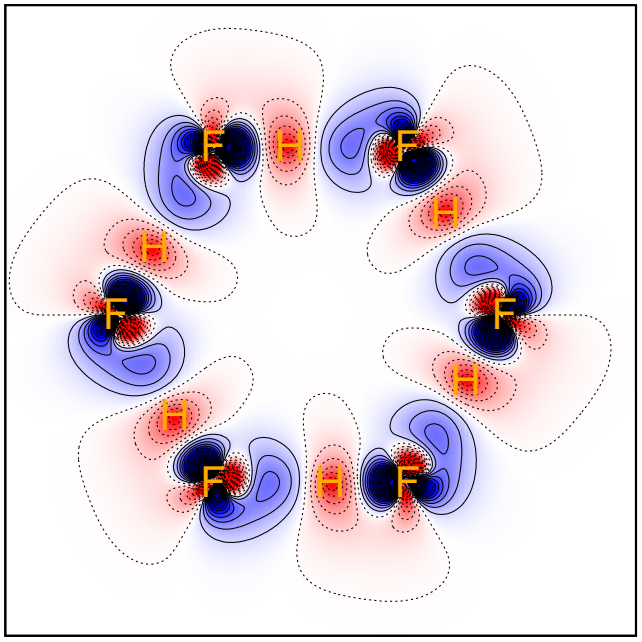
In contrast, for the H-bonded clusters, there are significant 3-body and ≥4-body contributions to the binding energies. These non-additive effects have been described in the past,[113,114] and are sometimes known as cooperative effects.[113] In an intermolecular perturbation framework,[23] these arise from many-body induction effects, and they usually (though not always[23]) stabilize the H-bonded network relative to the sum of the binding energies of the component dimers. The 3-body and ≥4-body energies are correlated, and increase in the sequence NH_3 <H_2O<HF. HF is an extreme case: for $(HF)_6$, the total 3-body contribution is almost as large as the 2-body for the clusters with $n > 3$.

Because of its behavior, its simple geometry, and the relatively small number of atoms, we will focus on the energy contributions to the HF clusters in the rest of the article to analyze functional performance in H-bonded networks. The individual contributions to the HF cluster binding energies per hydrogen bond are shown as a function of cluster size in Figure 2. The equivalent figure reporting total energies is shown as Figure 1 in the SI. The pairwise contribution per hydrogen bond is roughly constant, although it dips for $n = 3$, where the geometry is strained. However, the 3-body contribution per hydrogen bond, which is directly related to induction effects, grows rapidly with cluster size. The strength of the cooperative effects per monomer peaks at the hexamer and stabilizes for larger clusters. Figures 1 and 2 also show that the dispersion contribution to the 3-body energies in the H-bonded systems is smaller than in the dispersion-bound clusters, and negligible compared to the other energy contributions.[46,59,61,115,116]

Cooperative effects arise from the collective polarization of the monomers forming the H-bonded network, and have a characteristic signature in the electron density and density matrix. The polarization density, defined as the difference between the electron density in the hexamer minus the monomer densities, is shown in Figure 3. The effect of forming a H-bonded cluster is that the monomer polarizes so as to deplete the electron density of the hydrogen and populate the electron-pair region close to the fluorine, which is oriented toward the hydrogen of the adjacent molecule. The leading contribution to the polarization density in Figure 3 is the 2-body term, defined as in Equation 4, but with the energies replaced by densities. The 3-body and higher-order n-body contributions to the polarization density have essentially the same shape but decrease in magnitude with increasing n.

The deformation energy is also an important contribution to the binding energy in H-bonded clusters. Because it is determined by

Figure 3: Difference between the electron density for the HF hexamer and the sum of the monomer densities (BH&HLYP/aug-cc-pVTZ). Blue indicates density accumulation and red indicates density depletion in the hexamer, relative to the sum of the monomers. The colors scale goes from 0.03 (blue) to −0.03 (red). Positive and negative contours are represented by full and dotted lines, respectively.



the intramolecular potential-energy surface, the strength of the relaxation-energy effect correlates directly with the bond strength of the monomer. The deformation energy trend is NH_3 <H_2O<HF even though the water-dimer hydrogen bond is stronger than that in the HF dimer. The magnitude of the deformation energy is mirrored by changes in X-H bond length (X=N,O,F) between the complexes and relaxed monomers. The maximum change in bond length is 0.010 Å for NH_3, 0.021 Å for H_2O, and 0.044 Å for HF hydrogen-bond donors in the clusters.

Because the deformation and ≥4-body terms are effectively zero for the dispersion-bound clusters (Ne, Ar, N_2, CO, and CH_4), but large in magnitude for the H-bonded clusters (NH_3, H_2O, and HF), we will consider these two classes separately.

5 Density-functional results: Dispersion-bound clusters

The 2-body, 3-body, and total rigid-monomer binding energies for each set of dispersion-bound clusters are shown in Table 1 for CCSD(T) and the density functionals employed in this work. The deformation energies are negligible, so the relaxed and rigid binding energies are approximately equal. While the total binding energies are dominated by the pairwise contributions, the mean absolute errors (MAE) show that the 3-body errors with DFAs are large compared to the reference 3-body contribution to binding. They range from a quarter to a half of the 2-body errors and are much larger than the CCSD(T) 3-body energies.

To help analyze the source of the DFA errors in our dimers, we calculated the dimer binding energy components using symmetry-adapted perturbation theory (SAPT). Table 2 shows that the two significant contributions to binding in the dispersion-bound clusters are dispersion and exchange-repulsion. Since dispersion interactions are assumed to be treated by the dispersion correction, errors caused at the

Table 1: Accumulated binding-energy contributions for the dispersion-bound clusters (sum of the corresponding contributions for all n-body clusters, with n = 2–6) with CCSD(T) and the dispersion-corrected functionals used in this work. All values are kcal/mol.

	CCSD(T)	LC-ωPBE	B3LYP	B971	PBE0	BLYP	PW86PBE	PBE
				BE_{rigid}				
Ne	2.53	2.62	2.32	4.19	3.55	1.16	2.41	4.72
Ar	8.55	5.86	5.24	10.20	8.11	2.73	7.58	9.41
N_2	8.76	9.17	7.38	11.22	10.03	5.15	9.33	11.05
CO	10.67	10.89	10.11	14.51	12.74	7.78	12.57	14.03
CH_4	13.20	12.35	11.99	15.55	13.28	10.54	14.20	14.65
MAE		0.85	1.34	2.39	0.98	3.27	0.91	2.03
				$\Delta E^{(2)}$				
Ne	2.53	2.47	2.32	4.68	4.03	1.10	2.73	5.65
Ar	8.66	5.25	5.13	11.33	8.97	2.38	7.98	10.93
N_2	8.83	8.83	7.03	11.67	10.44	4.62	9.40	11.84
CO	10.64	10.26	9.61	15.26	13.32	6.95	12.69	15.18
CH_4	13.48	11.38	11.44	17.01	14.33	9.30	14.61	16.61
MAE		1.19	1.73	3.16	1.39	3.96	0.93	3.21
				$\Delta E^{(3)}$				
Ne	−0.01	0.09	0.07	−0.57	−0.56	0.19	−0.35	−1.04
Ar	−0.11	0.47	0.27	−1.20	−0.97	0.62	−0.42	−1.66
N_2	−0.08	0.33	0.39	−0.49	−0.45	0.57	−0.10	−0.86
CO	0.01	0.57	0.55	−0.83	−0.69	0.94	−0.18	−1.32
CH_4	−0.30	0.96	0.71	−1.61	−1.24	1.58	−0.48	−2.25
MAE		0.58	0.49	0.84	0.69	0.88	0.21	1.33

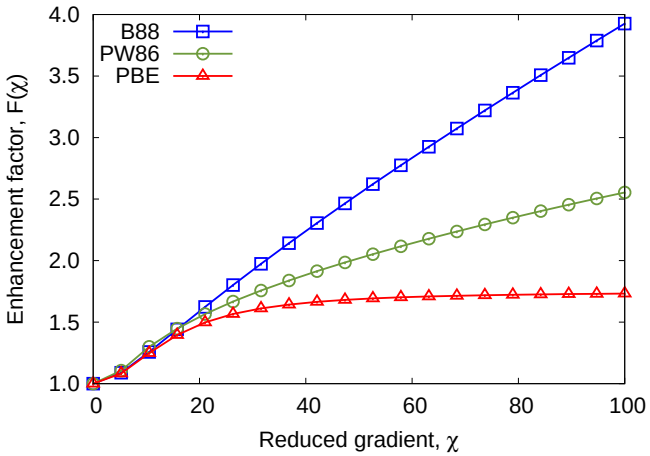
Table 2: Symmetry-adapted perturbation theory (SAPT) energy components (electrostatic, exchange, induction, dispersion) and total binding energies for the studied dimers. The FBr$\cdots$NCH dimer from the Kozuch and Martin XB18 set[110] is also included for comparison.

Dimer	Elec.	Exch.	Ind.	Disp.	Total	Ref.
Ne	0.02	−0.10	0.00	0.14	0.07	0.08
Ar	0.11	−0.37	0.02	0.51	0.26	0.27
N_2	0.04	−0.27	0.01	0.49	0.27	0.27
CO	0.27	−0.38	0.03	0.51	0.43	0.39
CH_4	0.18	−0.69	0.03	1.05	0.58	0.55
NH_3	5.03	−5.15	1.38	2.09	3.36	3.16
H_2O	8.10	−7.88	2.47	2.59	5.28	5.09
HF	6.75	−6.59	2.59	2.08	4.83	4.73
FBr$\cdots$NCH	14.94	−21.63	8.26	6.68	8.25	7.61

base functional level in these systems can be attributed to the exchange-repulsion component, and we will call this the "repulsion" error in the rest of the article.

