View Online / Journal Homepage / Table of Contents for this issue
Dynamic Article Links

Cite this: *Phys. Chem. Chem. Phys.*, 2012, **14**, 7679–7699

www.rsc.org/pccp

PAPER

Rapid computation of intermolecular interactions in molecular and ionic clusters: self-consistent polarization plus symmetry-adapted perturbation theory

John M. Herbert,* Leif D. Jacobson,† Ka Un Lao and Mary A. Rohrdanz‡

Received 20th December 2011, Accepted 13th March 2012
DOI: 10.1039/c2cp24060b

A method that we have recently introduced for rapid computation of intermolecular interaction energies is reformulated and subjected to further tests. The method employs monomer-based self-consistent field calculations with an electrostatic embedding designed to capture many-body polarization (the "XPol" procedure), augmented by pairwise symmetry-adapted perturbation theory (SAPT) to capture dispersion and exchange interactions along with any remaining induction effects. A rigorous derivation of the XPol + SAPT methodology is presented here, which demonstrates that the method is systematically improvable, and moreover introduces some additional intermolecular interactions as compared to the more heuristic derivation that was presented previously. Applications to various non-covalent complexes and clusters are presented, including geometry optimizations and one-dimensional potential energy scans. The performance of the XPol + SAPT methodology in its present form (based on second-order intermolecular perturbation theory and neglecting intramolecular electron correlation) is *qualitatively* acceptable across a wide variety of systems—and quantitatively quite good in certain cases—but the quality of the results is rather sensitive to the choice of one-particle basis set. Basis sets that work well for dispersion-bound systems offer less-than-optimal performance for clusters dominated by induction and electrostatic interactions, and *vice versa*. A compromise basis set is identified that affords good results for both induction and dispersion interactions, although this favorable performance ultimately relies on error cancellation, as in traditional low-order SAPT. Suggestions for future improvements to the methodology are discussed.

I. Introduction

Two of the most challenging problems at the frontier of contemporary electronic structure theory are accurate calculations of non-covalent interactions, and development of reduced-scaling algorithms applicable to large systems. To some extent, these two problems are antithetical, since *accurate* calculation of non-covalent interactions (especially when the interaction is dominated by dispersion) typically requires correlated, post-Hartree–Fock methods whose computational scaling with respect to system size precludes the application to large systems.[1,2] Nevertheless, to bring quantum chemistry into the condensed phase—that is, to perform accurate, first-principles, all-electron Born–Oppenheimer molecular dynamics simulations in liquids and solids under periodic boundary conditions—both of these difficult problems must be surmounted.

Of course, there are several pragmatic approaches to condensed-phase quantum chemistry that offer immediate and (in many cases) useful results. Mixed quantum mechanics/molecular mechanics (QM/MM) simulations, in conjunction with semi-empirical electronic structure methods that utilize a simplified electronic Hamiltonian that depends on adjustable parameters (allowing for relatively large QM regions and/or relatively long simulation time scales) are the workhorse methods in condensed-phase and macromolecular electronic structure theory, and will continue in that role for the foreseeable future.[3,4] Here, however, we wish to consider the possibility of performing all-electron calculations (no MM region!) in a condensed-phase system, using methods based on *ab initio* quantum chemistry.

If an all-electron and more-or-less *ab initio* approach is desired, then a variety of codes already exist for performing plane-wave density functional theory (DFT) calculations in periodic simulation cells. However, there are several reasons that one might want to look beyond such an approach. First, the use of plane-wave basis functions dramatically increases the cost of incorporating Hartree–Fock (HF) exchange into DFT calculations. Essentially, this means that most of the widely-used functionals that are known to be accurate for molecular properties are prohibitively expensive for condensed-phase calculations. Moreover, most DFT approaches fail to incorporate dispersion interactions properly, unless specifically

Department of Chemistry, The Ohio State University, Columbus, OH 43210, USA. E-mail: herbert@chemistry.ohio-state.edu
† Present address: Dept. of Chemistry, Yale University, New Haven, CT, USA.
‡ Present address: Dept. of Chemistry, Rice University, Houston, TX, USA.

parameterized to do so.[5–7] Finally, in the absence of further approximations, such calculations still scale as $\mathcal{O}(N^3)$ with respect to the system size, N, so that these methods are extremely expensive and are limited to short simulation time scales (tens of picoseconds) and small unit cells (*e.g.*, 32–64 water molecules). This precludes calculation of transport properties as well as use of specialized sampling algorithms that are needed in order to simulate rare events.[8]

What might a "wish list" for a condensed-phase, *ab initio* quantum chemistry platform look like? Ideally, such a method should:

(1) be free of adjustable parameters,

(2) exhibit a computational cost that scales linearly with system size, and

(3) be amenable to systematic approximations of increasing accuracy.

Moreover, such a method should also be affordable and accurate enough to do something useful! Our expeditionary inroads towards these goals are the subject of this work. Although we are still very far from our long-term goal of accurate and affordable all-electron *ab initio* calculations in liquids, we have developed a general platform that, with a few caveats (to be discussed below), satisfies all of the aforementioned criteria and may ultimately provide a theoretical foundation upon which to construct efficient *ab initio* simulation methods for condensed phases. At present, it appears to be a useful method for performing all-electron calculations in large molecular and ionic clusters.

The methodology introduced here is a reformulation and slight generalization of a method designed for fast computation of intermolecular interactions that two of us introduced recently.[9] This method combines one aspect of the *explicit polarization potential* (XPol) idea of Gao and co-workers,[10–13] which we use to describe many-body polarization effects, with the well-established methods of symmetry-adapted perturbation theory (SAPT),[14,15] which we use to describe other contributions to the intermolecular interactions, primarily dispersion and exchange repulsion.

Consider a supersystem composed of interacting monomers. In the original XPol method,[10,11] the intramolecular electronic structure is described at the self-consistent field (SCF) level, whereas many-body intermolecular polarization (induction) effects are described *via* an embedding scheme, wherein the monomer SCF equations are solved subject to the electrostatic potential due to all of the other monomer wave functions. This electrostatic potential could be the exact one, in principle, but is typically approximated by collapsing the other monomer electron densities onto point charges. The computational cost of the "dual SCF" procedure[9,10] required to converge the monomer SCF equations scales linearly with respect to the number of monomers, by virtue of the *ansatz* that each monomer's molecular orbitals are expanded in terms of only those atomic orbitals that are centered on the same monomer. Our method utilizes the XPol procedure to define monomer wave functions that are polarized in a manner that reflects their environment.

Intermolecular polarization is known to be an inherently many-body (not pairwise-additive) phenomenon, and our hope is that the dual-SCF XPol procedure incorporates the most important many-body effects, such that what remains of the intermolecular interactions can be accurately described in a pairwise-additive fashion using low-order intermolecular perturbation theory. At the same time, the XPol procedure does not allow exchange of electrons between monomers and thus does not capture exchange repulsion. In the original XPol method of Gao and co-workers,[10–13] the dual SCF is supplemented with empirical (Lennard-Jones) intermolecular potentials, which also serve to approximate intermolecular dispersion interactions. In our work,[9] we sought to eliminate these empirical terms in favor of an *ab initio* approach based on pairwise-additive SAPT (with a modified perturbation as compared to traditional SAPT, in order to avoid double-counting). The SAPT calculations build in dispersion and exchange-repulsion interactions, along with corrections to the XPol treatment of induction. We call this method XPol + SAPT, or XPS for short.

Originally, we introduced this XPol + SAPT methodology in a rather *ad hoc* way. Here, we reformulate this method in a more rigorous manner based on perturbative approximations starting from the supersystem Hamiltonian. Certain three-body induction corrections, which we did not anticipate in our intuitive formulation of the method, arise naturally from this new formulation, and these are shown to be qualitatively important in certain cases. Moreover, the new formulation provides a theoretical framework in which systematic improvements can be made. The developments discussed here are equally valid with or without charge embedding. The method is systematically improvable and is ideally suited for parallelization. Unlike numerous other fragmentation methods based upon the many-body expansion,[16,17] a class that includes the so-called fragment molecular orbital method,[16,18] our XPol + SAPT approach ultimately rests upon a well-defined supersystem wave function, so that in principle this approach provides a route to compute properties other than the energy. This is a potentially significant advantage relative to alternative fragmentation methods that are fundamentally combinatorial (rather than quantum-mechanical) in nature.[17]

II. Theory

A. Notation

In what follows we will use $A,B,C,\ldots$ to label monomers, of which there are N in total. The indices $i,j,k,\ldots$ will be used to label electrons and $a \in A$, $b \in B$, *etc.*, will label occupied molecular orbitals (MOs) belonging to fragments A and B, respectively. Greek indices ($\mu,\nu,\ldots$) will label atomic orbitals (AOs) and $I,J,\ldots$ will label nuclei. This will be sufficient notation to introduce the XPol method in §II B. The SAPT corrections in §II C require virtual MOs, and we will label these as $r \in A$, $s \in B$, $t \in C$, and $u \in D$. Only the closed-shell, spin-restricted case will be considered herein. All equations are written in atomic units.

B. Many-body polarization: XPol

The XPol method is based upon a direct-product *ansatz* for the wave function:

$$|\Psi_{\text{XPol}}\rangle = \prod_{A=1}^{N} |\Psi_A\rangle, \tag{1}$$

where $|\Psi_A\rangle$ represents the wave function for an individual fragment (monomer). The XPol energy expression is consistent with this *ansatz*, and is defined to be that resulting from the dual-SCF calculation described in §I, wherein the monomers interact with one another only *via* their electrostatic potentials. For closed-shell monomers described at the level of Hartree–Fock theory, this energy expression is[9]

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

The term in square brackets is the ordinary HF energy expression for fragment A (see ref. 9). The XPol method employs *absolutely localized* MOs (ALMOs),[19] meaning that the MO $\mathbf{c}_a$ is expanded using only AOs centered on monomer A. This partitioning of the basis set leads to a significant computational savings and a method whose cost is $\mathcal{O}(N)$ with respect to the number of monomers. Furthermore, basis-set superposition error (BSSE) is excluded by *ansatz*. Intermonomer charge transfer will also be absent in compact basis sets.

The embedding potential in eqn (2), E_{embed}, represents the electrostatic interactions between the monomers, which can be further decomposed into nuclear and electronic parts, $E_{\mathrm{embed}} = E_{\mathrm{embed}}^{\mathrm{nucl}} + E_{\mathrm{embed}}^{\mathrm{elec}}$. These two components could, in principle, be expressed in terms of the monomer densities:

$$E_{\mathrm{embed}}^{\mathrm{elec}} = \frac{1}{2}\sum_{A}\sum_{B\neq A}\int\frac{\rho_A(\vec{r}_1)\rho_B(\vec{r}_2)}{|\vec{r}_1-\vec{r}_2|}d\vec{r}_1 d\vec{r}_2 \tag{3a}$$

$$E_{\mathrm{embed}}^{\mathrm{nucl}} = -\sum_{A}\sum_{B\neq A}\sum_{J\in B}\int\frac{Z_J\rho_A(\vec{r})}{|\vec{r}-\vec{R}_J|}d\vec{r}. \tag{3b}$$

This "density embedding" provides the exact Coulomb interaction between monomers, and may be important for describing short-range induction interactions between strongly interacting monomers. In the present work, however, we rely upon an approximate "point-charge embedding" instead, in which the monomer densities ρ_B (for $B \neq A$) are collapsed onto a set of point charges, $\{q_J\}$, for the purpose of computing the embedding potential for monomer A. Thus, we express the embedding potential as

$$E_{\mathrm{embed}}^{\mathrm{elec}} = -\sum_{A}\sum_{B\neq A}\sum_{J\in B}\sum_{a} q_J \mathbf{c}_a^{\dagger}\mathbf{I}_J\mathbf{c}_a \tag{4a}$$

$$E_{\mathrm{embed}}^{\mathrm{nucl}} = \tfrac{1}{2}\sum_{A}\sum_{B\neq A}\sum_{I\in A}\sum_{J\in B} q_J L_{IJ}. \tag{4b}$$

Here, $L_{IJ} = Z_I|\vec{R}_I - \vec{R}_J|^{-1}$ and $\mathbf{I}_J$ is the matrix of one-electron integrals

$$(\mathbf{I}_J)_{\mu\nu} = \int\frac{\chi_\mu(\vec{r})\chi_\nu(\vec{r})}{|\vec{r}-\vec{R}_J|}d\vec{r}, \tag{5}$$

where χ_ν and χ_μ are AOs centered on A. In the original XPol method of Gao and co-workers,[11] the point charges q_J in eqn (4) were taken to be Mulliken charges, although other charge schemes are possible,[9] as discussed below. In some sense, the choice of embedding is systematically improvable in that one could imagine a procedure that switches seamlessly from density–density interactions at short range to a multipole (and, ultimately, point-charge) description of more distant interactions, as in the continuous fast multipole method.[20] We hope to develop a procedure along these lines in future work.

By requiring that the energy expression in eqn (2) be stationary with respect to variations in the MO coefficients, subject to the constraint that MOs *within* a fragment are orthonormal, one arrives at the XPol SCF equations:[11]

$$\mathbf{F}^A\mathbf{C}^A = \mathbf{S}^A\mathbf{C}^A\boldsymbol{\varepsilon}^A. \tag{6}$$

The one-electron energy levels are obtained by diagonalizing $\boldsymbol{\varepsilon}^A = (\mathbf{C}^A)^{\dagger}\mathbf{F}^A\mathbf{C}^A$. In the AO basis, the Fock matrix for monomer A has matrix elements

$$F_{\mu\nu}^{A} = f_{\mu\nu}^{A} - \tfrac{1}{2}\sum_{J\notin A}(\mathbf{I}_J)_{\mu\nu}q_J + \sum_{J\in A}(\boldsymbol{\Lambda}_J)_{\mu\nu}M_J \tag{7}$$

where $\mathbf{f}^A = 2\mathbf{h}^A + 2\mathbf{J}^A - \mathbf{K}^A$ is the Fock matrix for fragment A in isolation. The other two terms in $F_{\mu\nu}^A$ arise from the embedding potential, and in particular the last term accounts for the variation of the embedding charges with respect to the MOs, which ensures that the XPol method is variational within the ALMO *ansatz*. This variation is expressed in terms of energy derivatives

$$M_J = \frac{\partial E_{\mathrm{embed}}}{\partial q_J} \tag{8}$$

that are trivial to evaluate using eqn (4), and also charge derivatives

$$(\boldsymbol{\Lambda}_J)_{\mu\nu} = \frac{\partial q_J}{\partial P_{\mu\nu}} \tag{9}$$

that can be evaluated once one has specified how the embedding charges will be obtained from the monomer densities.

In previous work,[9] we investigated the Mulliken and Löwdin prescriptions but found them to be ill-behaved in some cases, leading to convergence difficulties in the dual-SCF procedure. Charges derived from the electrostatic potential ("CHELPG" charges[21–23]) were found to be stable and well-behaved in all cases examined so far,[9] and this is our preferred choice for the embedding charges. The derivatives in eqn (9) are worked out in the Appendix for the case of CHELPG embedding charges, where we also discuss a smoothing procedure that is employed to ensure that these charges, whose determination requires the evaluation of the electrostatic potential on a real-space grid, are smooth functions of the nuclear coordinates. In some of the calculations in §III, however, we do use Löwdin charges because they are much less expensive to compute. Mulliken charges are unstable in extended basis sets, and are not used here.

C. XPol + SAPT

1. Motivation and outline. The goal of our approach is to utilize the dual-SCF XPol procedure described above to provide an approximate but self-consistent description of many-body induction. In clusters of polar molecules, the polarization effect is non-negligible and is expected to constitute the leading-order many-body effect. In $(H_2O)_6$, for example, energy decomposition analysis suggests that many-body (non-pairwise-additive) polarization contributes anywhere from 9–13 kcal mol^{-1} to the binding energy, depending upon the particular isomer, but is the only component of the interaction energy that exhibits significant non-additivity.[24] Although some exchange non-additivity can be seen in SAPT calculations on $HO^-(H_2O)_n$, the non-additivity is dominated by induction.[25] Furthermore, in the context of many-body expansion methods, it is found that the MP2 energy for a cluster of polar molecules (water, ammonia, formaldehyde, formamide, *etc.*) can be accurately approximated using either a supersystem Hartree–Fock calculation[26,27] or else a polarizable force field calculation[28,29] to incorporate many-body polarization, followed by a pairwise-additive approximation to the correlation energy *via* dimer MP2 calculations.

Together, these observations indicate that three-body contributions to dispersion and exchange-repulsion are mostly negligible in polar systems. In non-polar systems, the leading many-body effect will *not* be induction, yet three-body effects are nevertheless found to be small in rare-gas trimers near the equilibrium geometries (*e.g.*, ~1% of the total binding energy in He_3 and Ne_3 and ~4% in Ar_3).[30] In small benzene clusters, three-body effects are mostly negligible,[31] and are estimated to be negligible as well in crystalline benzene.[32]

Our approach is therefore inspired by the hypothesis that if polarization is described accurately enough at the XPol level, then it may be reasonable to approximate the remaining intermolecular interactions in a pairwise-additive fashion. In what follows, we describe the application of SAPT to compute these pairwise-additive corrections to the XPol energy and wave function. At a given order in perturbation theory, the structure of the intermolecular SAPT energy corrections consists of certain "direct" terms that come from Rayleigh-Schrödinger perturbation theory (RSPT), along with a corresponding exchange correction for each direct term, which arises due to the antisymmetry requirement. The RSPT corrections, which include intramonomer electron correlation, are described in §II C2, and then in §II C3 we introduce symmetry adaptation to deal with intermonomer exchange interactions.

