An improved treatment of empirical dispersion and a many-body energy decomposition scheme for the explicit polarization plus symmetry-adapted perturbation theory (XSAPT) method

Ka Un Lao and John M. Herbert[a)]
Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, USA

(Received 3 April 2013; accepted 26 June 2013; published online 18 July 2013)

We recently introduced a low-cost quantum chemistry method for computing intermolecular interactions, combining a monomer-based self-consistent field calculation (the "explicit polarization" method, XPol) with pairwise-additive symmetry adapted perturbation theory (SAPT). The method uses Kohn-Sham (KS) orbitals in the SAPT formalism but replaces the SAPT dispersion and exchange-dispersion terms with empirical potentials ("+D"), and we called this method XPol+SAPT(KS)+D. Here, we report a second-generation version of this approach, XPol+SAPT(KS)+D2 or XSAPT(KS)+D2 for short, in which we have modified the form of the empirical atom–atom dispersion potentials. Accurate binding energies are obtained for benchmark databases of dimer binding energies, and potential energy curves are captured accurately for a variety of challenging systems. We suggest that using different asymptotic corrections for different monomers is necessary to get good binding energies in general, especially for hydrogen-bonded complexes. As compared to our original "+D" formulation, the second-generation "+D2" method accurately reproduces not only total binding energies but also the various components of the interaction energy, and on this basis we introduce an energy decomposition scheme that extends traditional SAPT energy decomposition to systems containing more than two monomers. For $(H_2O)_6$, the many-body contribution to the interaction energy agrees well with that obtained from traditional Kitaura-Morokuma energy decomposition analysis in a large basis set. © *2013 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4813523]

I. INTRODUCTION

Despite enormous progress in electronic structure theory over the past several decades, the "gold standard" of chemical accuracy, coupled cluster singles and doubles with perturbative triples, CCSD(T), remains out of reach for most systems. For calculation of non-covalent interactions, which is the topic of the present work, CCSD(T)/aug-cc-pVTZ calculations on $(H_2O)_{20}$ represent the present state-of-the-art,[1] and are only feasible on massively parallel architectures. Second-order Møller-Plesset perturbation theory (MP2) is tractable in larger systems but overestimates binding energies in cases where the binding is dominated by dispersion interactions.[2] Various strategies have been used to correct this deficiency, including spin-component scaled (SCS) MP2 methods,[3,4] sometimes with parameters fit specifically for non-covalent interaction energies,[5] and also MP2.X methods that combine MP2 and MP3 results in an empirical way.[6] These methods achieve a mean accuracy of $\lesssim 0.4$ kcal/mol with respect to complete-basis CCSD(T) binding energies,[7] but scale no better than $\mathcal{O}(N^5)$ with respect to total system size, N.

Density functional theory (DFT) is more affordable but many of the most popular functionals afford a poor description of dispersion interactions.[8,9] One strategy to circumvent this problem is to introduce a rather large number of empirical parameters into the functional,[10,11] then optimize these parameters using data sets that include weak interactions. The M06-2X functional[10] is a popular example of this approach. Alternatively, one might attempt to capture the dispersion interactions directly by means of classical atom–atom potentials, typically with r^{-6} dependence.[8,12,13] Such approaches fall under the moniker "DFT-D," and are often able to reproduce higher-level calculations rather well.[13–16] The ωB97X-D functional[11] is a popular example. Grimme *et al.*[14] have recently introduced a "third generation" (DFT-D3) correction that further improves the description of non-bonded interactions, in functionals such as ωB97X-D3.[16] Finally, dispersion can be introduced into DFT by means of non-local correlation functionals.[17–19] The LC-VV10 functional[19] is an example of such an approach, which has recently been shown to exhibit outstanding performance for many types of intermolecular interactions.[20] The description of weak interactions in DFT has thus come a long way in the past few years, yet none of these methods scales better than $\mathcal{O}(N^3)$. As such, these DFT methods are still not feasible for applications to molecular liquids or biomolecules.

Fragment-based approaches[21–26] do provide a relatively affordable route to computing binding energies in large systems, by partitioning the supersystem into subsystems (fragments). In particular, our group has developed a low-cost, monomer-based approach that we call explicit polarization (XPol) plus symmetry-adapted perturbation theory (SAPT), or XSAPT,[26–29] which we have previously abbreviated XPS.[30]

a)herbert@chemistry.ohio-state.edu

This approach starts from the variational XPol method,[31] which is used to capture many-body (MB) polarization effects by means of a charge-embedded, monomer-based self-consistent field (SCF) calculation. In a subsequent step, we apply a pairwise-additive form of SAPT.[32–34] The resulting XSAPT method generalizes traditional SAPT from two-body to many-body systems. Theoretical details can be found in our previous work,[27,28] including a recent review.[26] The wall time for an XSAPT calculation scales as $\mathcal{O}(n)$ with respect to the number of monomers, n, assuming that one has ${}_nC_2$ processors to run in "embarrassingly parallel" mode for the second-order SAPT calculations, and is $\mathcal{O}(n^3)$ even in serial mode.[29] When a Hartree-Fock (HF) description of the monomers is used in the XPol calculation, in conjunction with a suitable basis set, errors in dimer binding energies computed by XSAPT lie within 1 kcal/mol of high-level benchmarks.[27,28]

One might try to improve on this accuracy by using instead a Kohn-Sham (KS) description of intramolecular electron correlation, in what we have termed XSAPT(KS).[26] However, this approach suffers from the same problem as SAPT(KS),[35,36] namely, significant overestimation of dispersion energies.[28] Ironically, the dispersion and exchange-dispersion terms in SAPT(KS) are not only the least accurate ones, but also the most expensive to compute, scaling as $\mathcal{O}(N^4)$ and $\mathcal{O}(N^5)$, respectively, with respect to monomer size, N. For this reason, we recently introduced XSAPT(KS)+D,[29] in which the sum-over-states (uncoupled Hartree-Fock[37]) dispersion formula in second-order SAPT(KS) is replaced by empirical atom–atom potentials developed for this purpose by Hesselmann.[38] In conjunction with a double-ζ basis set, XSAPT(KS)+D exhibits mean errors in binding energies of <0.5 kcal/mol for the S22A data set[39] and the larger (and more balanced) S66 data set.[40] In this respect, XSAPT(KS)+D is superior to various MP2-type methods extrapolated to the complete basis-set (CBS) limit, and methods containing triple excitations are required in order to do better.[26]

In traditional SAPT, the interaction energy decomposes naturally according to

$$E_{\rm int}^{\rm SAPT} = E_{\rm elst}^{(1)} + E_{\rm exch}^{(1)} + E_{\rm ind}^{(2)} + E_{\rm exch\text{-}ind}^{(2)} + E_{\rm disp}^{(2)} + E_{\rm exch\text{-}disp}^{(2)} + \cdots . \tag{1}$$

All terms up to second order in the intermolecular interaction are listed explicitly, with subscripts that denote electrostatic (elst), exchange (exch), induction (ind), and dispersion (disp) contributions. This work is mainly focused on developing a similar interaction-energy decomposition scheme for XSAPT, and comparing the various energy components to dimer SAPT results for the S22A database.[41] In the course of this analysis, we discovered that the original XSAPT(KS)+D method does not do a good job of reproducing individual energy components, and thus its favorable performance for binding energies is due in no small part to error cancellation. This led us to pursue a "second generation" dispersion correction ("+D2"), using alternative dispersion potentials developed by Podeszwa *et al.*[42]

II. THEORY

A. Summary of XSAPT

The success of XSAPT is based on the fact that the many-body (non-pairwise-additive) contribution to cluster binding energies is dominated by induction interactions, whereas electrostatic, exchange-repulsion, and dispersion interactions are nearly pairwise additive.[43–48] An overview of the formal theory is provided in Ref. 26 and a thorough exposition is provided in Ref. 28; here, we provide only a brief summary.

Assuming closed-shell monomers for simplicity, the XPol energy is given by[26,27]

$$E_{\rm XPol} = \sum_{A=1}^{n} \left[2 \sum_a \mathbf{c}_a^\dagger \left(\mathbf{h}^A + \mathbf{J}^A - \tfrac{1}{2}\mathbf{K}^A \right) \mathbf{c}_a + E_{\rm nuc}^A \right] + E_{\rm embed}. \tag{2}$$

The term in square brackets is the ordinary HF energy for monomer A, but we assume that the MOs $\mathbf{c}_a$ for monomer A are expanded using atom-centered Gaussian basis functions located on monomer A only. This affords a method whose cost grows only linearly with respect to the number of monomers, n, and furthermore excludes basis set superposition error, by construction. The embedding potential $E_{\rm embed}$ is electrostatic only, and in applications to date (including the present work), this potential arises from charge–density interactions between the density of the monomer undergoing SCF iterations and atom-centered ChElPG[28,49] charges derived from the SCF electrostatic potentials of the other monomers.

Subsequent to the XPol step, we use direct products of XPol monomer wave functions as zeroth-order states for SAPT, with an appropriately modified perturbation that avoids double-counting any interactions that are already included in Eq. (2).[27] These SAPT calculations are performed in a pseudocanonicalized dimer basis set, which captures some intermolecular charge-transfer interactions.[27] The final XSAPT energy, including all terms through second order in the intermolecular interactions, can be expressed as

$$\begin{aligned} E_{\rm XSAPT} = {} & \sum_A \left(\sum_a \left[2\varepsilon_a^A - \mathbf{c}_a^\dagger \left(\mathbf{J}^A - \tfrac{1}{2}\mathbf{K}^A \right) \mathbf{c}_a \right] + E_{\rm nuc}^A \right) \\ & + \sum_A \sum_{B>A} \left[E_{\rm RSPT}^{[0;1_{AB}]} + E_{\rm exch}^{[0;1_{AB}]} + E_{\rm RSPT}^{[0;2_{AB}]} + E_{\rm exch}^{[0;2_{AB}]} \right. \\ & \left. + \sum_C{}' \sum_{D>C}{}' \left(E_{\rm RPST}^{[0;1_{AB},1_{CD}]} + E_{\rm exch}^{[0;1_{AB},1_{CD}]} \right) \right]. \end{aligned} \tag{3}$$

The superscripts in square brackets denote various orders in perturbation theory,[28] e.g., $E_{\rm RSPT}^{[0;1_{AB}]}$ is zeroth-order in the monomer fluctuation potentials (intramolecular electron correlation) and first-order in the intermolecular potential between monomers A and B. In terms of the more traditional SAPT notation of Eq. (1),

$$E_{\rm RSPT}^{[0;1_{AB}]} = E_{{\rm elst},A}^{(1)} + E_{{\rm elst},B}^{(1)}, \tag{4a}$$

$$E_{\rm exch}^{[0;1_{AB}]} = E_{{\rm exch},A}^{(1)} + E_{{\rm exch},B}^{(1)}, \tag{4b}$$

and

$$E_{\rm RSPT}^{[0;2_{AB}]} = E_{{\rm ind},A}^{(2)} + E_{{\rm ind},B}^{(2)} + E_{{\rm disp},AB}^{(2)}, \tag{5a}$$

$$E_{\rm exch}^{[0;2_{AB}]} = E_{\text{exch-ind},A}^{(2)} + E_{\text{exch-ind},B}^{(2)} + E_{\text{exch-disp},AB}^{(2)}. \tag{5b}$$

See Ref. 28 for details and an explanation of the notation, but note that there is a typographical error in Eq. (51) of the latter paper; Eq. (3) above is the corrected version.

In XSAPT(KS)+D,[29] we replace the terms $E_{{\rm disp},AB}^{(2)} + E_{\text{exch-disp},AB}^{(2)}$ [cf. Eq. (1)] with empirical atom–atom potentials. Furthermore, the monomers are described using KS-DFT, hence the various SAPT terms in Eqs. (4) and (5) do include some intramolecular electron correlation. To date, however, we have omitted the three-body induction couplings,[28] $E_{\rm RPST}^{[0;1_{AB},1_{CD}]} + E_{\rm exch}^{[0;1_{AB},1_{CD}]}$ in Eq. (3), from the XSAPT(KS)+D method, and we also neglect these terms in the present work. This reduces the cost of the method from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2)$ with respect to the number of monomers, yet accurate results can still be obtained for many-body systems, as demonstrated herein.

B. XSAPT energy decomposition

In the context of XPol-SAPT, it makes sense to define the MB contribution to the interaction energy, $E_{\rm int}^{\rm MB}$, according to

$$E_{\rm int}^{\rm MB} = E_{\rm int}^{\rm XSAPT} - \sum_A \sum_{B<A} E_{AB}^{\rm XSAPT}, \tag{6}$$

where $E_{\rm int}^{\rm XSAPT}$ is the overall XSAPT interaction energy and $E_{AB}^{\rm XSAPT}$ is the interaction energy for dimer $A\cdots B$. The SAPT part is pairwise additive by construction (we do not consider the three-body SAPT terms derived by Lotrich and Szalewicz[50,51]). Thus, the total SAPT interaction energy for a collection of monomers is

$$E_{\rm int}^{\rm SAPT} = \sum_A \sum_{B<A} E_{AB}^{\rm SAPT}. \tag{7}$$

Addition of Eqs. (6) and (7) affords

$$E_{\rm int}^{\rm MB} = E_{\rm int}^{\rm XSAPT} - E_{\rm int}^{\rm SAPT} - \sum_A \sum_{B<A} \left(E_{AB}^{\rm XSAPT} - E_{AB}^{\rm SAPT}\right). \tag{8}$$

The SAPT interaction energy can be decomposed as in Eq. (1) and substituted into Eq. (8), with the result rewritten as

$$\begin{aligned} E_{\rm int}^{\rm XSAPT} = {} & E_{\rm elst}^{(1)} + E_{\rm exch}^{(1)} + E_{\rm ind}^{(2)} + E_{\text{exch-ind}}^{(2)} \\ & + E_{\rm disp}^{(2)} + E_{\text{exch-disp}}^{(2)} + \cdots \\ & + \sum_A \sum_{B<A} \left(E_{AB}^{\rm XSAPT} - E_{AB}^{\rm SAPT}\right) + E_{\rm int}^{\rm MB}. \end{aligned} \tag{9}$$

Equation (9) is the interaction-energy decomposition scheme for XSAPT. This energy decomposition analysis (EDA) requires three sets of calculations. First, traditional SAPT calculations and also XSAPT calculations must be performed on all pairs of monomers, to obtain pairwise electrostatic, exchange, dispersion, and induction terms. Second, an XSAPT calculation must be performed on the entire system, which provides not only $E_{\rm int}^{\rm MB}$ but also the rest of the induction terms that are missing in the first set of calculations. The latter are defined by the double sum in Eq. (9).

