Accurate and Efficient Quantum Chemistry Calculations for Noncovalent Interactions in Many-Body Systems: The XSAPT Family of Methods[†]

Ka Un Lao and John M. Herbert*

Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210, United States

Ⓢ *Supporting Information*

ABSTRACT: We present an overview of "XSAPT", a family of quantum chemistry methods for noncovalent interactions. These methods combine an efficient, iterative, monomer-based approach to computing many-body polarization interactions with a two-body version of symmetry-adapted perturbation theory (SAPT). The result is an efficient method for computing accurate intermolecular interaction energies in large noncovalent assemblies such as molecular and ionic clusters, molecular crystals, clathrates, or protein–ligand complexes. As in traditional SAPT, the XSAPT energy is decomposable into physically meaningful components. Dispersion interactions are problematic in traditional low-order SAPT, and two new approaches are introduced here in an attempt to improve this situation: (1) third-generation empirical atom–atom dispersion potentials, and (2) an empirically scaled version of second-order SAPT dispersion. Comparison to high-level *ab initio* benchmarks for dimers, water clusters, halide–water clusters, a methane clathrate hydrate, and a DNA intercalation complex illustrate both the accuracy of XSAPT-based methods as well as their limitations. The computational cost of XSAPT scales as $O(N^3)$–$O(N^5)$ with respect to monomer size, N, depending upon the particular version that is employed, but the accuracy is typically superior to alternative *ab initio* methods with similar scaling. Moreover, the monomer-based nature of XSAPT calculations makes them trivially parallelizable, such that wall times scale linearly with respect to the number of monomer units. XSAPT-based methods thus open the door to both qualitative and quantitative studies of noncovalent interactions in clusters, biomolecules, and condensed-phase systems.
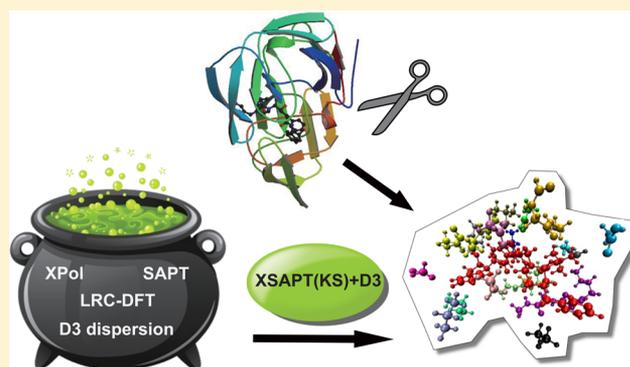
I. BACKGROUND

A. Quantum Chemistry for Noncovalent Interactions. Noncovalent or "nonbonded" interactions are responsible for the properties of a variety of complex systems ranging from the structures of both single- and double-stranded DNA,[1] drug binding to both proteins and DNA,[2,3] and also crystal engineering and crystal structure prediction.[4] Electronic structure calculations of noncovalent interactions have seen much progress in recent years, due to improvements in both algorithms and computer power. In particular, symmetry-adapted perturbation theory[5–10] (SAPT) provides a natural decomposition of noncovalent interactions into physical meaningful components (electrostatics, induction, and dispersion), along with a corresponding exchange term for each. The dispersion (van der Waals) interaction is particularly interesting as it is a purely quantum-mechanical effect arising solely from intermolecular electron correlation. Dispersion is therefore absent at the level of Hartree–Fock molecular orbital

(MO) theory and has historically been difficult to describe with density functional theory (DFT) as well, because popular semilocal functionals fail to account for long-range electron correlation.

Various strategies have been devised to incorporate dispersion into DFT, including highly parametrized meta-GGA[11] functionals where nonbonded interactions are included in the fitting set. The "Minnesota" family of functionals are prime examples of this approach.[12] Alternatively, explicit r^{-2k} dependence ($k = 3, 4, ...$) can be added to DFT *a posteriori*, via classical atom–atom potentials, in a "DFT+D" approach popularized by Grimme.[13–15] The ωB97X-D functional[16] is one of the best-performing examples of a DFT+D functional. Finally, "double hybrid" functionals that mix second-order Møller–Plesset (MP2) correlation with DFT,[17] and other nonlocal correlation functionals,[18,19] also do a better job of describing noncovalent interactions than traditional semilocal GGAs. However, the cost of these methods scales no better

---

[†]John Herbert was the recipient of the 2013 *The Journal of Physical Chemistry A* Lectureship, which was awarded at the 2013 National Meeting of the American Chemical Society in Indianapolis, IN.

than $O(N^3)$ with respect to total system size, N, which limits their routine application to systems with $\lesssim 100$ atoms.

