NRC Publications Archive
Archives des publications du CNRC

Predicting energetics of supramolecular systems using the XDM dispersion model
Otero-De-La-Roza, A.; Johnson, Erin R.

National Research Council Canada   Conseil national de recherches Canada

Canada

Predicting Energetics of Supramolecular Systems Using the XDM Dispersion Model

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

†National Institute for Nanotechnology, National Research Council of Canada, 11421 Saskatchewan Drive, Edmonton, Alberta, Canada T6G 2M9

‡Department of Chemistry, Dalhousie University, 6274 Coburg Road, Halifax, Nova Scotia, Canada B3H 4R2

ABSTRACT: In this article, we examine the ability of the exchange-hole dipole moment (XDM) model of dispersion to treat large supramolecular systems. We benchmark several XDM-corrected functionals on the S12L set proposed by Grimme, which comprises large dispersion-bound host−guest systems, for which back-corrected experimental and Quantum Monte Carlo (QMC) reference data are available. PBE-XDM coupled with the relatively economical and efficient pc-2-spd basis set gives excellent statistics (mean absolute error (MAE) = 1.5 kcal/mol), below the deviation between experimental and QMC data. When compared only to the (more accurate) QMC results, PBE-XDM/pc-2-spd (MAE = 1.2 kcal/mol) outperforms all other dispersion-corrected DFT results in the literature, including PBE-dDsC/QZ4P (6.2 kcal/mol), PBE-NL/def2-QZVP (4.7 kcal/mol), PBE-D2/def2-QZVP′ (3.5 kcal/mol), PBE-D3/def2-QZVP′(2.3 kcal/mol), M06-L/def2-QZVP (1.9 kcal/mol), and PBE-MBD (1.8 kcal/mol), with no significant bias (mean error (ME) = 0.04 kcal/mol). PBE-XDM/pc-2-spd gives binding energies relatively close to the complete basis-set limit and does not necessitate the use of counterpoise corrections, which facilitates its use. The dipole−quadrupole and quadrupole−quadrupole pairwise dispersion terms (C_8 and C_{10}) are critical for the correct description of the dimers. XDM-corrected functionals different from PBE that work well for small dimers do not yield good accuracy for the large supramolecular systems in the S12L, presenting errors that scale linearly with the dispersion contribution to the binding energy.
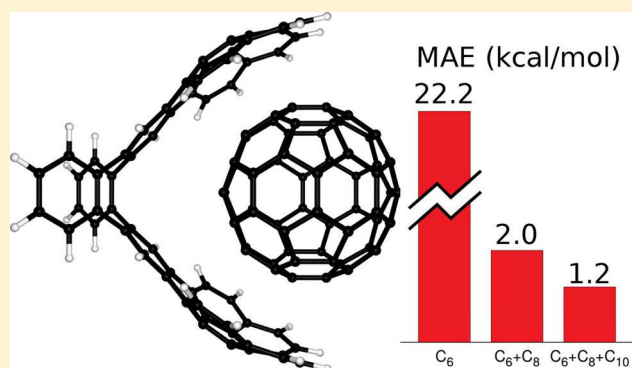
INTRODUCTION

Common exchange-correlation functionals in density-functional theory (DFT) lack the subtle long-range correlation effects necessary to model noncovalent interactions accurately. Several dispersion-energy corrections have been proposed in the past few years[1−5] that, when combined with common base functionals, permit the calculation of noncovalent binding energies with an accuracy that rivals higher-order wave function methods,[2,6−8] without negatively impacting their performance for thermochemistry.[6,8]

Dispersion-corrected density functionals have been thoroughly tested for two types of systems: (i) binding energies of small noncovalently bound dimers[2,6−13] and (ii) lattice energies and geometries of molecular crystals.[14−22] For the former, calculated binding energies are compared to high-level coupled-cluster reference data. For the molecular crystals, back-corrected experimental sublimation enthalpies are used.[15,21,22] Despite this wealth of information, little is known about how dispersion-corrected functionals behave for relatively large (but finite) supramolecular systems; this is a question that has recently come to the forefront.[19,23,24] Specifically, it is presently unknown how the error for noncovalent binding energies using dispersion-corrected density functionals scales with system size,

which is a question that is critical for applications of dispersion-corrected DFT to large complex systems,[25−28] as well as for *ab initio* molecular dynamics simulations.[29−31]