2. Rayleigh–Schrödinger corrections. We begin by writing the Hamiltonian for an arbitrary number of interacting monomers as the sum of Fock operators ($\hat{f}_A$), Møller–Plesset fluctuation operators ($\hat{W}_A$), and interaction operators ($\hat{V}_{AB}$):

$$\hat{H} = \sum_A \hat{f}_A + \sum_A \xi_A \hat{W}_A + \sum_A \sum_{B>A} \zeta_{AB} \hat{V}_{AB}. \tag{10}$$

The quantities ξ_A and ζ_{AB} are parameters to keep track of the order in perturbation theory. As in traditional SAPT,[14] it is convenient to write the interaction operator as

$$\hat{V}_{AB} = \sum_{i\in A} \sum_{j\in B} \hat{v}_{AB}(ij) \tag{11}$$

with

$$\hat{v}_{AB}(ij) = \frac{1}{|\vec{r}_i - \vec{r}_j|} + \frac{\hat{v}_A(j)}{N_A} + \frac{\hat{v}_B(i)}{N_B} + \frac{V_0}{N_A N_B}. \tag{12}$$

Here, $i \in A$ and $j \in B$ index electrons in monomers A and B, N_A and N_B denote the number of electrons in each monomer, and V_0 represents the nuclear interaction energy between A and B. The operator

$$\hat{v}_A(j) = -\sum_{I\in A} \frac{Z_I}{|\vec{r}_j - \vec{R}_I|} \tag{13}$$

represents the interaction of electron j with the nuclei in monomer A.

We take the zeroth-order wave function to be the direct product of XPol monomer wave functions, $|\Psi_0\rangle = |\Psi_{\text{XPol}}\rangle$ [see eqn (1)]. Each monomer wave function $|\Psi_A\rangle$ consists of a single determinant of MOs and is an eigenfunction of $\hat{f}_A$. As such, it makes sense to take

$$\hat{H}_0 = \sum_{A=1}^{N} \hat{f}_A \tag{14}$$

as the zeroth-order Hamiltonian. The zeroth-order energy is then equal to the sum of the occupied eigenvalues of each $\hat{f}_A$.

Given this notation, the time-independent Schrödinger equation can be written

$$\left(\hat{H}_0 + \sum_A \xi_A \hat{W}_A + \sum_A \sum_{B>A} \zeta_{AB} \hat{V}_{AB}\right)|\Psi\rangle = E|\Psi\rangle. \tag{15}$$

Taking the intermediate normalization,[33] one can left-multiply by $\langle\Psi_0|$ to obtain an expression for the interaction energy, $E_{\text{int}} = E - E_0$:

$$E_{\text{int}} = \sum_A \xi_A \langle\Psi_0|\hat{W}_A|\Psi\rangle + \sum_A \sum_{B>A} \zeta_{AB} \langle\Psi_0|\hat{V}_{AB}|\Psi\rangle. \tag{16}$$

We consider E_{int}, along with the exact wave function $|\Psi\rangle$, to be functions of all N parameters ξ_A and all $N(N-1)/2$ parameters ζ_{AB}. Next, we expand the interaction energy and wave function in terms of these variables:

$$E_{\text{int}} = \sum_{\substack{m,\ldots,p=0\\ q,\ldots,t=0}}^{\infty\;\downarrow} \xi_1^m \cdots \xi_N^p \zeta_{1,2}^q \zeta_{1,3}^r \cdots \zeta_{N-1,N}^t E^{(m,\ldots,p:q,r,\ldots,t)} \tag{17}$$

$$|\Psi\rangle = \sum_{\substack{m,\ldots,p=0\\ q,\ldots,t=0}}^{\infty} \xi_1^m \cdots \xi_N^p \zeta_{1,2}^q \zeta_{1,3}^r \cdots \zeta_{N-1,N}^t |\Psi^{(m,\ldots,p:q,r,\ldots,t)}\rangle. \tag{18}$$

The superscript "↓" on the summation in eqn (17) indicates that the first term in the sum (where all indices are zero) is excluded. Note that, according to our $(m,\ldots,p:q,\ldots,t)$ indexing convention, the N indices to the left of the colon correspond to corrections for intramonomer electron correlation whereas the $N(N-1)/2$ indices to the right of the colon correspond to corrections for intermolecular perturbations. Inserting eqns (17) and (18) into eqn (16) and collecting matching powers of the perturbation parameters affords

$$\begin{aligned} &E^{(m,\ldots,n,\ldots,p:q,\ldots,r,\ldots,s)} \\ &\quad =\sum_{A}\langle\Psi_0|\hat{W}_A|\Psi^{(m,\ldots,(n-1)_A,\ldots,p:q,\ldots,r,\ldots,s)}\rangle \\ &\qquad +\sum_{A}\sum_{B>A}\langle\Psi_0|\hat{V}_{AB}|\Psi^{(m,\ldots,n,\ldots,p:q,\ldots,(r-1)_{AB},\ldots,s)}\rangle. \end{aligned} \tag{19}$$

The superscript $(n - 1)_A$ in the wave function correction appearing in the first term indicates that it is the Ath index that has been decremented. Similarly, the ABth dimer index has been decremented in the second term. Eqn (19) is valid only when all indices are greater than or equal to zero.

Eqn (19) is a completely general expression for the intra- and intermolecular interactions that appear in Rayleigh–Schrödinger perturbation theory, but in this work we will only investigate a small number of these terms. The first class of terms that are of interest here are the first-order intramolecular corrections, in which one index to the left of the colon is unity and all other indices are zero. For brevity, we denote these terms as $E^{[1_A;0]} \equiv E^{(0,\ldots,0,1_A,0,\ldots,0,:0,\ldots,0)}$, and they are obtained from simple expectation values,

$$E^{[1_A;0]} = \langle\Psi_0|\hat{W}_A|\Psi_0\rangle, \tag{20}$$

where $|\Psi_0\rangle \equiv |\Psi^{(0,\ldots,0:0,\ldots,0)}\rangle$.

The other terms that we will consider are intermolecular in nature with all indices to the left of the colon equal to zero, which amounts to a neglect of intramolecular electron correlation. We consider all intermolecular terms through second order, *i.e.*, those in which the indices to the right of the colon sum to no more than two. (In the context of traditional SAPT calculations, this approximation is often termed "SAPT0".[2]) We consider several different types of intermolecular corrections, including the first-order corrections

$$E^{[0;1_{AB}]} = \langle\Psi_0|\hat{V}_{AB}|\Psi_0\rangle \tag{21}$$

along with corrections that are second-order in a given monomer,

$$E^{[0;2_{AB}]} = \langle\Psi_0|\hat{V}_{AB}|\Psi^{[0;1_{AB}]}\rangle \tag{22}$$

Here, $|\Psi^{[0;1_{AB}]}\rangle \equiv |\Psi^{(0,\ldots,0;0,\ldots,0,1_{AB},0,\ldots,0)}\rangle$. Lastly, we consider corrections that are first-order in each of two different intermolecular perturbations. We denote these as $E^{[0;1_{AB},1_{CD}]} = E^{(0,\ldots,0;0,\ldots,0,1_{AB},0,\ldots,0,1_{CD},0,\ldots,0)}$, and they are given by

$$E^{[0;1_{AB},1_{CD}]} = \langle\Psi_0|\hat{V}_{AB}|\Psi^{[0;1_{CD}]}\rangle + \langle\Psi_0|\hat{V}_{CD}|\Psi^{[0;1_{AB}]}\rangle. \tag{23}$$

As described below, eqn (21) gives rise to Coulomb interactions between monomers, eqn (22) gives rise to induction and dispersion interactions, and eqn (23) can be interpreted in terms of induction coupling between two pairs of monomers. These induction-coupling terms arise naturally when we include all terms through second order in the intermolecular perturbations, but they were absent in our original, intuitive formulation of XPol + SAPT.[9]

To complete the description of the many-body RSPT, we need to work out the wave function corrections. Only first-order wave function corrections are required for the second-order energy corrections considered here, but in the interest of providing a platform for future development we will present the corrections in a general form. They are obtained by subtracting E_0 from both sides of eqn (15), then substituting the expansions in eqns (17) and (18), and collecting like terms. The result is

$$(\hat{H}_0 - E_0)|\Psi^{(m,\ldots,p:q,\ldots,s)}\rangle \tag{24}$$

$$\begin{aligned} &+\sum_{A}\hat{W}_A|\Psi^{(m,\ldots,(n-1)_A,\ldots,p:q,\ldots,s)}\rangle \\ &+\sum_{A}\sum_{B>A}\hat{V}_{AB}|\Psi^{(m,\ldots,p:q,\ldots,r_{AB},\ldots,s)}\rangle \\ &= \sum_{m',n',\ldots=0}^{m,n,\ldots}{}^{\downarrow} E^{(m',\ldots,p':q',\ldots,r',\ldots,s')} \\ &\quad \times |\Psi^{(m-m',\ldots,p-p':q-q',\ldots,s-s')}\rangle. \end{aligned} \tag{25}$$

This expression can be simplified by explicitly writing out the final term in the sum on the right side of the equality, and then substituting eqn (19). The result is

$$\begin{aligned} |\Psi^{(m,\ldots,p:q,\ldots,s)}\rangle = &-\sum_{A}\hat{R}_0\hat{W}_A|\Psi^{(m,\ldots,(n-1)_A,\ldots,p:q,\ldots,s)}\rangle \\ &-\sum_{A<B}\hat{R}_0\hat{V}_{AB}|\Psi^{(m,\ldots,p:q,\ldots,(r-1)_{AB},\ldots,s)}\rangle \\ &+\sum_{m',n',\ldots=0}^{m,n,\ldots}{}^{\downarrow\uparrow} E^{(m',\ldots,p':q',\ldots,s')}\hat{R}_0 \\ &\times |\Psi^{(m-m',\ldots,p-p':q-q',\ldots,s-s')}\rangle. \end{aligned} \tag{26}$$

The "↓↑"on the final summation in this equation indicates that this sum does not include either the first term (where $m = n = \cdots = 0$) or the last term (where $m = m'$, $n = n',\ldots$). Again, only terms in this expression where all indices are non-negative are included. The reduced resolvent,[15] $\hat{R}_0$, that appears in eqn (26) can be expressed as

$$\hat{R}_0 = \sum_{i\neq 0}\frac{|\Psi_0^i\rangle\langle\Psi_0^i|}{E_i - E_0}. \tag{27}$$

Here, $|\Psi_0^i\rangle$ is a zeroth-order excited state whose energy is E_i. In this work, the zeroth-order excited states are replacement determinants and we will only require single and double excitations, since we consider only first- and second-order corrections. These require only the first-order corrections to the wave function, each of which has the form

$$|\Psi^{[0;1_{AB}]}\rangle = -\hat{R}_0\hat{V}_{AB}|\Psi_0\rangle. \tag{28}$$

3. Symmetry adaptation and exchange. The expressions above account for electrostatic interactions, induction, and dispersion but neglect the antisymmetry requirement of the wave function, which is important at short range. We follow the excellent review of Jeziorski *et al.*[15] and apply symmetrized RSPT, which is also known as *weak symmetry forcing*. In this approach, we accept non-antisymmetric corrections to the direct product wave function but then project away the Pauli-forbidden components in the energy corrections.

This amounts to a modification of the interaction energy formula in eqn (16). The (anti)symmetry-adapted formula is

$$
\begin{aligned}
E_{\text{int}} \;=\; & \left[\sum_{A} \xi_A \langle \Psi_0 | \hat{W}_A \hat{\mathcal{A}} | \Psi \rangle + \sum_{A} \sum_{B>A} \zeta_{AB} \langle \Psi_0 | \hat{V}_{AB} \hat{\mathcal{A}} | \Psi \rangle \right] \\
& \times \langle \Psi_0 | \hat{\mathcal{A}} | \Psi \rangle^{-1},
\end{aligned}
\tag{29}
$$

where $\hat{\mathcal{A}}$ is the antisymmetrizer (see below). To obtain expressions for the order-by-order energy corrections, we once again substitute the energy and wave function expansions, eqns (17) and (18), to afford

$$
\begin{aligned}
E^{(m,\ldots,p:q,\ldots,s)} \;=\; & -\mathcal{N}_{\mathcal{A}}^{-1} \sum_{m',n',\ldots}^{m,n,\ldots}{}^{\downarrow\uparrow} E^{(m',\ldots,p':q',\ldots,s')} \\
& \times \langle \Psi_0 | \hat{\mathcal{A}} | \Psi^{(m-m',\ldots,p-p':q-q',\ldots,s-s')} \rangle \\
& + \mathcal{N}_{\mathcal{A}}^{-1} \sum_{A} \langle \Psi_0 | \hat{W}_A \hat{\mathcal{A}} | \Psi^{(m,\ldots,(n-1)_A,\ldots,p:q,\ldots,s)} \rangle \\
& + \mathcal{N}_{\mathcal{A}}^{-1} \sum_{A} \sum_{B>A} \langle \Psi_0 | \hat{V}_{AB} \hat{\mathcal{A}} | \Psi^{(m,\ldots,p:q,\ldots,r_{AB},\ldots,s)} \rangle
\end{aligned}
\tag{30}
$$

where

$$
\mathcal{N}_{\mathcal{A}} = \langle \Psi_0 | \hat{\mathcal{A}} | \Psi_0 \rangle. \tag{31}
$$

As above, only corrections up to second order are considered here. There are three types of corrections,

$$
E^{[0;1_{AB}]} = \mathcal{N}_{\mathcal{A}}^{-1} \langle \Psi_0 | \hat{V}_{AB} \hat{\mathcal{A}} | \Psi_0 \rangle, \tag{32}
$$

$$
\begin{aligned}
E^{[0;1_{AB},1_{CD}]} \;=\; & \mathcal{N}_{\mathcal{A}}^{-1} [\langle \Psi_0 | \hat{V}_{AB} - E^{[0;1_{AB}]}) \hat{\mathcal{A}} | \Psi^{[0;1_{CD}]} \rangle \\
& + \langle \Psi_0 | (\hat{V}_{CD} - E^{[0;1_{CD}]}) \hat{\mathcal{A}} | \Psi^{[0;1_{AB}]} \rangle].
\end{aligned}
\tag{33}
$$

and

$$
E^{[0;2_{AB}]} = \mathcal{N}_{\mathcal{A}}^{-1} [\langle \Psi_0 | \hat{V}_{AB} \hat{\mathcal{A}} | \Psi^{[0;1_{AB}]} \rangle - E^{[0;1_{AB}]} \langle \Psi_0 | \hat{\mathcal{A}} | \Psi^{[0;1_{AB}]} \rangle], \tag{34}
$$

which are analogous to the RSPT corrections in eqns (21)–(23)

To proceed towards practical formulas for these energy corrections, we write the antisymmetrizer as

$$
\hat{\mathcal{A}} = \frac{\prod_A N_A! \hat{\mathcal{A}}_A}{(\sum_A N_A)!} (1 + \hat{\mathcal{P}}^{(2)} + \hat{\mathcal{P}}'), \tag{35}
$$

where $\hat{\mathcal{A}}_A$ is the antisymmetrizer for fragment A, $\hat{\mathcal{P}}^{(2)}$ is the sum of all pairwise electron exchanges between monomer A and the other monomers, and the operator $\hat{\mathcal{P}}'$ consists of all higher-order exchanges, which we will neglect. (In the SAPT literature, this is known as the *single-exchange approximation*.[15]) Lotrich and Szalewicz[34,35] have considered the three-body exchange non-additivity in SAPT by including terms up to $\hat{\mathcal{P}}^{(4)}$ in the antisymmetrizer, but we are interested in low-order terms that can be evaluated efficiently. In any case, the effects of multiple exchanges are thought to small except possibly at intermolecular distances much shorter than the van der Waals separation.[34–38]

Our aim is to generalize the symmetrized RSPT expansion[15] to the case of an arbitrary number of interacting monomers. We have initially explored only low-order perturbation theory, in the interest of efficiency. Each of the corrections appearing in eqns (32)–(34) consists of an RSPT correction, arising from $\hat{V}_{AB}$ or $\hat{V}_{CD}$, along with an exchange correction arising from $\hat{\mathcal{A}}$:

$$
E^{(0,\ldots,0;q,\ldots,s)} = E_{\text{RSPT}}^{(0,\ldots,0;q,\ldots,s)} + E_{\text{exch}}^{(0,\ldots,0;q,\ldots,s)}. \tag{36}
$$

The RSPT corrections correspond to eqns (21)–(23) whereas eqns (32)–(34) will generate both the RSPT terms and the exchange terms.