Table 1 illustrates some of the challenges facing the development of density-functional methods for non-covalent interactions. Pairwise dispersion corrections like XDM can only affect the 2-body energies. A great deal of effort has been put into calculating the 3-body dispersion contribution[3,22,112] but, although this term is non-negligible for certain intermolecular geometries, errors coming from the base functional in the representation of exchange-repulsion 3-body interactions are typically one order of magnitude larger.[111] This observation explains why previous efforts to include 3-body dispersion effects in a dispersion functional based on the asymptotic expression have not been successful,[3,39] even though the coefficients for the leading 3-body term (C_9) can be calculated with relatively high accuracy.[3] In the absence of base functional errors, the asymptotic Axilrod-Teller-Muto (ATM) term damped with a product of Tang-Toennies damping functions[117] successfully reproduces 3-body dispersion effects with an atom-based summation.[78] However, a base functional that gives an accurate representation of the short-range 3-body repulsion is necessary in order to make progress in the field.[77]

Figure 4: Enhancement factors for three GGA exchange functionals.



5.1 Exchange enhancement factor

Table 1 shows that PW86PBE performs best for both $\Delta E^{(2)}$ and $\Delta E^{(3)}$ and is second only to LC-ωPBE for accuracy of the total binding energies. This result is in stark contrast to the other two GGAs, BLYP and PBE, which give quite large errors. The behavior of these GGAs is convincingly explained by the argument of Lacks and Gordon[25,53,118,119] using the exchange enhancement factor, which was recently extended to 3-body and higher-order interactions by Gillan.[53] For GGA functionals, the exchange energy can be written as:

$$E^{\mathrm{GGA}} = \sum_{\sigma} \int F(\chi_\sigma) \varepsilon_\sigma^{\mathrm{LSDA}} d\mathbf{r}, \qquad (11)$$

where $\varepsilon_\sigma^{\mathrm{LSDA}}$ is the local spin-density approximation (LSDA) exchange-energy density:

$$\varepsilon_\sigma^{\mathrm{LSDA}} = c_X \rho_\sigma^{4/3}, \qquad (12)$$

and $F(\chi_\sigma)$ is the exchange enhancement factor, which depends on the reduced density gradient, $\chi_\sigma = |\nabla \rho_\sigma| / \rho_\sigma^{4/3}$.

The enhancement factors for three GGA exchange functionals considered in this work are plotted in Figure 4. They show very different behaviors in the limit of large reduced gradient, which is related to the intermolecular repulsion energy obtained from the base functional. In particular, the large-χ part of the enhancement factor determines the energy contribution coming from the exponentially-decaying tails of the electron density in the region of space surrounding a molecule. A portion of this energy contribution disappears upon dimer formation, because parts of those regions are now occupied by other monomers, which explains the importance of the large-χ enhancement factor in determining intermolecular repulsion.[53] The argument can be easily extended to 3-body and higher-order n-body contributions[53] to show that functionals that underestimate the 2-body intermolecular closed-shell repulsion overestimate the 3-body contribution, and *vice versa*. This argument also agrees with the picture

of exchange-repulsion arising from intermolecular wavefunction overlap.[23] Accurate values for all the n-body components of the exchange-repulsion energies can be obtained with a simple Hartree-Fock (HF) calculation.[53,111]

PW86PBE has the lowest 2-body and 3-body errors of all GGAs.[118,119] For BLYP, the large-χ enhancement factor is too high, relative to PW86, so the large gradient regions in the monomers are too stable and the 2-body term is too repulsive. For PBE, the enhancement factor is too small, so the large gradient regions are not stable enough and the 2-body term is too attractive. In this sense, BLYP and PBE can be considered extremes of repulsion behavior.[25,59,119] The 3-body errors follow the same pattern as the 2-body errors, but opposite in sign. The 3-body contribution is now too attractive with BLYP and too repulsive with PBE. Similar alternation of signs is seen if one considers the 4- and 5-body errors for the hexamers, although they are much smaller in magnitude.

5.2 Exact-exchange mixing

Next, we consider the effect of exact-exchange mixing. Plots of the 2-body and 3-body energies for the methane clusters are shown in Figure 5. As increasing fractions of exact exchange are used, the contribution of the GGA exchange term is reduced. Focusing only on the functionals with LYP correlation, which gives approximately correct 2-body and 3-body energies in the 100% exact-exchange limit, we see that PBE and PW86 2-body energies become less binding with increasing exact exchange, and the opposite happens for B88. The reverse trend is observed for the 3-body energies. This is consistent with the enhancement-factor effects discussed in previous studies[53,118] and shown in Figure 4.

A technique that is successfully used to address shortcomings of hybrid functionals is range-separation. Common range-separated functionals combine semilocal and exact exchange such that the semilocal part is used for the short-range interelectronic interactions, while exact exchange is used for the long-range interactions. The parameter that controls the extent of the short- and long-range terms (ω) can be tuned to improve the description of a number of properties, including band gaps, polarizabilities, and ionization potentials (see Ref. 120 and references therein).

Figure 6 shows the effect of varying ω on the 2-body and 3-body errors for the dispersion-bound methane clusters with the LC-ωPBE functional. The semilocal functional ($\omega = 0$) is overly repulsive for the 2-body energies and too attractive for the 3-body contributions. As ω increases and long-range exact exchange is incorporated, the error curves move toward the HFPBE result, with the $\omega = 1.0$ functional giving almost the same results as HFPBE. The $\omega = 0.2$ value provides good performance for both the 2-body and 3-body contributions (c.f. the $\omega = 0.4$ value in the usual definition of LC-ωPBE[93,94]).

5.3 Correlation effects

A comparison of the HFPBE and HFLYP results (the red curves in Figure 5) shows that the choice of correlation functional also impacts the calculation of the 2-body and 3-body energies. HF accurately describes repulsion effects, so errors in the HFPBE and HFLYP curves come exclusively from the correlation functional, which should have a zero contribution to the exchange-repulsion energy. Therefore, Figure 5 shows that the LYP correlation functional has superior performance compared to PBE. Moreover, the signature of opposing contributions to the 2-body and 3-body errors is again present. PBE correlation overestimates 2-body energies and underestimates 3-body energies, whereas LYP gives small errors in both cases.

The observation that exchange and correlation functionals have the same repulsion error behavior offers the possibility of using functionals that benefit from error cancellation between the exchange and the correlation part. In fact, the good performance of PW86PBE seen in Table 1 and in previous studies[118,119] can be attributed not to PW86 exchange giving a particularly accurate representation of exchange-

Figure 5: Errors for the different contributions to the $(CH_4)_n$ cluster binding energies against the number of molecules in the cluster using hybrid functionals built from combinations of B88 (left), PW86 (middle), and PBE (right) exchange with PBE (top) and LYP (bottom) correlation. The errors are for the 2-body energies ($\Delta E^{(2)}$, first block) and 3-body energies ($\Delta E^{(3)}$, second block). Pure GGA functionals are represented in blue, half-and-half functionals are shown in green, and HF plus correlation is in red. Positive values indicate overestimation of the corresponding energy contribution. All units are kcal/mol. Figure 2 in the SI contains the same information with 10% steps in the exact exchange fraction.

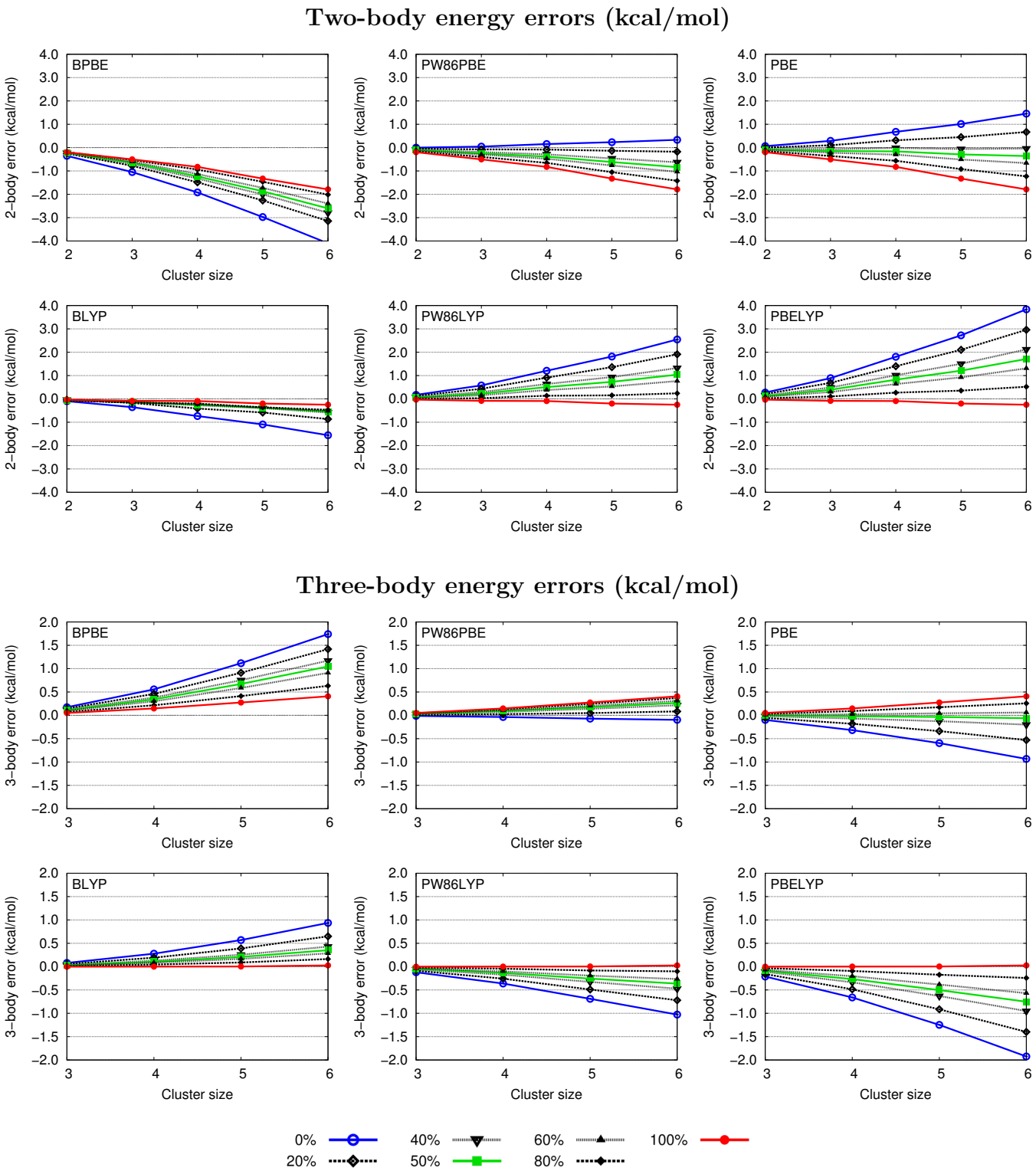
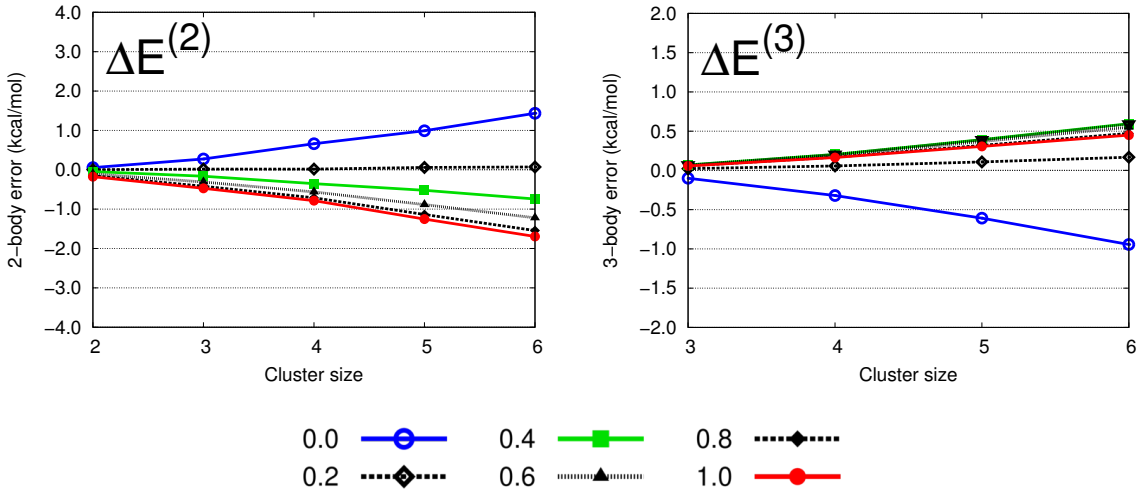