Obtaining high-quality reference data for large systems is difficult, because of the extremely poor scaling of wave function methods with system size. MP2, which is the most efficient correlated-wave function method, already has a steeper scaling than DFT while providing notoriously poor binding energies for noncovalently bound systems with low-lying excitation energies.[1,2,25,32,33] In a recent article, Risthaus and Grimme[23] screened a collection of density functionals against a new test set (S12L) composed of large supramolecular host−guest complexes.[34] The reference data for the S12L set was obtained by back-correcting experimental binding affinities for thermal and solvent effects. In a subsequent article, Ambrosetti et al.[24] performed Quantum Monte Carlo (QMC) calculations on a subset of these complexes and showed that there were significant differences with some of the experimental results, likely due to inaccuracies in the back-correction caused by the solvation model.[24,35,36]
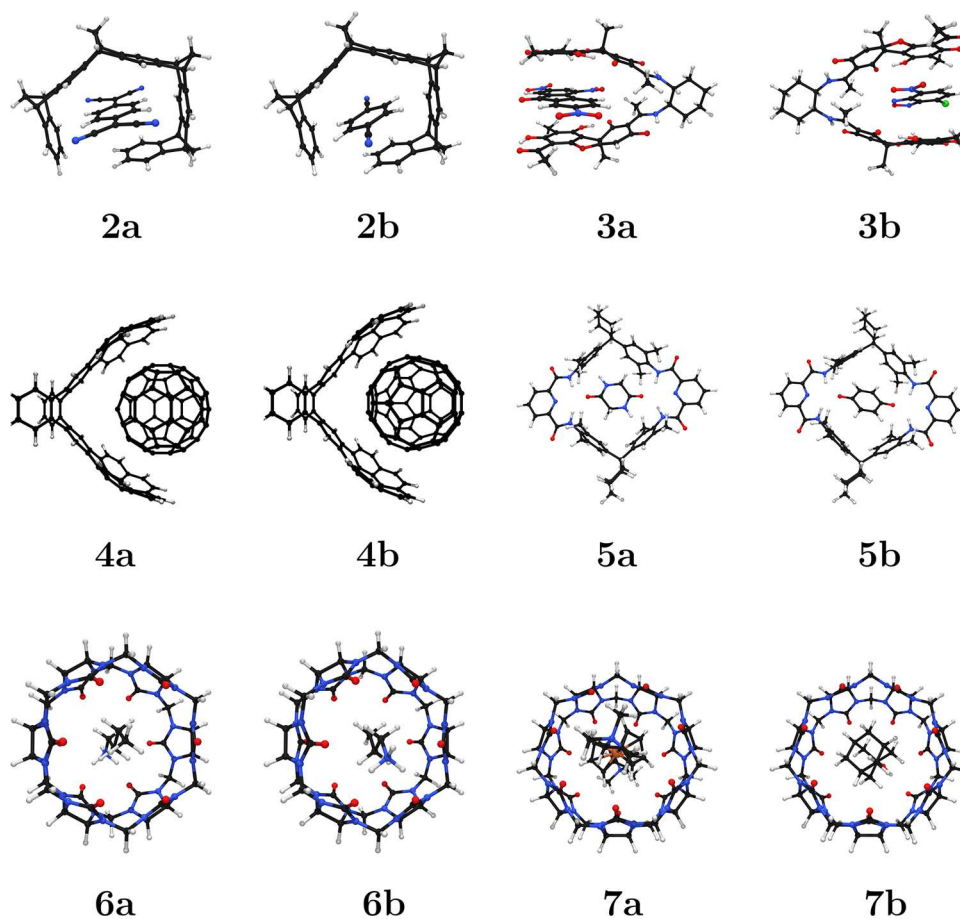
Figure 1. Structures of the dimers in the S12L benchmark set proposed by Grimme.[34] The labels are the same as in the original reference.

In this work, we explore the ability of the exchange-hole dipole moment (XDM) dispersion model[1,6,37−40] to treat large supramolecular systems by applying it to the S12L dataset. The adequacy of XDM to model these large host−guest systems has recently been put into question by Risthaus and Grimme.[23] However, in this article, we show that PBE-XDM performs excellently for the S12L benchmark when appropriate basis sets and damping-function parameters are used, with a mean absolute error (MAE) of 1.5 kcal/mol, compared to the best estimate for the reference binding energies (QMC where available, back-corrected experimental otherwise), and only 1.2 kcal/mol when compared to the subset for which more-accurate QMC data are available. This MAE is significantly smaller than that presented by Risthaus and Grimme[23] (whose XDM calculations were affected by basis-set incompleteness and a mismatch between the base functional and damping parameters). Our proposed method does not require counter-poise corrections, which reduces its complexity and facilitates its application to unexplored problems. We also demonstrate that the results are highly sensitive to the base functional and, hence, the choice of functional dramatically affects the performance of dispersion-corrected DFT for large supra-molecular systems. PBE-XDM presents a much better perform-ance than other XDM-corrected functionals, and the error shows a linear dependence on the dispersion contribution to the binding energy.

■ COMPUTATIONAL DETAILS

The S12L dimer structures are shown in Figure 1. Single-point energy calculations were performed on the complexes and separated monomers, using the literature geometries,[23,34] with the Gaussian 09 program.[41] XDM dispersion energies were obtained using the postg program[6,42] as a post-SCF correction to the base DFT energy:

$$E = E_{\text{base}} + E_{\text{disp}} \tag{1}$$

The XDM dispersion energy[6,37,38] involves a sum over atom pairs,

$$E_{\text{disp}} = - \sum_{n=6,8,10} \sum_{i<j} \frac{C_{n,ij} f_{n,ij}(R_{ij})}{R_{ij}^n} \tag{2}$$

where R_{ij} is the interatomic distance, $C_{n,ij}$ are the dispersion coefficients for each atomic pair, and the sum runs over the leading-order C_6 and higher-order C_8 and C_{10} contributions to the dispersion energy. These coefficients are obtained using a semilocal functional of the electron density and the kinetic energy density, based on the principle that dispersion arises from the interaction between the electron plus exchange-correlation hole dipole distribution on separate frag-ments.[37,38,43,44]

The parameter $f_{n,ij}(R)$ is a damping function that causes the dispersion energy to approach a (small) constant value at short interatomic distances. The damping function involves two atom-independent parameters: a_1 and a_2. The values of these parameters are fit for use with a given density functional and

basis set by minimizing the mean absolute percent error (MAPE) for the Kannemann−Becke set of 49 molecular dimers (KB49),[6] which is a subset of the original KB65 set,[40] excluding complexes involving rare-gas atoms. The remaining atomic dispersion coefficients can also be calculated using the XDM approach,[45] including the C_9 coefficient for the leading-order Axilrod−Teller−Muto (ATM) term,

$$E_{ATM} = - \sum_{i<j<k} C_{9,ijk} \left(\frac{3 \cos(\theta_i) \cos(\theta_j) \cos(\theta_k) + 1}{R_{ij}^3 R_{ik}^3 R_{jk}^3} \right) \quad (3)$$

which, in the remainder of this article, we will refer to simply as the "three-body" term.

In order to compare to the results reported by Risthaus and Grimme, calculations were performed using the PBE functional[46] and the TZVP and def2-TZVP basis sets, as in their work.[23] In a previous article,[47] we showed that an efficient basis set for intermolecular interactions necessitates at least a set of diffuse and polarization functions in order to represent the intermolecular region correctly. The pc-2-spd basis set[47−50] presents a good balance between accuracy and computational cost for intermolecular interactions in the gas phase,[47] and we use it for the majority of the calculations in this article.