By inserting eqn (35) into eqns (32)–(34) and keeping terms proportional to the square of the intermolecular overlap integral (corresponding to the aforementioned single-exchange approximation), one obtains the following expressions for the exchange corrections:

$$
E_{\text{exch}}^{[0;1_{AB}]} = \langle \hat{V}_{AB} \hat{\mathcal{P}}^{(2)} \rangle - \langle \hat{V}_{AB} \rangle \langle \hat{\mathcal{P}}^{(2)} \rangle, \tag{37}
$$

$$
\begin{aligned}
E_{\text{exch}}^{[0;1_{AB},1_{CD}]} \;=\; & \langle \Psi_0 | \hat{V}_{AB} \hat{\mathcal{P}}^{(2)} - \hat{V}_{AB} \langle \hat{\mathcal{P}}^{(2)} \rangle \\
& - \langle \hat{V}_{AB} \rangle \hat{\mathcal{P}}^{(2)} | \Psi^{[0;1_{CD}]} \rangle + \langle \Psi_0 | \hat{V}_{CD} \hat{\mathcal{P}}^{(2)} \\
& - \hat{V}_{CD} \langle \hat{\mathcal{P}}^{(2)} \rangle - \langle \hat{V}_{CD} \rangle \hat{\mathcal{P}}^{(2)} | \Psi^{[0;1_{AB}]} \rangle,
\end{aligned}
\tag{38}
$$

and

$$
E_{\text{exch}}^{[0;2_{AB}]} = \langle \Psi_0 | \hat{V}_{AB} \hat{\mathcal{P}}^{(2)} - \hat{V}_{AB} \langle \hat{\mathcal{P}}^{(2)} \rangle - \langle \hat{V}_{AB} \rangle \hat{\mathcal{P}}^{(2)} | \Psi^{[0;1_{AB}]} \rangle. \tag{39}
$$

(Except where indicated otherwise, angle brackets denote expectation values with respect to $|\Psi_0\rangle$). The corresponding RSPT corrections are relatively simple:

$$
E_{\text{RSPT}}^{[0;1_{AB}]} = \langle \hat{V}_{AB} \rangle, \tag{40}
$$

$$
E_{\text{RSPT}}^{[0;1_{AB},1_{CD}]} = \langle \Psi_0 | \hat{V}_{AB} | \Psi^{[0;1_{CD}]} \rangle + \langle \Psi_0 | \hat{V}_{CD} | \Psi^{[0;1_{AB}]} \rangle, \tag{41}
$$

and

$$
E_{\text{RSPT}}^{[0;2_{AB}]} = \langle \Psi_0 | \hat{V}_{AB} | \Psi^{[0;1_{AB}]} \rangle. \tag{42}
$$

In order to evaluate the second-order corrections, we need to specify the wave function corrections to first order, as implied by eqn (28). Taking the zeroth-order excited states to be single and double replacement determinants, we obtain

$$
\begin{aligned}
|\Psi^{[0;1_{AB}]}\rangle \;=\; & \sum_{r,a\in A} (t_{AB})_r^a |\Psi_1\rangle \cdots |\Psi_A\rangle_a^r \cdots |\Psi_B\rangle \cdots |\Psi_N\rangle \\
& + \sum_{s,b\in B} (t_{AB})_s^b |\Psi_1\rangle \cdots |\Psi_A\rangle \cdots |\Psi_B\rangle_b^s \cdots |\Psi_N\rangle \\
& + \sum_{r,a\in A} \sum_{s,b\in B} (t_{AB})_{rs}^{ab} |\Psi_1\rangle \cdots |\Psi_A\rangle_a^r \cdots |\Psi_B\rangle_b^s \cdots |\Psi_N\rangle.
\end{aligned}
\tag{43}
$$

The single and double excitation amplitudes, $(t_{AB})_r^a$ and $(t_{AB})_{rs}^{ab}$, give rise to induction and dispersion, respectively, in the second-order energy corrections. These amplitudes are given by

$$(t_{AB})_r^a = \sum_b \frac{(ra|\hat{v}_{AB}|bb)}{\varepsilon_a - \varepsilon_r} \tag{44a}$$

$$(t_{AB})_{rs}^{ab} = \frac{(ra|\hat{v}_{AB}|sb)}{\varepsilon_a + \varepsilon_b - \varepsilon_r - \varepsilon_s}, \tag{44b}$$

where we use "chemists' notation" for the two-electron integrals.[33]

Given the wave function corrections in eqn (43), we see that eqn (40) corresponds to a Coulomb interaction between fragments A and B. This term appears in the two-body version of RSPT and is generally denoted $E_{\text{elst}}^{(1)}$, *i.e.*, it is the first-order electrostatic correction. Eqn (42) is also familiar and can be identified with the induction ($E_{\text{ind}}^{(2)}$) and dispersion ($E_{\text{disp}}^{(2)}$) corrections that arise at second order in $\hat{V}_{AB}$ in a standard RSPT or SAPT treatment.[14,15] The energy correction in eqn (41), however, does not arise in two-body SAPT. This term vanishes due to orthogonality, unless the set $\{A,B,C,D\}$ consists of at most three distinct indices. Furthermore, only single excitations (induction amplitudes) contribute to this particular energy correction. As a concrete example, suppose that $B = D$. Then the first term on the right side of eqn (41), $\langle\Psi_0|\hat{V}_{AD}|\Psi^{[0;1_{CD}]}\rangle$, represents the contraction of CD induction amplitudes, representing single excitations out of monomer D, with interaction integrals over $\hat{V}_{AD}$. That is,

$$E_{\text{RSPT}}^{[0;1_{AB},1_{CD}]} = \sum_{d,u\in D}\left[\sum_a (du|\hat{v}_{AD}|aa)(t_{CD})_u^d + \sum_b (du|\hat{v}_{CD}|bb)(t_{AD})_u^d\right]. \tag{45}$$

Physically, this term represents the coupling between the electronic polarization of D by A and the polarization of D by C. We refer to this term as a *three-body induction coupling*. These terms were absent in our original formulation of XPol + SAPT.[9]

To simplify the expressions appearing in eqns (37)–(39), we introduce a *diagonal exchange approximation*.[9] First, we partition the pairwise exchange operator, $\hat{\mathcal{P}}^{(2)}$, into a sum of exchanges between all pairs of dimers,

$$\hat{\mathcal{P}}^{(2)} = \sum_A \sum_{B>A} \hat{\mathcal{P}}_{AB}. \tag{46}$$

The operator $\hat{\mathcal{P}}_{AB}$ generates all possible swaps of one electron in A with one electron in B. (Multiple exchanges, *i.e.*, terms higher than $\hat{\mathcal{P}}^{(2)}$ in the antisymmetrizer, have already been neglected.) The expression for $\hat{\mathcal{P}}^{(2)}$ in eqn (46) can be inserted into the energy correction formulas in eqns (37)–(39) and manipulated as was done previously by Lotrich and Szalewicz[34,35] in deriving expressions for the exchange non-additivity in three-body SAPT calculations. However, the resulting formulas are complicated and would be difficult to program absent further approximation.

If we imagine pairwise exchange as a simultaneous tunneling of two electrons,[15] then it stands to reason that this dynamical process will be most important when the two electrons are coupled by some interaction operator. For this reason, we expect that terms such as $\langle\hat{V}_{AD}\hat{\mathcal{P}}_{CD}\rangle$, where A, C, and D are distinct monomers, will be less important than the "diagonal" terms where $A = C$. (Furthermore, terms such as $\langle\hat{V}_{AB}\hat{\mathcal{P}}_{CD}\rangle$ with no indices in common should not contribute at all.) Neglecting all but the "diagonal" terms leaves us with

$$E_{\text{exch}}^{[0;1_{AB}]} \approx \langle\hat{V}_{AB}\hat{\mathcal{P}}_{AB}\rangle - \langle\hat{V}_{AB}\rangle\langle\hat{\mathcal{P}}_{AB}\rangle, \tag{47}$$

$$\begin{aligned} E_{\text{exch}}^{[0;1_{AB},1_{CD}]} \approx{}& \langle\Psi_0|\hat{V}_{AB}\hat{\mathcal{P}}_{AB} - \hat{V}_{AB}\langle\hat{\mathcal{P}}_{AB}\rangle \\ &- \langle\hat{V}_{AB}\rangle\hat{\mathcal{P}}_{AB}|\Psi^{[0;1_{CD}]}\rangle + \langle\Psi_0|\hat{V}_{CD}\hat{\mathcal{P}}_{CD} \\ &- \hat{V}_{CD}\langle\hat{\mathcal{P}}_{CD}\rangle - \langle\hat{V}_{CD}\rangle\hat{\mathcal{P}}_{CD}|\Psi^{[0;1_{AB}]}\rangle \end{aligned} \tag{48}$$

and

$$E_{\text{exch}}^{[0;2_{AB}]} \approx \langle\Psi_0|\hat{V}_{AB}\hat{\mathcal{P}}_{AB} - \hat{V}_{AB}\langle\hat{\mathcal{P}}_{AB}\rangle - \langle\hat{V}_{AB}\rangle\hat{\mathcal{P}}_{AB}|\Psi^{[0;1_{AB}]}\rangle. \tag{49}$$

We expect the diagonal exchange approximation to be most severe for the first-order exchange correction, eqn (47). For trimers, the terms that we have neglected in eqns (47)–(49) have been considered previously by Moszynski *et al.*[39] and by Lotrich and Szalewicz,[34,35] in the context of three-body SAPT calculations. There is no reason, in principle, why these terms could not be included in XPol + SAPT , but they would add additional computational cost and have not been programmed.

Eqn (47) is equivalent to the first-order exchange correction in two-body SAPT, $E_{\text{exch}}^{(1)}$. Eqn (49) incorporates the second-order exchange-induction ($E_{\text{exch-ind}}^{(2)}$) and exchange-dispersion ($E_{\text{exch-disp}}^{(2)}$) corrections, explicit expressions for which can be found in the literature.[14,40] By inserting the expressions for the first-order corrected wave function, it is easy to see that eqn (48) is analogous to eqn (45) but with a pairwise exchange operator accompanying the interaction matrix element. These integrals are identical to those required to evaluate the second-order exchange-induction interaction in two-body SAPT, but are contracted against different amplitudes in eqn (48).

4. Final XPol + SAPT formulas and remarks. The intermolecular perturbation indicated in eqn (13) is valid for non-interacting zeroth-order monomers, *i.e.*, for traditional SAPT calculations. When XPol is used to generate the zeroth-order wave functions and energies, one should remove from the perturbation those intermolecular interactions contained in the XPol single-particle eigenvalues. The appropriate replacement for the AB interaction is[9]

$$\hat{v}_A(j) = -\sum_{I\in A}\left(Z_I - \tfrac{1}{2}q_I\right)\hat{I}_I(j) - \sum_{J\in B} M_J^A\hat{\Lambda}_J(j), \tag{50}$$

with a similar expression for $v_B(i)$. The operator $\hat{\Lambda}_j$ is defined so that it has the matrix elements that appear in eqn (9), and M_J^A is an element of an "M-vector" [M_J in eqn (8)] that contains only contributions from monomer A.[9]

The total energy is constructed from the sum of zeroth-order fragment energies (XPol eigenvalues) plus the first-order intramolecular RSPT corrections and the sum of intermolecular RSPT and SAPT corrections to second order, including induction and exchange-induction couplings. The resulting energy expression can be written

$$
\begin{aligned}
E_{\mathrm{XPS}} &= \sum_{A}\left(\sum_{a}\left[2\varepsilon_a^A - \mathbf{c}_a^{\dagger}(\mathbf{J}^A - \tfrac{1}{2}\mathbf{K}^A)\mathbf{c}_a\right] + E_{\mathrm{nuc}}^A\right) \\
&\quad + \sum_{A}\sum_{B>A}\left(E_{\mathrm{RSPT}}^{[0;1_{AB}]} + E_{\mathrm{RSPT}}^{[0;2_{AB}]}\right) \qquad (51) \\
&\quad + \sum_{A,C}\sum_{B>A}\sum_{D>C}\left(E_{\mathrm{RSPT}}^{[0;1_{AB},1_{CD}]} + E_{\mathrm{exch}}^{[0;1_{AB},1_{CD}]}\right).
\end{aligned}
$$

It is to be understood that the final set of summations includes only those terms with least one index in common between AB and CD.

Without any formal justification, we can extend the XPol + SAPT methodology to a Kohn–Sham (KS) DFT description of the monomers simply by adding the DFT exchange–correlation matrix to the Fock matrix for each monomer, and possibly scaling the $\mathbf{K}^A$ terms, as appropriate. In the context of traditional SAPT calculations, this extension is known as "SAPT(KS)".[41,42] Although this represents a tempting way to incorporate intramolecular electron correlation at low cost, the perils of SAPT(KS) calculations—namely, severe overestimation of dispersion interactions due to incorrect asymptotic behavior of the exchange–correlation potential—are well documented.[47] These problems are inherited by the "XPS(KS)" generalization of SAPT(KS),[9] which we will discuss only briefly, in §III B.

D. Implementation details

The XPol and XPol + SAPT methods, as described above, have been implemented in a developers' version of the Q-CHEM software.[49] Our implementation of the dual-SCF XPol procedure and the intermolecular SAPT corrections was reported previously,[9] and the implementation of the exchange-induction couplings requires no new integrals. To evaluate the induction coupling corrections, we loop over all unique AB pairs once, and store the induction amplitudes on disk. A second loop over AB pairs is then performed to compute the interaction energies. The result is a $\mathcal{O}(N^2)$ algorithm with a large prefactor, as opposed to a $\mathcal{O}(N^3)$ algorithm that would not require any amplitudes to be stored.

The second-order SAPT procedure is MP2-like, and like MP2, the cost of this procedure can be reduced substantially by means of a resolution-of-identity (RI) approximation[50] that eliminates four-index integrals by expanding basis function products in an auxiliary basis set. The RI procedure, in conjunction with standard auxiliary basis sets,[51] is employed for most of the calculations in this work, and introduces negligible errors in energy differences as compared to the standard procedure.[50]

As discussed above, the XPol monomer wave functions are computed in the ALMO basis, meaning that only AOs centered on monomer A are allowed to contribute to $|\Psi_A\rangle$. In the language of SAPT calculations, this corresponds to using what is traditionally termed the *monomer-centered basis set*. This is not the preferred choice for two-body SAPT calculations because its use precludes the description of charge transfer between A and B, except as the monomer basis sets approach completeness. More often, SAPT calculations employ the *dimer-centered basis set* (DCBS), in which the combined AB basis set is used to compute the zeroth-order wave functions for both monomers.[52] The dimer-centered approach is attractive because it can potentially capture charge-transfer effects that would be absent in a monomer-centered calculation, but it may be significantly more expensive if the monomer basis sets are large, and moreover it is not clear how the dimer-centered approach can be generalized, in an efficient way, to systems with more than two monomers.

As an affordable alternative to the DCBS, we have proposed what we call a "projected" basis set.[9] For a given pair of monomers, $A + B$, this entails computing XPol wave functions $|\Psi_A\rangle$ and $|\Psi_B\rangle$ in the individual ALMO bases for monomers A and B, then performing a pseudocanonicalization[53] of the $A + B$ dimer basis, *i.e.*, a diagonalization of the occupied–occupied and virtual–virtual blocks of the dimer Fock matrix. The two-body SAPT calculation is then performed in the pseudocanonical dimer basis, which does not disturb the converged XPol MOs for the monomers but does provide a larger set of virtual orbitals—extending over both monomers—for the subsequent SAPT calculation. However, the pseudocanonical MOs are *not* eigenfunctions of the fragment Fock operators. Although we could, in principle, include a "non-Brillouin singles" term[19] of the form $\sum_{ar} F_{ar}^A/(\varepsilon_a - \varepsilon_r)$ for fragment A (similarly for B), we decline to do so because this would re-introduce BSSE that is absent, by construction, in the monomer-centered ALMO basis.

Finally, we call attention to two parts of the XPol + SAPT interaction energy that will be turned on or off in various calculations reported in §III. First, there are the three-body induction coupling terms in eqn (45), which were absent in our original work.[9] In certain cases we shall delete these terms from the energy expression, as a means to test their importance. Furthermore (as in traditional two-body SAPT), one may compute an infinite-order correction for the polarization of monomer A that results from a frozen density on monomer B, and *vice versa*. To do so, one replaces the second-order induction and exchange-induction energies with their "response" analogues, in which the the induction amplitudes $(t_{AB})_r^a$ in eqn (44a) are replaced by amplitudes obtained by solving coupled-perturbed SCF equations for either monomer. The perturbation in these equations is the partner's electrostatic potential. (See refs. 9 and 14 for additional details.) Consistent with the terminology used in SAPT,[14] we call this the "response" (resp) version of XPol + SAPT,[9] although we will frequently refer to this as the "CPHF correction" in what follows.

III. Numerical results

Below, we present some applications to benchmark systems including water clusters (§III A), the S22 database (§III B), some anion–water clusters (§III C), and several dispersion-bound dimers (§III D). Most studies of non-covalent clusters emphasize binding energies, but in addition we shall explore geometry optimizations and one-dimensional potential energy scans. In some cases, we will compare XPol + SAPT results to conventional SAPT within a pairwise-additive, single-exchange approximation.