Figure 6: Errors in the 2-body (left) and 3-body (right) contributions to the $(CH_4)_n$ cluster binding energies using range-separated functionals based on LC-ωPBE with different values of the range-separation parameter (ω) against the number of molecules in the cluster. The pure GGA functional is represented in blue. Positive values indicate overestimation of the corresponding energy contribution. Figure 3 in the SI contains the same information with 0.1 steps in the range separation parameter. All units are kcal/mol.



repulsion effects, but to its successful pairing with PBE correlation and consequent error cancellation. When PW86 exchange is incrementally replaced by exact exchange, the 2-body and 3-body errors increase and approach the erroneous HFPBE results.

In summary, the binding-energy errors for dispersion-bound clusters can be easily understood by the description of exchange-repulsion effects demonstrated by different functionals. These errors have a characteristic signature: if the functional overbinds the 2-body energy, it underbinds the 3-body energy, and *vice versa*. Exchange and correlation functionals, separately and in combination, show this behavior and, as it is the case for PW86PBE, it is possible to find a particular exchange-correlation combination that cancels both errors, resulting in a relatively accurate description of intermolecular repulsion at the semilocal functional level. However, we observed in previous attempts to use PW86PBE hybrid functionals that the performance in the XDM parametrization degrades with increasing exact-exchange fraction, which is explained by the loss of favorable error cancellation. A good functional for dispersion-bound systems could, in principle, be designed to avoid error cancellation between exchange and correlation, and hybrids based on this functional should be equally successful in describing intermolecular repulsion.

6 Density-functional results: Hydrogen-bonded clusters

Turning to the H-bonded clusters, the components of the total binding energies are shown in Table 3 with selected DFAs. Overall, LC-ωPBE gives the best performance for the rigid- and relaxed-monomer binding energies and for $\Delta E^{(2)}$ and $\Delta E^{(\geq 4)}$. B971 gives the lowest errors for $\Delta E^{(3)}$ and PBE0 the lowest errors for $\Delta E^{(1)}$. No single term dominates either the total binding energies or the errors and the sources of error for each component will be discussed separately.

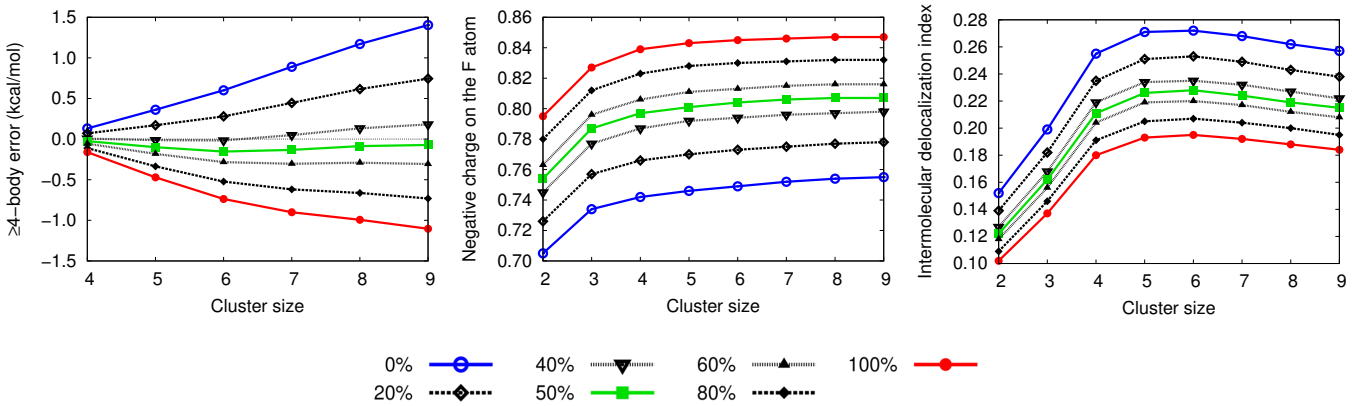
6.1 $\geq$4-body errors

Let us begin by analyzing the highest-order contribution to binding: the $\geq$4-body energies. The errors in Table 3 follow the general trend that range-separated < hybrid < GGA, regardless of the choice of semilocal functional. This

Table 3: Accumulated binding-energy contributions for the H-bonded clusters (sum of the corresponding contributions for all n-body clusters, with $n = 2$–6) with CCSD(T) and selected dispersion-corrected functionals. All values are kcal/mol.

	CCSD(T)	LC-ωPBE	B3LYP	B971	PBE0	BLYP	PW86PBE	PBE
				BE_{rigid}				
NH_3	76.65	76.86	81.56	81.94	84.33	83.65	87.65	87.35
H_2O	134.72	137.07	142.68	138.20	144.42	141.38	143.97	144.59
HF	154.78	158.91	166.80	161.06	170.29	160.23	163.05	167.06
MAE		2.23	8.30	5.02	10.96	6.37	9.51	10.95
				$\Delta E^{(2)}$				
NH_3	64.61	63.14	67.37	69.62	71.49	68.03	73.46	74.35
H_2O	101.25	100.73	105.17	103.69	109.48	100.91	105.44	108.41
HF	87.86	88.90	95.95	92.88	101.43	86.79	91.54	96.94
MAE		1.01	4.92	4.15	9.56	1.61	5.57	8.66
				$\Delta E^{(3)}$				
NH_3	11.12	12.76	13.25	11.25	11.72	14.63	13.01	11.70
H_2O	29.55	31.70	32.36	29.90	30.51	33.71	31.76	30.44
HF	58.19	60.98	61.61	58.67	59.40	63.68	61.34	59.68
MAE		2.19	2.78	0.32	0.92	4.38	2.41	0.98
				$\Delta E^{(\geq 4)}$				
NH_3	0.92	0.95	0.94	1.07	1.12	0.99	1.18	1.30
H_2O	3.91	4.64	5.15	4.62	4.43	6.76	6.77	5.74
HF	8.72	9.02	9.24	9.51	9.46	9.76	10.17	10.44
MAE		0.35	0.59	0.55	0.48	1.32	1.52	1.31
				$\Delta E^{(1)}$				
NH_3	−1.15	−1.98	−1.23	−0.90	−1.37	0.33	0.16	0.35
H_2O	−6.06	−5.27	−4.36	−5.08	−5.71	−1.31	−1.95	−2.18
HF	−24.20	−19.92	−18.19	−20.97	−22.61	−8.07	−8.81	−10.72
MAE		1.97	2.60	1.49	0.72	7.46	6.94	6.29
				BE_{relax}				
NH_3	75.50	74.88	80.34	81.04	82.97	83.98	87.81	87.70
H_2O	128.65	131.81	138.32	133.12	138.71	140.08	142.02	142.41
HF	130.58	138.98	148.61	140.09	147.67	152.15	154.24	156.34
MAE		4.06	10.85	6.51	11.54	13.83	16.45	17.24

Figure 7: Left: error in the ≥4-body binding-energy contribution for the $(HF)_n$ clusters, relative to the CCSD(T) reference data. Middle: monomer polarization measured using the Bader charge on the F atom. Right: intermolecular delocalization index. All plots use hybrid functionals based on the BLYP semilocal functional with varying fractions of exact exchange.



suggests that the errors in the ≥4-body energies are determined by delocalization error from the exchange functional,[79–82] and we argue in the following that this is indeed the case by employing two usual indicators for delocalization error in complex systems.[107,121]

The first indicator is that the error in the 4-body energies is almost exclusively determined by the amount of exact exchange in the functional, and is minimized for fractions of exact-exchange mixing close to 50%. Figure 7 (left) shows this behavior for BLYP-based functionals. Analogous plots for the other semilocal functionals examined (PBE, PW86PBE, and TPSS) are almost exactly the same, and are shown in Figure 4 of the SI. This behavior was also observed for halogen-bonded systems,[107] where delocalization error is expected to impact the binding energies to a greater extent because of the increased orbital-interaction component compared to H-bonded systems. In that work, we showed that only functionals that minimize delocalization error (range-separated hybrids and 50%-hybrids) are successful in treating halogen-bonded dimers. Delocalization error in halogen-bonded systems strongly influences the binding energy of simple dimers, where many body induction effects are absent, so there are reasonable grounds to believe that delocalization error also affects the 2-body and 3-body energies of the H-bonded clusters studied in this work.