Prior to the advent of these newer DFT-based approaches, the $O(N^5)$ MP2 method was considered the simplest way to incorporate dispersion in electronic structure calculations. MP2 performs well for hydrogen-bonded systems and thus continues to play a vital role in the study of water clusters[20] (and recently, bulk liquid water[21]). However, MP2 significantly overestimates π-stacking interactions and other dispersion-dominated interactions.[22,23] This behavior stems from poor effective C_6 coefficients,[24,25] which at the MP2 level correspond to an uncoupled Hartree–Fock (UCHF) description of the frequency-dependent polarizabilities for the monomers. Moreover, slow convergence of the MP2 correlation energy to the complete basis set (CBS) limit requires costly counterpoise correction[26] to eliminate basis set superposition error (BSSE). A self-consistent treatment of double excitations, i.e., the $O(N^6)$ coupled-cluster singles and doubles (CCSD) method, also underestimates $\pi \cdots \pi$ interactions,[27] and on average represents only a modest improvement upon MP2, with errors of 0.7–1.0 kcal/mol relative to converged CCSD(T) values.[28]

In short, the $O(N^7)$ CCSD(T) method remains the "gold standard" for noncovalent quantum chemistry, though there is some recent effort to explore quantum Monte Carlo techniques as an alternative.[29,30] CBS extrapolation is required to obtain converged CCSD(T) results, but higher-order electron correlation effects are consistently <0.1 kcal/mol.[31–33] The CCSD(T)/CBS limit can more affordably be obtained by adding a correction

$$\delta_{CCSD(T)} = E_{CCSD(T)} - E_{MP2} \tag{1}$$

to the MP2/CBS binding energy, as this correction is generally converged in triple-ζ basis sets,[34] whereas the MP2/CBS extrapolation requires a basis of at least aug-cc-pVQZ (aQZ) quality. Nevertheless, this approach remains prohibitively expensive except for small systems. For example, a recent CCSD(T)/aTZ calculation on $(H_2O)_{17}$ required 3.3 h on 120 000 processors simply for the "(T)" part of the calculation.[35]

The performance of various electronic structure methods that have been suggested for noncovalent interactions, and which scale better than $O(N^7)$, is summarized in Figure 1 for the S66 data set of noncovalent dimers.[36] Among these methods, the MP2/CBS results are actually the worst, and this is a direct result of severe overestimation of π-stacking interactions. The best-performing method is SCS-MI-CCSD (spin-component scaled CCSD for molecular interactions[41]), but its sixth-order scaling is also severely limiting. The MP2C approach[42–45] (MP2 with coupled dispersion) also affords very small errors, with only fifth-order scaling, but this method is formulated exclusively for dimers. Finally, Figure 1 shows selected DFT results using functionals that afford good results for S66; however, this good performance is not transferred to anionic systems such as halide–water clusters, $X^-(H_2O)_n$,[46] except for the ωB97X-V function as discussed below.

In view of these remarks, it is clear that quantum chemistry calculations with sub-kcal/mol accuracy remain out of reach for large noncovalent assemblies, such as the HIV protease + inhibitor system that is shown in Figure 2. With a binding pocket consisting of 16 nearby amino acids plus two crystallographic waters, the total system size for a reasonable quantum chemistry model system amounts to 323 atoms, or
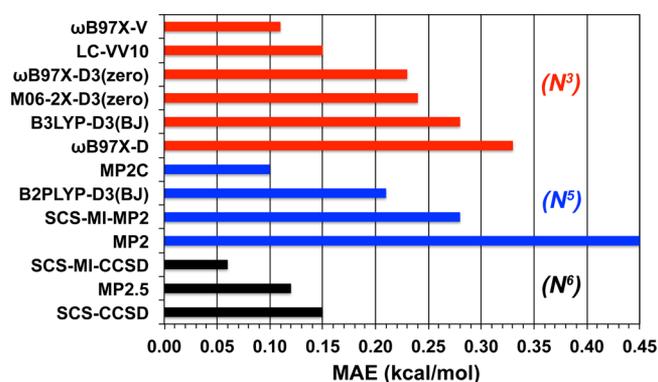
Figure 1. Mean absolute error (MAE) with respect to CCSD(T)/CBS benchmarks, for binding energies in the S66 database[36] of non-covalently bound dimers. The various methods are color-coded according to how their cost scales as a function of system size. All MP2- and CCSD-based results[36] are evaluated in the CBS limit. M06-2X-D3(zero), B3LYP-D3(BJ), and B2PLYP-D3(BJ) calculations employ the def2-QZVP basis.[37] ωB97X-V and LC-VV10 calculations use aug-cc-pVTZ.[38] The ωB97X-D and ωB97X-D3(zero) calculations use the 6-311++G(3df,3pd) basis.[39] [Here, "zero" indicates the "zero-damping" function of ref 40, which damps empirical dispersion to zero as $R \to 0$, whereas Becke–Johnson (BJ) damping damps it to a finite value.] The counterpoise correction is employed for all methods except M06-2X-D3(zero), B3LYP-D3(BJ), and B2PLYP-D3(BJ).
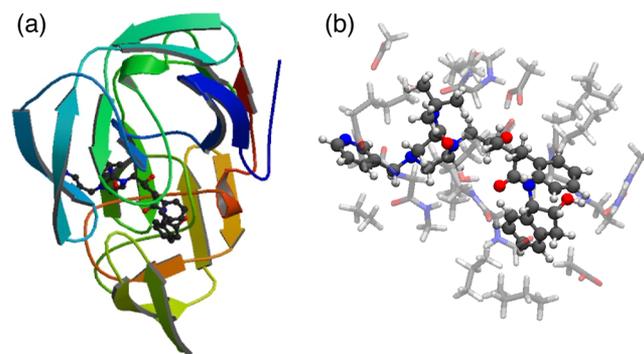
Figure 2. (a) Structure of the protease inhibitor indinavir bound to HIV protease, as obtained from PDB crystal structure 1HSG.[47] (b) An enlarged view of the binding pocket, consisting of indinavir (opaque ball-and-stick model) along with 16 amino acids and 2 crystallographic waters (translucent tubular models). [Panel b is reproduced with permission from ref 48. Copyright 2011 American Institute of Physics.]