Calculations were performed using the PW86PBE,[46,51] BLYP,[52,53] PBE0,[54] B3LYP,[53,55] and LC-ωPBE[56,57] functionals, all with pc-2-spd. A list of the XDM parameters used, in conjunction with each functional and basis set, is given in Table 1. In agreement with our previous work,[47] the table shows that pc-2-spd gives average errors only slightly higher than those obtained using an almost complete basis set such as aug-cc-pVTZ. Namely, with aug-cc-pVTZ[6] the mean absolute percent errors are 14.3 (PBE), 11.3 (PW86PBE), 9.8 (BLYP), 10.2 (PBE0), 6.7 (B3LYP), and 7.8 (LC-ωPBE).

Additional calculations were performed by systematically varying the fraction of Hartree−Fock exchange in PBE-based and BLYP-based hybrid functionals from 0 to 100%, in 10% increments, and the range-separation parameter in LC-ωPBE from 0 to 1, in 0.1 increments. In these calculations, the two XDM parameters were fit for each degree of exact-exchange mixing and ω, and the damping function parameters and the average errors are shown in Table 1.

RESULTS AND DISCUSSION

Role of the Basis Set and Damping Parameters. Binding energies and mean absolute errors (MAE) for the S12L set obtained with PBE-XDM and three basis sets (TZVP and def2-TZVP, previously used by Risthaus and Grimme,[23] and pc-2-spd) are given in Table 2. The complete-basis-set (CBS) estimate calculated using def2-QZVP′ values[23] is also given. PBE-XDM shows excellent performance with all three basis sets (MAE = 1.3, 1.4, and 1.5 kcal/mol with TZVP, def2-TZVP, and pc-2-spd, respectively) and an error of 2.0 kcal/mol for the CBS estimate. The average error is smaller than the deviation between QMC and back-corrected experimental results (2.4 kcal/mol). Better agreement is obtained with the QMC values than with experiment, probably because the reference data from experiment is affected by errors introduced by the back-correction procedure.[24,36]

Table 2 shows that, despite having additional polarization functions, def2-TZVP and pc-2-spd give slightly higher MAEs than TZVP with PBE-XDM, although this may be the result of errors in the extrapolated binding energies from the experimental data, as the trend is reversed when only QMC

Table 1. XDM Parameters Used with Each Base Functional[a] (Mean Absolute Errors and Mean Absolute Percent Errors for the KB49 Fit Set Are Also Given)

	a_1	a_2 (Å)	mean absolute error, MAE (kcal/mol)	mean absolute percent error, MAPE (%)
Functionals				
PBE/TZVP	0.0000	4.0222	0.70	20.9
PBE/def2-TZVP	0.0000	4.0314	0.66	19.7
PBE	0.2281	3.2443	0.57	16.7
PW86PBE	0.5709	2.0560	0.45	13.5
BLYP	0.7065	1.0274	0.35	11.6
PBE0	0.1980	3.3551	0.46	12.4
B3LYP	0.5166	1.8829	0.29	8.0
LC-ωPBE	0.5922	1.9441	0.24	6.7
PBE-Based Hybrids				
10%	0.2046	3.3232	0.51	14.4
20%	0.1862	3.3897	0.47	12.8
30%	0.1730	3.4431	0.45	12.0
40%	0.1674	3.4761	0.45	11.6
50%	0.1714	3.4833	0.46	11.6
60%	0.1892	3.4495	0.49	11.9
70%	0.2238	3.3675	0.54	12.6
80%	0.2787	3.2255	0.59	13.4
90%	0.3557	3.0186	0.64	14.4
HFPBE	0.4527	2.7539	0.72	15.8
BLYP-Based Hybrids				
10%	0.6327	1.3526	0.28	8.8
20%	0.5376	1.7486	0.25	7.2
30%	0.4207	2.2201	0.26	7.2
40%	0.2910	2.7400	0.30	8.0
50%	0.1537	3.2963	0.37	9.2
60%	0.0082	3.8919	0.45	10.9
70%	−0.1587	4.5746	0.56	13.1
80%	−0.3642	5.4025	0.68	15.6
90%	−0.6287	6.4498	0.81	18.4
HFLYP	−0.9763	7.8063	0.96	21.3
ωPBE-Based Range-Separated Hybrids				
0.0	0.2139	3.3557	0.59	16.8
0.1	0.2727	3.0786	0.51	14.3
0.2	0.3906	2.5603	0.33	9.1
0.3	0.5036	2.1792	0.25	6.7
0.5	0.6275	1.9055	0.23	7.0
0.6	0.6392	1.9268	0.27	8.3
0.7	0.6206	2.0272	0.37	10.0
0.8	0.5919	2.1458	0.45	11.4
0.9	0.5525	2.2871	0.51	12.5
1.0	0.5018	2.4580	0.56	13.3

[a]The basis set is pc-2-spd, unless noted otherwise. The value in the first column for the PBE and BLYP-based hybrids is the fraction of exact exchange. For ωPBE-based range-separated hybrids, it is the value of ω (the default for LC-ωPBE is ω = 0.4).

reference data are considered. The differences between the three basis sets are small, because of the absence of hydrogen bonds or strong donor−acceptor interactions in the S12L set, which makes it possible to obtain good results with relatively incomplete basis sets.[47] The excellent average errors achieved by PBE-XDM do not necessitate the use of counterpoise corrections, which would hinder its applicability in a practical context.

Table 2. Calculated and Reference Values for the Dimers in the S12L Test Set, Using PBE-XDM with C_6, C_8, and C_{10} Dispersion-Energy Terms[a]

name	Exp	QMC	TZVP Opt	TZVP Def	def2-TZVP	pc-2-spd	CBS
			Dispersion Energies (kcal mol⁻¹)				
2a	29.9	27.2	27.3	24.2	27.1	27.1	26.5
2b	20.5	17.2	17.6	15.4	17.6	17.5	16.8
3a	24.3		22.6	20.8	22.1	21.7	20.9
3b	20.4		19.5	18.6	18.6	18.4	17.4
4a	27.5	25.8	28.7	24.1	27.6	27.5	26.9
4b	28.7		30.5	25.3	29.3	29.4	27.7
5a	34.8	33.4	31.6	31.0	31.5	30.7	29.9
5b	21.3		21.2	20.1	20.7	20.2	19.2
6a	77.4	81.0	81.8	82.6	81.8	80.3	80.2
6b	77.0		78.4	79.2	78.1	76.8	76.0
7a	131.5		130.7	129.2	128.5	127.1	125.5
7b	22.6	24.1	26.7	25.6	26.9	25.9	24.7
			Errors				
MAE (Exp + QMC)			1.3	2.2	1.4	1.5	2.0
ME (Exp + QMC)			0.4	−1.3	−0.2	−0.8	−1.7
MAE (Exp)		2.4	2.1	3.4	2.3	2.3	2.8
ME (Exp)		−0.7	0.7	−1.6	−0.5	−1.1	−2.0