Due to delocalization error, GGA functionals overestimate the cooperative H-bonding effects that increase the polarization of each monomer involved in H-bonding.[47] This effect carries a characteristic signature in the electron density (fractional charges are over-stabilized[81,107,121] and monomers are spuriously over-polarized[47]) and in the spread of the density matrix (overly delocalized[107]). Intermolecular delocalization can be conveniently measured by calculating Bader's delocalization indices[122–126] (DIs) between neighboring molecules, and we have shown previously that DIs serve as excellent indicators for delocalization error.[77,107] Figure 7 shows the intramolecular charge transfer and intermolecular delocalization index in the HF cycles for BLYP-based hybrids. The plots show that the electron-density distribution and electron delocalization change dramatically with the fraction of exact exchange, much more so than with cluster size.

An interesting question regarding delocalization error is how to separate the direct energetic contributions from the form of the exchange functional and the changes in density resulting from the overpolarization of the monomers in the cluster.[78] To do this, we compare the n-body energies calculated using BLYP, BH&HLYP, and BLYP evaluated using the self-consistent BH&HLYP density (shown in Figure 7 of the SI). The difference between the BLYP energies using the BLYP and

BH&HLYP densities is much smaller than between the BLYP and BH&HLYP energies both using the BH&HLYP density. Since delocalization error depends only on the exact exchange fraction, this observation indicates that delocalization error comes directly from the exchange functional energy term.

The choice of semilocal exchange-correlation functional has no significant effect on the ≥4-body contributions to the H-bonded cluster binding energies, which are controlled by the exact-exchange content alone. This observation stands in contrast with our analysis in dispersion-bound clusters, where we saw that many-body errors coming from exchange-repulsion are strongly dependent on the underlying GGA. Since many-body dispersion contributions can not be modeled by the base functional, delocalization error is necessarily linked to errors in the ≥4-body induction terms. This argument is also intuitively reasonable, since induction energies arise from molecular charge rearrangements, and delocalization error operates by overstabilizing systems with fractional charges.[127]

It is interesting to point out that these results, and those shown below, imply that semilocal functionals, which are always affected by delocalization error, are inevitably unable to give an accurate description of systems where induction effects are important, and in particular H-bonding interactions. This is the case even if the semilocal functional is fitted to dispersionless SAPT interaction energies.[128]

6.2 Three-body errors

The errors in the 3-body terms for hybrids built using six GGA functionals are shown in Figure 8. The plots show that all of the hybrids exhibit the same trend for the 3-body energies: they become less binding as the exact exchange fraction is increased. This suggests that the 3-body energies are also affected by delocalization error (c.f. dispersion-bound clusters, Figure 5). However, unlike the ≥ 4-body energies, the extent of the overestimation of the 3-body energies is dependent on the identity of the semilocal exchange functional upon which the hybrid is built.

It is illustrative to compare the 3-body errors for HF clusters in Figure 8 to those for the methane clusters in Figure 5. For the methane clusters, the trends in the 3-body errors were explained as arising from the error in predicted repulsion by the semilocal functional. A combination of the overestimation of 3-body induction caused by delocalization (overly-attractive for all GGA functionals) and repulsion errors (attractive or repulsive depending on the functional form) can be used to explain the shape of the plots in Figure 8. With B88 exchange, the semilocal functional over-estimates the 3-body contribution in the methane clusters, and exact-exchange mixing decreases the binding. For the HF clusters, delocalization error adds to this effect, resulting in large variations in the 3-body energies. For the functionals based on PBE or PW86, more exact-exchange mixing leads to increased binding in the methane clusters. Therefore, in the HF clusters, cancellation occurs between delocalization and intermolecular-repulsion errors, and the impact of varying the exact-exchange mixing fraction on the 3-body errors is diminished.

These observations explain the results in Table 3, where we see the same trend in the GGA functionals as in the dispersion-bound clusters: the 3-body energies are most attractive with BLYP, followed by PW86, and then PBE. However, unlike the case for the dispersion-bound complexes, all the GGAs predict the 3-body terms to be too stabilizing and this can be attributed to delocalization error.

6.3 Two-body errors

Figure 9 shows the behavior of the six series of hybrid functionals for the pairwise contributions to the binding energies of the HF clusters. The 2-body errors are dominated by the fraction of exact exchange in each functional. However, unlike the 3-body and higher-order contributions to the binding energies, increased fractions of exact exchange result in strong overbinding. This observation is in strong contrast with the trend for the $n > 2$ terms in the same clusters and also with our results for halo-

Figure 8: Errors for the 3-body contributions to the cyclic $(HF)_n$ cluster binding energies against the number of molecules in the cluster using hybrid functionals built from combinations of B88 (left), PW86 (middle), and PBE (right) exchange with PBE (top) and LYP (bottom) correlation. Pure GGA functionals are represented in blue, half-and-half functionals are shown in green, and HF plus correlation is in red. Positive values indicate overestimation of the corresponding energy contribution. Figure 5 in the SI contains the same information with 10% steps in the exact exchange fraction. All units are kcal/mol.

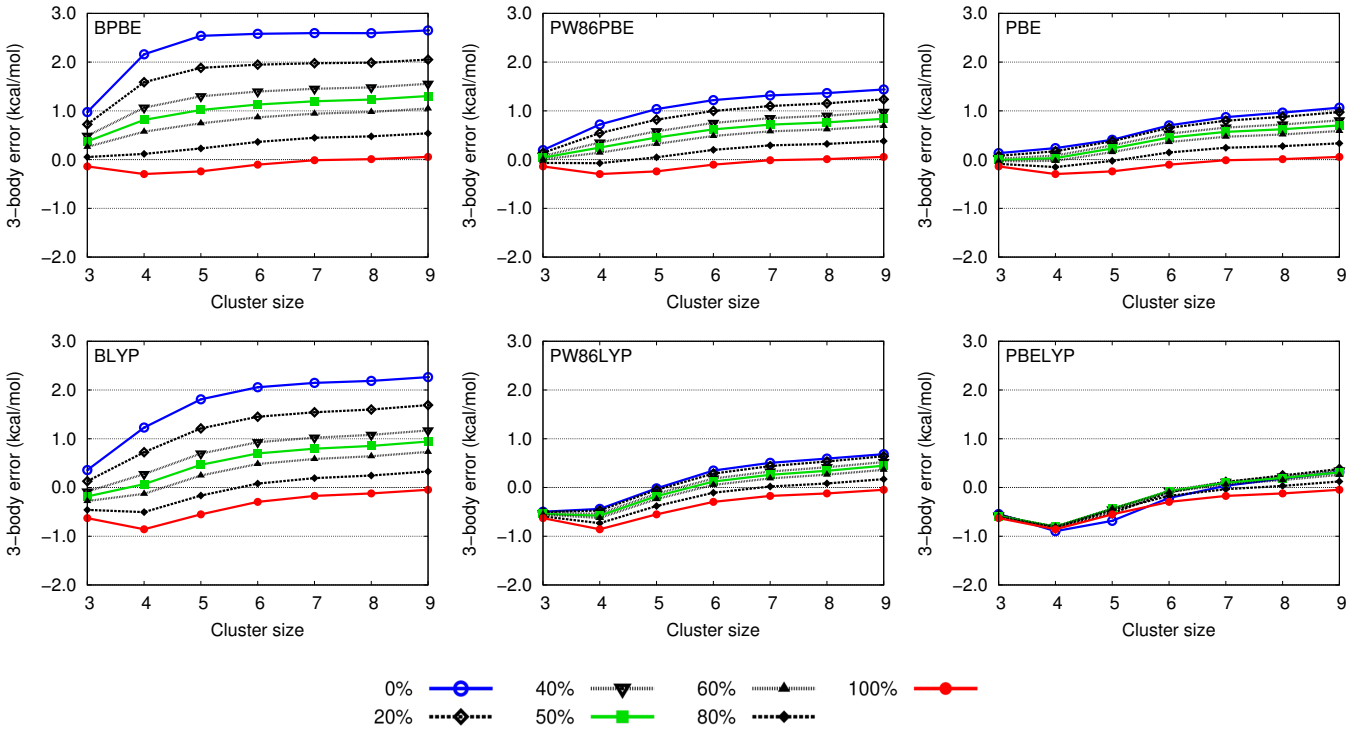
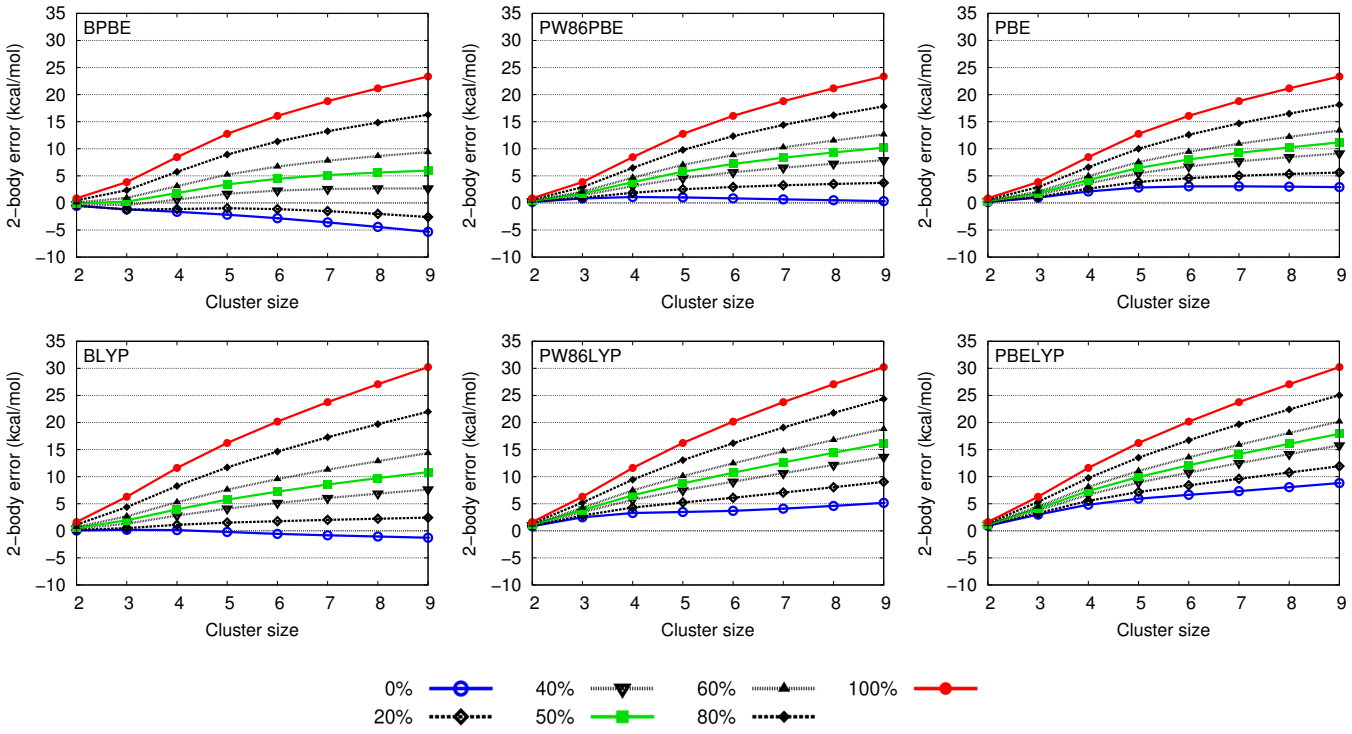