It is common in SAPT calculations to solve coupled-perturbed Hartree-Fock (CPHF) equations and thereby include an infinite-order response correction for polarization in the presence of a frozen partner density.[52] In contrast, the XSAPT method treats polarization self-consistently in the XPol step. As such, the infinite-order response correction for induction should be included exactly by the XPol part of the calculation (with further induction corrections vanishing), if the XPol calculation is performed using density embedding,[23] that is, if monomer SCF densities are used to compute the electrostatic interactions between the monomers in the self-consistent XPol calculation. We do not pursue this possibility here, but instead use atom-centered point charges to do the embedding, as in previous work.[27–29] This procedure is significantly less expensive (especially if the monomers are large), and implicitly includes some higher-order induction effects into the zeroth-order XPol monomer energies.[27] Therefore, the energy difference between XSAPT and SAPT for all pairs of dimers [the double sum in Eq. (9)] partly includes the infinite-order response correction for induction. In addition, some higher-order induction effects are captured by the $\delta E_{\rm int}^{\rm HF}$ correction[53,54] that is discussed below.

It is common to truncate the SAPT interaction energy at second order in Eq. (9) and incorporate higher-order polarization effects by adding a correction

$$\delta E_{\rm int}^{\rm HF} = E_{\rm int}^{\rm HF} - \left(E_{\rm elst}^{(1)} + E_{\rm exch}^{(1)} + E_{\text{ind,resp}}^{(2)} + E_{\text{exch-ind,resp}}^{(2)}\right) \tag{10}$$

to the interaction energy. The "response" (resp) subscripts indicate that the infinite-order response correction for induction is incorporated by solving CPHF equations, and $E_{\rm int}^{\rm HF}$ is the counterpoise-corrected HF binding energy for the dimer. It is recommended to include the $\delta E_{\rm int}^{\rm HF}$ term in SAPT calculations involving polar monomers, because induction corrections converge slowly for polar molecules.[53,54] Furthermore, we find that $\delta E_{\rm int}^{\rm HF}$ is necessary in traditional SAPT calculations, not only to obtain quantitative binding energies but also to obtain qualitatively correct potential energy surfaces for induction-dominated systems.[55] In the present work, we assume that this correction term is pairwise additive for many-body XSAPT calculations,

$$\delta E_{\rm int}^{\rm HF} = \sum_A \sum_{B<A} \delta E_{AB}^{\rm HF}. \tag{11}$$

The quality of the results presented herein indicates that this assumption is quite robust.

Adding the $\delta E_{\rm int}^{\rm HF}$ correction term to Eq. (9), the XSAPT interaction energy becomes

$$\begin{aligned} E_{\rm int}^{\rm XSAPT} = {} & E_{\rm elst}^{(1)} + E_{\rm exch}^{(1)} + E_{\rm disp}^{(2)} + E_{\text{exch-disp}}^{(2)} \\ & + \Bigg[E_{\rm ind}^{(2)} + E_{\text{exch-ind}}^{(2)} + \sum_A \sum_{B<A} \delta E_{AB}^{\rm HF} \\ & + \sum_A \sum_{B<A} \left(E_{AB}^{\rm XSAPT} - E_{AB}^{\rm SAPT}\right) + E_{\rm int}^{\rm MB} \Bigg]. \end{aligned} \tag{12}$$

Those terms in square brackets are regarded as the total induction energy, and the total dispersion energy is $E_{\rm disp}^{(2)} + E_{\rm exch\text{-}disp}^{(2)}$. One may argue that this procedure double-counts the higher-order induction terms since both the XPol SCF procedure and $\delta E_{\rm int}^{\rm HF}$ are included in those terms. In the charge-embedded, monomer-based SCF calculation for XPol, both infinite-order response and higher-order corrections for induction are only partly included. Furthermore, we assume that the $\delta E_{\rm int}^{\rm HF}$ correction term is pairwise for the many-body system. Thus, this approach relies partly on cancellation of errors to provide meaningful results. However, thorough comparisons to benchmark calculations will show that this interaction-energy decomposition scheme is accurate and robust.

C. Second-generation dispersion potential

Our original XSAPT(KS)+D method[29] was based on empirical atom–atom dispersion potentials developed for SAPT(KS) by Hesselmann.[38] Here, we test a second-generation ("+D2") version of XSAPT that employs an alternative dispersion potential developed by Podeszwa *et al.*:[42]

$$E_{\rm disp}^{\rm dlDF} = -\sum_{i\in A}\sum_{\substack{j\in B\\(B\neq A)}}\left[\frac{C_{ij,6}}{r_{ij}^6}\,f_6(\beta_{ij}r_{ij}) + \frac{C_{ij,8}}{r_{ij}^8}\,f_8(\beta_{ij}r_{ij})\right]. \tag{13}$$

In this equation, $C_{ij,\,6} = (C_{i,\,6}C_{j,\,6})^{1/2}$ (similar for $C_{ij,\,8}$) and $\beta_{ij} = (\beta_i\beta_j)^{1/2}$. This empirical potential has previously been used to correct the results of a "dispersionless" density functional (dlDF),[42,56] hence the notation $E_{\rm disp}^{\rm dlDF}$. The indices i and j in Eq. (13) represent nuclei located on different monomers and

$$f_n(r_{ij}) = 1 - \exp(-r_{ij})\sum_{m=0}^{n}\frac{r_{ij}^m}{m!} \tag{14}$$

is the Tang-Toennies damping function.[57] The quantities $C_{i,6}$, $C_{i,8}$, and β_i are parameters that are fit to reproduce SAPT(DFT) dispersion energies $(E_{\rm disp}^{(2)} + E_{\rm exch\text{-}disp}^{(2)})$ for a training set of dimers.[42] For hydrogen, the values of these parameters depend upon the identity of the nearest-neighbor atom.

The Hesselmann[38] ("+D") and Podeszwa *et al.*[42] ("+D2") dispersion potentials were parameterized in completely different ways. The latter was fit directly to SAPT dispersion potentials,

$$E_{\rm disp}(R) = E_{\rm disp}^{(2)}(R) + E_{\rm exch\text{-}disp}^{(2)}(R) \tag{15}$$

for a large training set of dimers, where the dispersion energies $E_{\rm disp}(R)$ were computed using SAPT(DFT).[58] Hesselmann's[38] dispersion potential, on the other hand, was parameterized to reproduce benchmark intermolecular interaction energies for the S22 data set.[38] In view of this, the dispersion potential of Podeszwa *et al.*[42] seems better suited for use in our energy decomposition scheme, since it constitutes a well-defined dispersion component. As in our previous work,[29] we refer to results using Hesselmann's[38] dispersion potential as XSAPT(KS)+D, whereas results using Eq. (13), which are presented here for the first time, will be called XSAPT(KS)+D2.

III. COMPUTATIONAL ASPECTS

A. Basis sets and functionals

Our tests indicate that triple-ζ basis sets, augmented with diffuse functions, are required in order to obtain accurate results for individual energy components. Our basis set of choice is the second-generation Ahlrichs triple-ζ one, def2-TZVPP,[59] as XSAPT calculations with this basis exhibit the best accuracy for S22A binding energies, amongst many basis sets that we have tested.[27,28] We augment def2-TZVPP with diffuse functions taken from Dunning's aug-cc-pVTZ (aTZ) basis set, and refer to the augmented basis set as aug-def2-TZVPP (aTZVPP). In addition to the computational cost of augmenting all atoms, however, the use of a large number of diffuse functions can sometimes lead to overpolarization in the context of charge-embedding schemes.[60] Thus, most of the XSAPT calculations reported here use "heavy augmented" (ha) basis sets that have diffuse functions only on the non-hydrogen atoms, and in particular we make extensive use of ha-def2-TZVPP (haTZVPP).

We use long-range-corrected (LRC) density functionals[61–64] to obtain correct asymptotic behavior, as in previous work.[29] Specifically, we employ two LRC functionals based on the short-range ωPBE exchange functional.[65] One of these (LRC-ωPBEh[64]) also contains 20% short-range Hartree-Fock exchange, whereas the other (LRC-ωPBE[61]) does not. The range separation parameter (ω) is determined by system-specific tuning[66] to satisfy the condition

$$\varepsilon_{\rm HOMO} = -{\rm IP}, \tag{16}$$

where "IP" denotes the lowest ionization potential. For clusters of monomers, we previously took ω for the supersystem to be the one corresponding to the lowest monomer IP,[29] on the assumption that the non-covalent interactions would not significantly affect the IPs (hence the lowest monomer IP equals the cluster IP). As such, the same value of ω was used for all monomers in the cluster calculations reported in Ref. 29. Subsequently, we discovered that the results could be significantly improved in certain cases by using different ω values for different monomers, in order to obtain an exact asymptotic correction (AC) for each monomer unit, as is done in the SAPT(DFT) method[58] and the DFT-SAPT method.[67] In the present work, we compare these two approaches, using the designation "(AC)" whenever different ω values are used for different monomers. Tuned ω values for the monomers considered here are provided in Table S1 of the supplementary material.[68]

Our tests indicate that errors for the exchange energy components increase with increasing fraction of short-range HF exchange in the LRC functional. Furthermore, we find that Podeszwa's[42] dispersion potential [Eq. (13)] gives lower errors for dispersion components as compared to Hesselmann's[38] potential. We expect that Podeszwa's[42] potential works well with pure functionals in the short range

since both exchange and dispersion components give lower errors. Actually, the LRC-ωPBE functional gives better results when used with Podeszwa's[42] dispersion potential and the LRC-ωPBEh functional is more suitable to use with Hesselmann's[38] dispersion force field, and for this reason we use LRC-ωPBEh for XSAPT(KS)+D and LRC-ωPBE for XSAPT(KS)+D2, with the haTZVPP basis set in either case. As in our previous work,[29] the parameter s_β in Hesselmann's[38] dispersion potential was optimized in a basis-set-specific way, using the S22A dimer binding energies[39] as benchmarks. The fitted values of s_β for several different basis sets are listed in Table S2 of the supplementary material.[68]

In consideration of computational cost, it is not feasible to use a triple-ζ basis set to do a $\delta E_{\rm int}^{\rm HF}$ calculation for large monomers, even with the pairwise-additive approximation used here. We find that the 6-31+G(3d,3pd) basis set affords $\delta E_{\rm int}^{\rm HF}$ estimates comparable to those obtained using triple-ζ basis sets. Comparing this basis set to def2-TZVPP augmented with Pople diffuse sp functions on non-hydrogen atoms, the unsigned difference in the $\delta E_{\rm int}^{\rm HF}$ corrections for S22A data set is only 0.02 kcal/mol. Thus, we use the smaller 6-31+G(3d,3pd) basis set to compute the $\delta E_{\rm int}^{\rm HF}$ corrections throughout this work.

Previously, we showed[29] that XSAPT(KS)+D combined with the partially augmented aug-cc-pVDZ′ (aDZ′) basis set[69,70] outperforms MP2/CBS and related SCS-MP2/CBS methods, so XSAPT(KS)+D/aDZ′ results are presented here as well, for comparison. The aDZ′ and TZVPP basis sets gave the best results for the XSAPT(HF) method,[27,28] so we also include them in this work.

All XSAPT calculations reported here use smooth ChElPG embedding charges[28] for the XPol calculations and "projected" (pseudocanonical dimer) basis sets[27] for the SAPT corrections. For supersystem DFT calculations, the def2-QZVP basis set is used. The Euler-Maclaurin-Lebedev (99,590) quadrature grid was used for all semi-local functionals and the (75,302) grid was used for LC-VV10 calculations, except in the case of the $F^-(H_2O)_{10}$ clusters where considerations of cost led us to use a (75,302) grid for the semi-local functionals and a (50,194) grid for LC-VV10.

Finally, for comparative purposes we will present some results computed using a recently proposed Hartree-Fock plus dispersion (HFD) method augmented with first-order correlation effects.[42] This method, which was termed "HFD$_{\rm as}$c$^{(1)}$" in Ref. 42, computes the interaction energy according to the formula

$$E_{\mathrm{HFD_{as}c^{(1)}}} = E_{\rm int}^{\rm HF} + E_{\rm disp}^{\rm dlDF} + \varepsilon^{(1)}. \tag{17}$$

Here, $E_{\rm int}^{\rm HF}$ is the Hartree-Fock interaction energy, $E_{\rm disp}^{\rm dlDF}$ is the empirical dispersion potential in Eq. (13), and

$$\varepsilon^{(1)} = \varepsilon_{\rm elst}^{(1)} + \varepsilon_{\rm exch}^{(1)} \tag{18}$$

is the first-order SAPT interaction energy. We compute $\varepsilon^{(1)}$ as the difference between the first-order SAPT(KS) energy (using the LRC-ωPBE functional[61]) and the first-order SAPT0 energy,

$$\varepsilon^{(1)} = E_{\rm SAPT(KS)} - E_{\rm SAPT0}. \tag{19}$$

The haTZVPP basis set was used for supersystem Hartree-Fock calculations and $\varepsilon^{(1)}$ was calculated using the projected (pseudocanonicalized) version of haTZVPP.