[a]The columns are the S12L back-corrected experimental reference binding energies proposed by Grimme[34] (Exp), the Quantum Monte Carlo values given by Ambrosetti et al.[24] (QMC), and the PBE-XDM binding energies calculated using TZVP basis set using Gaussian09/postg with optimized XDM parameters (Opt), TZVP using the default Q-Chem damping-function parameters (Def), def2-TZVP, pc-2-spd (both with the optimized damping parameters), and the complete-basis-set estimate using the def2-QZVPP results reported by Risthaus and Grimme,[23] and the complete-basis-set XDM correction. The mean absolute errors (MAE) and mean errors (ME) were calculated using the QMC data where available (Exp + QMC) and using only the experimental reference values (Exp).

Our results for the S12L benchmark differ from the data presented in a previous work by Risthaus and Grimme,[23] which reported a MAE for PBE-XDM of 4.5 kcal/mol, relative to the experimental data, to be compared to the values with TZVP (2.1 kcal/mol) and def2-TZVP (2.3 kcal/mol) in Table 2. This deviation can be traced to a combination of basis-set effects and an erroneous choice of damping function parameters, which contribute roughly the same to the MAE discrepancy.

In a previous paper, we showed that one set each of polarization and diffuse functions are essential in order to represent the intermolecular region properly.[47] The adequate basis-set representation of the intermolecular space has a large impact on the calculation of noncovalent interactions in dispersion-corrected DFT,[47] as well as with other methods.[35] The calculations in ref 23 used PBE-XDM with a combination of basis sets called mixedTZ:cc-pVTZ(seg-opt) for all dimers except 7a, for which TZV was used. Both of the basis sets that form the mixedTZ set lack diffuse primitives. The TZV basis used for the 7a complex lacks polarization functions as well, and, as a result, its binding energy is grossly overestimated (by 9.0 kcal/mol). The average error for the KB49 parametrization set is usually a good indicator of basis-set incompleteness,[47] and the MAPE values obtained for the fit with TZVP (20.9%) and def2-TZVP (19.7%) are clearly lower than with TZV (29.3%) and cc-pVTZ(seg-opt) (26.0%), although they are still far from the basis set limit (14.2%[6]). Risthaus and Grimme corrected for basis-set superposition error (BSSE) effects using the counterpoise method; this actually caused an increase in the MAE to 5.2 kcal/mol with PBE-XDM/mixedTZ/0.5CP* (0.5CP symbolizes half the counterpoise correction). However, the counterpoise correction (denoted by "CP*" in that article) was actually calculated for a different basis set (def2-TZVP) that contains additional polarization functions and, as seen above, has a completely different behavior.

We stress the importance of using the correct damping function parameters. The default Q-Chem parameters used in the work of Risthaus and Grimme[23] for the XDM damping function ($a_1 = 0.83$ and $a_2 = 1.55$ Å) were originally obtained for use with a functional composed of Hartree−Fock exchange plus Becke-Roussel dynamical correlation[58] and they should not be used in conjunction with PBE (or other local or hybrid functionals). XDM, similar to most dispersion corrections, employs the damping function not only to deactivate the dispersion correction at short range but also to offset errors in the treatment of the nondispersion part that come from the base functional.[2] Table 2 shows that, for TZVP, using the incorrect parameters greatly impacts the performance. Users of XDM in Q-Chem are advised to use parameters that have been optimized for use with the appropriate functional.[6,42]

Table 3. Binding Energies[a] for the S12L for Which QMC Data Are Available[24]

name	QMC	PBE-XDM pc-2-spd	PBE-XDM CBS	PBE-dDsC QZ4P	def2-QZVP PBE-NL	def2-QZVP PBE-D2	def2-QZVP PBE-D3	def2-QZVP M06-L	PBE-MBD
				Binding Energies (kcal/mol)					
2a	27.2	26.5	26.02	33.1	30.3	30.6	29.8	27.8	29.0
2b	17.2	16.8	16.79	21.8	19.6	20.3	19.4	18.1	18.8
4a	25.8	26.9	26.32	34.9	34.9	31.8	29.8	25.4	28.3
5a	33.4	29.9	29.82	38.1	35.4	35.9	33.3	31.4	33.8
6a	81.0	80.2	80.35	84.7	85.3	83.4	82.9	84.1	82.1
7b	24.1	24.7	24.73	33.1	31.5	27.8	26.8	28.4	27.4
				Errors (kcal/mol)					
MAE		1.2	1.2	6.2	4.7	3.5	2.3	1.9	1.8
ME		0.0	−0.6	6.2	4.7	3.5	2.2	1.1	1.8

[a]The columns show, in order, the reference QMC values, our calculated PBE-XDM/pc-2-spd results, the complete basis-set PBE-XDM estimate (using the previously reported def2-QZVP' results[23]), the values reported by Risthaus and Grimme[23] for other dispersion corrections with a relatively large basis set, and the PBE-MBD results given by Ambrosetti et al.[24]

Table 4. Calculated Binding Energies for the S12L Set with Selected XDM-Corrected Functionals and the pc-2-spd Basis Set[a]