Figure 9: Errors in the 2-body contributions to the cyclic $(HF)_n$ cluster binding energies against the number of molecules in the cluster using hybrid functionals built from combinations of B88 (left), PW86 (middle), and PBE (right) exchange with PBE (top) and LYP (bottom) correlation. Pure GGA functionals are represented in blue, half-and-half functionals are shown in green, and HF plus correlation is in red. Positive values indicate overestimation of the corresponding energy contribution. Figure 6 in the SI contains the same information with 10% steps in the exact exchange fraction.All units are kcal/mol.



gen bonded dimers,[107] where the observed trend was the reverse: additional exact exchange in the base functional decreases the binding energy.

To understand these results, we can use the fact that the two-body energy in HF clusters comes mostly from the hydrogen bonds. We can therefore use the dimer as a model to study the source of DFA errors in the 2-body term. Table 2 shows the SAPT breakdown of the binding energies in three H-bonded dimers (NH_3, H_2O, and HF) and one halogen-bonded dimer ($FBr\cdots NCH$). Our results show that the induction contribution in H-bonded systems is relatively small compared to the electrostatic term, whereas it is relatively much higher in the halogen-bonded dimer.

A simple explanation for the results in Figure 9 is that, in the case of the H-bonded dimers, delocalization error affects both the induction and the electrostatic contributions, although with errors of different sign. This is reasonable; given the wide variation in atomic charges caused by exact-exchange admixture shown in Figure 7, one would expect that the electrostatic energies will be affected by delocalization error, which is a function solely of the exact-exchange fraction. In the case of the three-body and higher-order energies, no electrostatic term is present (because electrostatic effects are additive[23]), and delocalization error causes an overestimation of the binding energy contribution through the induction term. In the case of two-body energies, however, both the induction and electrostatic energies are affected.

The SAPT results in Table 2 suggest that, since the induction term in H-bonded systems is small compared to the electrostatic term, a decrease in delocalization error causes charge localization and an increase in binding from the electrostatic term. When the exact-exchange fraction in the hybrid functionals increases above 50%, localization error sets in, causing the well-known overstabilization of ionic states seen from Hartree-Fock calculations. As an illustrative example, HFLYP and HFPBE overbind the LiF molecule relative to the separated Li^+ and F^- ions by 5.4 and 3.4 kcal/mol, respectively, compared to CCSD(T) reference data, despite the usual underbinding tendency of Hartree-Fock exchange regarding bond dissociation.[129] Figure 7 shows that the fluorine atomic charges in the HF hexamer become more negative and the H-F bond becomes more ionic as the exact-exchange mixing fraction is increased. This can be contrasted with halogen-bonded dimers (Table 2 and ref. 107), where induction is more important than in H-bonded systems, and the delocalization-error effect on induction dominates over electrostatics.[107]

In addition, Figure 9 shows that the 2-body errors also depend on the underlying GGA on which the hybrids are built, although to a lesser extent than the exact-exchange mixing in the particular case of HF clusters. As in the case of the 3-body errors, this is caused by the repulsion error discussed above. Let us focus only on the pure GGA functionals (all present delocalization error to the same extent) and compare to the repulsion error exhibited in the dispersion-bound clusters (Figure 5). For BLYP and BPBE, delocalization error overbinds the clusters, but this is offset by the strong underestimation of 2-body energies caused by repulsion error. As a result, BLYP performs particularly well for the 2-body energies of the H-bonded clusters in Table 3 and excellent performance of BLYP-D3 has been noted previously for ice.[130] Conversely, for PW86 and PBE, the repulsion error causes an overestimation of the 2-body energies, which is only made worse by delocalization error effects. Indeed, for hybrid functionals based on PW86 and PBE exchange, the base functional alone is already over-binding for any amount of exact exchange. This is clearly a problem for the use of dispersion-correction methods built on these functionals, since the dispersion-energy contribution is always attractive.

Similar trends to those in Figure 9 are observed for the two-body energy errors in the other H-bonded clusters (H_2O and NH_3). However, by comparing PBE and PBE0 or BLYP and B3LYP results in Table 3, one can readily see that the action of exact exchange on the total binding energies differs depending on the H-bonded system, and the relative importance of

exchange-repulsion, induction, and electrostatics. For water, as for HF, an increased fraction of exact exchange results in an increase in binding, whereas it is the contrary for ammonia. A succinct analysis of the n-body contributions in the other H-bonded clusters is presented in Section 6.6.

We note that PBE0, which is commonly chosen for modeling water and other H-bonded systems, behaves well for the 3-body energies, but grossly overestimates the 2-body energies, as shown in Figures 8 and 9 and in Table 3. This observation explains the systematic overestimation of water-cluster binding energies,[59] the lattice energies of ice,[115,116] and the density of liquid water with PBE0. Any dispersion correction coupled with PBE or PBE0 is incapable of modeling strongly-polarized H-bonded systems like the HF clusters because these base functionals overestimate the binding energies.

6.4 Monomer deformation

Finally, monomer deformation is also an important contribution to the binding energies in H-bonded clusters. Table 3 shows that GGA functionals give the largest errors in the deformation energies, while the errors are reduced for the hybrid and range-separated hybrid functionals. All methods shown overestimate rigid-monomer binding energies. As the stabilization coming from monomer deformation is overestimated, agreement with the reference relaxed-monomer binding energies is consequently worse than with rigid-monomer binding energies. Because benchmarking studies are often carried out using rigid geometries,[24,26,28] one should expect more overbinding in production DFT calculations (e.g. in the calculation of lattice energies[42]), where energies are always computed relative to relaxed-monomer geometries.

The trends in deformation energies can be explained based on the observed changes in bond lengths upon hydrogen bond formation. When a hydrogen bond is formed, the X-H bond of the hydrogen donor is stretched relative to the isolated monomer, and the degree of bond stretching increases with stronger cooperative H-bonding effects, as shown in the left panel of Figure 10 (c.f. Figure 2). Thus, the deformation errors are related to how well the functionals reproduce the reference CCSD(T) intramolecular X-H bond stretching potential-energy surface close to the equilibrium bond length.

The errors in the deformation energies for the HF clusters are shown in the right panel of Figure 10 for BLYP-based hybrid functionals with various fractions of exact exchange. Analogous plots for hybrids based on all other semilocal functionals considered (shown in Figure 8 of the SI) are essentially the same. The ability of a functional to predict an accurate X-H bond length for the H-bond donor determines its performance for $\Delta E^{(1)}$. The GGA functionals all overestimate the equilibrium bond length and consequently predict too little energy penalty for monomer deformation. Conversely, functionals with high percentage of exact exchange underestimate the equilibrium bond length and overestimate the deformation energies. Roughly 30% exact exchange gives the best agreement with CCSD(T) reference data, explaining the good performance of PBE0 (25% exact exchange[104]) in Table 3. Hybrid functionals that perform well for thermochemistry and bond-dissociation energies (with exact-exchange fractions close to 25%) provide reliable deformation energies.[101,104] It is also important to note that deformation energies in Figure 10 are calculated at a fixed cluster geometry. In practical applications, the cluster geometry is relaxed for each functional, which would introduce variations in the monomer bond stretching and, possibly, decrease the magnitude of the deformation error for approximate density functionals.

6.5 Range separation

Now we examine the performance of LC-ωPBE as a function of the range-separation parameter (the literature definition of LC-ωPBE[93,94] uses $\omega = 0.4$). Figure 11 shows the n-body contributions to the binding energy in the $(HF)_n$ clusters with ω. As in the case of hybrid functionals, the GGA ($\omega = 0.0$) overbinds the $\geq$4-body contribution. Increasing ω, which increases the weight of the long-range, exact-exchange part

Figure 10: Reference CCSD(T) deformation energies for the cyclic $(HF)_n$ clusters and H-F bond lengths (left) and relaxation-energy errors using hybrid functionals with various fractions of exact exchange built on BLYP (right) against the number of molecules in the cluster.