B. Benchmarks

For most of the systems considered in this work, we benchmark against CCSD(T)/CBS results, some of which are taken from the literature and some of which are computed here for the first time. In a few cases, alternative levels of theory are used as benchmarks due to the availability of EDA results. This section described the benchmarks that we use and how they are obtained.

For dimers, we use the CCSD(T)/CBS results in the S22A and S66 data sets.[39,40] The S22A database revises the energetics of the original S22 database,[71] and although a further revision (S22B) has been published,[72] the mean unsigned error (MUE) between the S22A and the S22B binding energies is only 0.035 kcal/mol. As such, we use the S22A binding energies here, to facilitate comparison to some other *ab initio* results in the literature. For benchmarks of individual energy *components* for the S22 data set, we use the SAPT2+(3)/aTZ results from Ref. 41. For S22A binding energies, this level of theory is found to be the most accurate SAPT method.[34]

We will also benchmark against CCSD(T)/CBS potential energy curves (PECs) for several systems. For Ar$\cdots$Ne, the CCSD(T)/CBS result was obtained here by combining the HF/aug-cc-pV6Z energy with a two-point extrapolation[73] (aug-cc-pV5Z and aug-cc-pV6Z) of the CCSD(T) correlation energy. Potential energy curves for $(C_6H_6)_2$, $F^-(H_2O)$, and $Cl^-(H_2O)$ are taken from the literature,[29,74] where they are computed using standard extrapolation techniques to obtain the MP2/CBS result combined with a triples correction

$$\delta_{\rm CCSD(T)} = E_{\rm CCSD(T)} - E_{\rm MP2}, \tag{20}$$

computed in a basis set where the CCSD(T) calculation is feasible.

For the $F^-(H_2O)_n$ and $Cl^-(H_2O)_n$ clusters (where $n \leq 6$), structures were optimized at the resolution-of-identity (RI) MP2/aTZ level. For the ten isomers of $F^-(H_2O)_{10}$, structures were taken from Ref. 25 and then re-optimized at the B3LYP/6-31G* level. We report unrelaxed binding energies for these halide–water clusters, i.e., the H_2O monomer energies are computed using their cluster geometries.

RI-MP2 correlation energies in the CBS limit were estimated using a two-point (aug-cc-pVTZ and aug-cc-pVQZ) extrapolation.[73] This extrapolated correlation energy was added to the HF/aug-cc-pVQZ energy to estimate the RI-MP2/CBS energy. To this RI-MP2/CBS energy, we then add a triples correction [see Eq. (20)] in which the MP2 and CCSD(T) energies are each computed within the RI approximation using the "haTZ" basis set, which was suggested for this purpose in Ref. 75.

The expensive $E_{\text{RI-CCSD(T)}}$ term needed in Eq. (20) is approximated efficiently by means of an electrostatically embedded many-body expansion (EE-MBE) of the correlation energy.[76] Specifically, we use a three-body truncation of the MBE in conjunction with gas-phase Mulliken embedding charges computed at the B3LYP/6-31G* level. (EE-MBE

results for water clusters and ion–water clusters are remarkably insensitive to the details of the embedding charges in many cases.[77])

These halide–water benchmarks are new, and both coordinates and benchmark energetics for them are available in the supplementary material.[68] The accuracy of the three-body approximation to $\delta_{\text{RI-CCSD(T)}}$ can be gauged by comparing it to a two-body approximation, and data for both are provided in the supplementary material.[68] To summarize: for $X^-(H_2O)_2$ (X = F, Cl), the difference between a two-body approximation to $\delta_{\text{RI-CCSD(T)}}$ and the exact triples correction defined by Eq. (20) is <0.1 kcal/mol. Furthermore, for the ten isomers of $F^-(H_2O)_{10}$ considered in this work, values of $\delta_{\text{RI-CCSD(T)}}$ computed by means of a two-body truncation range from −2.1 to −2.8 kcal/mol while values obtained using a three-body approximation range from −1.6 to −2.1 kcal/mol, for an average difference of 0.6 kcal/mol between the two- and three-body approximations to $\delta_{\text{RI-CCSD(T)}}$. It is therefore anticipated that a four-body approximation would modify the total binding energies by significantly less than 0.6 kcal/mol.

SAPT2+(3) calculations were performed using the SAPT 2008.2 program[78] with integrals generated by the ATMOL program.[79] The RI-CCSD(T) and RI-MP2 calculations necessary for the triples correction in Eq. (20) were computed using the ORCA program,[80] v. 2.9.1. Large-basis CCSD(T) calculations for Ar···Ne were computed using CFOUR.[81,82] All other calculations were performed using a locally modified version of Q-CHEM, v. 4.0.[83,84] All supersystem calculations are counterpoise corrected according to the Boys-Bernardi scheme,[85] with the exception of the EE-MBE triples correction in Eq. (20) and also the Ar···Ne calculations.

IV. RESULTS AND DISCUSSION

A. S22 and S66 data sets

Interaction energies calculated using XSAPT, traditional SAPT, the effective fragment potential (EFP) method,[21] the $\text{HFD}_{\text{as}}\text{c}^{(1)}$ approach,[42] and several post-HF methods are compared to benchmark CCSD(T)/CBS results for the S22A data set in Table I. The MUE for high-level SAPT2+(3)/aTZ calculations is 0.32 kcal/mol,[41] and is just a little bit worse than the SCS(MI)-MP2/CBS and SCS-CCSD/CBS results.[39] The SAPT2+(3) and XSAPT-based results are all better than the EFP and MP2/CBS results. As noted previously,[41] EFP results are quite poor (MUE = 1.79 kcal/mol) when S22 geometries are used, but the MUE is reduced to 0.91 kcal/mol at EFP geometries. This suggests that the XSAPT results might also improve if XSAPT-optimized geometries were used (an assertion that is supported by some limited finite-difference optimizations[27,28]), but we have not explored this possibility, as analytic gradients for XSAPT are not available.

The XPol procedure was originally developed as a "next generation" force field,[86,87] where it was combined with pairwise Lennard-Jones (LJ) potentials to account for short-range exchange repulsion and long-range dispersion interactions. LJ parameters were optimized in Ref. 87 at the B3LYP/6-31G(d) level using 105 hydrogen-bonded dimers. However, we find that this XPol-LJ method affords very poor binding energies for the S22A data set, with a MUE of 3.29 kcal/mol. (The method furthermore predicts that three of the dispersion-dominated dimers are not bound at all.) It was argued in Ref. 87 that the LJ parameters are optimized especially for the hydrogen-bonded systems and that XPol-LJ should provide a good description of hydrogen-bonded interactions. However, for the H-bonded subset of S22, the MUE remains large, 3.16 kcal/mol. This is the reason why we need to use SAPT to capture the rest of interaction following an XPol calculation.

The results for XSAPT(KS)+D are better than those for XSAPT(KS)+D2, but one should recall that in developing the original "+D" dispersion potential, we specifically optimized one parameter (s_β) to minimize errors for S22A binding energies. The main source of error in XSAPT(KS)+D2 comes from dispersion-dominated complexes, especially the π-stacked uracil dimer and adenine–thymine complexes, where the errors are 3 and 4 kcal/mol, respectively. The primary source of these errors is the dispersion component of the potential. For uracil dimer, the D2 dispersion potential overestimates the SAPT2+(3)/aTZ dispersion energy by 1.8 kcal/mol, and for adenine–thymine by 2.0 kcal/mol. These two systems also represent the outliers for the $\text{HFD}_{\text{as}}\text{c}^{(1)}$ method [Eq. (17)], which uses the same form of the dispersion potential as in XSAPT(KS)+D2, although the errors are somewhat larger for $\text{HFD}_{\text{as}}\text{c}^{(1)}$: 5 kcal/mol for uracil dimer and 3 kcal/mol for adenine–thymine. In fact, there is much formal similarity between the XSAPT(KS)+D2 and $\text{HFD}_{\text{as}}\text{c}^{(1)}$ methods, as will be discussed below.

We do not understand the origin of the large errors for the two π-stacked systems in any further detail, although it is noteworthy that of the 79 dimers used as a training set for the D2 dispersion potential [Eq. (13)], only two π-stacked complexes are included: the sandwich isomer of benzene dimer (for which our method performs very well, as discussed below), and the pyrazine dimer.[42] Thus, some refinement for larger π-stacked systems may be in order. If we eliminate the two problematic π-stacked systems from the S22 data set, then the MUE for XSAPT(KS)+D2/haTZVPP (AC) is reduced to 0.34 kcal/mol for the dispersion-dominated subset and 0.54 kcal/mol for the entire set of 20 complexes.

One curious feature of the data in Table I is that the "first generation" XSAPT(KS)+D exhibits a smaller MUE than does second-generation XSAPT(KS)+D2 for the dispersion-dominated subset of S22. This is puzzling insofar as the "+D2" dispersion potential is fit directly to dispersion energies ($E^{(2)}_{\text{disp}} + E^{(2)}_{\text{exch-disp}}$) computed using SAPT(DFT). To investigate this further, we have benchmarked the individual energy components (electrostatics, exchange, induction, and dispersion) against those obtained at the SAPT2+(3)/aTZ level,[41] which includes monomer electron correlation effects. These comparisons are listed in Table II, and provide a clue as to the main sources of error. [At first glance, one might wonder why the energy components other than dispersion should be different for XSAPT(KS)+D and XSAPT(KS)+D2, but one must recall that the two approaches employ different KS functionals, as discussed in Sec. III A.]

The MUE for the dispersion energy obtained using XSAPT(KS)+D2 is actually better than the corresponding MUE for XSAPT(KS)+D. As such, the larger errors in

TABLE I. Mean unsigned errors (MUEs), in kcal/mol, and percent errors (in parentheses), with respect to CCSD(T)/CBS benchmarks for the S22A data set along with subsets consisting of the hydrogen-bonded dimers, dispersion-dominated dimers, and dimers of mixed influence. All calculations were performed at S22 geometries except for one set of EFP results, as indicated.

Method	H-bonded	Disp.-bound	Mixed	All S22
XSAPT(HF)/aDZ′	0.60 (6.55)	0.85 (32.06)	0.30 (9.32)	0.60 (16.70)
XSAPT(HF)/TZVPP	0.40 (3.63)	0.78 (21.99)	0.37 (10.09)	0.53 (12.36)
XSAPT(KS)+D/aDZ′	0.73 (5.93)	0.38 (9.87)	0.52 (12.31)	0.53 (9.39)
XSAPT(KS)+D[a]	0.76 (6.54)	0.30 (7.34)	0.32 (7.26)	0.45 (7.06)
XSAPT(KS)+D (AC)[a]	0.62 (5.91)	0.41 (7.95)	0.40 (9.19)	0.47 (7.70)
XSAPT(KS)+D2[a]	0.88 (5.54)	0.99 (15.17)	0.35 (7.24)	0.75 (9.58)
XSAPT(KS)+D2 (AC)[a]	0.72 (4.57)	1.18 (15.88)	0.52 (11.82)	0.82 (10.99)
$HFD_{as}c^{(1)}$	0.59 (3.98)	1.44 (21.45)	0.59 (13.69)	0.90 (13.42)
SAPT2+(3)[b]	0.42 (4.09)	0.41 (8.79)	0.13 (4.64)	0.32 (5.98)
EFP (S22 geoms.)[c]	2.93 (19.53)	1.70 (42.62)	0.76 (17.43)	1.79 (27.26)
EFP (EFP geoms.)[d]	1.97 (14.51)	0.48 (13.10)	0.34 (7.39)	0.91 (11.73)
MP2[e]	0.24	1.69	0.61	0.88
SCS-MP2[e]	1.54	0.55	0.37	0.80
SCS(MI)-MP2[e]	0.30	0.37	0.17	0.28
SCS-CCSD[e]	0.40	0.23	0.08	0.24

[a]Using the haTZVPP basis set and the $\delta E_{\rm int}^{\rm HF}$ correction.
[b]Using the aug-cc-pVTZ basis set, from Ref. 41.
[c]From Ref. 41.
[d]Using EFP-optimized geometries, from Ref. 41.
[e]CBS limit, from Ref. 39.

XSAPT(KS)+D2 binding energies for dispersion-dominated dimers must arise from less satisfactory error cancellation between the different energy components, as compared to the original +D method. As such, XSAPT(KS)+D2 error statistics for S22A (Table I) are slightly worse than for XSAPT(HF). For XSAPT(KS)/haTZVPP with empirical dispersion, using the exact AC for different monomers is a little bit worse than using the same ω value for the whole system as compared to their errors with respect to CCSD(T)/CBS benchmark for the S22A data set. However, we will see that for other systems, the use of exact AC is essential, and in any case for S22A it does not significantly degrade the results.

For the XSAPT(HF)/aDZ′, XSAPT(HF)/TZVPP, and XSAPT(KS)+D/aDZ′ methods, the MUEs for the various energy components with respect to the SAPT2+(3) benchmarks are quite large as shown in Table II. Insofar as these methods provide good results for binding energies, it is clear that cancellation of errors must play a significant role. XSAPT(KS)/haTZVPP methods with empirical dispersion potentials and $\delta E_{\rm int}^{\rm HF}$ corrections give very good results for individual energy components, especially XSAPT(KS)+D2. The electrostatic and exchange energies predicted by XSAPT(KS)+D and XSAPT(KS)+D2 with exact AC are a bit worse than the corresponding results without using exact AC, or in other words, a bit farther from the SAPT2+(3) results.