The difference between these two approaches is simply whether the zeroth-order wave function is a direct product of Hartree–Fock or XPol wave functions for the monomers.

All calculations were performed using a developers' version of Q-CHEM.[49] Analytic energy gradients are not yet available for the XPol + SAPT methodology, so in cases where we report XPol + SAPT geometry optimizations, these were performed *via* finite difference, using displacements of 10^{-3} bohr. (As discussed in the Appendix, the CHELPG algorithm that we have implemented is designed to provide charges that are continuous functions of the nuclear coordinates.)

A. Water clusters

Previously,[9] we explored the accuracy of XPol + SAPT calculations for one-dimensional dissociation curves of four symmetry-distinct conformers of $(H_2O)_2$.[54] Reasonable potential energy curves can be obtained in a variety of basis sets, but the best results are obtained using projected basis sets. The dissociation curves shown in Fig. 1 were obtained using the aug-cc-pVDZ(proj) basis set, *i.e.*, the projected version of a monomer-centered aug-cc-pVDZ basis; these dissociation curves are in *quantitative* agreement with complete basis set (CBS) MP2 results.

Calculations on water dimer do not exploit the full many-body nature of the XPol + SAPT methodology, so we next examine the relative energies of eight different $(H_2O)_6$ isomers, for which CCSD(T)/CBS results are available.[55] Unlike in the $(H_2O)_2$ calculations, where geometries were relaxed self-consistently at each level of theory, here we utilize the MP2/haTZ geometries[56] from ref. 55.

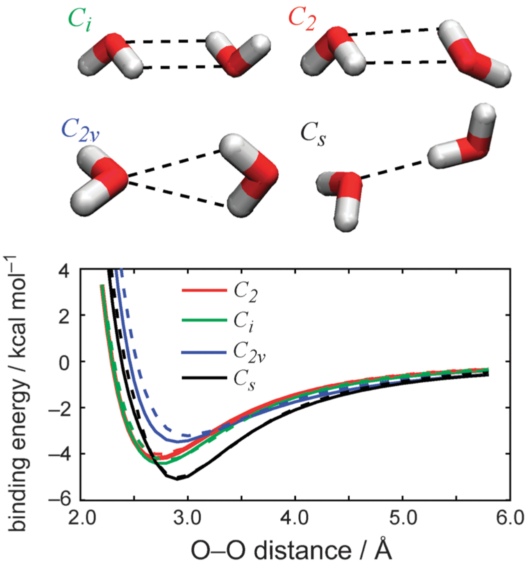

Fig. 1 Minimum-energy dissociation curves, along the O–O distance coordinate, for four symmetry-distinct isomers if $(H_2O)_2$. At each O–O distance, all other degrees of freedom have been relaxed at the MP2/aTZ level, subject to the constraint of point-group symmetry, and these MP2 geometries are used for all calculations. Solid curves represent single points computed at the XPS(HF)/aDZ(proj) level and dashed curves are MP2/CBS results. The XPS results employ CHELPG embedding charges, but nearly identical potential energy curves are obtained using Löwdin charges. The CPHF induction correction has not been employed.

In Fig. 2, we show XPol + SAPT results for $(H_2O)_6$ in comparison to the CCSD(T)/CBS benchmarks. Using CHELPG embedding charges, and with CPHF and three-body induction corrections turned on, we find that the aDZ(proj) and aTZ(proj) basis sets (where "aXZ" stands for aug-cc-pVXZ) correctly reproduce the trend in relative energies amongst

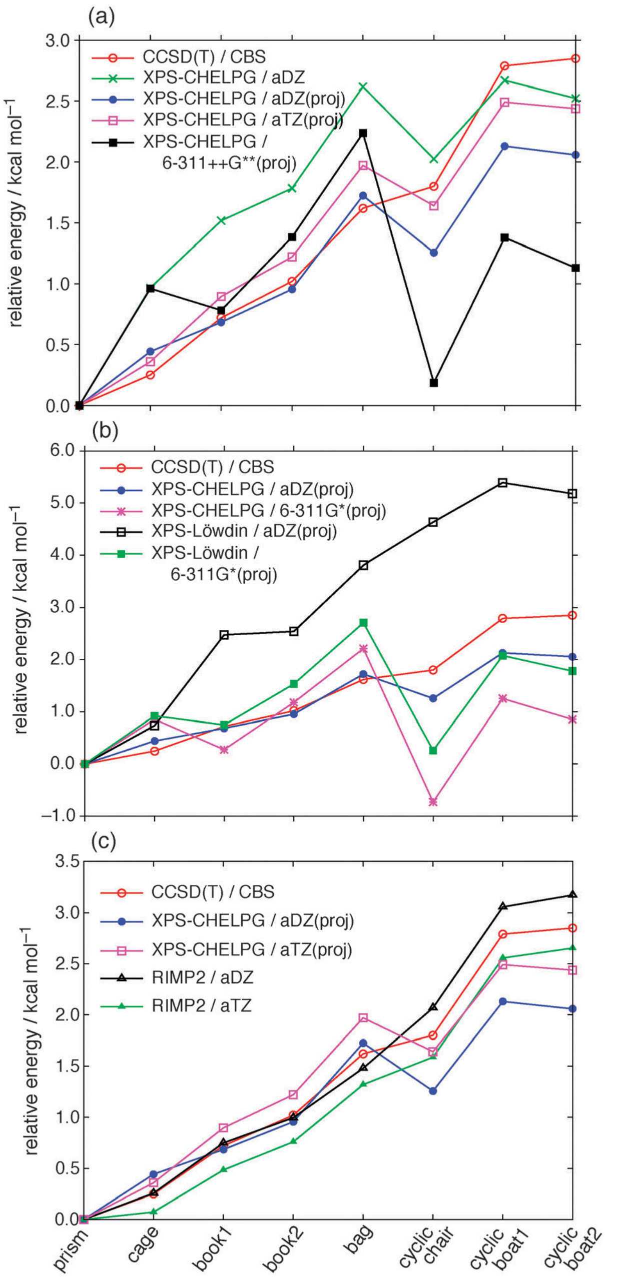

Fig. 2 Relative energies of $(H_2O)_6$ isomers as compared to benchmark CCSD(T)/CBS results from ref. 55, using MP2/haTZ geometries. The XPol + SAPT calculations in (a) all utilize CHELPG embedding charges, whereas (b) compares CHELPG and Löwdin embedding charges. (Note that the vertical energy scale is different in the two panels.) In (c), we compare one of the better XPS methods to supersystem (frozen core) RIMP2 results. All XPol + SAPT calculations include the infinite-order (frozen density) induction corrections obtained by solving monomer CPHF equations, and thus could be labeled "XPS-resp".

these isomers, with the exception that the "cyclic chair" isomer is overstabilized by about 0.5 kcal mol^{-1} [see Fig. 2(a)]. Overall, the mean absolute errors (MAEs) evaluated over seven isomers (excluding the prism isomer that defines $E = 0$) are 0.34 and 0.24 kcal mol^{-1}, respectively, for the aDZ(proj) and aTZ(proj) basis sets. Basis-set projection is essential to obtain results of this quality; use of the unprojected aDZ basis set affords a MAE of 1.58 kcal mol^{-1}. For this reason, we will use projected basis sets almost exclusively in subsequent calculations. On the other hand, basis-set projection is necessary but not sufficient for accurate results; the relative energies obtained with the 6-311G*(proj) and 6-311++G**(proj) basis sets (the latter being the Pople-style analogue of Dunning's aTZ) are poor, with MAEs of 1.12 and 0.93 kcal mol^{-1}, respectively.

Fig. 2(b) compares XPol + SAPT results using either CHELPG or Löwdin embedding charges. Although the general conclusion in ref. 9 was that XPol + SAPT binding energies computed with CHELPG embedding charges were superior to those obtained using Löwdin charges, the latter can be computed at negligible cost, and are therefore worth exploring further in comparison to the CHELPG charges that do add a non-trivial amount of overhead, owing to the need to evaluate the electrostatic potential over a grid. Use of Löwdin charges, however, seriously degrades the quality of results obtained in the aDZ(proj) basis set, as compared to the quite good results obtained with CHELPG charges using the same basis set. The difference is less pronounced in the 6-311G*(proj) basis.

In Fig. 2(c), we compare the two XPS methods that were found to provide the best results for $(H_2O)_6$ with supersystem RIMP2 results using the same basis sets. For the prism, cage, bag, and book isomers, the RIMP2 and XPS results are quite comparable (and accurate), but as mentioned above the XPS methods exhibit a tendency to overstabilize the cyclic structures, so that RIMP2 is notably more accurate in these cases. It is not clear to us why this is the case, although energy decomposition analysis of $(H_2O)_6$ at the MP2/aug-cc-pV5Z level does indicate that the three cyclic isomers exhibit the largest many-body contributions to the binding energy.[24] Note that the pairwise SAPT calculations that we employ are quite similar to MP2, in terms of the excitations that are included, and the primary differences between XPS and supersystem MP2 are a different SCF procedure (supersystem HF *versus* XPol), and also the fact that MP2 incorporates some monomer electron correlation, whereas XPol + SAPT neglects this entirely. MP2 also incorporates some three-body correlation effects insofar as a double excitation can couple together three monomers, but based on the rather small magnitudes of the XPol + SAPT three-body induction couplings (as documented below), we suspect that these effects are small in this particular case.

All of the XPol + SAPT results depicted in Fig. 2 include both the CPHF and three-body induction couplings. Since both of these items add extra cost to the calculation, it is worth exploring whether they are necessary. Of the basis sets tested for this system, aTZ(proj) affords the best results, so in Fig. 3 we explore XPol + SAPT calculations of $(H_2O)_6$ using this basis set but turning off either the CPHF induction correction, the three-body induction couplings, or both. Each of these new variants affords a larger MAE than does the "full" XPol + SAPT calculation, although the trends in the relative isomer energies are basically the same in each case, with the exception that the "3-body only" version (no CPHF) significantly overstabilizes the cage isomer. For best results, the CPHF correction appears to be important, which is not altogether surprising since the use this correction is recommended in SAPT calculations of polar molecules.[14]

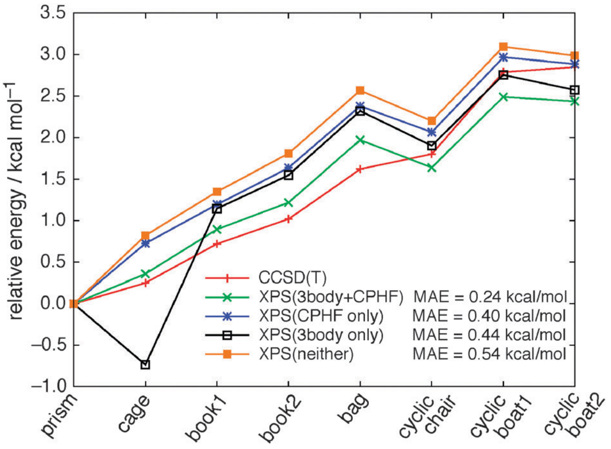

Fig. 3 Relative energies of $(H_2O)_6$ isomers computed at the XPS(HF)-CHELPG/aTZ(proj) level, as compared to CCSD(T)/CBS benchmarks from ref. 55, using MP2/haTZ geometries. Results from different variants of XPS(HF) are compared, depending on whether the three-body induction couplings and/or the infinite-order frozen-density induction corrections (based on solving CPHF equations) are included.

Whereas all of the XPol + SAPT calculations in Fig. 3 use CHELPG embedding charges, we observe that the CPHF and three-body induction corrections can be more significant when Löwdin charges are used instead—as large as 2 kcal mol^{-1} for certain isomers, as shown in Fig. 4. Because these corrections push the relative energies closer to the benchmark results, we suspect that in the case of Löwdin embedding these corrections serve to compensate for deficiencies in the description of the embedding potential. [Recall that the perturbation used in the SAPT part of the calculation is the full intermolecular interaction potential minus the embedding potential; see eqn (50).] If so, then the comparative smallness of the CPHF and three-body induction corrections in the CHELPG case offers additional evidence that the CHELPG charges provide a better description of the molecular electrostatic potential.

Our motivation for considering Löwdin charges lies in the fact that these charges are far less costly to compute, hence XPS-Löwdin calculations are more amenable to finite-difference geometry optimizations. (These optimizations are still quite costly, however!) In Fig. 4, we also plot the $(H_2O)_6$ relative energies following geometry optimization at two different XPol + SAPT levels of theory. In all cases where the CPHF and three-body induction corrections are included, geometry optimization serves to move all of the relative isomer energies closer to the benchmark values, and furthermore preserves the energetic ordering of the isomers. This ordering is basically correct with respect to the benchmark calculations, aside from reversing the order of the nearly-degenerate book1/book2 and cyclic boat1/cyclic boat2 isomer pairs, which differ

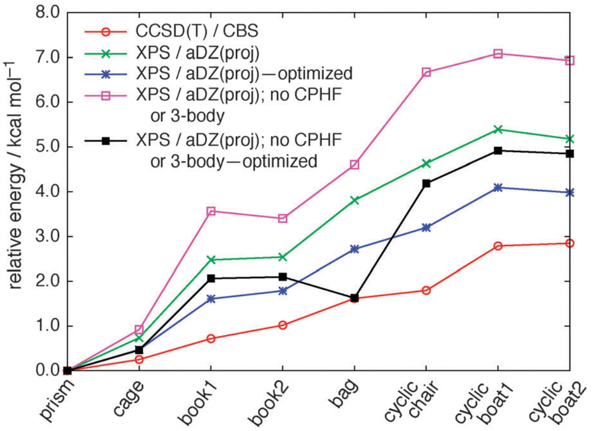

Fig. 4 Relative energies of $(H_2O)_6$ isomers computed at the XPS(HF)-Löwdin/aDZ(proj) level. Two sets of XPS calculations employ both the CPHF and three-body induction corrections, whereas the other two sets of XPS calculations employ neither. XPS results are compared at benchmark (MP2/haTZ) geometries and also at geometries that are optimized at the same (XPS) level of theory that is used to compute the binding energy.

only in the orientation of the dangling O–H moieties. Absent the CPHF and three-body corrections, optimization moves the energies closer to benchmark values in every case except that of the "bag" isomer, which is strongly overstabilized following optimization.

Examining the optimized geometries of each isomer, some examples of which are depicted in Fig. 5, we find that the various levels of theory afford optimized structures that are essentially indistinguishable to the eye, except in the case of the "bag" isomer. In that particular case, omission of the CPHF and three-body corrections causes two of the H_2O molecules to orient in a qualitatively incorrect fashion, as compared to the MP2 geometry. This fact is clearly evident if we compute the root-mean-square deviations (RMSDs) with respect to MP2 geometries,[57] which are listed in Table 1. All of the RMSDs are less than 0.2 Å except for the bag structure optimized without CPHF or three-body corrections, for which the RMSD > 0.5 Å. As such, the bag structure represents the lone exception to the trend in relative energies observed in Fig. 4 and is also an outlier with respect to the structural RMSDs.

Table 1 also lists the average deviations in hydrogen bond lengths and angles with respect to MP2/haTZ results.[59] Examining the entire data set of H-bond lengths and angles (not shown), we find that the results *with* the CPHF and three-body induction corrections are almost universally more accurate as compared to results where these corrections are omitted, which speaks to the systematic improvability of the method. For the more accurate "XPS + CPHF + 3-body" results, H-bond lengths are consistently overestimated by 0.09–0.15 Å, whereas in the absence of these corrections the MADs in H-bond lengths are ≈0.05–0.06 Å larger still. For the H-bond lengths, the XPS calculations err in every case toward longer bond lengths. Hydrogen bond angles are accurate to within a few degrees, and often (but not always) err toward smaller H-bond angles. (Geometric constraints occasionally cause a few of the XPol + SAPT H-bond angles to be larger than benchmark results.)

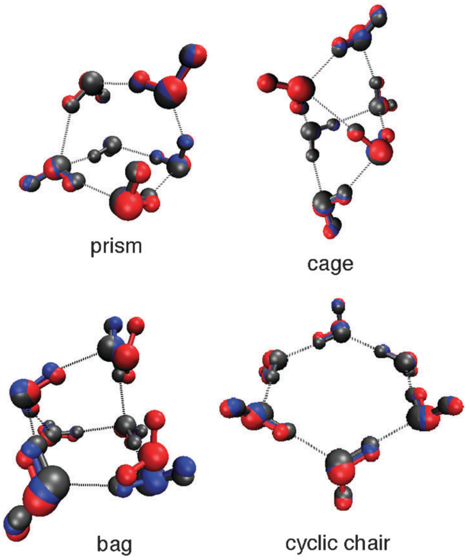

Fig. 5 Comparison of optimized geometries for several isomers of $(H_2O)_6$. In each case, the geometry rendered in grey has been optimized at the MP2/haTZ level while the remaining two geometries are optimized at the XPS(HF)-Löwdin/aDZ(proj) level, either with CPHF and three-body induction corrections (geometries in blue) or without these corrections (geometries in red).