	Expt	QMC	PBE	PW86PBE	BLYP	PBE0	B3LYP	LC-ωPBE
				Binding Energies (kcal/mol)				
2a	29.9	27.2	27.1	29.2	32.6	30.4	31.8	35.8
2b	20.5	17.2	17.5	19.0	21.2	19.7	20.6	23.5
3a	24.3		21.7	23.7	26.2	23.4	25.6	27.0
3b	20.4		18.4	19.2	20.4	19.5	20.8	21.8
4a	27.5	25.8	27.5	30.7	34.6	31.7	32.5	40.0
4b	28.7		29.4	32.6	37.1	33.6	34.7	41.5
5a	34.8	33.4	30.7	32.4	34.8	33.8	34.9	36.5
5b	21.3		20.2	22.1	24.1	22.1	23.8	23.6
6a	77.4	81.0	80.3	83.0	84.9	83.5	85.9	87.1
6b	77.0		76.8	79.0	80.6	79.8	81.6	82.9
7a	131.5		127.1	131.4	136.0	133.6	137.7	142.5
7b	22.6	24.1	25.9	28.6	30.3	28.0	30.4	31.7
				Errors (kcal/mol)				
MAE (Exp + QMC)			1.5	2.1	4.2	2.6	4.0	6.8
ME (Exp + QMC)			−0.8	1.6	4.2	2.6	4.0	6.8
MAE (Exp)		2.4	2.3	2.3	3.9	2.5	3.7	6.5
ME (Exp)		−0.7	−1.1	1.2	3.9	1.9	3.7	6.5

[a]The mean absolute errors (MAE) and mean errors (ME) were calculated using the QMC data where available (Exp + QMC) and using only the experimental reference values (Exp).

For the S12L, the mean error (ME) obtained using PBE-XDM with TZVP is practically zero (no bias) and the MAE is 2.1 kcal/mol. This outperforms almost all methods reported by Risthaus and Grimme for which TZVP was used: 1.6 kcal/mol (PBE-D2), 2.3 kcal/mol (PBE-NL), 2.3 kcal/mol (PBE-D3), and 2.5 kcal/mol (PBE-dDsC). In particular, PBE-XDM outperforms PBE-dDsC, contradicting previous claims to the contrary in the literature.[59] Table 3 compares our PBE-XDM/pc-2-spd results (and the corresponding PBE-XDM complete-basis-set estimate) to those obtained with different dispersion-corrected functionals by Risthaus and Grimme[23] and Ambrosetti et al.[24] PBE-XDM/pc-2-spd clearly outperforms all other dispersion-corrected PBE methods, with a MAE (1.2 kcal/mol) that is significantly smaller than PBE-MBD (1.8 kcal/mol), M06-L (1.9 kcal/mol), and the other alternatives. In addition, there is no significant bias in our XDM results (ME = 0.04 kcal/mol), in contrast with the overbinding behavior of all the other examined methods. The complete-basis-set PBE-XDM results confirm the efficient nature of the pc-2-spd basis set, showing the same MAE but a slight underbinding behavior (ME = −0.6 kcal/mol).

Role of the Dipole–Quadrupole, and Quadrupole–Quadrupole Two-Body Dispersion Terms. An ongoing debate in the field of dispersion-corrected DFT is the relative importance of including three-body terms (C_9, eq 3) and higher-order pairwise terms (C_8 and C_{10} in eq 2) in the asymptotic dispersion energy expression, as well as other effects missing from the usual asymptotic pairwise description of dispersion.[23,24,45,60,61] For most systems, the ATM term results in a repulsive contribution to the binding energy; this is also the case for the dimers in the S12L.[24] However, in a previous article,[45] we showed how the addition of three-body dispersion corrections is generally deleterious to the application of XDM and the choice of a suitable damping function is highly ambiguous.

Using the S12L benchmark results obtained for PBE-XDM/pc-2-spd (Table 2), one can advocate that the C_8 and C_{10} pairwise terms are essential to obtain a good description of the S12L binding energies. The C_8 and C_{10} terms generally account for roughly 40% of the dispersion energy.[58] If PBE-XDM/pc-2-spd is reparametrized for use with just the C_6 term, then the MAPE for the KB49 parametrization set is 20.6% (compared to 16.7% in Table 1). If PBE-XDM/pc-2-spd with just the C_6 term is then applied to the S12L set, the binding energies of all dimers are strongly overestimated, with a MAE of 23.4 kcal/mol (compared to 1.3 kcal/mol in Table 2). As discussed previously,[45] omission of the higher-order pairwise contributions leads to the C_6 term representing the dispersion energy corresponding to the C_8 and C_{10} terms, which should decay faster with distance. This results in a damping function that generally gives systematic overbinding for dimers in which the dispersion contribution to binding is dominant.

Role of the Base Functional. In previous dispersion-corrected DFT benchmark studies on the S12L, only PBE was used as the base functional.[23,24] This is an odd choice because dispersion-corrected PBE, generally, and PBE-XDM, in particular, are inferior to dispersion-corrected hybrids (e.g., B3LYP or PBE0) and range-separated hybrids (e.g., LC-ωPBE) for small dimers,[6] as evidenced by the results in Table 1. This is particularly true for hydrogen-bonded systems for which, in many cases, uncorrected PBE already overestimates the binding energy (e.g., PBE gives a binding energy of 4.7 kcal/mol for the HF dimer, compared to the reference value of 4.6 kcal/mol). This is caused by the additive combination of two effects: the underestimation of intermolecular Pauli repulsion[6,14] and an overestimation of the electrostatic contribution to binding caused by delocalization error.[10] The systematic overestimation of hydrogen-bonded binding energies during the parametrization causes PBE-XDM to dampen the dispersion contribution more strongly, typically underestimating the binding energies for the dispersion-bound dimers in the KB49 set. For example, PBE-XDM predicts binding energies of 1.4 and 2.5 kcal/mol for the stacked and T-shaped benzene dimers, compared to the reference values of 1.7 and 2.7 kcal/mol, respectively. This behavior causes PBE-XDM to be a poor choice for across-the-board noncovalent interactions in small dimers, as indicated by the relatively high values of the average percent error in Table 1.