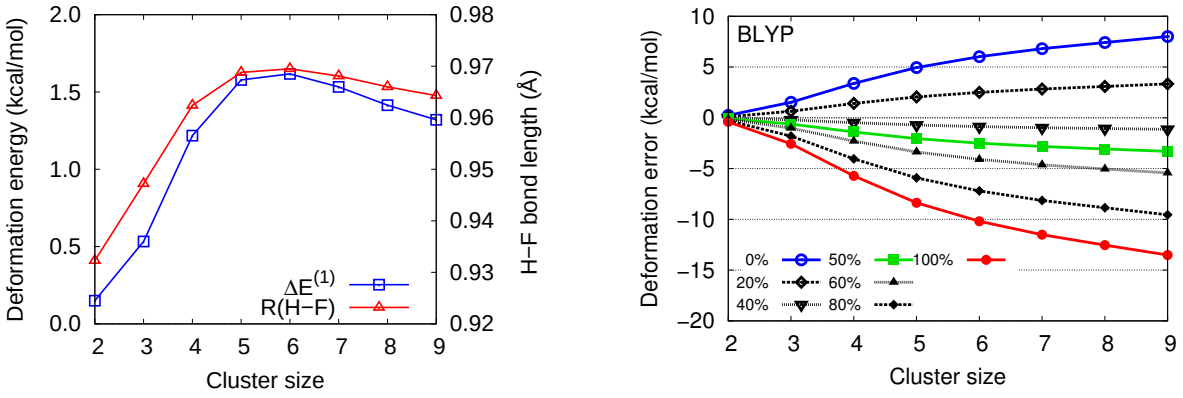
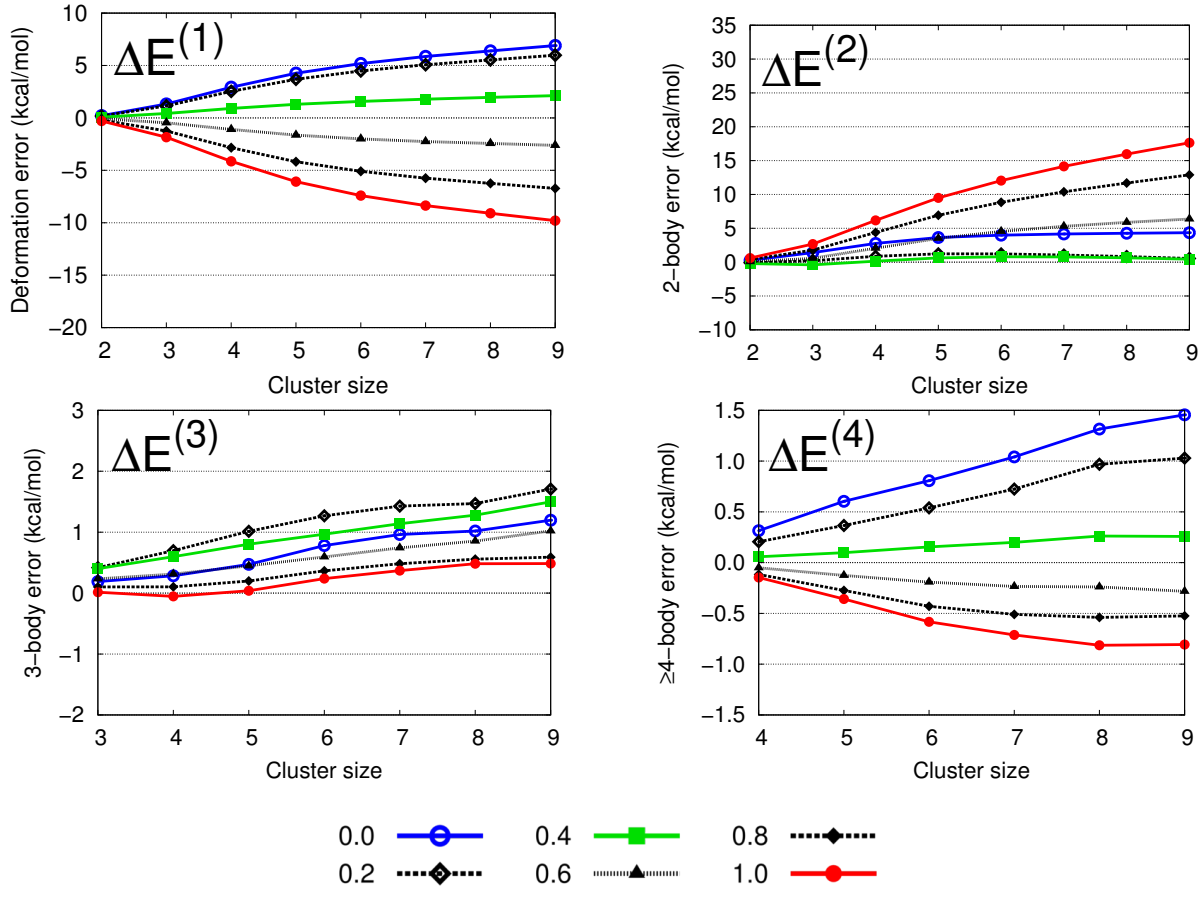


Figure 11: Errors in the deformation ($\Delta E^{(1)}$), 2-body, 3-body, and ≥4-body ($\Delta E^{(4)}$) contributions to the $(HF)_n$ cluster binding energies using range-separated functionals based on LC-ωpbe as a function of the range-separation parameter (ω) and cluster size. The pure GGA is represented in blue. Positive values indicate overestimation of the corresponding energy contribution. All units are kcal/mol.



of the functional, decreases the value of this contribution, minimizing delocalization error at about $\omega = 0.5$. Interestingly, the same trend is observed for the deformation energies, with the same optimal value of the range separation parameter. This is in contrast to the observed behavior of hybrid functionals for which a single fraction of exact exchange can not minimize delocalization error (50%) and deformation errors (20–25%).

The 3-body errors, as in the case of the hybrids, are affected by a mixture of delocalization and repulsion error. The error plot is similar to PBE (Figure 8), which is no surprise given the similarity of the underlying GGA functionals. However, the variation of ω induces a more complex behavior than in the case of PBE hybrids. With higher ω, the 3-body error first increases (more binding) and peaks at $\omega = 0.3$, at which point the error decreases and approaches zero. This behavior is likely the result of the interplay between delocalization error (as evidenced by the $\Delta E^{(4)}$ plot) and repulsion error as exact exchange enters the functional definition. The 2-body plot shows a similar behavior, where increasing ω first decreases the binding (and the error) up to $\omega = 0.3$, at which point overbinding begins to occur. Minimal 2-body error is found at $\omega = 0.2$.

In view of the preceding discussion, it is clear that the underlying GGA functional in LC-ωPBE is affected by non-negligible repulsion error. However, provided a range-separated functional could be built on a GGA with zero repulsion error, Figure 11 suggests that a range-separation parameter can be found that gives good deformation energies and minimizes delocalization error. Although this may be a good recipe for building functionals with accurate treatment of the non-dispersive parts of non-covalent interactions in small systems, range-separated functionals are afflicted by strong size-dependence in their delocalization error behavior.[131–133]

6.6 Other hydrogen-bonded systems

The conclusions from the preceding analysis also apply to the other hydrogen-bonded systems in this study—water and ammonia—provided one takes into account the relative importance of each n-body term, as shown in Figure 1. Figures 8 and 9 of the SI show the n-body contribution errors in water and ammonia clusters, respectively. The 2-body, 3-body, $\geq$4-body, and deformation errors are plotted as a function of cluster size (up to the hexamer), for hybrids built on BLYP and PBE with varying amounts of exact exchange.

For all n-body errors, the trends regarding the fraction of exact exchange, and the relative behavior of PBE and BLYP, are the same those presented in figures 7, 8, 9, and 10 for the HF clusters. The $\geq$4-body contribution trend is ammonia $<$ water $<$ HF (Figure 1), and the same is observed for the corresponding errors. The $\geq$4-body error is functional independent, except in the case of ammonia, where the delocalization contribution is so small that the 4-body repulsion error becomes noticeable.

The spread of the three-body errors caused by variations in the exact exchange fraction in water and ammonia clusters is higher in BLYP than in PBE, and has the same behavior as in HF clusters (Figure 8). The invariance of PBE hybrids to exact exchange composition can be explained by cancellation between repulsion (PBE gives an overly repulsive 3-body term) and delocalization error (overly attractive). The 2-body and deformation error plots are also consistent with the HF results in Figure 9 and 10, respectively.

7 Conclusions

In this article, we analyzed the way in which different density-functional approximations affect the description of the non-dispersion components in non-covalent interactions. This was accomplished by first developing a database of reference CCSD(T) many-body energy terms for a set of small molecular clusters bound by non-

covalent interactions of different types. The ability of various density-functional approximations to reproduce the different n-body terms was assessed, including the effect of the semilocal functional, exact-exchange admixture, and range separation. The reference n-body energy decomposition shows that only pairwise contributions to the binding energy are significant for non-hydrogen-bonded (dispersion-bound) clusters. However, for hydrogen-bonded clusters, 3-body, monomer deformation, and higher-order effects all contribute to the total binding energies.

Our analysis shows that there are three main contributors to the error in the binding energy from different density-functional approximations:

1. *Repulsion error* arises from the incorrect treatment of intermolecular repulsion originating from the overlap between closed-shell wavefunctions. This error, related to the exchange-repulsion term in perturbation theory, is controlled by the semilocal functional, primarily by the exchange term, and to a lesser extent by the correlation functional. The 2-body repulsion error is positive or negative depending on the functional, and the sign alternates with the order (n) associated with higher-order n-body terms. The repulsion error from the exchange and the correlation functionals behaves in similar ways, which allows for error cancellation if they are correctly paired (for instance, PW86PBE). However, the presence of such error cancellation in the semilocal functional implies that hybrid functionals built on it will not represent intermolecular repulsion as accurately. Repulsion error affects all types of non-covalent interactions, since the exchange-repulsion contribution is always present.

2. *Delocalization error* arises from the tendency of semilocal functionals to overstabilize fractional charges. Delocalization error is controlled by the amount of exact exchange and is independent of the semilocal functional employed. In perturbation theory terms, it affects the induction energy, which results in an overstabilization of all n-body induction terms, and an underestimation of the additive (2-body) electrostatic energy. Delocalization error is only present in systems with significant intra- or intermolecular charge transfer such as hydrogen bonds or halogen bonds. While the 3-body and higher order error results in spurious attraction, the effect on the 2-body electrostatic term depends on the balance between induction and electrostatics, and hence on the type of non-covalent contact. From the systems examined in this work and in our previous study,[107] evidence indicates that delocalization error results in the overestimation of 2-body energies in halogen bonds and underestimation of 2-body energies in hydrogen bonds.

3. *Deformation error* is related to the ability of the functional to model the energetics of bond stretching upon the formation of a non-covalent interaction. This error affects deformation energies, and has the same behavior as errors in the calculation of bond dissociation energies, which have been extensively studied in the past.[101,104,106] Semilocal functionals overestimate deformation energies, and a 25% fraction of exact exchange is optimal. This error only appears if the non-covalent interaction causes significant monomer distortion, for instance, in hydrogen bonds and halogen bonds.