We will have more to say about these data in §III C, but for now the conclusion seems to be the following. For water clusters, the CPHF correction offers consistent (albeit modest) improvements to geometries and relative energies in most cases. In some cases, use of this correction avoids an egregious outlier.

Finally, let us consider binding energies in some larger $(H_2O)_n$ clusters. As benchmarks, we take MP2/CBS results for n = 2, 3, 4, 5, 6, 8, 11 and 20.[61–67] Multiple isomers are available for certain of these cluster sizes, but the differences amongst their BEs are small relative to the differences between variants of XPS, so we consider only one isomer at each cluster size except for n = 20, where the energy differences are much larger. In this case, we consider one isomer from each family of low-lying minima (dodecahedron, fused cubes, face-sharing pentagonal prisms, and edge-sharing pentagonal prisms).

XPS/aDZ(proj) binding energies (BEs) for these species, relative to relaxed H_2O monomers, are plotted against MP2/CBS benchmark values in Fig. 6. This plot includes two sets of XPS-CHELPG and XPS-Löwdin results: one in which the cluster geometries (and the relaxed H_2O monomer geometry) are the MP2 geometries used to compute the MP2/CBS benchmarks,[61–67] and another in which the geometry has been optimized using the same XPS method that is used for the single-point BE calculation.[60] Errors are observed to increase with cluster size in all cases. In previous work,[9] we noted that XPol + SAPT calculations in various basis sets achieve a roughly constant BE error per hydrogen bond, for clusters containing more than about five hydrogen bonds, which suggests that the total error should be an extensive quantity and thus explains the trend seen in Fig. 6. For homologous clusters in general, this same trend is likely to be observed insofar as the XPol + SAPT method achieves a consistent, if approximate, description of each intermolecular interaction.

Table 1 Errors in XPS-optimized geometries for isomers of $(H_2O)_6$, as compared to MP2/haTZ geometries. Several structural parameters are listed that characterize the deviation from the MP2/haTZ geometry. These parameters are: the root-mean-square deviation (RMSD) of the atomic Cartesian coordinates; the mean absolute deviation (MAD) in the hydrogen-bond distances (H···O); and the MAD in the hydrogen-bond (O–H···O) angles. All XPS calculations are XPS(HF)-Löwdin/aDZ(proj); the comparison is whether or not the CPHF and three-body induction corrections are included

	XPS (including CPHF and 3-body)			XPS (without CPHF or 3-body)		
Isomer	Coordinate RMSD/Å	H-bond Distance, MAD/Å	H-bond angle, MAD (°)	Coordinate RMSD/Å	H-bond Distance, MAD/Å	H-bond angle, MAD (°)
prism	0.069	0.090	3.4	0.109	0.143	5.1
cage	0.087	0.090	2.5	0.136	0.136	4.0
book1	0.089	0.117	1.6	0.140	0.174	2.7
book2	0.085	0.115	1.8	0.134	0.173	2.3
bag	0.121	0.120	1.3	0.533	0.193	9.1
cyclic chair	0.111	0.149	0.4	0.178	0.217	2.9
cyclic boat1	0.125	0.145	0.8	0.165	0.214	2.0
cyclic boat2	0.118	0.147	0.4	0.163	0.214	2.1

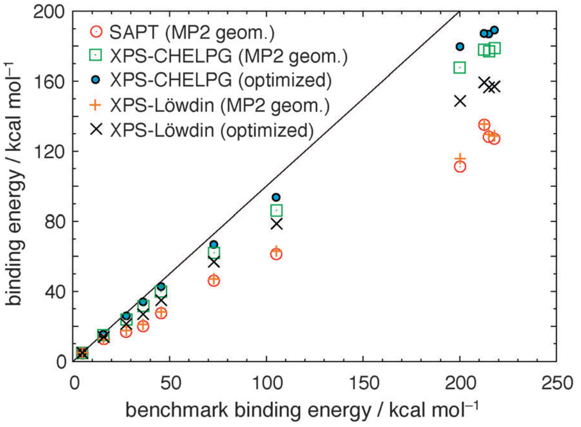

Fig. 6 SAPT(HF)/aDZ(proj) and XPS(HF)/aDZ(proj) binding energies for $(H_2O)_n$ clusters, as compared to MP2/CBS benchmarks. In the XPS cases, results are shown both at the benchmark MP2 geometries and also at geometries that have been self-consistently optimized (or at least, relaxed[60]) on the XPS potential energy surface. In all XPS cases, three-body induction couplings are neglected and we do not solve CPHF equations. The cluster isomers include four different $n = 20$ isomers along with one isomer each for $n = 2, 3, 4, 5, 6, 8$, and 11. The oblique line indicates where the XPS or SAPT binding energy would coincide with the benchmark.

When MP2 geometries are used, we note from Fig. 6 that the XPS-Löwdin method is only *slightly* more accurate than a pairwise-additive SAPT calculation with no embedding whatsoever,[68] a method that entirely neglects many-body effects! Given that XPS-CHELPG results at the same geometries are *significantly* more accurate that pairwise-additive SAPT results, this result is a strong indictment of the Löwdin charge scheme, which does not appear to be appropriate for large clusters.

One should bear in mind that the MP2 and XPol + SAPT methods predict somewhat different monomer geometries, since the latter (in its present form) neglects monomer electron correlation. As such, one should anticipate that XPol + SAPT errors in BEs will decrease if XPS-optimized geometries are employed instead, and the data in Fig. 6 demonstrate that this is indeed the case. Self-consistent optimization significantly improves the accuracy of BEs computed at both the XPS-Löwdin and XPS-CHELPG levels. (BEs computed using the XPS-CHELPG method with CPHF and three-body induction corrections are nearly indistinguishable from the XPS-CHELPG results presented in Fig. 6, although we are unable to optimize geometries in the presence of these corrections, owing to the tremendous computational expense.)

B. S22 database

The S22 data set, originally assembled by Hobza and co-workers[69] but whose energetics were subsequently revised by Sherrill and co-workers,[70] consists of estimates of the CCSD(T)/CBS binding energy for 22 dimers ranging in size from species like $(NH_3)_2$ and $(CH_4)_2$ up to adenine–thymine and indole–benzene. This database offers a convenient way to screen a large number of different basis sets and other variations of the XPol + SAPT methodology. Because our aim is to develop a low-cost method applicable to large clusters, we focus primarily on double-ζ basis sets. Because projected basis sets afford such excellent results for the water dimer potential energy surface, the calculations described below all use projected basis sets.

Whereas all of the calculations in §III A used Hartree–Fock theory to obtain the monomer wave functions, there is nothing in principle to stop us from using KS-DFT in the XPol procedure, and subsequently using the KS determinant as the zeroth-order monomer wave function. As mentioned in §II C 4, this constitutes a many-body extension of what is usually called SAPT(KS).[41,42] The SAPT(KS) approach was originally deemed unsuccessful,[47] in that the electrostatic and induction energies failed to reproduce traditional SAPT(HF) values. Discrepancies between SAPT(HF) and SAPT(KS) calculations were ultimately traced to the incorrect asymptotic behavior of typical exchange–correlation (XC) functionals used in DFT, and an asymptotic correction to the XC potential was found to improve the agreement with benchmark values.[47] Even following asymptotic correction, however, SAPT(KS) dispersion energies are still poor, which ultimately led to the development of alternative "SAPT(DFT)" methods.[42–46,48] These methods are not considered here, although they represent a promising direction for the XPol + SAPT methodology, as discussed in §IV.

Fig. 7 compares error statistics for S22 binding energies at the SAPT(HF) and SAPT(KS) levels, and we note that BE

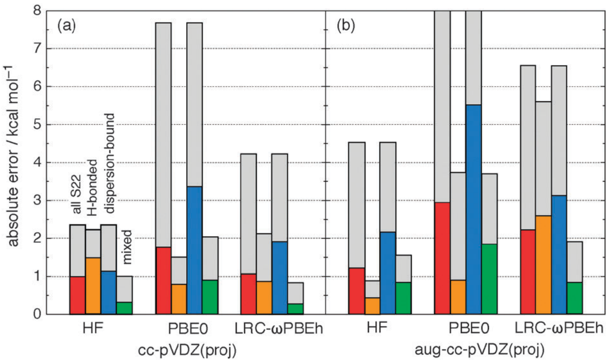

Fig. 7 Mean absolute errors in SAPT(KS)-resp/cc-pVDZ(proj) calculations for the S22 database,[69] comparing Hartree–Fock theory to results from two different density functionals, using (a) the cc-pVDZ(proj) basis and (b) the aDZ(proj) basis. In each case, we plot the mean absolute error (colored bars) and maximum absolute errors (gray bars) evaluated over the entire S22 set as well as three different subsets. From left to right, the data sets for each functional are: the full S22 data set (red), the subset of seven dimers dominated by hydrogen bonding (orange), the subset of eight dimers dominated by dispersion (blue), and the subset of seven dimers whose interactions are of mixed influence (green), as classified by Sherrill and co-workers.[70] For the SAPT(PBE0)/aDZ(proj) calculations, the maximum error in the dispersion-bound subset (11.5 kcal mol^{-1}) is off scale in the figure.

errors in SAPT(KS) calculations are indeed quite large for the dispersion-dominated subset of S22. This is a result of the MP2-like sum-over-states dispersion formula that is used in SAPT0, in association with the fact that KS-DFT methods tend to predict smaller HOMO/LUMO gaps than HF theory. Long-range corrected (LRC) density functionals tend to widen the HOMO/LUMO gap,[71,78] and furthermore provide HOMO eigenvalues that better approximate ionization potentials, indicating an improved description of the asymptotic part of the XC potential. Previously, we found that the LRC-ωPBEh functional[72,76] offered the best results among the limited set of functionals that we tested,[9] and from Fig. 7 we see that this functional does indeed offer a significant improvement over the global hybrid PBE0 functional,[73] although the errors in BEs for dispersion-dominated dimers remain significantly larger than SAPT(HF) results. (A system-specific tuning of ω, as suggested for other DFT applications,[74] might improve the results, but we have not pursued such an approach here.) We also note that the errors for dispersion-dominated dimers tend to increase in augmented basis sets, even in the SAPT(HF) case, which is consistent with an overestimate of the dispersion interaction along with the general observation that the dispersion energy increases in extended basis sets.[52] This is a fundamental limitation of the second-order (SAPT0) treatment of dispersion; when higher-order terms are included in the SAPT calculation, very accurate results are obtained for both $(CH_4)_2$ and $(C_6H_6)_2$, even in large, diffuse basis sets.[2]

Whereas both SAPT(HF) and SAPT(KS) tend to overbind the dispersion-dominated dimers, these methods typically underbind the strongly H-bonded dimers. Like the overestimation of the dispersion interaction, this effect is more pronounced in augmented basis sets, so that while SAPT(KS) tends to be more accurate than SAPT(HF) for H-bonded systems in small basis sets, this trend is reversed when diffuse basis functions are present. We have previously suggested[9] that this behavior may result from delocalization error[75] in DFT, which leads to SAPT(KS) exchange-repulsion energies that are too large, and therefore BEs that are too small.

One final trend that is evident from the data in Fig. 7 is that for SAPT(HF) calculations, increasing the quality of the basis set tends to improve the results for systems dominated by hydrogen bonding but degrades the accuracy for systems dominated by dispersion. As a result of this contrasting behavior, much of the rest of this work is devoted to a study of the basis-set dependence of XPS results, with the aim to identify (if possible) a basis set that affords reasonable results across a wide variety of intermolecular interactions.

Owing to the large errors encountered in SAPT(KS) calculations for dispersion-dominated systems, we focus exclusively on SAPT(HF) and XPS(HF) calculations in the remainder of this work. Fig. 8 compares S22 error statistics for SAPT(HF) and XPS(HF) calculations using various basis sets. The SAPT(HF) and XPS(HF) error statistics are quite comparable, and because SAPT(HF) is a reasonable approach for these two-body systems (unlike in larger water clusters, where this method misses important many-body polarization effects), this favorable comparison between SAPT(HF) and XPS(HF) serves to demonstrate that the XPS-CHELPG charge-embedding scheme has not introduced any significant new errors.

The comparison of different double-ζ basis sets in Fig. 8 proves to be quite interesting, and confirms several of the observations made above. For the subset of S22 consisting of strongly H-bonded dimers, the mean accuracy of the BEs is seen to improve in the presence of diffuse basis functions, which is generally in line with the results for water clusters discussed above. However, full augmentation degrades the accuracy of the binding energies for the dispersion-dominated

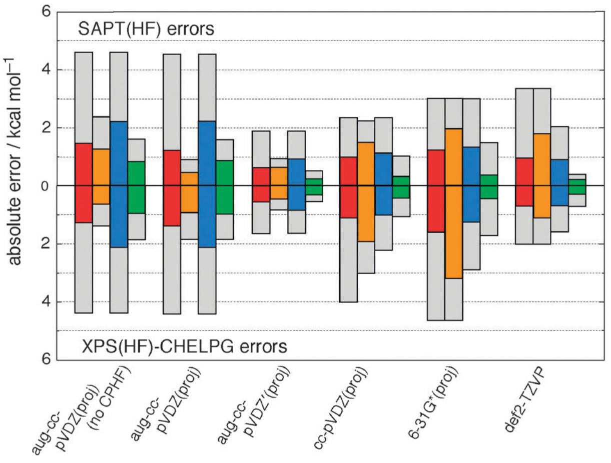

Fig. 8 Mean absolute errors in SAPT(HF) and XPS(HF) calculations for the S22 database,[69] evaluated for various basis sets. As in Fig. 7, the mean absolute error (colored bars) and maximum absolute errors (gray bars) are shown, evaluated over both the entire S22 set (red bars) and also three different subsets: H-bonded dimers (orange bars), dispersion-bound dimers (blue bars), and dimers of mixed influence (green bars). Except in the leftmost data set, all calculations solve CPHF equations to compute the infinite-order frozen-density induction correction.

dimers, and (to a lesser extent) reduces the accuracy for the subset of dimers of "mixed influence" interactions as well.

Observations along these lines have led Sherrill and co-workers[2,50,79,80] to recommend a partially-augmented basis set for use in MP2 and SAPT0 calculations of dispersion-dominated systems. This basis, which they call aug-cc-pVDZ′ (and which we abbreviate as aDZ′) consists of cc-pVDZ for the hydrogen atoms and aDZ for the heavy atoms, except that the diffuse *d* functions are removed from the latter. Of the basis sets that we have tested, aDZ′(proj) is clearly the best for the S22 data set, as shown in Fig. 8. Errors in dispersion-dominated complexes are greatly reduced, relatively to aDZ(proj) results, but this basis also reduces both the mean and maximum error for the strongly H-bonded subset of S22.

At least in these S22 examples, and for the low-order forms of SAPT and XPS that are used here, induction and dispersion interactions appear to have competing needs in terms of selection of the monomer basis set. In order to examine this competition in more detail, we consider in §III C some anion–water clusters, which ought to exhibit even larger induction effects than the charge-neutral H-bonded systems that we have considered thus far. Then, as a counterpoint, in §III D we will examine a few dispersion-bound systems in more detail.

C. Anion–water clusters

Results presented above for water clusters and for the S22 database suggest that, for systems bound primarily by electrostatics and induction, binding energies typically improve as basis-set quality improves. Next, we consider the $F^-(H_2O)$ dimer, which is not included in the S22 data set and whose CCSD(T)/CBS binding energy (25.4 kcal mol^{-1}) is considerably larger than that of any of the S22 dimers. Several one-dimensional potential energy scans along the $F^-\cdots HOH$ coordinate, computed at the SAPT(HF) level using various projected basis sets, are depicted in Fig. 9. (Since our previous results indicate that projected basis sets are necessary to obtain high-accuracy potential energy surfaces for water clusters,[9] we shall use projected basis sets exclusively in all remaining calculations.) Little difference is observed amongst these basis sets, and the accuracy is no better than 1 kcal mol^{-1} in any case. This is consistent with—if perhaps a bit larger than—what one might have expected based on S22 results for hydrogen-bonded dimers.