10 626 basis functions using aug-cc-pVTZ. Fragment-based quantum chemistry methods offer a way to surmount this predicament.[49,50] This article describes a family of fragmentation methods that we have developed in an attempt to achieve sub-kcal/mol accuracy for noncovalent interactions yet remain affordable enough to be applied to systems such as the one in Figure 2b, where the monomers naturally form fragments but the protease inhibitor molecule is 92 atoms by itself. As such, any method that aims to describe this system must be efficient both for large fragments and for systems composed of a large number of fragments.

The XSAPT family of methods that is described in this article has been developed by our group over the past several years.[46,50–54] These methods employ the variational explicit polarization (XPol) method of Xie et al.[55] to generate monomer wave functions that include many-body polarization

effects and then exploit these XPol wave functions as zeroth-order states for various forms of SAPT. Whereas SAPT has traditionally been limited to dimers, XSAPT extends this methodology (including its energy decomposition analysis) to many-body systems, in an affordable way.

B. Symmetry-Adapted Perturbation Theory (SAPT). It is useful to review the original SAPT methodology.[5−9] SAPT is a direct, perturbative expansion of the intermolecular (dimer) interaction energy based on noninteracting monomer wave functions, and BSSE is avoided because subtraction of monomer energies is not required. A double perturbation expansion is employed, in which *intra*molecular electron correlation (Møller−Plesset fluctuation potentials[5,56] or a cluster *ansatz*[57−59]) is one perturbation, and the intermolecular Coulomb operators comprise the other. Methods that include intramolecular correlation, such as SAPT2+, SAPT2+(3), SAPT2 + 3, and SAPT(CCSD), are generally quite accurate[9,60] but scale no better than $O(N^7)$, the same as CCSD(T). However, an accurate SAPT calculation may be able to use a smaller basis set than CCSD(T), owing to SAPT's intrinsic lack of BSSE, and furthermore SAPT comes with an informative energy decomposition.

A comparatively low-cost means to introduce intramolecular electron correlation is to swap Kohn−Sham (KS) MOs into the SAPT formalism, an approach that has been called SAPT(KS).[61] This approach was considered and rejected more than a decade ago, however, because it fails to yield accurate intermolecular interaction energies due to inaccurate asymptotic behavior of the exchange−correlation (XC) potentials, $v_{xc}(r)$, that are obtained from standard functionals.[61−63] The asymptotic (large r) behavior should be[64]

$$v_{xc}(r) \sim -\frac{1}{r} + \Delta_\infty \tag{2}$$

where the limiting ($r \to \infty$) value is[64,65]

$$\Delta_\infty = \text{IP} + \varepsilon_{\text{HOMO}} \tag{3}$$

In eq 3, IP denotes the lowest ionization potential and $\varepsilon_{\text{HOMO}}$ is the KS eigenvalue for the highest occupied molecular orbital (HOMO). The failure of standard density-functional approximations to satisfy eq 2 leads to large errors in SAPT(KS) dispersion energies,[53] even though the energetics of strongly hydrogen-bonded systems are somewhat improved as compared to calculations that use HF wave functions for the monomers.[53]

This failure of SAPT(KS) is partially ameliorated by using an asymptotically correct v_{xc} to compute KS orbitals for the monomers,[54,63,66,67] though dispersion energies remain poor[54,62,68,69] for the same reason that MP2 dispersion energies are poor. It is possible to solve coupled KS equations to obtain frequency-dependent density susceptibilities for the monomers, and this improved approach was developed independently by Heßelmann and Jansen,[66,67,69−71] who named the method DFT-SAPT, and by Misquitta et al.,[63,68,72−75] who called it SAPT(DFT). This approach, which is closely related to MP2C, scales as $O(N^6)$ but can be reduced to $O(N^5)$ using resolution-of-identity techniques.[71,73−75]

C. "Extended" SAPT (XSAPT). Most electronic structure methods for intermolecular interactions either function as supersystem calculations or (like SAPT) are designed to compute pairwise interactions only. Benchmark noncovalent data sets have also largely focused on dimers. However, many-body contributions to the interaction energy are often quite

significant, especially in polar systems where nonadditive interactions are dominated by induction (i.e., polarization).[76,77] For example, the many-body contribution to polarization in isomers of $(H_2O)_6$ is about 10 kcal/mol, whereas electrostatic, exchange−repulsion, and dispersion interactions are nearly pairwise additive.[46,77] Likewise, electron correlation effects have often been found to be largely pairwise-additive in small clusters of small molecules, provided that many-body induction effects are incorporated self-consistently,[78] but many-body dispersion is more important in systems with a large number of monomers.[79]