In this section, we consider the performance of XDM-corrected base functionals other than PBE for the supra-molecular dimers in the S12L set. Binding energies and statistics with selected functionals are given in Table 4. From the results in the table, we note that there is a limited number of functionals that can be used to predict the S12L binding energies reliably: PBE-XDM (MAE = 1.5 kcal/mol) and, perhaps, PW86PBE-XDM (MAE = 2.1 kcal/mol) and PBE0-XDM (MAE = 2.6 kcal/mol). All functionals except PBE systematically overbind, on average. LC-ωPBE performs particularly poorly with a MAE of 6.8 kcal/mol. This observation is in stark contrast with the behavior of the same functionals for smaller dimers, for which the performance of hybrid functionals (with ~25% exact exchange) and range-separated functionals is systematically better[6] (see Table 1). In particular cases where the charge-transfer contribution to binding is dominant, even amounts of exact exchange close to 50%, which minimize delocalization error, are needed.[10]

Similar results as those in Table 4 have been reported by Tkatchenko et al.[62] for the buckyball-catcher complex, which is very similar to the host−guest complexes 4a and 4b in the S12L (Figure 1). The reference binding energy for the catcher complex is 26.0 kcal/mol (±2 kcal/mol, calculated with QMC), compared with the reported values of 43.1 kcal/mol (B97-D), 41.2 kcal/mol (B97-D3), and 37.5 kcal/mol (TPSS-D3). Several Minnesota functionals give binding energies much closer to the correct result, although this may be caused by error cancellation from the missing long-range dispersion effects.[62] In sharp contrast, Hesselmann has shown in a recent study that the Symmetry-Adapted Perturbation Theory (SAPT) breakdown of the total energy for the S12L complexes does not have a signature that is particularly different from those of smaller dimers.[35] Then, what is the reason for the results shown in Table 4?

Let us examine another benchmark also composed of dispersion-bound dimers: the L7 set proposed by Sedlak et al.,[63] with reference values revised by Grimme.[19,64] The L7 set comprises stacks of aromatic molecules. Its dimers are smaller and less strongly bound than those in the S12L set. The benchmark results with a subset of the XDM-corrected functionals used for the S12L are shown in Table 5. Even though the nature of the dimers is apparently similar to the

S12L, the statistics resemble those for the KB49 fit set and it is LC-ωPBE-XDM that achieves the best performance, with a MAE of 1.1 kcal/mol. For the L7, all functionals are underbinding, but the ordering of functionals for the mean error mirrors that observed for the S12L: PBE0 and particularly PBE are much less binding than BLYP, B3LYP, or LC-ωPBE. However, the binding energies and the dimers in the L7 set are smaller and more similar to those in the parametrization set than to the S12L, so LC-ωPBE-XDM, instead of PBE-XDM, minimizes the MAE.
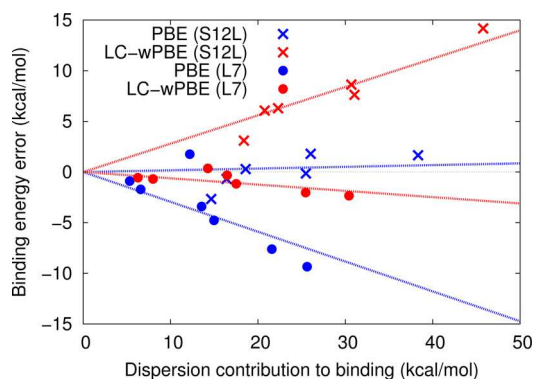
A second item of information is shown in Figure 2, where the MAE for the S12L set is plotted as a function of either the
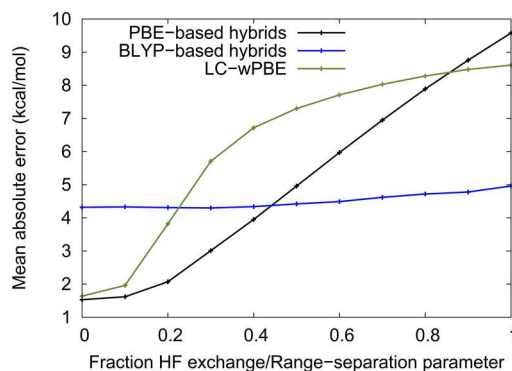
Figure 2. Evolution of the mean absolute error (MAE) on the S12L set with fraction of exact exchange for PBE- and BLYP-based functionals, and with the range-separation parameter for LC-ωPBE. The reference data corresponds to Quantum Monte Carlo, where available, and back-corrected experimental otherwise.

percentage of exact exchange in PBE-based and BLYP-based hybrid functionals or the range-separation parameter (ω) for the LC-ωPBE functional. For the PBE hybrids and LC-ωPBE, adding more exact exchange causes the MAE to increase and the complexes to be overbound. In BLYP-based hybrids, the average error remains constant and all BLYP-based hybrids are strongly overbinding. This behavior is unusual in that it is the opposite of that observed for the parametrization set (Table 1). Because PBE underestimates Pauli repulsion (and over-estimates the electrostatic stabilization of hydrogen bonds), the addition of a moderate fraction of exact exchange normally results in an improvement of the average error. The same effect is observed for charge-transfer dimers.[10] Figure 2 can be explained by relating the observed average error to the amount of Pauli repulsion introduced by the functional or, in other words, to the total dispersion contribution to the binding energies. PBE underestimates Pauli repulsion, so PBE-based hybrids with more exact exchange, and higher ω values in LC-ωPBE (whose base functional is similar to PBE regarding Pauli repulsion), require a larger dispersion contribution to the binding energy; this results in an overbinding behavior. BLYP, which gives repulsion values closer to the correct Hartree−Fock results,[2] yields hybrids with roughly the same overbinding error, regardless of the amount of exact exchange.