However, one of the potential energy curves depicted in Fig. 9 is not accurate at all, and severely overbinds the $F^-(H_2O)$ complex. This is the calculation in which the so-called $\delta E_{\rm int}^{\rm HF}$ correction is added to the SAPT(HF) binding energy.[2,15] This correction is defined as

$$\delta E_{\rm int}^{\rm HF} = E_{\rm int}^{\rm HF} - (E_{\rm elst}^{(1)} + E_{\rm exch}^{(1)} + E_{\rm ind,resp}^{(2)} + E_{\rm exch\text{-}ind,resp}^{(2)}), \quad (52)$$

where $E_{\rm int}^{\rm HF}$ is the dimer energy computed at the Hartree–Fock level. Addition of this correction helps to build in higher-order induction effects that would otherwise be absent in a SAPT0 calculation.[2] As such, the use of $\delta E_{\rm int}^{\rm HF}$ is recommended for polar systems.[2,81] Clearly, this correction is *disadvantageous* for the $F^-(H_2O)$ system, at least at the SAPT(HF)-resp/aTZ(DCBS) level considered in Fig. 9. The reasons for this have to do with a subtle imbalance between competing interactions when the $\delta E_{\rm int}^{\rm HF}$ correction is applied; as we discuss detail elsewhere,[38] this can be rectified by including higher-order terms in the SAPT calculation. The details are not particularly important here, and we include this example merely to point out that one cannot always appeal to $\delta E_{\rm int}^{\rm HF}$ to correct the deficiencies of a SAPT0 calculation, even when polar molecules are involved. More importantly in the present context, we wish to avoid the use of $\delta E_{\rm int}^{\rm HF}$ for the simple reason that this correction is no longer well-defined for a system comprised of more than two monomers.[9]

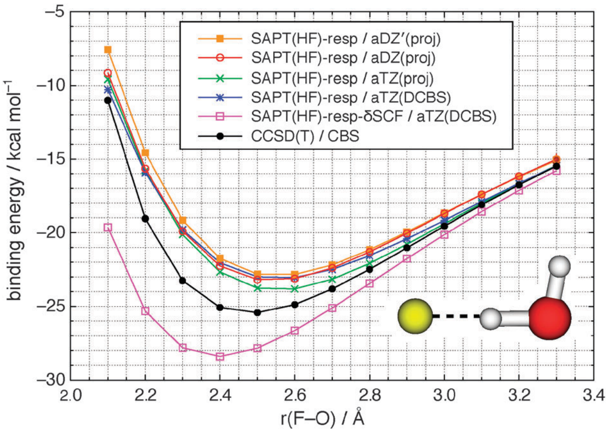

Fig. 9 One-dimensional potential energy scans for $F^-(H_2O)$ along the F–O distance coordinate, at a fixed H_2O geometry, computed using several variants of SAPT(HF) as described in the text. Results of a CCSD(T)/CBS benchmark calculation are also shown. The δSCF term is the correction defined in eqn (52).

Fig. 10 shows various XPS(HF) results for the same one-dimensional potential energy curve for $F^-(H_2O)$, in comparison to both CCSD(T)/CBS results and also to the most accurate SAPT(HF) curve from Fig. 9. When the CPHF induction correction is included, the XPS result is more accurate than SAPT, whether the embedding charges are of the Löwdin or the CHELPG variety, and the latter method [XPS(HF)-resp-CHELPG/aTZ(proj)] is extremely accurate, consistent with the notion that high-quality monomer basis sets are preferred for describing induction and electrostatics. In addition, we note that XPS with CHELPG charges is more accurate than XPS with Löwdin charges even when the former method does *not* include the CPHF correction. This provides another strong argument in favor of CHELPG charges.

Fig. 11 presents a thorough comparison of basis sets for XPS(HF) calculations of the $F^-(H_2O)$ potential curve. Comparing results obtained with Pople-style basis sets [Fig. 11(a)] to those obtained using Dunning-style basis sets [Fig. 11(b)], a few general remarks can be made. First, diffuse basis functions are crucial. In their absence, the complex is overbound (in a vertical sense) by 5–8 kcal mol^{-1} at the benchmark geometry, and errors in the adiabatic BE (measured from the XPS potential minimum) are even larger, up to 10.4 kcal mol^{-1}. The minimum-energy $F\cdots O$ distance is also too short, by 0.1–0.2 Å, in the absence of diffuse basis functions. Addition of diffuse functions shifts the whole potential energy curve upward, and shifts the minimum to longer $F\cdots O$ distance,

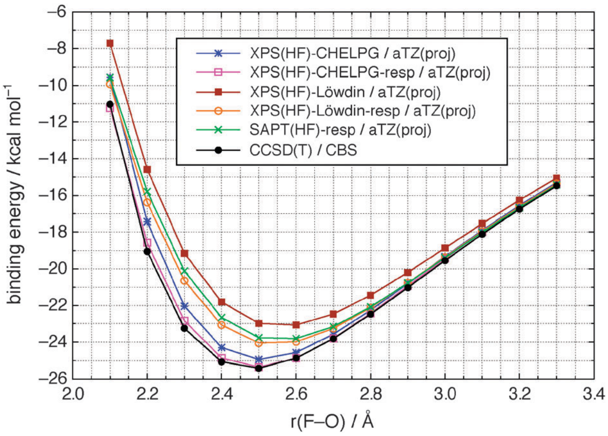

Fig. 10 One-dimensional potential energy scans for $F^-(H_2O)$ along the F–O distance coordinate, at a fixed H_2O geometry, computed using several variants of XPS(HF) as described in the text. Also shown are a CCSD(T)/CBS benchmark calculation, along with the most accurate SAPT(HF) potential energy scan from Fig. 9.

although in smaller basis sets both effects overshoot in the opposite direction. The particular case of aDZ′ [Fig. 11(b)] is especially interesting in light of its excellent performance for S22, along with the fact that XPS(HF)/aDZ′(proj) calculations afford accurate potential energy curves for $(C_6H_6)_2$.[9] For $F^-(H_2O)$, this limited set of diffuse basis functions eliminates the severe (7 kcal mol^{-1}) overbinding observed at the XPS/cc-pVDZ(proj) level but results in a small (3 kcal mol^{-1}) *underbinding*, and a minimum-energy F···O distance that is too long by about 0.1 Å. In contrast, full augmentation (*i.e.*, aDZ) provides a more accurate potential energy curve, with the minimum in the right place and a BE error of < 2 kcal mol^{-1}.

The aTZ basis set, on the other hand, affords a potential energy curve that agrees quantitatively with the CCSD(T)/CBS benchmark, and this leads to a second general statement, that improving the quality of the basis set generally improves the quality of the $F^-(H_2O)$ potential energy curve. This is consistent with results obtained above for other systems dominated by electrostatic and induction effects. However, quantitative results are obtained not only using aTZ but also using the Pople-style 6-311++G(3df,3pd) basis, which is a better approximation to Dunning's aTZ than is the more standard 6-311++G(d,p).

As we have already seen in the context of diffuse functions, however, improving the basis set does not afford monotonic convergence toward the benchmark result, due to several competing effects. Increasing the number of valence basis functions (*i.e.*, switching from double-ζ to triple-ζ) increases the $F^-(H_2O)$ binding energy by 1–2 kcal mol^{-1}, and adding additional polarization functions tends to have the same effect [see Fig. 11(a)]. In contrast, addition of diffuse basis functions engenders a dramatic *decrease* in the binding energy.

The analysis above holds for the Dunning- and Pople-style basis sets, where the results are fairly systematic. Results for the second-generation ("def2") Karlsruhe basis sets [Fig. 11(c)] are not quite as systematic in their convergence toward the benchmark result, especially with regard to the role of increasing the "ζ number" of the basis as well as the number of polarization functions. Some of the trends observed above can still be seen, however. For example, the double-ζ def2-SVP basis affords a BE that is more than 6 kcal mol^{-1} too large and a minimum-energy distance that is 0.1 Å too short, while augmentation (def2-SVPD) reduces the BE error to about 2 kcal mol^{-1} but in the opposite direction, and affords a minimum-energy F···O distance that is 0.1 Å too long. None of the Karlsruhe triple-ζ bases, however, achieves a quantitative result. The best result is obtained using the def2-TZVP basis set, which predicts the correct minimum-energy F···O distance, and overbinds the complex by only 1.1 kcal mol^{-1}.

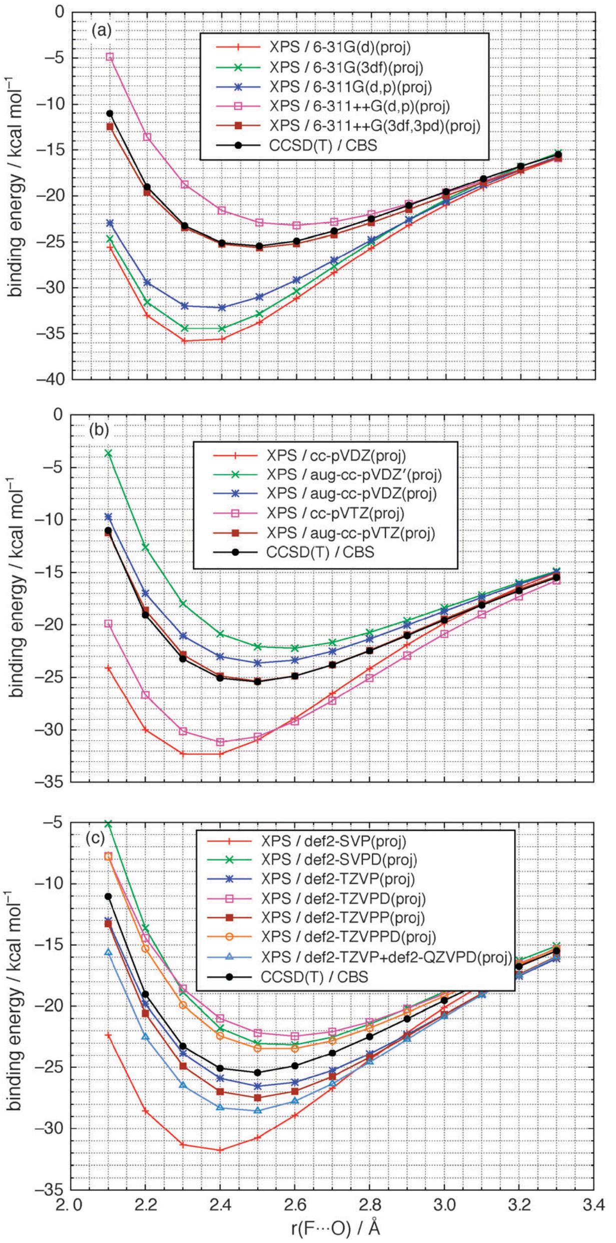

Fig. 11 One-dimensional potential energy scans for $F^-(H_2O)$ along the F–O distance coordinate, at a fixed H_2O geometry, computed using the XPS(HF)-resp-CHELPG method with (a) Pople basis sets, (b) Dunning basis sets, and (c) Karlsruhe basis sets. In the latter case, note that basis sets ending in "D" include diffuse functions. A CCSD(T)/CBS benchmark result is also shown.

This basis set will ultimately emerge as our best compromise choice, after considering dispersion-bound systems in §III D.

In the calculations on water clusters in §III A, we noted that XPol + SAPT binding energies improve substantially, with respect to benchmark values, when cluster geometries are optimized self-consistently at the XPol + SAPT level. Moreover, we have just seen that the choice of basis set shifts the location of the $F^- \cdots HOH$ potential energy minimum by as much as 0.2 Å even when the geometry of H_2O is fixed. Therefore, we next wish to examine some fully XPS-optimized anion–water cluster geometries. Fully-optimized results, computed at the XPS-resp-CHELPG level using a variety of basis sets, are presented in Table 2. A CCSD(T)/CBS benchmark binding energy is available for this system,[82] but differs by only 0.16 kcal mol^{-1} (0.6%) from the MP2/aTZ result, so we take the latter as a suitable benchmark since we have access to geometrical parameters for the MP2/aTZ calculation.

One interesting point to note from these data is that the XPS-resp-CHELPG/aTZ binding energy reported in Table 2 differs from the best available benchmark by 1.6 kcal mol^{-1}, whereas previously (in the context of Fig. 11) we noted that the XPS-resp-CHELPG/aTZ potential curve was in quantitative agreement with CCSD(T)/CBS results. The difference is that the value reported in Table 2 corresponds to a fully-optimized dimer geometry, whereas in constructing the potential curves in Fig. 11 we used H_2O geometries obtained using correlated wave functions. In a sense, Fig. 11 "cheats" a little bit by using an H_2O geometry computed at a correlated level of theory, which will modify the H_2O multipole moments relative to the fully-relaxed XPS geometry, where monomer correlation is absent.

Because analytic gradients for XPol + SAPT are not available, in efforts to optimize larger clusters we must be quite judicious in our choice of basis set. Thus, the comparisons for $F^-(H_2O)$ in Table 2 include a variety of Karlsruhe basis sets, as we have found that these are relatively cost-effective as compared to Dunning- or Pople-style basis sets that afford a similar level of accuracy. Comparing the XPS results in Table 2 to the MP2/aTZ benchmark, we see that even in very large basis sets, the XPS method consistently overestimates the $F \cdots O$ distance by 0.11–0.25 Å and underestimates the $F \cdots H–O$ angle by a few degrees, similar to results obtained for optimized $(H_2O)_6$ geometries.

Table 2 Comparison of optimized XPS(HF)-resp-CHELPG geometries and binding energies (BEs) for $F^-(H_2O)$, using a variety of basis sets. Benchmark MP2 and CCSD(T) values are also reported

Method	$r(F \cdots H)$ Å	$\angle(F \cdots H–O)$ (°)	BE/ kcal mol^{-1}
MP2/CBS			26.93[a]
CCSD(T)/CBS			27.20[a]
MP2/aug-cc-pVTZ	1.375	177.2	27.04[b]
XPS/6-311++G(3df,3pd)	1.513	172.6	25.80
XPS/aug-cc-pVTZ	1.508	173.2	25.61
XPS/def2-SVPD	1.591	169.6	23.50
XPS/def2-TZVP	1.539	171.4	26.81
XPS/def2-TZVPD	1.622	171.3	22.60
XPS/def2-TZVPP	1.512	172.3	27.71
XPS/def2-TZVPPD	1.578	172.2	23.64
XPS/def2-QZVPD,TZVPP[c]	1.484	174.8	28.84

[a] From ref. 82. [b] Average of counterpoise-corrected and uncorrected results. [c] Basis consists of def2-QZVPD for H and def2-TZVPP for O and F.

The fact that XPol + SAPT calculations produce hydrogen-bond angles that are typically a few degrees less linear than benchmark results may be due to orbital overlap effects, present in supersystem Hartree–Fock calculations but not in XPol calculations, that provide a driving force for linear $X \cdots H–O$ bonds.[83] However, this effect amounts to only a few degrees, and in our view the more significant problem is the error in H-bond lengths. Because this artifact persists across a wide variety of basis sets (including Karlsruhe basis sets that are lacking in diffuse functions), we suspect that it is not a deficiency of the intermolecular perturbation theory *per se*, but rather originates mostly in the lack of monomer correlation. When the $F^-(H_2O)$ geometry is optimized at the MP2/aTZ level, the O–H bond length for the $F \cdots H–O$ moiety is 0.06 Å shorter than when the dimer geometry is optimized at the HF/aTZ level, hence one might attribute ≈0.06 Å of the overly-long $F \cdots O$ distances computed at the XPS(HF) level simply to the Hartree–Fock description of the H_2O geometry, with the remaining error attributable to changes in the multipole moments when a correlated H_2O geometry is used, as well as inherent deficiencies (*e.g.*, absence of charge transfer) in the XPol procedure itself.

An immediately practical result that we glean from the basis-set comparison in Table 2 is that the def2-TZVP and def2-TZVPP basis sets affording binding energies that are within 0.5 kcal mol^{-1} (2%) of the best available benchmark, and are relatively economical [74 and 90 basis functions, respectively, for $F^-(H_2O)$, as compared to 114 basis functions for 6-311++G(3df,3pd) and 138 for aTZ, which were the basis sets one might have recommended based on the results for H-bonded systems in §III A and §III B].

To further economize the optimizations, we must decide which embedding charges to use and whether to employ the CPHF induction correction. Although we find, in general, that CHELPG charges are more accurate than Löwdin charges, they are also considerably more expensive to compute. After some experimentation, we found that the def2-TZVPP basis set, in conjunction with Löwdin embedding charges and *without* the CPHF correction affords a good BE for $F^-(H_2O)$: 27.16 kcal mol^{-1} as compared to 27.04 kcal mol^{-1} computed at the MP2/aTZ level. Undoubtedly, there is substantial cancellation of errors at work in these results, but at a practical level, this level of theory is affordable enough for finite-difference geometry optimizations in somewhat larger clusters.

Results of these optimizations, for several different $F^-(H_2O)_n$, $Cl^-(H_2O)_n$, and $OH^-(H_2O)_n$ clusters, are summarized in Tables 3 and 4. Whereas we experimented with turning off the three-body induction couplings in §III A, before ultimately concluding that this is not a good idea, here we include these terms in all cluster calculations, since they are a natural part of the second-order intermolecular perturbation theory for systems composed of three or more monomers. (The CPHF correction, on the other hand, is "extra", in the sense that it is an infinite-order correction. It is *not* used here, in order to make the geometry optimizations tractable.)