Although the equations for three-body SAPT have been written down,[80−82] their cost scales as $O(N^7)$ with respect to the size of the largest trimer, whereas the present work will focus on methods that scale as $O(N^3)−O(N^5)$ with respect to the size of the largest *dimer*. In recognition of the qualitative observations above, our group has recently developed an "extended" version of SAPT in which many-body induction and polarization effects are incorporated into the zeroth-order wave functions by means of the XPol method.[55] Other components of the intermolecular interaction are included via pairwise SAPT. The result is a monomer-based method that we call XSAPT,[46,50−53] which is aimed at doing fast calculations for noncovalent assemblies, including molecular and ionic clusters but also protein−ligand binding systems such as the one depicted in Figure 2. The XPol procedure starts with a charged-embedded self-consistent field (SCF) calculation on each monomer, whose cost is therefore $O(n)$ with respect the number of monomers, n. This is followed by an embarrassingly parallelizable $O(n^2)$ pairwise SAPT calculation. SAPT-style energy decomposition analysis is available, including a term that directly measures the many-body contribution to the interaction energy.[46]

II. OVERVIEW OF XSAPT

We briefly review the theory behind XSAPT; see ref 53 for a detailed derivation. We assume that covalent bonds remain intact in the fragmentation of the system, as appropriate for noncovalent assemblies.

A. Many-Body Polarization: XPol. XPol is a fragment-based MO method that has been put forward as a means to obtain the polarization term in a "next-generation" force field.[55,83−86] Upon partitioning the system into fragments, the XPol wave function is written as a direct product of fragment wave functions, $|\Psi\rangle = |\Psi_A\rangle|\Psi_B\rangle|\Psi_C\rangle\cdots$. The XPol energy is[51]

$$E_{\text{XPol}} = \sum_{A=1}^{n} \left(2 \sum_{a \in A} \mathbf{c}_a^\dagger \mathbf{f}^A \mathbf{c}_a + E_{\text{nuc}}^A \right) + E_{\text{embed}} \tag{4}$$

where the term in parentheses is the ordinary HF energy expression for monomer A, whose Fock matrix is $\mathbf{f}^A = \mathbf{h}^A + \mathbf{J}^A - (1/2)\mathbf{K}^A$. Crucially, the MOs $\mathbf{c}_a$ for fragment A are represented using only those atomic orbitals (AOs) that are centered on atoms in fragment A. This partitioning of the basis leads to $O(n)$ scaling and furthermore excludes BSSE, by construction. (Charge transfer between fragments is also excluded, at least in small basis sets.) The quantity E_{embed} in eq 4 is an electrostatic embedding potential that could in principle be obtained from the monomer electron densities, $\rho_A = |\Psi_A|^2$, but more often consists of the charge−density interactions that arise once each ρ_A is collapsed onto some set of atom-centered point charges.[51,53]

237

dx.doi.org/10.1021/jp5098603 | *J. Phys. Chem. A* 2015, 119, 235−252

Upon variational minimization of eq 4 with respect to the c_a, one obtains a set of monomer SCF equations that involve a modified Fock matrix $\mathbf{F}^A$ for monomer A:[51,53]

$$F^A_{\mu\nu} = f^A_{\mu\nu} - \frac{1}{2} \sum_{J \notin A} q_J (\mathbf{I}_J)_{\mu\nu} + \sum_{J \in A} M_J (\mathbf{\Lambda}_J)_{\mu\nu} \tag{5}$$

The second term on the right is the interaction of fragment A with the embedding charge q_J, where $(\mathbf{I}_J)_{\mu\nu}$ is a charge–density Coulomb integral involving shell pair $\mu\nu \in A$. In the final term, $M_J = \partial E_{embed} / \partial q_J$ is easy to compute (see ref 53), and

$$(\mathbf{\Lambda}_J)_{\mu\nu} = \frac{\partial q_J}{\partial P_{\mu\nu}} \tag{6}$$

Calculation of the $\mathbf{\Lambda}$ tensor does require some nontrivial overhead when the embedding charges q_J are "CHELPG" charges[87] that are fit to reproduce the monomer electrostatic potentials.[53,88] This is our preferred choice, as it seems physically sound and moreover the use of Mulliken or Löwdin charges often leads to convergence failure in the XPol procedure, when nonminimal basis sets are employed.[51,88] In any case, the monomer XPol SCF equations ($\mathbf{F}^A\mathbf{C}^A = \mathbf{S}^A\mathbf{C}^A\varepsilon^A$) are iterated to self-consistency using a "dual SCF" procedure consisting of an outer loop over monomers and an inner loop over SCF cycles for a particular monomer, updating the embedding charges as each $|\Psi_A\rangle$ is converged.