While LC-ωPBE-XDM gives the lowest overall errors, none of the studied common functionals is able to provide sufficiently accurate results for large molecular clusters. However, the insights listed above provide some guidelines for developing functionals that accurately describe the non-dispersive part of a non-covalent interaction. In order to minimize repulsion error, semilocal exchange and correlation functionals should be chosen such that they independently and correctly describe intermolecular repulsion. To our knowledge, there is no commonly used hybrid functional that simulta-

neously minimizes delocalization error (50% exact exchange is optimal) and deformation error (25%). LC-ωPBE with $\omega = 0.5$ seems to minimize both errors at the same time, so a range-separated functional based on a repulsion-error-free semilocal functional could be a good candidate for accurate, across-the-board performance for non-covalent interactions in small systems. However, it should be kept in mind that delocalization error increases with system size when range-separation is applied.[131–133]

8 Acknowledgements

AOR thanks the Spanish Malta/Consolider initiative (no. CSD2007-00045). GAD thanks the Natural Sciences and Engineering Research Council of Canada and the University of British Columbia for funding, and Compute Canada for a generous allocation of computing resources. ERJ thanks the Natural Sciences and Engineering Research Council of Canada for funding.

References

(1) DiLabio, G. A.; Otero-de-la Roza, A. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Ed.; Wiley-VCH, Hoboken, NJ, 2016; (in press, arxiv:1405.1771).

(2) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2007**, *127*, 154108.

(3) Otero-de-la Roza, A.; Johnson, E. R. *J. Chem. Phys.* **2013**, *138*, 054103.

(4) Otero-de-la Roza, A.; Johnson, E. R. *J. Chem. Phys.* **2013**, *138*, 204109.

(5) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

(6) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

(7) Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D.; Lundqvist, B. *Phys. Rev. Lett.* **2004**, *92*, 246401.

(8) Thonhauser, T.; Cooper, V.; Li, S.; Puzder, A.; Hyldgaard, P.; Langreth, D. *Phys. Rev. B* **2007**, *76*, 125112.

(9) Langreth, D. C.; Lundqvist, B. I.; Chakarova-Kck, S. D.; Cooper, V. R.; Dion, M.; Hyldgaard, P.; Kelkkanen, A.; Kleis, J.; Kong, L.; Li, S.; Moses, P. G.; Murray, E.; Puzder, A.; Rydberg, H.; Schrder, E.; Thonhauser, T. *J. Phys.: Condens. Matter* **2009**, *21*, 084203.

(10) Klimeš, J.; Bowler, D. R.; Michaelides, A. *J. Phys.: Condens. Matter* **2010**, *22*, 022201.

(11) Lee, K.; Murray, E. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. *Phys. Rev. B* **2010**, *82*, 081101.

(12) Torres, E.; DiLabio, G. A. *J. Phys. Chem. Lett.* **2012**, *3*, 1738–1744.

(13) DiLabio, G. A.; Koleini, M.; Torres, E. *Theor. Chem. Acc.* **2013**, *132*, 1–13.

(14) Vydrov, O. A.; Van Voorhis, T. *Phys. Rev. Lett.* **2009**, *103*, 063004.

(15) Vydrov, O. A.; Van Voorhis, T. *J. Chem. Phys.* **2009**, *130*, 104105.

(16) Vydrov, O. A.; Voorhis, T. V. *J. Chem. Phys.* **2010**, *133*, 244103.

(17) Vydrov, O. A.; Van Voorhis, T. *J. Chem. Theory Comput.* **2012**, *8*, 1929–1934.

(18) Steinmann, S. N.; Corminboeuf, C. *J. Chem. Phys.* **2011**, *134*, 044117.

(19) Steinmann, S. N.; Corminboeuf, C. *J. Chem. Theory Comput.* **2011**, *7*, 3567–3577.

(20) Tkatchenko, A.; Scheffler, M. *Phys. Rev. Lett.* **2009**, *102*, 73005.

(21) von Lilienfeld, O. A.; Tkatchenko, A. *J. Chem. Phys.* **2010**, *132*, 234109.

(22) Tkatchenko, A.; DiStasio Jr, R.; Car, R.; Scheffler, M. *Phys. Rev. Lett.* **2012**, *108*, 236402.

(23) Stone, A. J. *The theory of intermolecular forces*; Clarendon Press, Oxford, 1996.

(24) Jurečka, P.; Šponer, J.; Černỳ, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

(25) Kannemann, F. O.; Becke, A. D. *J. Chem. Theory Comput.* **2010**, *6*, 1081–1088.

(26) Gráfová, L.; Pitoňák, M.; Řezáč, J.; Hobza, P. *J. Chem. Theory Comput.* **2010**, *6*, 2365–2376.

(27) Takatani, T.; Hohenstein, E. G.; Malagoli, M.; Marshall, M. S.; Sherrill, C. D. *J. Chem. Phys.* **2010**, *132*, 144104.

(28) Řezáč, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2011**, *7*, 2427–2438.

(29) Řezáč, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2011**, *7*, 3466–3470.

(30) Marshall, M. S.; Burns, L. A.; Sherrill, C. D. *J. Chem. Phys.* **2011**, *135*, 194102.

(31) Goerigk, L.; Grimme, S. *J. Chem. Theory Comput.* **2011**, *7*, 291–309.

(32) Faver, J. C.; Benson, M. L.; He, X.; Roberts, B. P.; Wang, B.; Marshall, M. S.; Kennedy, M. R.; Sherrill, C. D.; Merz, K. M. *J. Chem. Theory Comput.* **2011**, *7*, 790–797.

(33) Grimme, S. *Chem. Eur. J.* **2012**, *18*, 9955–9964.

(34) Sedlak, R.; Janowski, T.; Pitoňák, M.; Řezáč, J.; Pulay, P.; Hobza, P. *J. Chem. Theory Comput.* **2013**, *9*, 3364–3374.

(35) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46–53.

(36) Bates, D. M.; Smith, J. R.; Janowski, T.; Tschumper, G. S. *J. Chem. Phys.* **2011**, *135*, 044123.

(37) Richard, R. M.; Herbert, J. M. *J. Chem. Phys.* **2012**, *137*, 064113.

(38) Richard, R. M.; Lao, K. U.; Herbert, J. M. *J. Chem. Phys.* **2013**, *139*, 224102.

(39) Risthaus, T.; Grimme, S. *J. Chem. Theory Comput.* **2013**, *9*, 1580–1591.

(40) Ambrosetti, A.; Alfé, D.; DiStasio Jr, R. A.; Tkatchenko, A. *J. Phys. Chem. Lett.* **2014**, *5*, 849–855.

(41) Otero-de-la Roza, A.; Johnson, E. R. *J. Chem. Theory Comput.* **2015**, *11*, 4033–4040.

(42) Otero-de-la Roza, A.; Johnson, E. R. *J. Chem. Phys.* **2012**, *137*, 054103.

(43) Dahlke, E. E.; Olson, R. M.; Leverentz, H. R.; Truhlar, D. G. *J. Phys. Chem. A* **2008**, *112*, 3976–3984.

(44) Bates, D. M.; Tschumper, G. S. *J. Phys. Chem. A* **2009**, *113*, 3555–3559.

(45) Bryantsev, V. S.; Diallo, M. S.; van Duin, A. C. T.; Goddard III, W. A. *J. Chem. Theory Comput.* **2009**, *5*, 1016–1026.

(46) Chen, Y.; Li, H. *J. Phys. Chem. A* **2010**, *114*, 11719–11724.

(47) Wang, F. F.; Jenness, G.; Al-Saidi, W. A.; Jordan, K. D. *J. Chem. Phys.* **2010**, *132*, 134303.

(48) Temelso, B.; Archer, K. A.; Shields, G. C. *J. Phys. Chem. A* **2011**, *115*, 12034–12046.

(49) Miliordos, E.; Aprà, E.; Xantheas, S. S. *J. Chem. Phys.* **2013**, *139*, 114302.

(50) Pruitt, S. R.; Leang, S. S.; Xu, P.; Fedorov, D. G.; Gordon, M. S. *Comp. Theor. Chem.* **2013**, *1021*, 70–83.

(51) Leverentz, H. R.; Qi, H. W.; Truhlar, D. G. *J. Chem. Theory Comput.* **2013**, *9*, 995–1006.

(52) Anacker, T.; Friedrich, J. *J. Comp. Chem.* **2014**, *35*, 634–643.

(53) Gillan, M. J. *J. Chem. Phys.* **2014**, *141*, 224106.

(54) Gadre, S. R.; Yeole, S. D.; Sahu, N. *Chem. Rev.* **2014**, *114*, 12132–12173.

(55) Axilrod, B. M.; Teller, E. *J. Chem. Phys.* **1943**, *11*, 299.

(56) Muto, Y. *J. Phys. Math. Soc. Japan* **1943**, *17*, 629.

(57) Bell, R. J. *J. Phys. B: At. Mol. Phys.* **1970**, *3*, 751.

(58) Gillan, M.; Manby, F.; Towler, M.; Alfè, D. *J. Chem. Phys.* **2012**, *136*, 244105.

(59) Gillan, M. J.; Alfè, D.; Bartók, A. P.; Csàyi, G. *J. Chem. Phys.* **2013**, *139*, 244504.

(60) Medders, G. R.; Babin, V.; Paesani, F. *J. Chem. Theory Comput.* **2013**, *9*, 1103–1114.

(61) Gillan, M. J.; Alfè, D.; Bygrave, P. J.; Taylor, C. R.; Manby, F. R. *J. Chem. Phys.* **2013**, *139*, 114101.

(62) Christie, R. A.; Jordan, K. D. *Struct. Bond.* **2005**, *116*, 27–41.

(63) Góra, U.; Podeszwa, R.; Cencek, W.; Szalewicz, K. *J. Chem. Phys.* **2011**, *135*, 224102.

(64) Tauer, T. P.; D, S. C. *J. Phys. Chem. A* **2005**, *109*, 10475–10478.

(65) Pitoňák, M.; Neogrády, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 1369–1378.

(66) Beran, G. J. O. *J. Chem. Phys.* **2009**, *130*, 164115.