SAPT2+(3)/aTZ was selected as a benchmark for the individual energy components because it is the most accurate version of SAPT for the S22A binding energies.[34] It is possible, however, that SAPT2+(3)/aTZ energy *components* do not accurately represent the exact values, owing to truncation of either the basis set or the perturbation series. For example, intramolecular electron correlation contributions to the electrostatic and exchange energies in SAPT2+(3) are truncated

TABLE II. Mean unsigned errors (MUEs), in kcal/mol, and percent errors (in parentheses), for individual energy components of the S22 data set, with respect to benchmarks computed at the SAPT2+(3)/aTZ level.[41] All calculations were performed at S22 geometries.

	Energy components			
Method	Electrostatic	Exchange	Induction	Dispersion
XSAPT(HF)/aDZ′	0.69 (10.74)	2.84 (18.87)	1.92 (61.80)	1.25 (21.95)
XSAPT(HF)/TZVPP	0.41 (7.35)	2.16 (13.43)	1.70 (58.56)	0.60 (9.04)
XSAPT(KS)+D/aDZ′	0.64 (13.37)	3.34 (25.10)	1.86 (60.20)	1.37 (18.08)
XSAPT(KS)+D[a]	0.35 (7.07)	0.81 (6.16)	0.25 (14.22)	0.55 (7.96)
XSAPT(KS)+D (AC)[a]	0.37 (7.54)	0.98 (7.77)	0.20 (11.30)	0.62 (8.52)
XSAPT(KS)+D2[a]	0.32 (6.31)	0.57 (4.81)	0.23 (12.01)	0.38 (6.68)
XSAPT(KS)+D2 (AC)[a]	0.36 (7.16)	0.72 (6.46)	0.23 (11.11)	0.38 (6.68)
EFP[b]	2.03 (34.26)	2.29 (16.53)	1.71 (50.73)	0.80 (12.61)

[a]Using the haTZVPP basis and the $\delta E_{\rm int}^{\rm HF}$ correction.
[b]From Ref. 41.

TABLE III. Mean unsigned errors (MUEs), in kcal/mol, and percent errors (in parentheses), with respect to CCSD(T)/CBS benchmarks for the S66 data set along with subsets consisting of the hydrogen-bonded dimers, dispersion-dominated dimers, and dimers of mixed influence. All calculations except the EFP ones were performed at S66 geometries.

Method	H-bonded	Disp.-bound	Mixed	All S66
XSAPT(HF)/aDZ′	0.26 (3.46)	0.83 (27.66)	0.45 (11.32)	0.54 (15.27)
XSAPT(HF)/TZVPP	0.36 (6.07)	0.44 (14.51)	0.36 (7.48)	0.39 (9.86)
XSAPT(KS)+D/aDZ′	0.28 (4.00)	0.22 (9.06)	0.34 (8.01)	0.27 (7.04)
XSAPT(KS)+D[a]	0.46 (5.13)	0.38 (13.46)	0.35 (7.50)	0.40 (9.11)
XSAPT(KS)+D (AC)[a]	0.37 (3.98)	0.34 (12.25)	0.34 (7.75)	0.35 (8.28)
XSAPT(KS)+D2[a]	0.49 (5.55)	0.34 (11.13)	0.47 (8.33)	0.42 (8.51)
XSAPT(KS)+D2 (AC)[a]	0.34 (3.15)	0.31 (10.06)	0.53 (9.69)	0.38 (7.56)
$HFD_{as}c^{(1)}$	0.39 (4.50)	0.41 (11.76)	0.65 (12.10)	0.46 (9.31)
EFP[b]	0.79	0.65	0.35	0.61
B2PLYP-D3[c]	0.50	0.13	0.15	0.26
M06-2X[c]	0.24	0.35	0.25	0.28
ωB97X-D[d]	0.16	0.58	0.24	0.33
MP2[e]	...	...	...	0.45
SCS-MP2[e]	...	...	...	0.74
SCS(MI)-MP2[e]	...	...	...	0.28
CCSD[e]	...	...	...	0.62
SCS-CCSD[e]	...	...	...	0.15

[a]Using haTZVPP and the $\delta E_{\rm int}^{\rm HF}$ correction.
[b]Using EFP-optimized geometries, from Ref. 41.
[c]Using def2-QZVP, from Ref. 13.
[d]Using 6-311++G(3df,3pd), from Ref. 16.
[e]CBS limit, from Ref. 40.

at

$$\epsilon^{(1)}_{\rm elst,resp}(3) = E^{(12)}_{\rm elst,resp} + E^{(13)}_{\rm elst,resp} \tag{21}$$

and

$$\epsilon^{(1)}_{\rm exch}(2) = E^{(11)}_{\rm exch} + E^{(12)}_{\rm exch}, \tag{22}$$

respectively. However, the convergence behavior is improved if these are replaced by $\epsilon^{(1)}_{\rm elst,resp}$(CCSD) and $\epsilon^{(1)}_{\rm exch}$(CCSD), computed at the CCSD level.[88–90] Due to the high computational cost for some of the larger dimers in the S22 data set, we do not consider this possibility in the present work, and treat SAPT2+(3)/aTZ as a good benchmark.

Examining the EFP results in Table II (which are taken from Ref. 41), we observe that the MUEs for the individual energy components are very large. For two of the dispersion-dominated complexes (π-stacked benzene dimer and π-stacked indole-benzene), the electrostatic interactions are even predicted to be repulsive by EFP.[41] On average, EFP underestimates electrostatic and induction energies in almost all strongly hydrogen-bonded complexes, by several kcal/mol, probably owing to insufficient capture of charge penetration by the screening function that is applied to the multipolar electrostatics.[41] Furthermore, neglect of the charge-transfer term in the EFP potentials is another source of error for induction energies, especially for hydrogen-bonded complexes.[41] The exchange energies are always underestimated by EFP, which may be caused by neglect of intramolecular correlation effects that are captured by SAPT.[41] Of the various EFP energy components, the dispersion energy best agrees with SAPT2+(3) results, although Podeszwa's[42] (+D2) dispersion potentials provide better agreement.

Another energy decomposition scheme based on the supermolecular method was proposed long ago by Kitaura and Morokuma,[91] and is conventionally known as EDA. Although the definitions of different energy components in the SAPT and EDA methods are different, they share a common term: electrostatic interaction. Given the same molecular geometry and level of theory, the electrostatic term calculated by SAPT should be similar to the corresponding term computed by EDA. For the S22A data set, the MUE of the electrostatic term calculated by EDA at the BLYP-D3/TZ2P level is 0.32 kcal/mol,[92] quite similar to XSAPT(KS)/haTZ results with empirical dispersion. In short, our interaction-energy decomposition scheme is accurate and can be extended to many-body systems that are not amenable to traditional SAPT energy decomposition.

Table III presents errors in binding energies, as compared to CCSD(T)/CBS benchmarks, for the S66 data set.[40] This set includes 66 weakly bound dimers representing binding motifs commonly found in biomolecular structures, and is thought to be more balanced than S22 with respect to different types of interactions. Although the original S66 binding energies[40] have been revised,[93] the revised values are only marginally different; the MUE between the original and the revised binding energies is 0.08 kcal/mol. We use the original binding energies here, to facilitate comparison to some published *ab initio* and DFT results.[13,40] As shown in previous work,[29] the XSAPT(KS)+D/aDZ′ method affords good cancellation of errors and gives very good results for S66 (MUE = 0.27 kcal/mol). A few supersystem DFT methods have been identified that yield comparable MUEs,[13] including M06-2X, ωB97X-D, and the double-hybrid B2PLYP-D3 method, and S66 error statistics for these methods are listed

in Table III as well. (Note, however, that a quadruple-ζ basis set was used in most of these DFT benchmarks,[13] so these DFT calculations are considerably more expensive than the XSAPT(KS)+D calculations.) For the -D and -D3 supersystem DFT methods, the classical dispersion potential is crucial to achieving MUEs of ∼0.3 kcal/mol. In the case of B2PLYP, for example, the MUE is reduced from 1.60 to 0.26 kcal/mol when the D3 correction is added.[13]

For XSAPT(KS)/haTZVPP in both its -D and -D2 form, we obtain MUEs of ∼0.4 kcal/mol when the $\delta E_{\rm int}^{\rm HF}$ correction is applied. This is superior to MP2, SCS-MP2, and CCSD results in the CBS limit, and superior also to EFP results. The maximum error for the XSAPT(KS)+D2/haTZVPP (AC) method is about 3 kcal/mol, for the π-stacked uracil dimer. That same system affords the maximum error when considering the S22 data set as well, and above we suggested that the +D2 potential could benefit from re-parameterization using a data set in which π-stacked complexes are better represented. For S66, the MUE for XSAPT(KS)+D2/haTZVPP (AC) is reduced to 0.33 kcal/mol if we eliminate the problematic π-stacked uracil dimer.

Note also that the use of exact AC (i.e., monomer-specific values of ω) reduces the MUE only marginally, from 0.42 to 0.38 kcal/mol for S66. For the S22A data set, exact AC actually *increases* the MUE, albeit by a similarly tiny amount (0.07 kcal/mol; see Table I). Thus, on the basis of the S22 and S66 results alone, there is really nothing to recommend the use of monomer-specific ω values, although there is also no compelling reason *not* to use the exact AC, other than the minor trouble of determining the optimized ω value for each monomer. However, use of the exact AC proves to be very important for binding energies of $Cl^-(H_2O)_n$ clusters, as discussed below in Sec. IV D.

Finally, the $\text{HFD}_{\text{as}}\text{c}^{(1)}$ results are worth considering in a bit more detail. This method exhibits a MUE of 0.46 kcal/mol for the S66 data set, and the maximum error (3.5 kcal/mol) also comes from the π-stacked uracil dimer. Its performance for S66 is thus very similar to that of XSAPT(KS)+D2, which was the case for the S22 data set as well. This is likely no accident, as the two methods bear much formal similarity. In particular, the $\delta E_{\rm int}^{\rm HF}$ correction is included in both, and the key difference is that whereas the $\text{HFD}_{\text{as}}\text{c}^{(1)}$ method includes the (infinite-order) CPHF response to the frozen, Hartree-Fock monomer density, the XSAPT(KS)+D2 method includes higher-order induction effects via the point-charge embedding of the iterative XPol procedure. As noted in Sec. II B, this means that XSAPT(KS)+D2 double-counts some higher-order induction effects, yet this evidently does not have deleterious effects on binding energies, even in strongly hydrogen-bonded complexes, and is considerably cheaper than solving CPHF equations. We therefore view this small formal inconsistency as a worthwhile price to pay for the ability to include electron correlation effects in the monomer properties.

B. Potential energy curves

The error statistics for S22A and S66 binding energies demonstrate that XSAPT(KS) methods with empirical dispersion potentials provide a good description of a wide variety of non-covalent interactions at their van der Waals (vdW) minima. It is important also to know how these methods perform across the whole range of intermolecular distances, so in this section we examine some one-dimensional PECs.

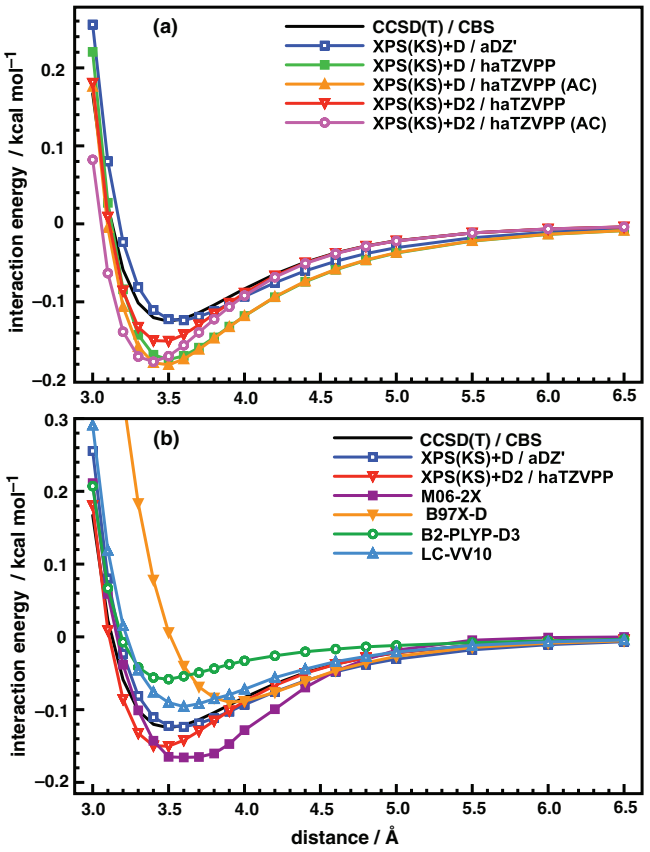

FIG. 1. Potential energy curves for Ar···Ne: (a) comparison of various XSAPT methods and (b) comparison of XSAPT and DFT methods. The $\delta E_{\rm int}^{\rm HF}$ correction is included in the XSAPT(KS)/haTZVPP methods with empirical dispersion potentials. For DFT methods, the def2-QZVP basis set is used with counterpoise correction.

The Ar···Ne interaction potential is thought to be a difficult case since several DFT methods, which are recommended for non-covalent interactions and which predict an accurate value of the Ar···Ne binding energy at the CCSD(T) vdW minimum, afford qualitatively *incorrect* PECs for this system.[29,56] Figure 1 compares the Ar···Ne interaction potentials given by various methods. XSAPT(KS)+D/aDZ′ provides a very good description of the whole range PEC for this system and is better than the XSAPT(KS)/haTZVPP + $\delta E_{\rm int}^{\rm HF}$ methods with empirical dispersion potentials. Furthermore, using the same ω for both monomers leads to slightly better results across the PEC.