The final term in eq 5 is equal to $\partial E_{embed}/\partial P_{\mu\nu}$ and ensures that the converged XPol energy is fully variational.[50,51,55] In contrast, the fragment Fock matrix that is traditionally used in the fragment molecular orbital (FMO) method,[89,90] and also in the electrostatically embedded many-body expansion,[91,92] is

$$F^A_{\mu\nu} = f^A_{\mu\nu} - \sum_{J \notin A} q_J (\mathbf{I}_J)_{\mu\nu} \tag{7}$$

This Fock matrix does *not* afford a variational method, because it omits the response of the embedding charges to changes in the fragment wave functions. As a result, analytic gradients for FMO and other methods that use eq 7 with self-consistent embedding charges *should*, in principle, require solution of coupled-perturbed equations,[93,94] although these response terms have often been neglected.[95–97] In addition, eq 7 omits the polarization work that diminishes the middle term in eq 5 by a factor of 2.[98,99]

B. Symmetry-Adapted Perturbation Theory. In the original XPol method of Xie et al.,[55,85] intermolecular dispersion (van der Waals) and exchange (Pauli repulsion) interactions are incorporated using empirical Lennard-Jones or Buckingham potentials. We find, however, that this does not afford benchmark-quality results for noncovalent interactions,[46] so we instead choose to use the XPol fragment wave functions as zeroth-order states for application of second-order SAPT.

In SAPT, the Hamiltonian for the dimer $A\cdots B$ is partitioned according to[5,6]

$$\hat{H} = \hat{F}^A + \hat{F}^B + \xi\hat{W}^A + \eta\hat{W}^B + \zeta\hat{V} \tag{8}$$

where $\hat{W}^A$ and $\hat{W}^B$ are the Møller–Plesset fluctuation potentials for monomers A and B, and $\hat{V}$ consists of all Coulomb operators that couple particles on A to particles on B. Application of (anti)symmetrized Rayleigh–Schrödinger perturbation theory affords an energy expansion

$$E^{SAPT}_{int} = \sum_{\zeta=1}^{\infty} \sum_{\kappa=0}^{\infty} (E^{(\zeta\kappa)}_{pol} + E^{(\zeta\kappa)}_{exch}) \tag{9}$$

where $\kappa = \xi + \eta$. The terms $E^{(\zeta\kappa)}_{pol}$ constitute the *polarization expansion*,[100,101] which neglects exchange of electrons between monomers. (The term "polarization expansion" is historical and should not be confused with the way in which we use the term "polarization" below, namely, to mean induction.) To correct this, each term in eq 9 has a corresponding exchange term $E^{(\zeta\kappa)}_{exch}$ arising from the antisymmetrizer that is used to project out the Pauli-forbidden components of the interaction energy. Of these exchange terms, it has historically only been possible to evaluate $E^{(10)}_{exch}$ exactly,[102] whereas other exchange terms are evaluated within the *single-exchange approximation*, in which permutations involving more than one pair of electrons are neglected. (The resulting formulas involve the square of the overlap matrix, $\mathbf{S}^2$, and for this reason the single-exchange approximation is often called the "S^2 approximation".[5,6,82]) Recently, an analytic form for $E^{(20)}_{exch}$ has been reported,[103,104] but its implementation is not yet widely available. The single-exchange approximation is expected to be accurate at or beyond the van der Waals contact distance,[6] although problems for anionic systems necessitate some rescaling of the higher-order exchange interactions.[105,106]

Neglecting *intra*molecular electron correlation but treating $\hat{V}$ to second order (the so-called SAPT0 approximation[9]), we have

$$E^{SAPT0}_{int} = E^{(1)}_{elst} + E^{(1)}_{exch} + E^{(2)}_{ind} + E^{(2)}_{exch-ind} + E^{(2)}_{disp} + E^{(2)}_{exch-disp} \tag{10}$$

(We have dropped the index κ, because $\kappa = 0$.) Explicit expressions for these terms can be found in ref 5 (MO basis) or ref 71 (AO basis). The dispersion and exchange–dispersion terms are MP2-like in both cost and accuracy.

Finally, it is common to incorporate polarization effects beyond second order by adding a correction term

$$\delta E^{HF}_{int} = E^{HF}_{int} - (E^{(1)}_{elst} + E^{(1)}_{exch} + E^{(2)}_{ind,resp} + E^{(2)}_{exch-ind,resp}) \tag{11}$$

to the SAPT interaction energy.[9] Here, E^{HF}_{int} is the counterpoise-corrected HF binding energy for the dimer.

C. Combining XPol with SAPT. The partition of the dimer Hamiltonian in eq 8 can be generalized to an arbitrary number of monomers,[53]

$$\hat{H} = \sum_A (\hat{F}^A + \xi_A \hat{W}^A) + \sum_A \sum_{B>A} \zeta_{AB} \hat{V}_{AB} \tag{12}$$

with zeroth-order wave functions taken to be direct products of XPol monomer wave functions. Modification of the SAPT perturbation is required to avoid double-counting, because some part of electrostatics and polarization is already included at the XPol level, but this modification is straightforward.[51,53] The resulting XSAPT energy expression, including all terms through second order in the intermolecular interactions, is[50,53]

$$E_{XSAPT} = \sum_{A=1}^{n} \Big[\sum_{a \in A} (2\varepsilon^A_a - \mathbf{c}^{\dagger}_a \mathbf{f}^A \mathbf{c}_a) + E^A_{nuc} \Big]$$
$$+ \sum_A \sum_{B>A} (E^{[0;1_{AB}]}_{RSPT} + E^{[0;1_{AB}]}_{exch} + E^{[0;2_{AB}]}_{RSPT}$$
$$+ E^{[0;2_{AB}]}_{exch}) + E_{3B} \tag{13}$$