Therefore, the average errors in S12L and L7 seem to be directly related to the dispersion energy contribution to the dimer binding energies. Figure 3 confirms this relation, which does not exist if one uses the total binding energy instead. For instance, the value on the far right of the plot in Figure 3 (PBE dispersion energy = 38.3 kcal/mol) corresponds to dimer 4a, with a binding energy of only only 25.8 kcal/mol (cf, the PBE

Table 5. Calculated Binding Energies[a] for the L7 Set by Sedlak et al.[63]

	ref	PBE	BLYP	PBE0	B3LYP	LC-ωPBE
Binding Energies (kcal/mol)						
C2C2PD	24.36	16.75	21.49	18.32	20.57	22.33
C3A	18.19	13.42	16.62	14.56	16.22	17.03
C3GC	31.25	21.91	28.57	23.92	27.47	28.91
CBH	11.06	12.82	13.21	12.17	13.18	11.42
GCGC	14.37	10.97	14.41	11.50	13.83	14.05
GGG	2.40	1.51	2.02	1.35	1.86	1.84
PHE	25.76	24.05	24.92	25.24	25.55	25.07
Errors (kcal/mol)						
MAE		4.21	1.51	3.22	1.85	1.07
ME		−3.71	−0.88	−2.90	−1.24	−0.96

[a]The reference values are DLPNO−CCSD(T) (with counterpoise correction and basis-set extrapolation), calculated by Grimme.[19,64] All functionals are XDM-corrected. We used the pc-2-spd basis set for all calculations. The naming convention is the same as that used in Sedlak et al.[63].

Figure 3. Error in the binding energies of the S12L sets (cross symbols, ✕) and L7 sets (circle symbols, ●), using the PBE (blue) and LC-ωPBE (red) functionals against the dispersion contribution to those binding energies, calculated using the respective functional. Only the S12L dimers for which there is a QMC reference energy available are shown.

dispersion contribution to the binding in 6a is 16.4 kcal/mol, out of a total binding energy of 80.3 kcal/mol). Hence, dimers that are greatly stabilized by dispersion interactions (such as 4a and 4b in the S12L), even if they are not strongly bound, are harder to model accurately. The error is linearly dependent on the dispersion energy, and the L7 and S12L sets give different slopes, which is reasonable, in view of the very different geometries of their dimers.

In conclusion, our results show that, for purely dispersion-bound complexes, the error in the total binding energy is proportional to its dispersion contribution. The slope of this relation is dependent on the particular geometry of the noncovalent contacts, and it is difficult to predict which functional must be used for arbitrary supramolecular systems. LC-ωPBE-XDM seems to outperform PBE-XDM for stacks (L7 set), but PBE-XDM offers excellent results for the S12L set, and LC-ωPBE-XDM is strongly overbinding. In consequence, regarding the choice of base functional for supramolecular chemistry studies,[65] the following considerations apply:

(i) Hybrid functionals with ~25% exact exchange generally work well for hydrogen-bonded systems and small dimers;[6,9]

(ii) For charge-transfer dimers, half-and-half hybrids or range-separated functionals should be used;[10]

(iii) For small stacks, LC-ωPBE-XDM gives accurate binding energies (see Table 5); and

(iv) For larger stacks, PBE-XDM should be used (see Table 4).

The systems in the S12L are "easy" for GGA functionals such as PBE-XDM, in the sense that they do not involve charge transfer or hydrogen-bonded interactions. However, in many applications (e.g., biological supermolecules), this will likely not be the case, and PBE-XDM will probably misrepresent donor–acceptor interactions. Thus, a base functional that gives uniformly high accuracy for all types of noncovalent interactions is still missing.

■ CONCLUSIONS

In this article, we have shown that PBE-XDM is an excellent choice for studying large dispersion-bound supramolecular complexes. When applied to the S12L set proposed by

Grimme, PBE-XDM coupled with the pc-2-spd basis set gives a mean absolute error (MAE) of 1.5 kcal/mol, which is lower than the deviation between Quantum Monte Carlo (QMC) and back-corrected experimental data (2.4 kcal/mol). When compared to other dispersion-corrected functionals at the same basis-set level (TZVP), PBE-XDM obtains a MAE against the back-corrected experimental data of 2.1 kcal/mol, which is higher than that for PBE-D2 (MAE = 1.6 kcal/mol), but outperforms more-sophisticated functionals such as PBE-NL (MAE = 2.3 kcal/mol), PBE-D3 (MAE = 2.3 kcal/mol), and PBE-dDsC (MAE = 2.5 kcal/mol).[23] Previous MAE statistics for PBE-XDM reported by Risthaus and Grimme were misleading, because of a combination of basis-set incompleteness and the use of unoptimized damping parameters.

When compared to the subset for which Ambrosetti et al.[24] reported QMC data, which are probably more accurate than the back-corrected experimental values, PBE-XDM/pc-2-spd also yields more-accurate results (MAE = 1.2 kcal/mol) than PBE-MBD (MAE = 1.8 kcal/mol[24]) and any of the other dispersion-corrected functionals reported by Risthaus and Grimme,[23] using the largest basis set available: PBE-dDsC/QZ4P (6.2 kcal/mol), PBE-NL/def2-QZVP (4.7 kcal/mol), PBE-D2/def2-QZVP′ (3.5 kcal/mol), PBE-D3/def2-QZVP′ (2.3 kcal/mol), and M06-L/def2-QZVP (1.9 kcal/mol). PBE-XDM/pc-2-spd shows negligible bias (mean error (ME) = 0.04 kcal/mol), whereas the rest of the methods systematically overbind, on average.

In addition, PBE-XDM/pc-2-spd does not require the use of counterpoise corrections, which simplifies its application to large systems. The higher-order dipole−quadrupole (C_8) and quadrupole−quadrupole (C_{10}) pairwise terms are essential, and their absence leads to a gross overestimation of all binding energies, with a MAE of 23.4 kcal/mol after reparametrization. We note, in passing, that the proper application of XDM necessitates that the damping function be parametrized to match the dispersion energy expression being used.

The choice of the base functional affects the performance of the dispersion-corrected functional dramatically for the dimers in the S12L set. Only PBE-XDM gives a reasonable average error, whereas LC-ωPBE and BLYP-based hybrids strongly overbind. This is the opposite of the behavior for smaller dispersion-bound dimers (such as the stacks in the L7 set), as well as for the dimers in the parametrization set. The error in the binding energy prediction depends linearly on the dispersion energy contribution, with different slopes for the S12L and the L7 sets. A functional that provides uniformly good across-the-board performance for all types of noncovalent interactions presently is not available, which hinders the applicability of dispersion-corrected DFT to complex systems.

■ AUTHOR INFORMATION

Corresponding Authors
*E-mail: alberto.oterodelaroza@nrc-cnrc.gc.ca (A. Otero-de-la-Roza).
*E-mail: erin.johnson@dal.ca (E. R. Johnson).