In Fig. 1(b), we also compare XSAPT(KS) to DFT results obtained using the def2-QZVP basis set. The DFT methods selected for this comparison have been shown to yield accurate binding energies for the dispersion-dominated subset of S66.[13,20] Of these DFT methods, M06-2X overestimates the well depth at the minimum-energy geometry, as demonstrated previously.[56] B2PLYP-D3, which has the smallest MUE of any DFT-based method for the dispersion-dominated subset

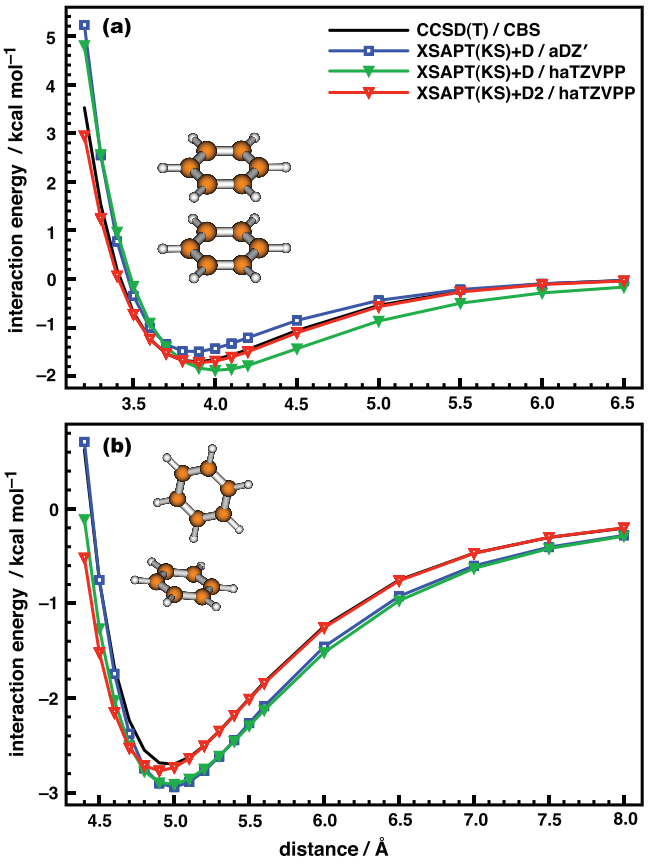

FIG. 2. Potential energy curves for (a) the "sandwich" and (b) the "T-shaped" isomer of $(C_6H_6)_2$. The distance coordinate in either case is the center-to-center distance between the benzene rings. Benchmark CCSD(T)/CBS results are taken from Ref. 74. The $\delta E_{\rm int}^{\rm HF}$ correction is included in the two XSAPT(KS)/haTZVPP calculations that include empirical dispersion.

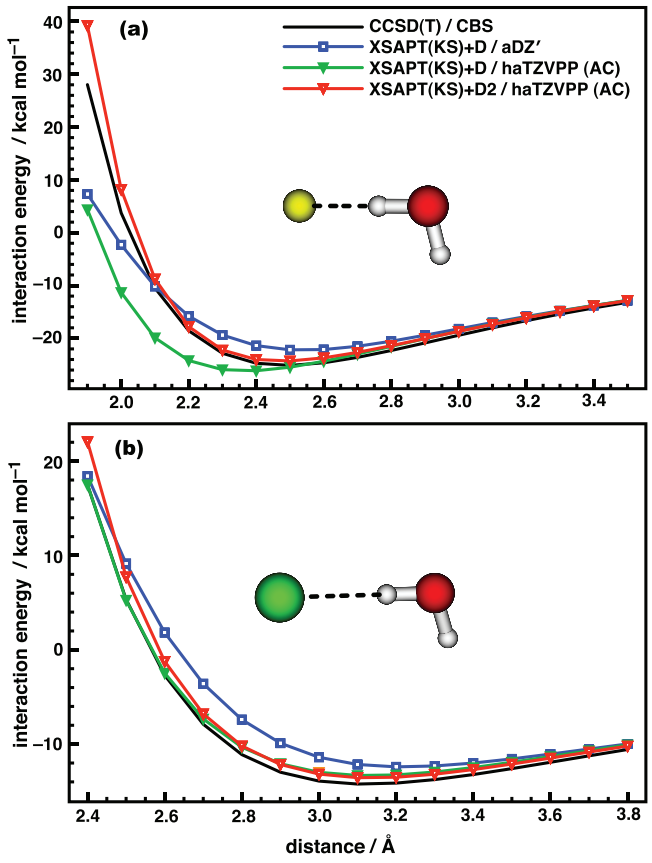

FIG. 3. Potential energy curves for (a) $F^-(H_2O)$ and (b) $Cl^-(H_2O)$ at a fixed H_2O geometry. The distance coordinate is the halide–oxygen distance and the benchmark is CCSD(T)/CBS. The $\delta E_{\rm int}^{\rm HF}$ correction is included in the XSAPT(KS)/haTZVPP methods with empirical dispersion potentials.

of S66 (see Table III), affords an Ar···Ne well that is much too shallow. The popular ωB97X-D functional[11] shifts the minimum-energy distance significantly as compared to other methods; the ωB97X-D minimum is located at a distance that is 0.4 Å too large, and the well depth is too shallow as well. The non-local correlation functional LC-VV10 yields a more accurate PEC than either M06-2X or ωB97X-D, consistent with results for the S66 data set, where the MUE for the dispersion-dominated subset is just 0.1 kcal/mol using LC-VV10.[20] However, XSAPT(KS)+D/aDZ′ is clearly superior across the entire PEC.

For the remaining homo-monomer systems considered in this work (benzene dimer and water clusters), the XSAPT(KS)+D2/haTZVPP methods with and without exact AC are the same. Furthermore, XSAPT(KS)+D/haTZVPP with and without exact AC affords very similar results, since these two methods differ only by way of slightly different s_β parameters in the dispersion potential.[68] For brevity, we thus limit our discussion to XSAPT(KS)/haTZVPP results without exact AC for homo-monomer systems. For the halide–water clusters (hetero-monomer systems), however, we will only show the results *with* exact AC, since this improves the binding energies, as will be seen in Sec. IV D.

Benzene dimer is considered to be a stringent test of dispersion interactions and we consider both "sandwich" and "T-shaped" isomers, as shown in Fig. 2. For the sandwich isomer, XSAPT(KS)+D2/haTZVPP reproduces the whole CCSD(T)/CBS potential curve almost quantitatively. For the T-shaped isomer, XSAPT(KS)+D2/haTZVPP slightly underestimates the binding energy at short intermolecular distance but is very accurate beyond the vdW minimum. The new +D2 dispersion potential outperforms the old one for this system.

In Fig. 3, we plot PECs for $F^-(H_2O)$ and $Cl^-(H_2O)$, which are known to be challenging cases for both XSAPT as well as traditional SAPT,[29,55] even when third-order corrections are included in the latter.[55] Using the exact AC improves the binding energies by about 0.1 kcal/mol for $F^-(H_2O)$ and 0.3 kcal/mol for $Cl^-(H_2O)$. Values of ω, optimized according to Eq. (16), are 0.500, 0.475, and 0.375 bohr^{-1} for H_2O, F^-, and Cl^-, respectively, and the similarity between the H_2O and F^- values explains why use of exact AC has a smaller effect for $F^-(H_2O)$ than for $Cl^-(H_2O)$. This provides further evidence that the exact AC afforded by the IP condition in Eq. (16) is necessary, especially for H-bonded systems.

For $F^-(H_2O)$, the XSAPT(KS)+D/aDZ′ and XSAPT(KS)+D/haTZVPP (AC) methods greatly overestimate the interaction energy at short distance. The second-generation method is significantly better, and in fact is in nearly quantitative agreement with CCSD(T)/CBS

results except at monomer separations much smaller than the vdW distance, where it becomes a bit too repulsive. In previous work,[55] we hypothesized that the poorly shaped PECs for $F^-(H_2O)$ that were obtained using XSAPT(KS)+D might originate in a low-energy $FH\cdots OH^-$ diabatic state[94] that simply cannot be captured by monomer-based methods. While such a state may ultimately lead to other problems, we see in Fig. 3 that most of the error in the shape of the PEC is eliminated when the +D dispersion potential is replaced by +D2, and the $\delta E_{\rm int}^{\rm HF}$ is included as well.

C. Water hexamer

Results presented above demonstrate the excellent performance of XSAPT(KS)+D2/haTZVPP (AC) for dimers, including some challenging ones. In this section and in Sec. IV D, we investigate whether these favorable results extend to large clusters composed of polar monomers, where many-body polarization effects are important.

In this section, we examine the widely studied $(H_2O)_6$ cluster, considering the eight low-lying isomers identified in Ref. 95. Although CCSD(T)/CBS binding energies are available for these isomers,[95,96] we have chosen to base our analysis instead on MP2/a5Z-h energies from Ref. 48. (The notation "a5Z-h" means aug-cc-pV5Z with h functions removed.) The reason for this choice is that traditional EDA has been performed at this level of theory,[48] the results of which can be used to validate our XSAPT version of EDA.

On the other hand, we have made binding energies a major focus of this work and therefore we should at least examine the $(H_2O)_6$ binding energies and make some effort to improve upon the MP2/a5Z-h//MP2/aTZ benchmarks from Ref. 48, which we will do using the CCSD(T)/CBS//MP2/haTZ results from Ref. 95. We have attempted to correct the former benchmarks in two ways, as detailed in the supplementary material.[68] In the first case, we simply add a CCSD(T) correction to the counterpoise-corrected MP2/a5Z-h binding energies in Ref. 48; this correction is computed as the difference between CCSD(T)/CBS and MP2/CBS binding energies from Ref. 95. The *slightly* different cluster geometries used in Ref. 95 versus Ref. 48 (MP2/haTZ versus MP2/aTZ) are unlikely to make much difference, hence the major source of error in what we will call "MP2/a5Z-h + $\delta_{\rm CCSD(T)}$" binding energies is basis-set incompleteness at a5Z-h level. This causes MP2/a5Z-h binding energies to differ from MP2/CBS values by 0.7–0.8 kcal/mol.[68] In view of this, a second attempt was made to obtain better binding energies simply by using the CCSD(T)/CBS//MP2/haTZ values[95] at MP2/aTZ geometries. In the latter case, we have added MP2/a5Z-h monomer relaxation energies (from Ref. 48), since monomer relaxation was included in the calculations of Ref. 95 but is not included in the binding energies that we report here, since we want to isolate intermolecular interactions.

Figure 4 compares the binding energies of eight $(H_2O)_6$ isomers, computed with various methods, to these benchmarks. Of the XSAPT methods, XSAPT(KS)+D/aDZ′ significantly overestimates the binding energies and furthermore makes rather large errors in the relative energies and the energetic ordering of the isomers. (Note that this is true regardless of which of the benchmarks from the previous paragraph we choose for comparison.) We have seen previously that triple-ζ basis sets are often important for H-bonded systems, but using XSAPT(KS)+D with the haTZVPP basis set leads to binding energies that are underestimated rather than overestimated, although the relative energies are improved. In all, the +D results cannot be said to be quantitative for $(H_2O)_6$.

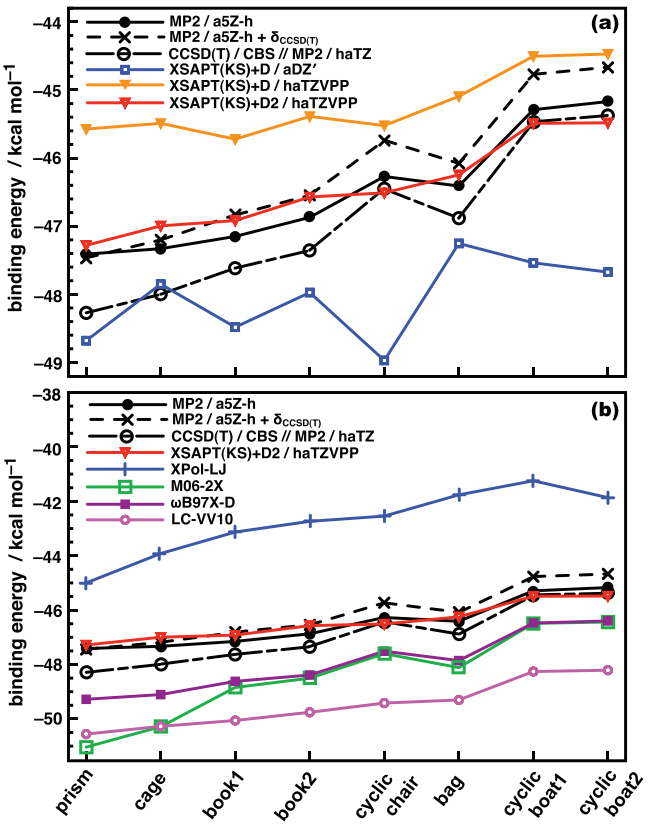

FIG. 4. Binding energies for eight isomers of $(H_2O)_6$, computed using (a) XSAPT methods and (b) XSAPT, XPol-LJ, and DFT methods. MP2/a5Z-h //MP2/aTZ results are taken from Ref. 48 and corrected, as described in the text, to obtain the "MP2/a5Z-h + $\delta_{\rm CCSD(T)}$" benchmarks. Alternatively, we can use CCSD(T)/CBS results at MP2/haTZ geometries, since the latter are very similar to the MP2/aTZ geometries. LJ parameters for XPol-LJ are taken from Ref. 87 where they were optimized for H-bonded systems. The def2-QZVP basis set with counterpoise correction is used for all DFT methods, and the $\delta E_{\rm int}^{\rm HF}$ correction is included in the XSAPT(KS)/haTZVPP methods with empirical dispersion.