The superscript $[0; n_{AB}]$ indicates a term that is zeroth-order in the monomer fluctuation potentials but nth order in the intermolecular perturbation, $\hat{V}_{AB}$.[50,53] The connection to traditional two-body SAPT is that

$$E_{\text{RSPT}}^{[0;1_{AB}]} = E_{\text{elst},A}^{(1)} + E_{\text{elst},B}^{(1)} \tag{14a}$$

$$E_{\text{exch}}^{[0;1_{AB}]} = E_{\text{exch},A}^{(1)} + E_{\text{exch},B}^{(1)} \tag{14b}$$

and

$$E_{\text{RSPT}}^{[0;2_{AB}]} = E_{\text{ind},A}^{(2)} + E_{\text{ind},B}^{(2)} + E_{\text{disp},AB}^{(2)} \tag{15a}$$

$$E_{\text{exch}}^{[0;2_{AB}]} = E_{\text{exch-ind},A}^{(2)} + E_{\text{exch-ind},B}^{(2)} + E_{\text{exch-disp},AB}^{(2)} \tag{15b}$$

Two perturbations can couple three monomers, and second-order XSAPT thus contains three-body induction couplings that have no analogues in dimer SAPT.[53] This is the meaning of the E_{3B} term in eq 13:

$$E_{3B} = \sum_{A,C} {\sum_{B>A}}' {\sum_{D>C}}' \left(E_{\text{RSPT}}^{[0;1_{AB},1_{CD}]} + E_{\text{exch}}^{[0;1_{AB},1_{CD}]} \right) \tag{16}$$

The primed summations indicate that these terms vanish unless no more than three of the indices A, B, C, and D are distinct. Except for some exploratory calculations in ref 53, these terms have been neglected in our previous work on XSAPT, and we will neglect them here unless stated otherwise. Results for water clusters and halide–water clusters, however, will demonstrate that these terms are important in large clusters of polar monomers. To compute the E_{3B} term, one must calculate all $n - 2$ fragments for each unique dimer pair, which increases the computational cost from $O(n^2)$ to $O(n^3)$. Alternatively, $O(n^2)$ scaling can be recovered (with a large prefactor) by storing induction amplitudes on disk.[53]

One of the attractive features of traditional SAPT is its energy decomposition analysis, and XSAPT extends this to many-body systems in a largely analogous way. We include a $\delta E_{\text{int}}^{\text{HF}}$ correction of the form given in eq 11, whose goal is to incorporate higher-order induction effects, and for many-body systems we assume that this correction is pairwise additive:

$$\delta E_{\text{int}}^{\text{HF}} = \sum_A \sum_{B>A} \delta E_{AB}^{\text{HF}} \tag{17}$$

This assumption appears to be robust.[46] The SAPT interaction energy can be decomposed as in eq 10, and the resulting XSAPT energy decomposition is[46]

$$E_{\text{int}}^{\text{XSAPT}} = E_{\text{int}}^{\text{SAPT}} + \delta E_{\text{int}}^{\text{HF}} + E_{\text{int}}^{\text{MB}} + \sum_A \sum_{B>A} (E_{AB}^{\text{XSAPT}} - E_{AB}^{\text{SAPT}}) \tag{18}$$

where the total SAPT interaction energy for a collection of monomers is

$$E_{\text{int}}^{\text{SAPT}} = \sum_A \sum_{B>A} E_{AB}^{\text{SAPT}} \tag{19}$$

The XSAPT interaction energy can be rewritten as

$$E_{\text{int}}^{\text{XSAPT}} = E_{\text{elst}}^{(1)} + E_{\text{exch}}^{(1)} + E_{\text{disp}}^{(2)} + E_{\text{exch-disp}}^{(2)}$$
$$+ \left[E_{\text{ind}}^{(2)} + E_{\text{exch-ind}}^{(2)} + \sum_A \sum_{B>A} \delta E_{AB}^{\text{HF}} \right.$$
$$+ \sum_A \sum_{B>A} (E_{AB}^{\text{XSAPT}} - E_{AB}^{\text{SAPT}}) + E_{\text{int}}^{\text{MB}} \bigg] \tag{20}$$

Here, the terms $E_{\text{elst}}^{(1)}$, $E_{\text{exch}}^{(1)}$, etc., represent the sum of these energy components over all pairs of dimers, and the many-body contribution to the interaction energy is[46]

$$E_{\text{int}}^{\text{MB}} = E_{\text{int}}^{\text{XSAPT}} - \sum_A \sum_{B>A} E_{AB}^{\text{XSAPT}} \tag{21}$$

The term in square brackets in eq 20 is regarded as the total induction energy, which includes a many-body contribution.