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Grimme for kindly providing us with revised reference values for the L7 set and for detailed technical information about the calculations in their article.

■ REFERENCES

(1) Johnson, E. R.; Mackie, I. D.; DiLabio, G. A. *J. Phys. Org. Chem.* **2009**, *22*, 1127.

(2) DiLabio, G. A.; Otero-de-la-Roza, A. *Rev. Comput. Chem.* **2015**, in press.

(3) Grimme, S. *WIREs: Comput. Mol. Sci.* **2011**, *1*, 211−228.

(4) Klimeš, J.; Michaelides, A. *J. Chem. Phys.* **2012**, *137*, 120901.

(5) DiStasio, R. A., Jr; Gobre, V. V.; Tkatchenko, A. *J. Phys.: Condens. Matter* **2014**, *26*, 213202.

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

(7) Goerigk, L.; Grimme, S. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670−6688.

(8) Hujo, W.; Grimme, S. *J. Chem. Theory Comput.* **2011**, *7*, 3866−3871.

(9) DiLabio, G. A.; Johnson, E. R.; Otero-de-la Roza, A. *Phys. Chem. Chem. Phys.* **2013**, *15*, 12821−12828.

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

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

(12) Marom, N.; Tkatchenko, A.; Rossi, M.; Gobre, V. V.; Hod, O.; Scheffler, M.; Kronik, L. *J. Chem. Theory Comput.* **2011**, *7*, 3944−3951.

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

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

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

(16) Carter, D. J.; Rohl, A. L. *J. Chem. Theory Comput.* **2014**, *10*, 3423−3437.

(17) Moellmann, J.; Grimme, S. *J. Phys. Chem. C* **2014**, *118*, 7615−7621.

(18) Turner, M. J.; Grabowsky, S.; Jayatilaka, D.; Spackman, M. A. *J. Phys. Chem. Lett.* **2014**, *5*, 4249−4255.

(19) Brandenburg, J. G.; Hochheim, M.; Bredow, T.; Grimme, S. *J. Phys. Chem. Lett.* **2014**, *5*, 4275−4284.

(20) Kronik, L.; Tkatchenko, A. *Acc. Chem. Res.* **2014**, *47*, 3208−3216.

(21) Reilly, A. M.; Tkatchenko, A. *J. Phys. Chem. Lett.* **2013**, *4*, 1028−1033.

(22) Reilly, A. M.; Tkatchenko, A. *J. Chem. Phys.* **2013**, *139*, 024705.

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

(24) Ambrosetti, A.; Alfè, D.; DiStasio, R. A., Jr.; Tkatchenko, A. *J. Phys. Chem. Lett.* **2014**, *5*, 849−855.

(25) Ehrlich, S.; Moellmann, J.; Grimme, S. *Acc. Chem. Res.* **2013**, *46*, 916−926.

(26) Ehrlich, S.; Bettinger, H. F.; Grimme, S. *Angew. Chem., Int. Ed.* **2013**, *52*, 10892−10895.

(27) Isborn, C. M.; Tang, C.; Martini, A.; Johnson, E. R.; Otero-de-la Roza, A.; Tung, V. C. *J. Phys. Chem. Lett.* **2013**, *4*, 2914−2918.

(28) Otero-de-la Roza, A.; Luaña, V.; Tiekink, E. R. T.; Zukerman-Schpector, J. *J. Chem. Theory Comput.* **2014**, *10*, 5010−5019.

(29) Tkatchenko, A.; Rossi, M.; Blum, V.; Ireta, J.; Scheffler, M. *Phys. Rev. Lett.* **2011**, *106*, 118102.

(30) DiStasio, R. A.; Santra, B.; Li, Z.; Wu, X.; Car, R. *J. Chem. Phys.* **2014**, *141*, 084502.

(31) Guevara-García, A.; Ireta, J.; Galván, M. *J. Chem. Phys.* **2015**, *142*, 014504.

(32) Janowski, T.; Ford, A. R.; Pulay, P. *Mol. Phys.* **2010**, *108*, 249−257.

(33) Tkatchenko, A.; DiStasio, R. A., Jr.; Head-Gordon, M.; Scheffler, M. *J. Chem. Phys.* **2009**, *131*, 094106.

(34) Grimme, S. *Chem.—Eur. J.* **2012**, *18*, 9955−9964.

(35) Heßelmann, A.; Korona, T. *J. Chem. Phys.* **2014**, *141*, 094107.

(36) Le, V. H.; Yanney, M.; McGuire, M.; Sygula, A.; Lewis, E. A. *J. Phys. Chem. B* **2014**, *118*, 11956−11964.

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

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

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

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

(41) Frisch, M. J. et al. *Gaussian 09, Revision A.1*; Gaussian, Inc.: Wallingford, CT, 2009.

(42) See http://gatsby.ucmerced.edu/wiki/.

(43) Ángyán, J. G. *J. Chem. Phys.* **2007**, *127*, 024108.

(44) Ayers, P. W. *J. Math. Chem.* **2009**, *46*, 86−96.

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

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

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

(48) Jensen, F. *J. Chem. Phys.* **2001**, *115*, 9113−9125.

(49) Jensen, F. *J. Chem. Phys.* **2002**, *116*, 7372−7379.

(50) Jensen, F.; Helgaker, T. *J. Chem. Phys.* **2004**, *121*, 3463−3470.

(51) Perdew, J. P.; Wang, Y. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 8800.

(52) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098−3100.

(53) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.

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

(55) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

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

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

(58) Johnson, E. R.; Becke, A. D. *J. Chem. Phys.* **2006**, *124*, 174104.

(59) Corminboeuf, C. *Acc. Chem. Res.* **2014**, *47*, 3217−3224.

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

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

(62) Tkatchenko, A.; Alfè, D.; Kim, K. S. *J. Chem. Theory Comput.* **2012**, *8*, 4317−4322.

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

(64) Grimme, S. Personal communication, 2014.

(65) Antony, J.; Sure, R.; Grimme, S. *Chem. Commun.* **2015**, *51*, 1764.