The +D2 potential, however, affords good results for both absolute binding energies and relative isomer energies, except for a slight overstabilization of the cyclic-chair isomer. For the most part, the XSAPT(KS)+D2 (AC)/haTZVPP binding energies lie between the two sets of CCSD(T) benchmarks, though slightly (<1 kcal/mol) underestimated with respect to the CCSD(T)/CBS//MP2/haTZ values that we consider the more reliable. The contribution from $\delta E_{\rm int}^{\rm HF}$ is about 10 kcal/mol and is essential for obtaining good results. This is consistent with recommendations to use $\delta E_{\rm int}^{\rm HF}$ for SAPT calculations involving polar monomers.[34,53,54] However, the $\delta E_{\rm int}^{\rm HF}$ correction is not suitable for use with XSAPT(KS)+D/aDZ′ since this method's success rests on cancellation of errors.

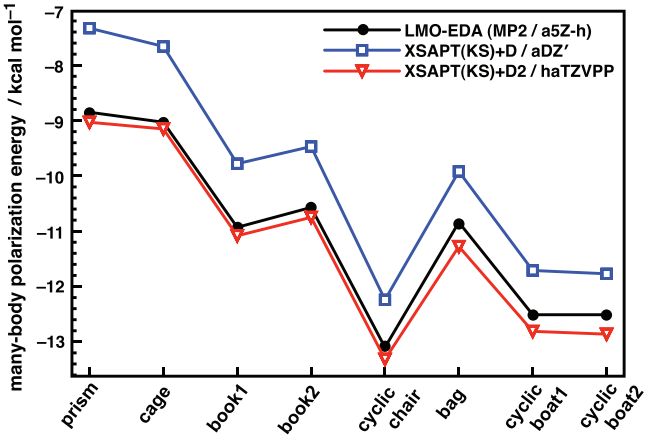

FIG. 5. Many-body polarization energies for eight $(H_2O)_6$ isomers. MP2/a5Z-h results are taken from Ref. 48. The $\delta E_{\rm int}^{\rm HF}$ correction is included in the XSAPT(KS)+D2/haTZVPP calculations.

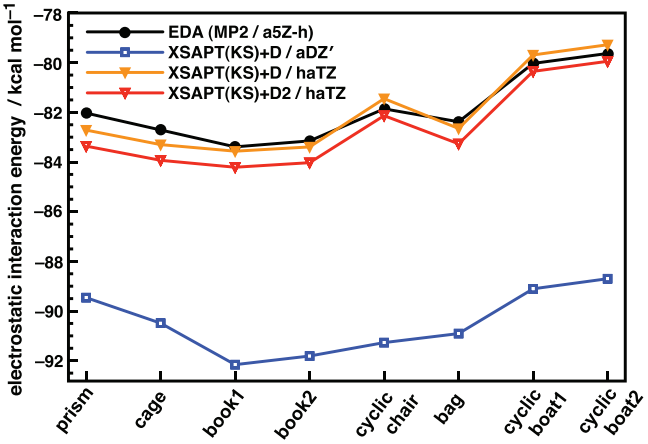

FIG. 6. Electrostatic interaction energy for eight $(H_2O)_6$. The MP2/a5Z-h results are taken from Ref. 48.

In Fig. 4(b), we compare XSAPT to M06-2X and ωB97X-D results, as these two functionals performed best for the H-bonded subset of S66, with MUEs of ≈0.2 kcal/mol (see Table III). For $(H_2O)_6$, however, the errors are much larger. M06-2X overestimates the binding energies by 1–3 kcal/mol, and ωB97X-D by about 1 kcal/mol, relative to CCSD(T)/CBS//MP2/haTZ benchmarks. The LC-VV10 functional also consistently overestimates the binding energies by a larger amount (≈3 kcal/mol), although the relative energies are quite good. Despite being parameterized for H-bonded complexes,[87] the XPol-LJ results are quite poor, with binding energies that are underestimated by up to about 5 kcal/mol. Unlike these other methods, the performance of XSAPT(KS)+D2/haTZVPP for dimers does extend to this larger system.

Turning now to analysis of energy components in $(H_2O)_6$, the importance of many-body effects in this system has been quantified recently by Chen and Li,[48] at the MP2/a5Z-h level, using localized molecular orbital energy decomposition analysis (LMO-EDA), which is a simplified version of the Kitaura-Morokuma EDA scheme.[91] These authors point out that the many-body effects are dominated by polarization interactions, whereas the other energy components are strictly or nearly pairwise additive.[48] Figure 5 compares the many-body polarization energies for the $(H_2O)_6$ isomers considered here, computed using LMO-EDA at the MP2/a5Z-h level[48] or using our XSAPT-based energy-decomposition scheme, the latter of which assumes that the many-body energy arises exclusively from polarization. The +D and +D2 versions of XSAPT(KS)/haTZVPP give very similar results for the many-body contribution to the energy, so only the latter are shown, for clarity. These results are in good agreement with the LMO-EDA results, whereas the XSAPT(KS)+D/aDZ′ method consistently underestimates the many-body energy.

Electrostatic energies from LMO-EDA and from XSAPT(KS)-based methods are shown in Fig. 6. The XSAPT(KS)/aDZ′ method greatly overestimates the electrostatic energies, as compared to values extracted from MP2/a5Z-h calculations, whereas XSAPT(KS)/haTZVPP methods with empirical dispersion potentials show good agreement with LMO-EDA for the electrostatic energy. This is consistent with our previous conclusions that triple-ζ basis sets are necessary for hydrogen-bonded systems.[28,29]

D. Halide–water clusters

The SAPT method does not afford chemical accuracy for strongly interacting systems, especially ions at short intermolecular distances, where the perturbation series converges slowly or may even diverge.[55,97] First-generation XSAPT(KS)+D/aDZ′ results for $F^-(H_2O)$ or $Cl^-(H_2O)$ also show qualitatively incorrect PECs at short distance, as demonstrated in Fig. 3 and also Ref. 29. As demonstrated above, however, this problem is resolved by the D2 dispersion potential, and the halide–water PECs have reasonable shapes and afford accurate binding energies. In this section, we extend these anion systems from two-body to many-body systems, investigating binding energies for $X^-(H_2O)_n$ up to $n = 6$, for X = F and Cl. All cluster geometries were optimized at the RI-MP2/aTZ level of theory, and we consider one geometry per cluster size. Since traditional SAPT is limited to dimers, we first compare the binding energies for the binary $X^-(H_2O)$ complexes in Table IV.

The difference between RI-CCSD(T)/CBS and RI-MP2/CBS binding energies is only 0.1–0.3 kcal/mol for these dimers, although the difference is larger in the larger halide–water clusters, as demonstrated below. Examining the DFT results in Table IV, we see that both M06-2X and LC-VV10 overestimate the binding energies, by ∼3 kcal/mol for $F^-(H_2O)$ and ∼1 kcal/mol for $Cl^-(H_2O)$. Addition of the D3 correction proposed by Grimme[14] (M06-2X-D3) leads to even worse results. The ωB97X-D binding energy for $Cl^-(H_2O)$ is quite accurate but that for $F^-(H_2O)$ is overestimated by about 1 kcal/mol. SAPT0/aDZ′ binding energies are reasonable (errors ∼1 kcal/mol) and are significantly better than SAPT0/aTZ results. This is consistent with SAPT0 results for other systems, where the aDZ′ basis set leads to favorable error cancellation,[34,69] which is why this basis set was

TABLE IV. Binding energies (in kcal/mol) for $F^-(H_2O)$ and $Cl^-(H_2O)$.

Method	$F^-(H_2O)$	$Cl^-(H_2O)$
RI-CCSD(T)/CBS	−32.32	−15.48
RI-MP2/CBS	−32.25	−15.75
M06-2X[a]	−35.67	−16.30
ωB97X-D[a]	−33.73	−15.48
LC-VV10[a]	−35.01	−16.49
SAPT0/aDZ′	−33.93	−14.67
SAPT0/aTZ	−37.57	−16.52
SAPT2+(3)/aTZ	−34.07	−15.53
XSAPT(KS)+D/aDZ′	−27.44	−13.52
XSAPT(KS)+D[b]	−36.22	−14.63
XSAPT(KS)+D (AC)[b]	−36.20	−14.79
XSAPT(KS)+D2[b]	−33.01	−14.69
XSAPT(KS)+D2 (AC)[b]	−33.15	−15.05

[a]Using the def2-QZVP basis set with counterpoise correction.
[b]Using the haTZVPP basis set and the $\delta E_{\rm int}^{\rm HF}$ correction.

proposed in the first place, albeit in the context of dispersion-bound complexes.

The high-level SAPT2+(3)/aTZ method affords almost the same binding energy as RI-CCSD(T)/CBS for $Cl^-(H_2O)$ but overestimates the $F^-(H_2O)$ binding energy by about 1.8 kcal/mol, consistent with previous results where methods beyond SAPT0 were employed.[55] In conjunction with the observation that supersystem DFT errors are consistently larger for $F^-(H_2O)$ than they are for $Cl^-(H_2O)$, these results suggest that fluoride–water is an especially challenging test of monomer-based quantum chemistry, as we have observed in previous work.[25,26,28,55,98]

Examining the XSAPT results in Table IV, we observe that XSAPT(KS)+D/aDZ′ underestimates the $X^-(H_2O)$ binding energy, especially for X = F, but the new XSAPT(KS)+D2/haTZVPP method performs much better, with errors <1 kcal/mol when exact AC is used. Once again, this is consistent with the need for triple-ζ basis sets for H-bonded systems.[28,29] In terms of the exact AC (versus using the same ω value for all monomers), results with and without exact AC differ by only 0.14 kcal/mol for $F^-(H_2O)$, but for $Cl^-(H_2O)$ exact AC improves the binding energy by 0.36 kcal/mol.

Figure 7 compares XSAPT(KS)+D2 binding energies for $X^-(H_2O)_n$ clusters (up to $n = 6$) to RI-CCSD(T)/CBS benchmarks. Also included in this comparison are binding energies computed using three different DFT methods. As cluster size increases from $n = 1$ to $n = 6$, we observe that the magnitude of the triples correction [Eq. (20)] increases from −0.07 to −1.46 kcal/mol for fluoride–water clusters and from 0.27 to 0.46 kcal/mol for chloride–water clusters. (Actual numerical data are given in the supplementary material.[68]) For $F^-(H_2O)_n$ clusters, the triples correction is −1.0 kcal/mol already for $n = 3$, and is as large as −2.1 kcal/mol for $F^-(H_2O)_{10}$. As such, the RI-MP2/CBS method cannot be considered a sub-kcal/mol benchmark level of theory for $F^-(H_2O)_n$ binding energies, although for $Cl^-(H_2O)_n$ the triples correction is no larger than 0.5 kcal/mol up to $n = 6$.

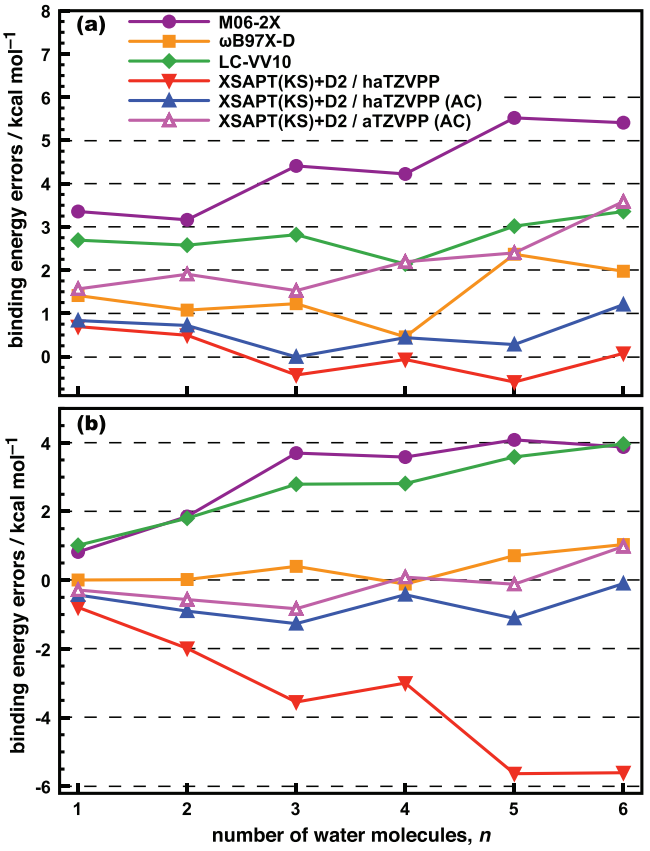

FIG. 7. Errors in binding energies with respect to RI-CCSD(T)/CBS benchmarks for (a) $F^-(H_2O)_n$ and (b) $Cl^-(H_2O)_n$, up to $n = 6$. The def2-QZVP basis set and counterpoise correction was used for all DFT calculations. The $\delta E_{\rm int}^{\rm HF}$ correction is included in the XSAPT calculations. Positive and negative errors imply that the binding energies are over- and underestimated, respectively.

Oddly, XSAPT(KS)+D2/haTZVPP results for $F^-(H_2O)_n$ are slightly better when the same ω value is used for both F^- and H_2O. Differences between this approach and the exact AC approach range up to about 1 kcal/mol. These discrepancies may be coincidence or cancellation of errors, because for $Cl^-(H_2O)_n$, where the ω values optimized for Cl^- and H_2O are quite different,[68] use of the exact AC proves to be crucial. Exact AC reduces the errors from ≈6 kcal/mol for $Cl^-(H_2O)_5$ and $Cl^-(H_2O)_6$ down to 1.1 and 0.1 kcal/mol, respectively.

In view of these results, we consider the use of exact AC (i.e., different ω values for different monomers) to be very important for XSAPT calculations, despite the small differences observed for the S22 and S66 data sets, which include no ions and whose interaction energies are relatively small as compared to those encountered for $X^-(H_2O)_n$ complexes. The S22 and S66 results likely understate the importance of monomer-specific AC, because many of the interaction energies in those data sets are quite small and thus their ω-dependence is small on an energy scale of, say, 1 kcal/mol. Moreover, several of the strongly H-bonded examples in those data sets, such as $(H_2O)_2$ and $(HCO_2H)_2$ from S22, are homodimers; in such cases, there is no distinction between the "exact" (monomer-specific) AC, and the tuning of ω based on using the lowest monomer IP in Eq. (16). In SAPT(DFT) and

TABLE V. Mean unsigned errors with respect to RI-CCSD(T)/CBS benchmarks, for binding energies of $X^-(H_2O)_n$, $n = 1$–6.