In dimer SAPT calculations, an infinite-order polarization correction (in the presence of a frozen partner density) can be included by solving coupled-perturbed equations.[5] However, XSAPT treats polarization self-consistently and the infinite-order response correction for induction should be included exactly, via the XPol procedure, if density embedding is used.[107] We prefer CHELPG embedding for reasons of cost, however. The pairwise difference between XSAPT and SAPT in eq 20 partly includes the infinite-order response correction for induction. An infinite-order polarization correction is still included in $\delta E_{\text{int}}^{\text{HF}}$ by solving coupled-perturbed equations,[108,109] which is the meaning of the "resp" (response) subscripts in eq 11.

In the language of traditional dimer SAPT, our XPol monomer wave functions are computed in a monomer-centered basis set, which largely excludes the description of charge transfer between monomers and is typically less accurate than if the zeroth-order wave functions are computed using the dimer-centered basis set.[110] The correct choice of a dimer basis is ambiguous in a many-body system, however, so we choose instead to converge the XPol wave functions in the monomer-centered basis and then compute the pairwise SAPT corrections in a "projected"[51] (pseudocanonicalized[111,112] monomer-centered) basis set.

D. XSAPT(KS)+D. Especially for strongly hydrogen-bonded systems, inclusion of intramolecular electron correlation effects may be important[53,113] but involves methods whose cost scales as $O(N^7)$ within the wave function-based SAPT formalism.[9] SAPT(KS) represents a low-cost way to include such effects, though asymptotic correction of the XC potential is required.[61−63] Various "splicing" schemes have been used in this context[114−116] but result in "stray" XC potentials[117] that do not correspond to any well-defined energy functional, $v_{\text{xc}}^{\text{AC}} \neq \delta E_{\text{xc}}/\delta \rho$. This is potentially problematic in the context of geometry optimizations,[117] and fatal to any attempt to derive analytic energy gradients.

We sidestep this problem using long-range corrected (LRC) density functionals,[118−121] which correctly reproduce the asymptotic $\sim r^{-1}$ behavior of v_{xc}. To achieve the proper limiting value, $\Delta_\infty = 0$ in eq 2, we apply a monomer-specific "tuning" procedure, as suggested by Baer and co-workers,[122,123] in which the range-separation parameter, ω, is adjusted to satisfy the condition

$$\varepsilon_{\text{HOMO}}(\omega) = -\text{IP}(\omega) \tag{22}$$

In the context of SAPT(KS), this significantly improves the quality of the various interaction energy components as compared to benchmark results[54] and represents a promising

alternative to other AC schemes. Dispersion energies are still not of benchmark quality, however, owing to problems with the uncoupled KS description of dispersion.[54]

To correct the latter problem, we replace the second-order dispersion and exchange−dispersion terms in SAPT with empirical atom−atom dispersion potentials,[46,52] following along the lines of the "SAPT(KS)+D" method introduced by Heßelmann.[124] This has the added benefit of reducing the scaling from $O(N^5)$ to $O(N^3)$.[124] At first glance, this approach seems similar in spirit to dispersion-corrected DFT,[15] but the separation of dispersion from other parts of the energy is much cleaner in SAPT, whereas in DFT+D there is a potential double-counting problem for midrange intermolecular distances, where the short-range DFT correlation may not have decayed completely to zero as the (damped) long-range dispersion potentials are turning on. Indeed, Grimme[125] suggests that dispersion in DFT+D is a model-dependent quantity with no real physical meaning.

Our original version[52] of XSAPT(KS)+D used Heßelmann's SAPT(KS)+D dispersion potential,[124] which was fit to reproduce S22 benchmark binding energies, and that "first generation" (+D1) approach affords an impressive MAE of only 0.3 kcal/mol for the S66 data set.[52] However, XSAPT(KS)+D1 benefits from favorable error cancellation and does *not* accurately reproduce individual energy components.[46] In subsequent work, we avoided fitting directly to binding energies and instead pursued a second-generation (+D2) method using alternative dispersion potentials developed by Podeszwa et al.,[126] which were fit to reproduce distance-dependent dispersion potentials, $E_{disp} = E_{disp}^{(2)} + E_{exch-disp}^{(2)}$, obtained from SAPT(DFT) calculations. XSAPT(KS)+D2 accurately reproduces not only total binding energies but also individual energy components.[46]

Tests on the S22 and S66 data sets reveal that the primary source of errors in XSAPT(KS)+D2 calculations comes from π-stacked complexes, where in some cases the dispersion energy is overestimated by ∼2 kcal/mol as compared to SAPT2+(3)/aTZ results.[46] Such systems are underrepresented in the training set used to parametrize the D2 potentials,[126] and here we report for the first time a third-generation dispersion potential for XSAPT. (It should be stressed that our "D3" dispersion potential is unrelated to Grimme's "D3" correction[14] for DFT.)