For each $X^-(H_2O)_n$ cluster that we consider, the XPS-optimized geometry is structurally quite similar to the MP2-optimized

Table 3 Mean absolute deviations between $X^-(H_2O)_n$ cluster geometries optimized at the XPS(HF)-Löwdin/def2-TZVP level, relative to MP2/aTZ benchmark geometries

		Angle (°)	
Cluster	$r(X\cdots H)$/Å	$X\cdots H–O$	$H\cdots X\cdots H$
$F^-(H_2O)_1$	0.163	5.9	—
$F^-(H_2O)_2$	0.139	9.3	16.2
$F^-(H_2O)_3$	0.104	4.6	1.1
$F^-(H_2O)_6$	0.075	1.7	1.8
$Cl^-(H_2O)_1$	0.172	14.5	—
$Cl^-(H_2O)_2$	0.145	8.6	1.0
$Cl^-(H_2O)_3$	0.117	3.3	0.7
$HO^-(H_2O)_1$	0.268	7.3	—
$HO^-(H_2O)_2$	0.193	10.7	42.4
$HO^-(H_2O)_3$	0.189	16.1	24.8

Table 4 Comparison of binding energies (in kcal mol^{-1}) for $X^-(H_2O)_n$ clusters. All XPS(HF) results used the def2-TZVP basis set with Löwdin embedding charges

		XPS(HF)	
Cluster	$MP2^a$/aTZ	MP2 geom.	XPS geom.
$F^-(H_2O)_1$	27.04	23.70	26.02
$F^-(H_2O)_2$	47.90	45.31	47.73
$F^-(H_2O)_3$	66.01	65.62	67.50
$F^-(H_2O)_6$	111.68	108.00	112.20
$Cl^-(H_2O)_1$	15.03	14.03	14.85
$Cl^-(H_2O)_2$	30.04	28.52	29.85
$Cl^-(H_2O)_3$	45.99	44.25	45.89
$HO^-(H_2O)_1$	26.51	21.85	26.70
$HO^-(H_2O)_2$	48.09	45.59	49.64
$HO^-(H_2O)_3$	66.72	65.40	70.82

[a] Average of counterpoise-corrected and uncorrected results.

geometry, although certainly the monomer geometries will differ since they are correlated at the MP2 level but not at the XPS level. A close examination of the *intermolecular* geometrical parameters (Table 3) demonstrates that the XPS results do not degrade with increasing cluster size; in fact, the geometrical parameters are slightly more accurate, on average, in the larger clusters, with the exception of the $X\cdots H–O$ angles in $HO^-(H_2O)_3$.

Binding energies at the XPS-optimized geometries are reported in Table 4, where they are compared to MP2 benchmarks as well as XPS binding energies computed at MP2-optimized geometries. As seen previously for $(H_2O)_n$ clusters, the XPS binding energies are typically much more accurate when the geometry is optimized self-consistently rather than using the MP2 geometry. Upon self-consistent optimization, XPS binding energy errors do not exceed 1.5 kcal mol^{-1} (or 0.5 kcal mol^{-1}/water molecule), except for the $HO^-(H_2O)_3$ system. For this particular system, the water molecules tend to aggregate together upon XPS optimization, which enhances the binding energy between the water molecules and overstabilizes the cluster, relative to the MP2 benchmark.

D. Dispersion-dominated complexes

Table 5 Binding energies (BEs) for several of the S22 dimers, computed at the XPS(HF)-Löwdin/def2-TZVP level and at the CCSD(T)/CBS level

	BE/kcal mol^{-1}	
Dimer	XPS	CCSD(T)
$CH_4\cdots CH_4$ (D_{3d})	0.24	0.53
$C_2H_4\cdots C_2H_4$ (D_{2d})	0.97	1.48
$C_2H_4\cdots C_2H_2$ (C_{2v})	1.48	1.50
$NH_3\cdots NH_3$ (C_{2h})	2.88	3.15
$H_2O\cdots H_2O$ (C_s)	5.02	5.07

To contrast the results for strongly H-bonded systems presented in §III A and §III C, we next examine in detail a few dimers that are bound primarily by dispersion. We have seen that a meaningful comparison to benchmark BEs typically requires geometry optimization, which at present is quite expensive for XPol + SAPT calculations. Thus, we have selected three of the smallest dispersion-bound systems from the S22 database [$(CH_4)_2$, $(C_2H_4)_2$, and $C_2H_4\cdots C_2H_2$] for geometry optimization and subsequent calculation of the binding energy. The results are presented in Table 5, where we make comparison to two small H-bonded dimers, $(H_2O)_2$ and $(NH_3)_2$, where geometry optimization is also possible. The basis set (def2-TZVP) was selected, after some experimentation, to provide reasonable binding energies for all five systems, and we observe that in an absolute sense, the BEs for the dispersion-bound dimers are not substantially less accurate than they are for the H-bonded dimers, although the fractional errors are much larger (up to 55%) for the dispersion-bound systems.

One-dimensional potential energy scans for $(C_2H_4)_2$ and $(C_6H_6)_2$ reveal errors in the minimum-energy geometries of no more than 0.1 Å and 0.04 Å, respectively, as compared to high-level *ab initio* results. These errors are comparable to the errors in H-bond distances that we observed for $(H_2O)_6$ and for $X^-(H_2O)_n$. The CPHF correction is found to be entirely negligible in these nonpolar systems, and the choice of embedding charges also makes very little difference: less than 0.25 kcal mol^{-1} in the BE of $(C_6H_6)_2$, for example. The choice of basis set is crucial, however, since the second-order dispersion formula is very sensitive to this choice.

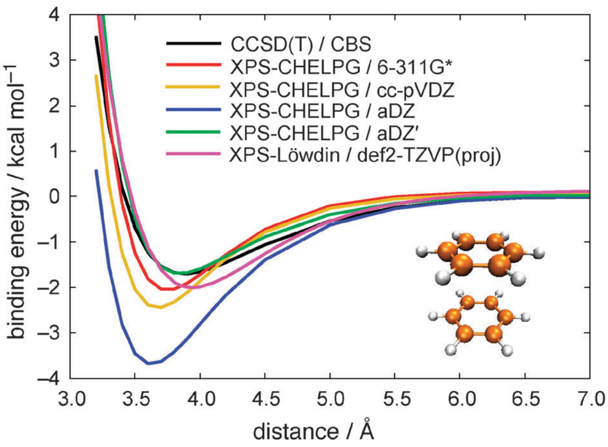

Fig. 12 Potential energy curves (for fixed monomer geometries) of the "sandwich" isomer of $(C_6H_6)_2$ computed at various XPS levels of theory. (The CPHF correction is not employed, as it is negligible for this system.) Benchmark CCSD(T)/CBS results are taken from ref. 84. Figure adapted from ref. 9 with additional new data; copyright 2011 American Institute of Physics.

The aDZ′(proj) basis set, with its reduced set of diffuse functions (see §III B), affords a nearly quantitative XPS(HF) potential energy curve for the D_{6h} "sandwich" isomer of $(C_6H_6)_2$, as shown in Fig. 12. (Results for the other two isomers are also quite good, albeit not quantitative.[9]) These results are consistent with the favorable error cancellation reported by Sherrill and co-workers in SAPT0/aDZ′ calculations for a larger set of dispersion-dominated systems.[2,50,80] Unfortunately, the $F^-(H_2O)$ complex is underbound by at least 3 kcal mol^{-1} when the aDZ′(proj) basis is used (see §III C). The def2-TZVP(proj) basis performs much better, and affords the best results for $F^-(H_2O)$ of any basis except for the fully-augmented ones that perform poorly for $(C_6H_6)_2$. As such, the def2-TZVP(proj) basis appears to be an affordable comprise for XPS(HF) calculations that provides a semi-quantitative description of both strongly H-bonded systems such as $F^-(H_2O)$ and dispersion-dominated systems such as $(C_6H_6)_2$.

IV. Summary and conclusions

In this work, we have reformulated the XPol + SAPT formalism first introduced in ref. 9, to obtain a systematically-improvable methodology for rapid, first-principles energy computations in large systems composed of small molecules. The method starts with a monomer-based SCF calculation in which a supersystem embedding potential is added in a variational way, based on the XPol idea of Gao and co-workers.[10] The cost of this step is $\mathcal{O}(N)$ with respect to the number of monomers, with a small prefactor.[9] On top of this modified XPol calculation, we add second-order intermolecular perturbation theory, based on a modified version of the two-body SAPT formalism.[15] This combination results in a method cost scales as $\mathcal{O}(N^2)$, although we have argued[9] that with parallelization and distance-based thresholding, the cost in wall time can be reduced to $\mathcal{O}(N)$ given a sufficient number of processors, since the bottleneck step is $N(N-1)/2$ pairwise SAPT calculations that are completely independent of one another.

The present version of XPol + SAPT represents a many-body generalization of the "SAPT0" methodology[2] and could, in principle, be extended to include intramolecular electron correlation (neglected in the version that is currently implemented) as well as higher-order intermolecular terms and certain exchange interactions that have been neglected here owing to their complexity. The inclusion of higher-order correlation terms amounts to a many-body generalization of the existing hierarchy of two-body SAPT methods [SAPT0, SAPT2, SAPT2+, SAPT2+(3), SAPT2+3].[2,50] Three-body SAPT contributions that go beyond the single-exchange approximation[34,35] could also be incorporated, although all of these additions would increase the cost of the method relative to the version described here.

We are ultimately interested in developing an affordable *ab initio* method for molecular simulations, and as such we have sought to keep the cost low for the pairwise SAPT calculations. The cost of each such calculation grows as $\mathcal{O}(N_{\rm bf}^5)$, where $N_{\rm bf}$ represents the number of monomer basis functions. The prefactor in this scaling expression has been reduced by means of an RI approximation,[50] nevertheless cost considerations dictate that we be very judicious regarding the choice of monomer basis set. In this work, we have explored a wide range of basis sets for a variety of non-covalent clusters ranging from $(H_2O)_n$ clusters with binding energies in excess of 200 kcal mol^{-1}, binary anion–water complexes with binding energies as large as 25 kcal mol^{-1}, and dispersion-bound complexes with binding energies of $<$ 1 kcal mol^{-1}. From these calculations, we conclude that although the XPol + SAPT methodology is parameter-free in principle, the accuracy of the low-order version that has been implemented so far is quite sensitive to the choice of monomer basis set. As with the analogous two-body SAPT0 method,[50] these low-order XPS(HF) calculations rely on error cancellation for accuracy and therefore selection of an appropriate basis set is crucial and, in effect, functions as an adjustable parameter.

Our results indicate that systems with very different non-covalent interactions place very different demands on the monomer basis set. When the binding energy is dominated by electrostatics and polarization, high-quality basis sets (*e.g.*, aug-cc-pVTZ) perform best, but such basis sets substantially overestimate the binding energies of dispersion-bound systems. The origin of this artifact traces ultimately to the MP2-like dispersion formula used in SAPT0, which significantly overestimates the dispersion interaction in diffuse basis sets. A partially-augmented basis (aug-cc-pVDZ′) has been suggested for SAPT0 calculations in dispersion-bound systems,[2,50,80] and while this basis works well for the S22 database, results are less accurate for anion–H_2O complexes. Fortunately, we have identified an affordable, compromise basis set, def2-TZVP, that appears to provide semi-quantitative results for both strongly H-bonded and dispersion-bound complexes. This offer a promising route to realistic applications in the near future, with the existing version of the methodology. The main technical hurdles that remain are the need to parallelize the pairwise SAPT calculations, which is not fundamentally difficult, and the need to implement analytic energy gradients, which is more challenging but in some sense resembles a more complicated version of the RIMP2 analytic gradient.[85]

Regarding future improvements to the theory, the most pressing problem (in our view) is the tremendous sensitivity to the choice of monomer basis set. Lack of intramolecular electron correlation is also an important shortcoming, even in cases where its impact on binding energies is small, because geometries and vibrational frequencies suffer from this defect as well. Both of these issues might share a common solution, however. First of all, we note that when electrostatics and induction dominate the intermolecular interactions, XPol + SAPT calculations behave in a generally systematic way with respect to the choice of basis set, whereas for dispersion-bound systems this is not the case. This observation suggests to us that the SAPT0 sum-over-states dispersion formula is a prime target for improvement. At the same time, a DFT description of the monomers would be a relatively low-cost way to incorporate intramolecular electron correlation, except that this approach further exacerbates problems with the description of dispersion interactions. Therefore, we suggest that a promising way forward is to incorporate so-called SAPT(DFT) methods[43–46,48] within the XPol + SAPT formalism. In SAPT(DFT), the sum-over-states dispersion formula is replaced by a generalized Casimir-Polder formula involving frequency-dependent density susceptibilities for the monomers that are obtained by solving coupled-perturbed Kohn–Sham equations. This eliminates energy denominators in the dispersion formula, which are the primary source of the basis-set dependence, and it seems reasonable that the basis-set convergence of the density susceptibilities might be more systematic. We hope to report on such an extension in future work.

Appendix: CHELPG charges

Here, we provide explicit working equations for the CHELPG charges and for the derivatives of these charges, $(\boldsymbol{\Lambda}_J)_{\mu\nu}$, that are required in the XPol procedure [see eqn (9)]. This material also serves to document the manner in which the CHELPG algorithm is implemented in the Q-CHEM program,[49] which is important in view of the fact that our implementation utilizes a weighted least-squares approach that is not described in the original CHELP or CHELPG papers.[21–23] By applying a smoothing function to the weights, this procedure ensures that charges are continuous functions of the nuclear coordinates, despite the reliance on a real-space grid. Some of the material in this Appendix was discussed already in the Supplementary Material that accompanies ref. 9, but in hindsight the notation in that reference is slightly misleading. This Appendix is offered as a clarification.

1. Charge derivatives

By definition,[21,22] the CHELPG atomic charges are the set of charges $\{q_J\}$ whose electrostatic potential, $\phi(\vec{R})$, is the closest (in a least-squares sense) to the true molecular electrostatic potential, $\Phi(\vec{R})$, subject to the constraint that the CHELPG charges must sum to the total charge of the system, Q. Both electrostatic potentials are evaluated on a real-space grid. Let $\Phi_k = \Phi(\vec{R}_k)$ denote the true molecular electrostatic potential, evaluated at the kth grid point:

$$\Phi_k = \sum_J^{N_{\text{atoms}}} \frac{Z_J}{|\vec{R}_k - \vec{R}_J|} - \sum_{\mu\nu} (\mathbf{I}_k)_{\mu\nu} P_{\mu\nu}. \tag{A.1}$$

The electrostatic potential $\phi_k = \phi(\vec{R}_k)$ generated by the charges q_J is

$$\phi_k = \sum_J^{N_{\text{atoms}}} \frac{q_J}{|\vec{R}_k - \vec{R}_J|}. \tag{A.2}$$

The CHELPG charges are defined as the ones that minimize the Lagrangian

$$\mathcal{L} = \sum_k^{N_{\text{grid}}} w_k(\Phi_k - \phi_k)^2 + \lambda\left(\sum_K^{N_{\text{atoms}}} q_K - Q\right), \tag{A.3}$$

where λ is a Lagrange multiplier and w_k is the weight given to the kth grid point, as defined below.

A formula for q_J is obtained by setting $\partial\mathcal{L}/\partial\lambda = 0$ and $\partial\mathcal{L}/\partial q_J = 0$. Solving the resulting $N_{\text{atoms}} + 1$ linear equations and eliminating λ affords

$$q_J = g_J + \alpha \sum_K (\mathbf{G}^{-1})_{JK}, \tag{A.4}$$

where the matrix $\mathbf{G}$ is defined by

$$G_{IJ} = \sum_k^{N_{\text{grid}}} w_k |\vec{R}_k - \vec{R}_I|^{-1} |\vec{R}_k - \vec{R}_J|^{-1} \tag{A.5}$$

and the vector $\mathbf{g} = \mathbf{G}^{-1}\mathbf{e}$ is defined in terms of a vector $\mathbf{e}$ whose elements are

$$e_J = \sum_k^{N_{\text{grid}}} \frac{w_k \Phi_k}{|\vec{R}_k - \vec{R}_J|}. \tag{A.6}$$

Finally,

$$\alpha = \frac{Q - \sum_J g_J}{\sum_{J,K} (\mathbf{G}^{-1})_{JK}}. \tag{A.7}$$

Note that $\mathbf{G}$ is independent of the density matrix elements, $P_{\mu\nu}$, but e_J (and therefore α) depends on $P_{\mu\nu}$ *via* the electrostatic potential, eqn (A.1)

Eqn (A.4) represents a formal solution to the so-called normal equations that define the least-squares problem. Especially in large molecules, the design matrix of the CHELPG least-squares problem may be rank-deficient, such that the effective dimensionality of the data set is less than N_{atoms}, and therefore N_{atoms} statistically-significant charges cannot be determined.[86,87] In such cases, an alternative procedure based on singular value decomposition,[87] which does not entail direct solution of the normal equations, may be desirable. We have not found it necessary to use such a procedure in the examples considered here or in ref. 9, so this has not been implemented.