	MUE/kcal mol^{-1}	
Method	X = F	X = Cl
RI-MP2/CBS	0.94	0.32
M06-2X[a]	4.35	2.99
ωB97X-D[a]	1.42	0.38
LC-VV10[a]	2.77	2.66
XSAPT(KS)+D2[b]	0.39	3.43
XSAPT(KS)+D2 (AC)[b,c]	0.58	0.71
XSAPT(KS)+D2 (AC)[c,d]	2.20	0.48

[a]Using the def2-QZVP basis set with counterpoise correction.
[b]Using the haTZVPP basis set.
[c]Using the $\delta E_{\rm int}^{\rm HF}$correction.
[d]Using the aTZVPP basis set.

TABLE VI. Mean unsigned errors in the binding energies of ten isomers of $F^-(H_2O)_{10}$, with respect to RI-CCSD(T)/CBS benchmarks.

	MUE/kcal mol^{-1}	
Method	Binding energy	Relative energy
RI-MP2/CBS	1.79	0.27
M06-2X[a]	8.69	0.78
ωB97X-D[a]	3.68	0.35
LC-VV10[a]	6.77	0.44
XSAPT(KS)+D2[b]	1.07	0.59
XSAPT(KS)+D2 (AC)[b]	1.34	0.49

[a]Using the def2-QZVP basis set with counterpoise correction.
[b]Using the haTZVPP basis set and the $\delta E_{\rm int}^{\rm HF}$ correction.

DFT-SAPT calculations, monomer-specific AC has been recommended even for charge-neutral monomers.[36,99]

Also interesting are the results when the basis set is increased from haTZVPP to the fully augmented aTZVPP. As summarized in Table V, the MUE for the fluoride–water clusters increases from 0.6 to 2.2 kcal/mol while that for chloride–water clusters *decreases* from 0.7 to 0.5 kcal/mol. The difference in the latter case is not substantial, but for fluoride–water clusters we think this is a possible indication of overpolarization in the XPol step, owing to the presence of a large number of diffuse basis functions. These results are the reason that we primarily use haTZVPP in this work.

Beran[100] has suggested that errors in fragment-based calculations of molecular clusters ought to be extensive (proportional to the number of monomer units). In applications to water clusters, our group has indeed observed some numerical evidence of a roughly constant error per hydrogen bond, for clusters with $\gtrsim$5 hydrogen bonds.[27] The slight uptick in the errors for $F^-(H_2O)_n$ as a function of increasing cluster size [see Fig. 7(a)] reminds us of these observations. To investigate this further, we performed calculations on a set of ten isomers of $F^-(H_2O)_{10}$, whose initial structures were taken from Ref. 25 but then optimized at the B3LYP/6-31G* level of theory. MUEs for these $F^-(H_2O)_{10}$ binding energies, with respect to RI-CCSD(T)/CBS benchmarks, are listed in Table VI. Results from several supersystem DFT methods are included in this comparison, and in these cases the def2-QZVP basis set and the Boys-Bernardi counterpoise correction were used, consistent with other DFT results presented herein.

As compared to the RI-CCSD(T)/CBS benchmarks, the MUE obtained at the RI-MP2/CBS level is about 1.8 kcal/mol, providing further demonstration that the many-body triples correction is significant for $F^-(H_2O)_n$. The three DFT methods that we examine (M06-2X, ωB97X-D, and LC-VV10) each afford very large errors in total binding energies, although errors for relative isomer energies are much smaller. M06-2X is the worst of the bunch, with a MUE of almost 9 kcal/mol in the binding energies. Errors in the absolute binding energies are much larger than those observed in a recent study of DFT methods applied to $SO_4^{2-}(H_2O)_6$ clusters,[101] confirming the challenging nature of fluoride–water clusters, and none of the three functionals that worked so well for S22 even comes close to achieving "chemical accuracy" for absolute binding energies. The MUEs for relative isomer energies, on the other hand, are <1 kcal/mol for all three functionals. A plot of the relative energies of our ten $F^-(H_2O)_{10}$ clusters, at various levels of theory and compared to RI-CCSD(T)/CBS benchmarks, can be found in Fig. S1 of the supplementary material.[68]

In contrast to these supersystem DFT results, the XSAPT(KS)+D2/haTZVPP method (either with or without exact AC) exhibits a much smaller MUE ($\sim$1 kcal/mol) for the absolute binding energies of these $F^-(H_2O)_{10}$ clusters. Thus, the absolute binding errors come close to the "chemical accuracy" standard of 1 kcal/mol, and are certainly good enough to be useful in practical applications (especially considering that errors in relative energies are smaller still), even though the errors are larger than the $\sim$0.5 kcal/mol errors in binding energies for $n \leq 6$. Further work is presently underway in our group to examine the issue of size extensivity in fragment-based methods and to test these methods in significantly larger clusters.

V. CONCLUSIONS

A second-generation ("+D2") version of our XPol+SAPT(KS) method with empirical dispersion has been introduced and tested. An exact asymptotic correction (AC) scheme, in which the AC is optimized separately for each monomer according to the criterion $\varepsilon_{\rm HOMO} = -{\rm IP}$, is found to be necessary in general to obtain good binding energies, especially for ion–water clusters and other strongly hydrogen-bonded complexes. The XSAPT(KS)+D2/haTZVPP method with monomer-specific AC exhibits MUEs of 0.82 and 0.38 kcal/mol for S22A and S66 binding energies, respectively, although two outliers (the π-stacked uracil dimer and π-stacked adenine–thymine complex) suggest that some further refinement of the empirical dispersion potentials may be in order, using data sets that contain additional π-stacked complexes.

A variety of other challenging systems have been considered as well, including Ar$\cdots$Ne, $(C_6H_6)_2$, $(H_2O)_6$, $Cl^-(H_2O)_n$, and $F^-(H_2O)_n$. The XSAPT(KS)+D2/haTZVPP method affords accurate one-dimensional potential energy scans for the dimers and accurate relative energies for the

larger clusters. In particular, this method corrects certain qualitative problems in the short-range description of halide–water interaction potentials that were observed in the "first generation" version of the method.[29]

The accuracy of XSAPT(KS)+D2/haTZVPP (AC) is superior to that of popular DFT approaches for non-covalent interactions, including M06-2X, ωB97X-D, and LC-VV10, which are selected as exemplars here because of their favorable accuracy for S22 binding energies.

For the very challenging halide–water clusters, XSAPT(KS)+D2/haTZVPP (AC) binding energies are of moderate accuracy, with $\sim$1 kcal/mol errors observed with respect to RI-CCSD(T)/CBS results. On the other hand, the DFT methods fare even worse, with errors of several kcal/mol in absolute binding energies. These supersystem DFT approaches are not only less accurate as compared to XSAPT(KS)+D2, but because the latter is a monomer-based method, the supersystem DFT calculations are much more expensive as well.[29]

The halide–water clusters examined here prove to be very challenging examples for both supersystem and monomer-based methods designed for non-covalent interactions, and we suggest that these examples should probably be considered routinely whenever evaluating the performance of various methods for non-covalent interactions. Coordinates for these clusters along with benchmark RI-CCSD(T)/CBS binding energies are available in the supplementary material.[68]

Finally, we have introduced an interaction-energy decomposition scheme for XSAPT that extends SAPT energy decomposition analysis to many-body systems. The different energy components (electrostatic, exchange, induction, and dispersion) for the S22A data set are in very good agreement with benchmark SAPT2+(3)/aug-cc-pVTZ results, demonstrating that our energy decomposition scheme is robust. Using this energy-decomposition scheme in conjunction with XSAPT(KS)+D2, the many-body contributions to the binding energies of $(H_2O)_6$ isomers are reproduced almost perfectly as compared to benchmark calculations. Therefore, XSAPT(KS)+D2 (AC) not only yields good binding energies for different non-covalent systems, but furthermore we can decompose these binding energy into physical meaningful energy components for many-body systems, which is not possible in traditional SAPT.

ACKNOWLEDGMENTS

This work was supported by the (U.S.) Department of Energy (DOE), Office of Basic Energy Sciences, Division of Chemical Sciences, Biosciences, and Geosciences (Award No. DE-SC0008550). Calculations were performed at the Ohio Supercomputer Center under projects PAS-0291 and PAA-0003. J. M. H. is an Arthur P. Sloan Foundation Fellow and a Camille Dreyfus Teacher-Scholar.

[1] E. Aprà, A. P. Rendell, R. J. Harrison, V. Tipparaju, W. A. de Jong, and S. S. Xantheas, "Liquid water: Obtaining the right answer for the right reasons," in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis, SC '09* (ACM, New York, 2009), pp. 66:1–7.
[2] D. Cremer, WIREs Comput. Mol. Sci. **1**, 509 (2011).
[3] S. Grimme, L. Goerigk, and R. F. Fink, WIREs Comput. Mol. Sci. **2**, 886 (2012).
[4] K. E. Riley, J. A. Platts, J. Řezác, P. Hobza, and J. G. Hill, J. Phys. Chem. A **116**, 4159 (2012).
[5] R. A. DiStasio, Jr., and M. Head-Gordon, Mol. Phys. **105**, 1073 (2007).
[6] K. E. Riley, J. Řezác, and P. Hobza, Phys. Chem. Chem. Phys. **13**, 21121 (2011).
[7] P. Hobza, Acc. Chem. Res. **45**, 663 (2012).
[8] S. Grimme, WIREs Comput. Mol. Sci. **1**, 211 (2011).
[9] J. Klimeš and A. Michaelides, J. Chem. Phys. **137**, 120901 (2012).
[10] Y. Zhao and D. G. Truhlar, Theor. Chem. Acc. **120**, 215 (2008).
[11] J.-D. Chai and M. Head-Gordon, Phys. Chem. Chem. Phys. **10**, 6615 (2008).
[12] S. Grimme, J. Comput. Chem. **27**, 1787 (2006).
[13] L. Goerigk, H. Kruse, and S. Grimme, ChemPhysChem **12**, 3421 (2011).
[14] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).
[15] L. A. Burns, A. V. Mayagoitia, B. G. Sumpter, and C. D. Sherrill, J. Chem. Phys. **134**, 084107 (2011).
[16] Y.-S. Lin, G.-D. Li, S.-P. Mao, and J.-D. Chai, J. Chem. Theory Comput. **9**, 263 (2013).
[17] K. Lee, É. D. Murray, L. Kong, B. I. Lundqvist, and D. C. Langreth, Phys. Rev. B **82**, 081101 (2010).
[18] O. A. Vydrov and T. Van Voorhis, Phys. Rev. Lett. **103**, 063004 (2009).
[19] O. A. Vydrov and T. Van Voorhis, J. Chem. Phys. **133**, 244103 (2010).
[20] O. A. Vydrov and T. Van Voorhis, J. Chem. Theory Comput. **8**, 1929 (2012).
[21] M. S. Gordon, L. V. Slipchenko, H. Li, and J. H. Jensen, Annu. Rep. Comput. Chem. **3**, 177 (2007).
[22] M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, Chem. Rev. **112**, 632 (2012).
[23] D. G. Fedorov, T. Nagata, and K. Kitaura, Phys. Chem. Chem. Phys. **14**, 7562 (2012).
[24] S. Wen, K. Nanda, Y. Huang, and G. J. O. Beran, Phys. Chem. Chem. Phys. **14**, 7578 (2012).
[25] R. M. Richard and J. M. Herbert, J. Chem. Phys. **137**, 064113 (2012).
[26] L. D. Jacobson, R. M. Richard, K. U. Lao, and J. M. Herbert, "Efficient monomer-based quantum chemistry methods for molecular and ionic clusters," Annu. Rep. Comput. Chem. (in press).
[27] L. D. Jacobson and J. M. Herbert, J. Chem. Phys. **134**, 094118 (2011).
[28] J. M. Herbert, L. D. Jacobson, K. U. Lao, and M. A. Rohrdanz, Phys. Chem. Chem. Phys. **14**, 7679 (2012).
[29] K. U. Lao and J. M. Herbert, J. Phys. Chem. Lett. **3**, 3241 (2012).
[30] In previous work, we have used the acronym "XPS" instead of XSAPT, but this risks confusion with X-ray photoelectron spectroscopy.
[31] W. Xie, L. Song, D. G. Truhlar, and J. Gao, J. Chem. Phys. **128**, 234108 (2008).
[32] B. Jeziorski, R. Moszynski, and K. Szalewicz, Chem. Rev. **94**, 1887 (1994).
[33] K. Szalewicz, WIREs Comput. Mol. Sci. **2**, 254 (2012).
[34] E. G. Hohenstein and C. D. Sherrill, WIREs Comput. Mol. Sci. **2**, 304 (2012).
[35] H. L. Williams and C. F. Chabalowski, J. Phys. Chem. A **105**, 646 (2001).
[36] A. J. Misquitta and K. Szalewicz, Chem. Phys. Lett. **357**, 301 (2002).
[37] S. M. Cybulski and M. L. Lytle, J. Chem. Phys. **127**, 141102 (2007).
[38] A. Hesselmann, J. Phys. Chem. A **115**, 11321 (2011).
[39] T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall, and C. D. Sherrill, J. Chem. Phys. **132**, 144104 (2010).
[40] J. Řezác, K. E. Riley, and P. Hobza, J. Chem. Theory Comput. **7**, 2427 (2011).
[41] J. C. Flick, D. Kosenkov, E. G. Hohenstein, C. D. Sherrill, and L. V. Slipchenko, J. Chem. Theory Comput. **8**, 2835 (2012).
[42] R. Podeszwa, K. Pernal, K. Patkowski, and K. Szalewicz, J. Phys. Chem. Lett. **1**, 550 (2010).
[43] G. Chałasiński, M. M. Szczęśniak, and R. A. Kendall, J. Chem. Phys. **101**, 8860 (1994).
[44] N. Turki, A. Milet, A. Rahmouni, O. Ouamerali, R. Moszynski, E. Kochanski, and P. E. S. Wormer, J. Chem. Phys. **109**, 7157 (1998).
[45] J. M. Pedulla, K. Kim, and K. D. Jordan, Chem. Phys. Lett. **291**, 78 (1998).
[46] T. P. Tauer and C. D. Sherrill, J. Phys. Chem. A **109**, 10475 (2005).
[47] A. L. Ringer and C. D. Sherrill, Chem. Eur. J. **14**, 2542 (2008).
[48] Y. Chen and H. Li, J. Phys. Chem. A **114**, 11719 (2010).
[49] C. M. Breneman and K. B. Wiberg, J. Comput. Chem. **11**, 361 (1990).
[50] V. F. Lotrich and K. Szalewicz, J. Chem. Phys. **106**, 9668 (1997).