(67) Beran, G. J. O.; Nanda, K. *J. Phys. Chem. Lett.* **2010**, *1*, 3480–3487.

(68) Wen, S.; Nanda, K.; Huang, Y.; Beran, G. J. O. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7578–7590.

(69) Bygrave, P. J.; Allan, N. L.; Manby, F. R. *J. Chem. Phys.* **2012**, *137*, 164102.

(70) Kennedy, M. R.; McDonal, A. R.; DePrince III, A. E.; Marshall, M. S.; Podeszwa, R.; Sherril, C. D. *J. Chem. Phys.* **2014**, *140*, 121104.

(71) Yang, J.; Hu, W.; Usvyat, D.; Matthew, D.; Schüz, M.; Chan, G. K.-L. *Science* **2014**, *345*, 640–643.

(72) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *122*, 154104.

(73) Otero-de-la Roza, A.; Mallory, J. D.; Johnson, E. R. *J. Chem. Phys.* **2014**, *140*, 18A504.

(74) Otero-de-la Roza, A.; Johnson, E. R. *J. Chem. Phys.* **2012**, *136*, 174109.

(75) Otero-de-la Roza, A.; Cao, B. H.; Price, I. K.; Hein, J. E.; Johnson, E. R. *Angew. Chem. Int. Ed.* **2014**, *53*, 7879–7882.

(76) Szalewicz, K.; Leforestier, C.; Van Der Avoird, A. *Chem. Phys. Lett.* **2009**, *482*, 1–14.

(77) Rezac, J.; Huang, Y.; Hobza, P.; Beran, G. J. *J. Chem. Theory Comput.* **2015**, *11*, 3065–3079.

(78) Huang, Y.; Beran, G. J. *J. Chem. Phys.* **2015**, *143*, 044113.

(79) Zhang, Y. K.; Yang, W. T. *J. Chem. Phys.* **1998**, *109*, 2604–2608.

(80) Ruzsinszky, A.; Perdew, J. P.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E. *J. Chem. Phys.* **2006**, *125*, 194112.

(81) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. *Science* **2008**, *321*, 792.

(82) Kim, M.-C.; Sim, E.; Burke, K. *Phys. Rev. Lett.* **2013**, *111*, 073003.

(83) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, .; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09 Revision A.1. Gaussian Inc. Wallingford CT 2009.

(84) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553–566.

(85) Liedl, K. R. *J. Chem. Phys.* **1998**, *108*, 3199–3204.

(86) Halkier, A.; Klopper, W.; Helgaker, T.; Jorgensen, P.; Taylor, P. R. *J. Chem. Phys.* **1999**, *111*, 9157–9167.

(87) Dunning Jr, T. H. *J. Phys. Chem. A.* **2000**, *104*, 9062–9080.

(88) Mackie, I. D.; DiLabio, G. A. *J. Chem. Phys.* **2011**, *135*, 134318.

(89) Burns, L. A.; Marshall, M. S.; Sherrill, C. D. *J. Chem. Theory Comput.* **2014**, *10*, 49–57.

(90) DiLabio, G. A.; Johnson, E. R.; Otero-de-la Roza, A. *Phys. Chem. Chem. Phys.* **2013**, *15*, 12821.

(91) Johnson, E. R.; Otero-de-la Roza, A.; Dale, S. G.; DiLabio, G. A. *J. Chem. Phys.* **2013**, *139*, 214109.

(92) Friedrich, J.; Perlt, E.; Roatsch, M.; Spickermann, C.; Kirchner, B. *J. Chem. Theory Comput.* **2011**, *7*, 843–851.

(93) Vydrov, O. A.; Scuseria, G. E. *J. Chem. Phys.* **2006**, *125*, 234109.

(94) Vydrov, O. A.; Heyd, J.; Krukau, A. V.; Scuseria, G. E. *J. Chem. Phys.* **2006**, *125*, 074106.

(95) Torres, E.; DiLabio, G. A. *J. Chem. Theory Comput.* **2013**, *9*, 3342–4439.

(96) TURBOMOLE V6.5 2013, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from `http://www.turbomole.com`.

(97) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(98) Perdew, J.; Yue, W. *Phys. Rev. B* **1986**, *33*, 8800.

(99) Perdew, J.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(100) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(101) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(102) Becke, A. D. *J. Chem. Phys* **1997**, *107*, 8554–8560.

(103) Hamprecht, F.; Cohen, A.; Tozer, D.; Handy, N. *J. Chem. Phys.* **1998**, *109*, 6264.

(104) Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

(105) Vydrov, O. A.; Scuseria, G. E. *J. Chem. Phys.* **2006**, *125*, 234109.

(106) Otero-de-la Roza, A.; Johnson, E. R. *J. Chem. Phys.* **2013**, *138*, 204109.

(107) Otero-de-la Roza, A.; Johnson, E. R.; DiLabio, G. A. *J. Chem. Theory Comput.* **2014**, *10*, 5436–5447.

(108) Turney, J. M.; Simmonett, A. C.; Parrish, R. M.; Hohenstein, E. G.; Evangelista, F. A.; Fermann, J. T.; Mintz, B. J.; Burns, L. A.; Wilke, J. J.; Abrams, M. L.; Russ, N. J.; Leininger, M. L.; Janssen, C. L.; Seidl, E. T.; Allen, W. D.; Schaefer, H. F.; King, R. A.; Valeev, E. F.; Sherrill, C. D.; Crawford, T. D. *WIREs: Comput. Mol. Sci.* **2012**, *2*, 556–565.

(109) Hohenstein, E. G.; Sherrill, C. D. *WIREs: Comput. Mol. Sci.* **2012**, *2*, 304–326.

(110) Kozuch, S.; Martin, J. M. L. *J. Chem. Theory Comput.* **2013**, *9*, 1918–1931.

(111) Tkatchenko, A.; Von Lilienfeld, O. A. *Phys. Rev. B* **2008**, *78*, 045116.

(112) DiStasio, R. A.; von Lilienfeld, O. A.; Tkatchenko, A. *Proc. Natl. Acad. Sci.* **2012**, *109*, 14791–14795.

(113) Mó, O.; Yáñez, M.; Elguero, J. *J. Chem. Phys.* **1992**, *97*, 6628–6638.

(114) Mhin, B. J.; Kim, J.; Lee, S.; Lee, J. Y.; Kim, K. S. *J. Chem. Phys.* **1994**, *100*, 4484–4486.

(115) Santra, B.; Klimes, J.; Alfe, D.; Tkatchenko, A.; Slater, B.; Michaelides, A.; Car, R.; Scheffler, M. *Phys. Rev. Lett.* **2011**, *107*, 185701.

(116) Santra, B.; Klimeš, J.; Tkatchenko, A.; Alfè, D.; Slater, B.; Michaelides, A.; Car, R.; Scheffler, M. *J. Chem. Phys.* **2013**, *139*, 154702.

(117) Tang, K.; Toennies, J. *J. Chem. Phys.* **1984**, *80*, 3726.

(118) Lacks, D. J.; Gordon, R. G. *Phys. Rev. A* **1993**, *47*, 4681.

(119) Kannemann, F. O.; Becke, A. D. *J. Chem. Theory Comput.* **2009**, *5*, 719–727.

(120) Whittleton, S. R.; Vazquez, X. A. S.; Isborn, C. M.; Johnson, E. R. *J. Chem. Phys.* **2015**, *142*, 184106.

(121) Johnson, E. R.; Otero-de-la Roza, A.; Dale, S. G. *J. Chem. Phys.* **2013**, *139*, 184116.

(122) Bader, R.; Stephens, M. *J. Am. Chem. Soc.* **1975**, *97*, 7391–7399.

(123) Fulton, R. L. *J. Phys. Chem.* **1993**, *97*, 7516–7529.

(124) Bader, R. F.; Streitwieser, A.; Neuhaus, A.; Laidig, K. E.; Speers, P. *J. Am. Chem. Soc.* **1996**, *118*, 4959–4965.

(125) Fradera, X.; Austen, M. A.; Bader, R. F. *J. Phys. Chem. A* **1999**, *103*, 304–314.

(126) Fradera, X.; Poater, J.; Simon, S.; Duran, M.; Solà, M. *Theor. Chem. Acc.* **2002**, *108*, 214–224.

(127) Johnson, E. R.; Salamone, M.; Bietti, M.; DiLabio, G. A. *J. Phys. Chem. A* **2013**, *117*, 947–952.

(128) Pernal, K.; Podeszwa, R.; Patkowski, K.; Szalewicz, K. *Phys. Rev. Lett.* **2009**, *103*, 263201.

(129) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab initio molecular orbital theory*; John Wiley & Sons, Inc., 1986; pp 270 – 324.

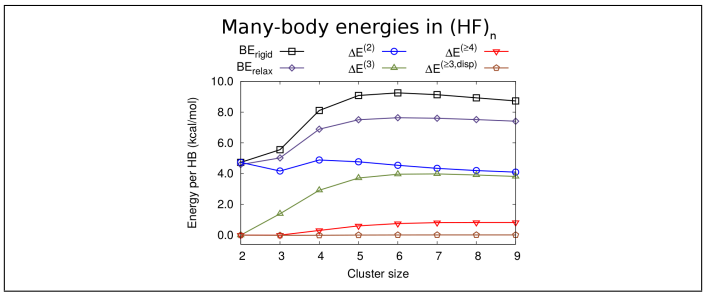
(130) Brandenburg, J. G.; Maas, T.; Grimme, S. *J. Chem. Phys.* **2015**, *142*, 124104.

(131) Mori-Sánchez, P.; Cohen, A. J.; Yang, W. *Phys. Rev. Lett.* **2008**, *100*, 146401.

(132) Koerzdoerfer, T.; Sears, J. S.; Sutton, C.; Bredas, J.-L. *J. Chem. Phys.* **2011**, *135*, 204107.

(133) Karolewski, A.; Kronik, L.; Kuemmel, S. *J. Chem. Phys.* **2013**, *138*, 204115.

Graphical TOC Entry



Density-functional approximations present errors in the treatment of the non-dispersive part of non-covalent interactions, which originate from repulsion, delocalization, and deformation errors. We analyze each source of error via a molecular many-body expansion and propose ways to improve functional performance for non-covalent interactions.