It is straightforward to evaluate the derivative of eqn (A.4) with respect to $P_{\mu\nu}$. The result is a formula for the charge derivatives, $(\boldsymbol{\Lambda}_K)_{\mu\nu}$

$$\begin{aligned}(\boldsymbol{\Lambda}_K)_{\mu\nu} = &-\sum_L (\mathbf{G}^{-1})_{KL}(\Xi_L)_{\mu\nu} \\ &+ \left(\frac{\sum_J (\mathbf{G}^{-1})_{KJ}}{\sum_{I,J} (\mathbf{G}^{-1})_{IJ}}\right) \sum_{LM} (\mathbf{G}^{-1})_{LM}(\Xi_M)_{\mu\nu}\end{aligned} \tag{A.8}$$

where

$$(\Xi_M)_{\mu\nu} = \sum_k^{N_{\text{grid}}} \frac{w_k (\mathbf{I}_k)_{\mu\nu}}{|\vec{R}_k - \vec{R}_M|}. \tag{A.9}$$

2. Smooth implementation

In our implementation, the weights $\{w_k\}$ associated with the grid points $\{\vec{R}_k\}$ are chosen to ensure that the CHELPG charges are continuous functions of the nuclear coordinates. We set

$$w_k = w_k^{\text{LR}} \prod_J^{N_{\text{atoms}}} A_k^J, \tag{A.10}$$

where w_k^{LR} is a long-range weighting function that is discussed below, and each A_k^J is an atomic switching function defined as

$$A_k^J = \begin{cases} 0 & \text{if } |\vec{R}_k - \vec{R}_J| < R_{\text{cut},J}^{\text{short}} \\ \tau(|\vec{R}_k - \vec{R}_J|;\, R_{\text{cut},J}^{\text{short}}, R_{\text{on},J}) & \text{if } R_{\text{cut},J}^{\text{short}} \le |\vec{R}_k - \vec{R}_J| < R_{\text{on},J}\,. \\ 1 & \text{otherwise} \end{cases} \tag{A.11}$$

The cutoff parameters $R_{\mathrm{cut},J}^{\mathrm{short}}$ and $\mathrm{R}_{\mathrm{on},J}$ are given below. The tapering function, τ, is taken from ref. 88:

$$\tau(R; R_{\mathrm{cut}}, R_{\mathrm{off}}) = \frac{(R - R_{\mathrm{cut}})^2(3R_{\mathrm{off}} - R_{\mathrm{cut}} - 2R)}{(R_{\mathrm{off}} - R_{\mathrm{cut}})^3}. \quad \text{(A.12)}$$

This function changes smoothly from $\tau = 0$ at $R = R_{\mathrm{cut}}$ to $\tau = 1$ at $R = R_{\mathrm{off}}$, thus the parameters $R_{\mathrm{cut},J}^{\mathrm{short}}$ and $\mathrm{R}_{\mathrm{on},J}$ in eqn (A.11) function as short- and long-range cutoffs, respectively, for the grid points $\vec{R}_k$, with respect to the position of atom J.

To determine the long-range weight, w_k^{LR}, we first find the minimum distance from the grid point $\vec{R}_k$ to any nucleus:

$$R_k^{\min} = \min_J |\vec{R}_k - \vec{R}_J|. \quad \text{(A.13)}$$

We then define

$$w_k^{\mathrm{LR}} = \begin{cases} 1 & \text{if } R_k^{\min} < R_{\mathrm{cut}}^{\mathrm{long}} \\ 0 & \text{if } R_k^{\min} > R_{\mathrm{off}} \\ 1 - \tau(R_k^{\min}; R_{\mathrm{cut}}^{\mathrm{long}}, R_{\mathrm{off}}) & \text{otherwise} \end{cases}. \quad \text{(A.14)}$$

To evaluate the weights, we set the short-range cutoff $R_{\mathrm{cut},J}^{\mathrm{short}}$ equal to the Bondi radius[89,90] for atom J. (Essentially identical results are obtained if we instead use radii obtained from the UFF force field.[91] These have the advantage that they are defined for the entire periodic table.) We set $R_{\mathrm{off}} = 3.0$ Å, $R_{\mathrm{on},J} = R_{\mathrm{cut},J}^{\mathrm{short}} + \Delta r$, and $R_{\mathrm{cut}}^{\mathrm{long}} = R_{\mathrm{off}} - \Delta r$, where the quantity Δr controls how rapidly a grid point's weight is scaled to zero by the tapering function. We use a fairly small value, $\Delta r = 0.1$ bohr, due to concerns about possible discontinuities during finite-difference geometry optimizations. We have not encountered any problems when using these values, although it is possible that they might need to be modified to ensure sufficient smoothness for molecular dynamics applications.

Acknowledgements

This work was supported by a National Science Foundation CAREER award (CHE-0748448) to J.M.H. Calculations were performed at the Ohio Supercomputer Center under project no. PAS-0291. J.M.H. is an Alfred P. Sloan Foundation Fellow and a Camille Dreyfus Teacher-Scholar. L.D.J. acknowledges support from a Presidential Fellowship awarded by The Ohio State University.

References

1 G. S. Tschumper, in *Reviews in Computational Chemistry*, ed. K. B. Lipkowitz and T. R. Cundari, Wiley-VCH, 2009, vol. 26, ch. 2, pp. 39–90.
2 E. G. Hohenstein and C. D. Sherrill, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 304.
3 M. Elstner, *Theor. Chem. Acc.*, 2006, **116**, 316.
4 D. Riccardi, P. Schaefer, Y. Yang, H. Yu, N. Ghosh, X. Prat-Resina, P. Koenig, G. Li, D. Xu, H. Guo, M. Elstner and Q. Cui, *J. Phys. Chem. B*, 2006, **110**, 6458.
5 E. R. Johnson, I. D. Mackie and G. A. Di Labio, *J. Phys. Org. Chem.*, 2009, **22**, 1127.
6 M. E. Foster and K. Sohlberg, *Phys. Chem. Chem. Phys.*, 2010, **12**, 307.
7 S. Grimme, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 211.
8 P. G. Bolhuis and C. Dellago, in *Reviews in Computational Chemistry*, ed. K. B. Lipkowitz, Wiley-VCH, 2011, vol. 27, ch. 3, pp. 111–210.
9 L. D. Jacobson and J. M. Herbert, *J. Chem. Phys.*, 2011, **134**, 094118.
10 W. Xie and J. Gao, *J. Chem. Theory Comput.*, 2007, **3**, 1890.
11 W. Xie, L. Song, D. G. Truhlar and J. Gao, *J. Chem. Phys.*, 2008, **128**, 234108.
12 L. Song, J. Han, Y.-L. Lin, W. Xie and J. Gao, *J. Phys. Chem. A*, 2009, **113**, 11656.
13 A. Cembran, P. Bao, Y. Wang, L. Song, D. G. Truhlar and J. Gao, *J. Chem. Theory Comput.*, 2010, **6**, 2469.
14 B. Jeziorski, R. Moszynski, A. Ratkiewicz, S. Rybak, K. Szalewicz and H. L. Williams, in *Methods and Techniques in Computational Chemistry METECC-94*, ed. E. Clementi, STEF, Cagliari, 1993, vol. B, ch 3, pp. 79–129.
15 B. Jeziorski, R. Moszynski and K. Szalewicz, *Chem. Rev.*, 1994, **94**, 1887.
16 M. S. Gordon, D. G. Fedorov, S. R. Pruitt and L. V. Slipchenko, *Chem. Rev.*, 2012, **112**, 632.
17 R. M. Richard and J. M. Herbert, (in preparation).
18 D. G. Fedorov and K. Kitaura, *J. Phys. Chem. A*, 2007, **111**, 6904.
19 R. Khaliullin, M. Head-Gordon and A. T. Bell, *J. Chem. Phys.*, 2006, **124**, 204105.
20 C. A. White, B. G. Johnson, P. M. W. Gill and M. Head-Gordon, *Chem. Phys. Lett.*, 1994, **230**, 8.
21 S. R. Cox and D. E. Williams, *J. Comput. Chem.*, 1981, **2**, 304.
22 L. E. Chirlian and M. M. Francl, *J. Comput. Chem.*, 1987, **8**, 894.
23 C. M. Breneman and K. B. Wiberg, *J. Comput. Chem.*, 1990, **11**, 361.
24 Y. Chen and H. Li, *J. Phys. Chem. A*, 2010, **114**, 11719.
25 N. Turki, A. Milet, A. Rahmouni, O. Ouamerali, R. Moszynski, E. Kochanski and P. E. S. Wormer, *J. Chem. Phys.*, 1998, **109**, 7157.
26 E. E. Dahlke and D. G. Truhlar, *J. Chem. Theory Comput.*, 2007, **3**, 46.
27 E. E. Dahlke and D. G. Truhlar, *J. Chem. Theory Comput.*, 2007, **3**, 1342.
28 G. J. O. Beran, *J. Chem. Phys.*, 2009, **130**, 164115.
29 A. Sebetci and G. J. O. Beran, *J. Chem. Theory Comput.*, 2010, **6**, 155.
30 G. Chałasiński, M. M. Szczęśniak and R. A. Kendall, *J. Chem. Phys.*, 1994, **101**, 8860.
31 T. P. Tauer and C. D. Sherrill, *J. Phys. Chem. A*, 2005, **109**, 10475.
32 A. L. Ringer and C. D. Sherrill, *Chem.–Eur. J.*, 2008, **14**, 2542.
33 A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry*, Macmillan, New York, 1982.
34 V. F. Lotrich and K. Szalewicz, *J. Chem. Phys.*, 1997, **106**, 9668.
35 V. F. Lotrich and K. Szalewicz, *J. Chem. Phys.*, 2000, **112**, 112.
36 G. Chałasiński, and B. Jeziorski, *Mol. Phys.*, 1976, **32**, 81.
37 V. F. Lotrich and K. Szalewicz, *J. Chem. Phys.*, 1997, **106**, 9688.
38 K. U. Lao and J. M. Herbert, *J. Phys. Chem. A*, 2012, **116**, 3042.
39 R. Moszynski, P. E. S. Wormer, B. Jeziorski and A. van der Avoird, *J. Chem. Phys.*, 1995, **103**, 8058.
40 P. S. Zuchowski, R. Podeszwa, R. Moszynski, B. Jeziorski and K. Szalewicz, *J. Chem. Phys.*, 2008, **129**, 084101.
41 H. L. Williams and C. F. Chabalowski, *J. Phys. Chem. A*, 2001, **105**, 646.
42 The SAPT(KS) approach should not be confused with "SAPT(DFT)", where the dispersion interaction is evaluated in terms of a Casimir-Polder-type expression involving frequency-dependent density susceptibilities for the monomers, as opposed to the MP2-like sum-over-states formula used in ordinary second-order SAPT. See refs. 43–46 for a discussion of SAPT(DFT).
43 A. J. Misquitta, B. Jeziorski and K. Szalewicz, *Phys. Rev. Lett.*, 2003, **91**, 033201.
44 A. J. Misquitta, R. Podeszwa, B. Jeziorski and K. Szalewicz, *J. Chem. Phys.*, 2005, **123**, 214103.
45 A. Heßelmann and G. Jansen, *Chem. Phys. Lett.*, 2003, **367**, 778.
46 A. Heßelman, G. Jansen and M. Schutz, *J. Chem. Phys.*, 2005, **122**, 014103.
47 A. J. Misquitta and K. Szalewicz, *Chem. Phys. Lett.*, 2002, **357**, 301.
48 A. J. Misquitta, in *Handbook of Computational Chemistry*, ed. J. Leszczynski, Springer Science + Business Media, 2012, ch. 6, pp. 157–193.

49 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. D. 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.*, 2006, **8**, 3172.

50 E. G. Hohenstein and C. D. Sherrill, *J. Chem. Phys.*, 2010, **132**, 184111.

51 F. Weigend, A. Köhn and C. Hättig, *J. Chem. Phys.*, 2002, **116**, 3175.

52 H. L. Williams, E. M. Mas and K. Szalewicz, *J. Chem. Phys.*, 1995, **103**, 7374.

53 M. Head-Gordon, D. Maurice and M. Oumi, *Chem. Phys. Lett.*, 1995, **246**, 114.

54 C. J. Burnham and S. S. Xantheas, *J. Chem. Phys.*, 2002, **116**, 1479.

55 D. M. Bates and G. S. Tschumper, *J. Phys. Chem. A*, 2009, **113**, 3555.

56 The "haTZ" basis designation indicates a "heavy augmented" basis consisting of aug-cc-pVTZ for the heavy atoms and cc-pVTZ for the hydrogen atoms.

57 That is, we compute the RMSD for the optimal (in a least-squares sense) superposition of the XPol + SAPT and MP2 geometries. This superposition was computed using the "superpose" module of the Tinker program[58].

58 Tinker, version 4.2 (http://dasher.wustl.edu/tinker).

59 For the purpose of this calculation, we define a hydrogen bond to exist whenever the MP2-optimized geometry exhibits a distance $r(\mathrm{H}\cdots\mathrm{O}) < 3.0$ Å and an H···O–H angle within 35° of linearity.

60 XPS-optimized $(H_2O)_n$ geometries are taken from our previous work.[9] Due to the considerable expense of these finite-difference optimizations, in some cases the geometry is not quite relaxed all the way to a proper local minimum, at least not according to Q-CHEM default geometry optimization thresholds.[49] This is especially true in the larger clusters, but in these cases we estimate that further optimization would increase the BEs by no more than a few kcal mol^{-1},[9] or in other words a few percent of the benchmark BE. This increase would move the XPol + SAPT results slight closer to the MP2/CBS benchmarks.

61 S. S. Xantheas, *J. Chem. Phys.*, 1996, **104**, 8821.

62 S. S. Xantheas, C. J. Burnham and R. J. Harrison, *J. Chem. Phys.*, 2002, **116**, 1493.

63 S. S. Xantheas and E. Aprà, *J. Chem. Phys.*, 2004, **120**, 823.

64 G. S. Fanourgakis, E. Aprà and S. S. Xantheas, *J. Chem. Phys.*, 2004, **121**, 2655.

65 S. Bulusu, S. Yoo, E. Aprà, S. S. Xantheas and X. C. Zeng, *J. Phys. Chem. A*, 2006, **110**, 11781.

66 S. Yoo, E. Aprà, X. C. Zeng and S. S. Xantheas, *J. Phys. Chem. Lett.*, 2010, **1**, 3122.

67 MP2-optimized geometries from refs. 61–66 were kindly provided by Sotiris Xantheas.

68 For many-body systems, what we mean by "SAPT" is simply a pairwise-additive version of ordinary two-body SAPT. In particular, we carry out the theory indicated in §II C, but with $|\Psi_0\rangle$ equal to a direct product of gas-phase Hartree–Fock wave functions for each of the monomers, rather than XPol wave functions that have been iterated to self-consistency in the presence of embedding charges. Furthermore, the intermolecular perturbation is given by eqn (13) rather than eqn (50).

69 P. Jurečka, J. Šponer, J. Černý and P. Hobza, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1985.

70 T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall and C. D. Sherrill, *J. Chem. Phys.*, 2010, **132**, 144104.

71 H. Iikura, T. Tsuneda, T. Yanai and K. Hirao, *J. Chem. Phys.*, 2001, **115**, 3540.

72 M. A. Rohrdanz, K. M. Martins and J. M. Herbert, *J. Chem. Phys.*, 2009, **130**, 54112.

73 C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158–6170.

74 R. Baer, E. Livshits and U. Salzner, *Annu. Rev. Phys. Chem.*, 2010, **61**, 85–109.

75 A. J. Cohen, P. Mori-Sánchez and W. Yang, *Science*, 2008, **321**, 792–794.

76 We use the optimal LRC-ωPBEh parameters suggested in ref. 72, namely, ω = 0.2 bohr^{-1} and 20% short-range Hartree–Fock exchange. The notation for this functional is that suggested in refs. 72 and 77.

77 R. M. Richard and J. M. Herbert, *J. Chem. Theory Comput.*, 2011, **7**, 1296.

78 H. Iikura, T. Tsuneda, T. Yanai and K. Hirao, *J. Chem. Phys.*, 2001, **115**, 3540.

79 M. O. Sinnokrot and C. D. Sherrill, *J. Phys. Chem. A*, 2006, **110**, 10656.

80 E. G. Hohenstein and C. D. Sherrill, *J. Chem. Phys.*, 2010, **133**, 14101.

81 K. Patkowski, K. Szalewicz and B. Jeziorski, *J. Chem. Phys.*, 2006, **125**, 154107.

82 P. Weis, P. R. Kemper, M. T. Bowers and S. S. Xantheas, *J. Am. Chem. Soc.*, 1999, **121**, 3531.

83 A. E. Reed, L. A. Curtiss and F. Weinhold, *Chem. Rev.*, 1988, **88**, 899.

84 C. D. Sherrill, T. Takatani and E. G. Hohenstein, *J. Phys. Chem. A*, 2009, **113**, 10146.

85 R. A. DiStasio Jr., R. P. Steele, Y. M. Rhee, Y. Shao and M. Head-Gordon, *J. Comput. Chem.*, 2007, **28**, 839.

86 T. R. Stouch and D. E. Williams, *J. Comput. Chem.*, 1993, **14**, 858.

87 M. M. Francl, C. Carey, L. E. Chirlian and D. M. Gange, *J. Comput. Chem.*, 1996, **17**, 367.

88 O. Steinhauser, *Mol. Phys.*, 1982, **45**, 335.

89 A. Bondi, *J. Phys. Chem.*, 1964, **68**, 441.

90 R. S. Rowland and R. Taylor, *J. Phys. Chem.*, 1996, **100**, 7384.

91 A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard III and W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024.