[51] V. F. Lotrich and K. Szalewicz, J. Chem. Phys. **112**, 112 (2000).
[52] B. Jeziorski, R. Moszynski, A. Ratkiewicz, S. Rybak, K. Szalewicz, and H. L. Williams, "SAPT: A program for many-body symmetry-adapted perturbation theory calculations of intermolecular interaction energies," in *Methods and Techniques in Computational Chemistry: METECC-94*, edited by E. Clementi (STEF, Cagliari, 1993), Vol. B, Chap. 3, pp. 79–129.
[53] K. Patkowski, K. Szalewicz, and B. Jeziorski, J. Chem. Phys. **125**, 154107 (2006).
[54] K. Patkowski, K. Szalewicz, and B. Jeziorski, Theor. Chem. Acc. **127**, 211 (2010).
[55] K. U. Lao and J. M. Herbert, J. Phys. Chem. A **116**, 3042 (2012).
[56] K. Pernal, R. Podeszwa, K. Patkowski, and K. Szalewicz, Phys. Rev. Lett. **103**, 263201 (2009).
[57] K. T. Tang and J. P. Toennies, J. Chem. Phys. **80**, 3726 (1984).
[58] A. J. Misquitta, R. Podeszwa, B. Jeziorski, and K. Szalewicz, J. Chem. Phys. **123**, 214103 (2005).
[59] F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).
[60] B. Wang and D. G. Truhlar, J. Chem. Theory Comput. **6**, 3330 (2010).
[61] A. W. Lange and J. M. Herbert, J. Am. Chem. Soc. **131**, 3913 (2009).
[62] A. W. Lange, M. A. Rohrdanz, and J. M. Herbert, J. Phys. Chem. B **112**, 6304 (2008).
[63] M. A. Rohrdanz and J. M. Herbert, J. Chem. Phys. **129**, 034107 (2008).
[64] M. A. Rohrdanz, K. M. Martins, and J. M. Herbert, J. Chem. Phys. **130**, 054112 (2009).
[65] T. M. Henderson, B. G. Janesko, and G. E. Scuseria, J. Chem. Phys. **128**, 194105 (2008).
[66] R. Baer, E. Livshits, and U. Salzner, Annu. Rev. Phys. Chem. **61**, 85 (2010).
[67] A. Heßelmann, G. Jansen, and M. Schütz, J. Chem. Phys. **122**, 014103 (2005).
[68] See supplementary material at http://dx.doi.org/10.1063/1.4813523 for s_β and ω parameters, a discussion of $(H_2O)_6$ benchmarks, and benchmark data for halide–water clusters.
[69] E. G. Hohenstein and C. D. Sherrill, J. Chem. Phys. **132**, 184111 (2010).
[70] E. G. Hohenstein and C. D. Sherrill, J. Chem. Phys. **133**, 014101 (2010).
[71] P. Jurečka, J. Šponer, J. Černý, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).
[72] M. S. Marshall, L. A. Burns, and C. D. Sherrill, J. Chem. Phys. **135**, 194102 (2011).
[73] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory* (Wiley, New York, 2000).
[74] C. D. Sherrill, T. Takatani, and E. G. Hohenstein, J. Phys. Chem. A **113**, 10146 (2009).
[75] D. M. Bates, J. R. Smith, T. Janowski, and G. S. Tschumper, J. Chem. Phys. **135**, 044123 (2011).
[76] E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **3**, 1342 (2007).
[77] H. R. Leverentz and D. G. Truhlar, J. Chem. Theory Comput. **5**, 1573 (2009).
[78] R. Bukowski, W. Cencek, P. Jankowski, B. Jeziorski, M. Jeziorska, S. A. Kucharski, V. F. Lotrich, A. J. Misquitta, R. Moszynski, K. Patkowski, R. Podeszwa, S. Rybak, K. Szalewicz, H. L. Williams, R. J. Wheatley, P. E. S. Wormer, and P. S. Żuchowski, *SAPT2008: An Ab Initio Program for Many-Body Symmetry-Adapted Perturbation Theory Calculations of Intermolecular Interaction Energies* (University of Delaware, 2008).
[79] V. Saunders and M. Guest, *ATMOL Program Package* (SERC Daresbury Laboratory, Daresbury, UK).
[80] F. Neese, WIREs Comput. Mol. Sci. **2**, 73 (2012).
[81] CFOUR, a quantum chemical program package written by J. F. Stanton, J. Gauss, M. E. Harding, and P. G. Szalay with contributions from A. A. Auer, R. J. Bartlett, U. Benedikt, C. Berger, D. E. Bernholdt, Y. J. Bomble, O. Christiansen, M. Heckert, O. Heun, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, K. Klein, W. J. Lauderdale, D. A. Matthews, T. Metzroth, D. P. O'Neill, D. R. Price, E. Prochnow, K. Ruud, F. Schiffmann, S. Stopkowicz, A. Tajti, J. Vázquez, F. Wang, J. D. Watts and the integral packages MOLECULE (J. Almlf and P. R. Taylor), PROPS (P. R. Taylor), ABACUS (T. Helgaker, H. J. A. Jensen, P. Jørgensen, and J. Olsen), and ECP routines by A. V. Mitin and C. van Wllen. For the current version, see http://www.cfour.de.
[82] M. E. Harding, T. Metzroth, and J. Gauss, J. Chem. Theory Comput. **4**, 64 (2008).
[83] Y. Shao, L. Fusti-Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio, Jr., R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock III, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer III, J. Kong, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon, Phys. Chem. Chem. Phys. **8**, 3172 (2006).
[84] A. I. Krylov and P. M. W. Gill, WIREs Comput. Mol. Sci. **3**, 317 (2013).
[85] S. F. Boys and F. Bernardi, Mol. Phys. **19**, 553 (1970).
[86] W. Xie and J. Gao, J. Chem. Theory Comput. **3**, 1890 (2007).
[87] J. Han, D. G. Truhlar, and J. Gao, Theor. Chem. Acc. **131**, 1161 (2012).
[88] R. Moszynski, B. Jeziorski, S. Rybak, K. Szalewicz, and H. L. Williams, J. Chem. Phys. **100**, 5080 (1994).
[89] T. Korona, R. Moszynski, and B. Jeziorski, Mol. Phys. **100**, 1723 (2002).
[90] A. J. Misquitta and K. Szalewicz, J. Chem. Phys. **122**, 214109 (2005).
[91] K. Kitaura and K. Morokuma, Int. J. Quantum Chem. **10**, 325 (1976).
[92] W. Gao, H. Fen, X. Xuan, and L. Chen, J. Mol. Model. **18**, 4577 (2012).
[93] J. Řezác, K. E. Riley, and P. Hobza, J. Chem. Theory Comput. **7**, 3466 (2011).
[94] J. M. Herbert and M. Head-Gordon, J. Am. Chem. Soc. **128**, 13932 (2006).
[95] D. M. Bates and G. S. Tschumper, J. Phys. Chem. A **113**, 3555 (2009).
[96] U. Góra, R. Podeszwa, W. Cencek, and K. Szalewicz, J. Chem. Phys. **135**, 224102 (2011).
[97] P. Matczak, J. Phys. Chem. A **116**, 8731 (2012).
[98] R. M. Richard and J. M. Herbert, J. Chem. Theory Comput. **9**, 1408 (2013).
[99] A. Heßelmann and G. Jansen, Chem. Phys. Lett. **357**, 464 (2002).
[100] G. J. O. Beran, J. Chem. Phys. **130**, 164115 (2009).
[101] N. Mardirossian, D. S. Lambrecht, L. McCaslin, S. S. Xantheas, and M. Head-Gordon, J. Chem. Theory Comput. **9**, 1368 (2013).

Erratum: "An improved treatment of empirical dispersion and a many-body energy decomposition scheme for the explicit polarization plus symmetry-adapted perturbation theory (XSAPT) method" [J. Chem. Phys. 139, 034107 (2013)]

Ka Un Lao and John M. Herbert[a)]
Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, USA

(Received 18 February 2014; accepted 13 March 2014; published online 21 March 2014)

[http://dx.doi.org/10.1063/1.4869543]

In a recent paper,[1] we used high-level symmetry-adapted perturbation theory (SAPT) results for the S22 data set, which were obtained from Ref. 2, to benchmark a dispersion-corrected version of the "extended SAPT" method[3] that we called XSAPT+D2. Since then, we have discovered that the data in the supplementary material to Ref. 2 were computed at the SAPT2+/aug-cc-pVDZ level but were mislabeled as SAPT2+(3)/aug-cc-pVTZ. We have since made the SAPT2+(3)/aug-cc-pVTZ results available in the supplementary material that accompanies Ref. 4. Since we expect these new benchmarks to be superior in quality, it is useful to discuss how the XSAPT+D2 method performs against these SAPT2+(3)/aug-cc-pVTZ benchmarks.

For comparison to SAPT2+(3)/aug-cc-pVTZ benchmarks, Table II in Ref. 1 should be replaced by Table II shown below. As compared to the results reported in Ref. 1, the XSAPT+D2 errors for non-dispersion components of the energy improve by small but noticeable amounts. For example, the mean unsigned errors (MUEs) for the electrostatic, exchange, and induction components obtained with the XSAPT(KS)+D2 (AC) method[1] decrease to 0.20, 0.53, and 0.17 kcal/mol, respectively, whereas we previously reported 0.36, 0.72, and 0.23 kcal/mol.

TABLE II. Mean unsigned errors (MUEs), in kcal/mol, and percent errors (in parentheses), for individual energy components of the S22 data set, with respect to benchmarks computed at the SAPT2+(3)/aug-cc-pVTZ level.[4] All calculations were performed at S22 geometries.

	Energy Components			
Method	Electrostatic	Exchange	Induction	Dispersion
XSAPT(HF)/aDZ′	0.71 (9.25)	2.49 (16.22)	2.02 (62.54)	1.42 (24.17)
XSAPT(HF)/TZVPP	0.28 (5.71)	1.83 (10.94)	1.80 (59.50)	0.76 (11.42)
XSAPT(KS)+D/aDZ′	0.55 (11.21)	3.00 (22.90)	1.96 (60.98)	1.55 (20.72)
XSAPT(KS)+D[a]	0.16 (3.17)	0.49 (3.99)	0.19 (13.53)	0.68 (8.98)
XSAPT(KS)+D (AC)[a]	0.17 (3.19)	0.64 (5.21)	0.15 (10.56)	0.75 (9.60)
XSAPT(KS)+D2[a]	0.20 (3.14)	0.48 (4.78)	0.17 (11.29)	0.39 (5.70)
XSAPT(KS)+D2 (AC)[a]	0.20 (3.00)	0.53 (4.91)	0.17 (10.36)	0.39 (5.70)
EFP[b]	1.77 (32.66)	2.07 (14.87)	1.81 (51.53)	0.95 (14.20)

[a]Using the haTZVPP basis and the $\delta E_{\text{int}}^{\text{HF}}$ correction.
[b]EFP data are from Ref. 2.

In Ref. 5, electrostatic energies for the S22 data set were computed via supermolecular energy decomposition analysis at the BLYP-D3/TZ2P level. The MUE with respect to the old SAPT2+/aug-cc-pVDZ benchmarks was 0.32 kcal/mol,[1] but increases to 0.48 kcal/mol in comparison to the new SAPT2+(3)/aug-cc-pVTZ benchmarks. In comparison, the MUE for XSAPT *decreases* from 0.36 to 0.20 kcal/mol, hence XSAPT+D2 (AC) outperforms traditional energy decomposition analysis, at least for the electrostatic term.

In conclusion, when the benchmark energy components are computed at the SAPT2+(3)/aug-cc-pVTZ level, as we claimed they were in our original work,[1] the XSAPT-based energy decomposition analysis looks like an even more attractive way to understand the nature of intermolecular interactions in non-covalent clusters.

We thank Professor David Sherrill for assistance in resolving these issues.

[1]K. U. Lao and J. M. Herbert, J. Chem. Phys. **139**, 034107 (2013).
[2]J. C. Flick, D. Kosenkov, E. G. Hohenstein, C. D. Sherrill, and L. V. Slipchenko, J. Chem. Theory Comput. **8**, 2835 (2012).
[3]K. U. Lao and J. M. Herbert, J. Phys. Chem. Lett. **3**, 3241 (2012).
[4]K. U. Lao and J. M. Herbert, J. Chem. Phys. **140**, 044108 (2014).
[5]W. Gao, H. Fen, X. Xuan, and L. Chen, J. Mol. Model. **18**, 4577 (2012).

a)herbert@chemistry.ohio-state.edu