The new D3 dispersion potential uses the same functional form as the D2 potential,[126]

$$E_{disp}^{D3} = -\sum_{i \in A} \sum_{\substack{j \in B \\ (B \neq A)}} \left[\frac{C_{ij,6}}{r_{ij}^6} f_6(\beta_{ij} r_{ij}) + \frac{C_{ij,8}}{r_{ij}^8} f_8(\beta_{ij} r_{ij}) \right] \tag{23}$$

where

$$f_n(r_{ij}) = 1 - \exp(-r_{ij}) \sum_{m=0}^{n} \frac{r_{ij}^m}{m!} \tag{24}$$

is the Tang−Toennies damping function[127] and i and j represent nuclei located on different monomers. We take $C_{ij,6} = (C_{i,6} C_{j,6})^{1/2}$, $C_{ij,8} = (C_{i,8} C_{j,8})^{1/2}$, and $\beta_{ij} = (\beta_i \beta_j)^{1/2}$, where $C_{i,6}$, $C_{i,8}$, and β_i are parameters fit to reproduce SAPT2+(3)/aTZ dispersion energies,

$$E_{disp}^{SAPT2+(3)} = E_{exch-disp}^{(20)} + E_{disp}^{(20)} + E_{disp}^{(21)} + E_{disp}^{(22)}(SDQ)$$
$$+ E_{disp}^{(22)}(T) + E_{disp}^{(30)} \tag{25}$$

For hydrogen, these parameters depend upon the identity of the nearest-neighbor atom. This is the similar procedure used in ref 126 to obtain the D2 potential, but we have expanded the training set to include additional π-stacked systems as well as the ionic systems $F^-(H_2O)$ and $Cl^-(H_2O)$. (The list of systems can be found in the Supporting Information.) For the latter two systems, the halide−water distance is short and dispersion is especially important. In these two cases, we use benchmarks from the highest level SAPT theory, SAPT2 + 3(CCD)/aTZ:[128,129]

$$E_{disp}^{SAPT2+3(CCD)} = E_{exch-disp}^{(20)} + E_{disp}^{(2)}[CCD]$$
$$+ E_{disp}^{(22)}[S(CCD)] + E_{disp}^{(22)}[T(CCD)]$$
$$+ E_{disp}^{(30)} + E_{exch-disp}^{(30)} + E_{ind-disp}^{(30)}$$
$$+ E_{exch-ind-disp}^{(30)} \tag{26}$$

For each dimer, we used five different geometries corresponding to intermolecular separations ranging from 0.9 to 2.0 times the equilibrium separation, for a total of 370 training geometries. Values obtained for the parameters $C_{i,6}$, $C_{i,8}$, and β_i are provided in the Supporting Information.

The LRC-ωPBE functional[130] with monomer-specific, tuned values of ω is used in all of our XSAPT+D3 calculations, as we have previously observed that errors in the exchange energy components increase as short-range HF exchange is added to LRC functionals.[46] Tuned values of ω are listed in the Supporting Information and differ, in some cases, from values reported previously.[46] In the case of thymine, for example, the optimally-tuned value changes from $\omega = 0.625\ a_0^{-1}$ to 0.275 a_0^{-1}. The large discrepancy comes from the fact that the condition in eq 22 cannot be satisfied for some monomers and basis sets. In such cases, we selected the closest point of approach between the $\varepsilon_{HOMO}(\omega)$ and $-IP(\omega)$ curves. The binding energy errors for the π-stacked uracil dimer and the adenine−thymine dimer were 3 and 4 kcal/mol, respectively, at the XSAPT(KS)+D2 level,[46] and these were the outliers among the S22 dimers. Simply using these newly-tuned values of ω determined in this study, the XSAPT(KS)+D2 errors for these two systems are reduced to 2.0 and 2.5 kcal/mol, respectively. We therefore recommend these new ω values.

E. sd-XSAPT(KS). Introduction of empirical dispersion into XSAPT calculations reduces the scaling from fifth-order to third-order with respect to dimer size and is further motivated by the fact that second-order, uncoupled dispersion energies are not nearly of benchmark quality. An alternative approach, which scales as $O(N^4)$, is to omit the fifth-order exchange−dispersion term in SAPT0 and then scale the fourth-order dispersion term by an empirical factor. This method was introduced recently by Ochsenfeld and co-workers,[131] who called it sd-SAPT0. We have implemented the corresponding sd-XSAPT(KS) method, and following ref 131, we neglect the δE_{int}^{HF} correction in this approach.

Using the S22B binding energies[34] to fit the dispersion scaling parameter, the best-performing combination was found to be the LRC-ωPBEh functional[121] (20% short-range HF exchange), 6-31G(d,2p) basis set, and a scaling parameter c_{disp} = 0.657, in which case the root-mean-square deviation (RMSD)

for S22 binding energies is 0.366 kcal/mol. That $c_{disp} < 1$ can be understood in terms of the neglect of the repulsive exchange–dispersion interaction and the fact that second-order perturbation theory tends to overestimate dispersion in the first place. (The double-ζ basis set also helps in this regard, as dispersion interactions converge slowly to the basis-set limit, and scaling back the basis set is a well-established way to reduce errors in second-order dispersion energies,[9] albeit by error cancellation.) Because the sd-XSAPT(KS) method is based on fitting to obtain accurate total binding energies, it is not recommended as a means to do energy decomposition analysis.

III. PERFORMANCE BENCHMARKS

In this section, we document timing data, the validity of the new D3 dispersion potential, and basis-set convergence.

A. Timings. So as not to obfuscate the fact that the primary purpose of XSAPT is *efficient* calculation of intermolecular interactions, we lead off with data illustrating the efficiency of the method. Figure 3 plots timings for XSAPT(KS)+D
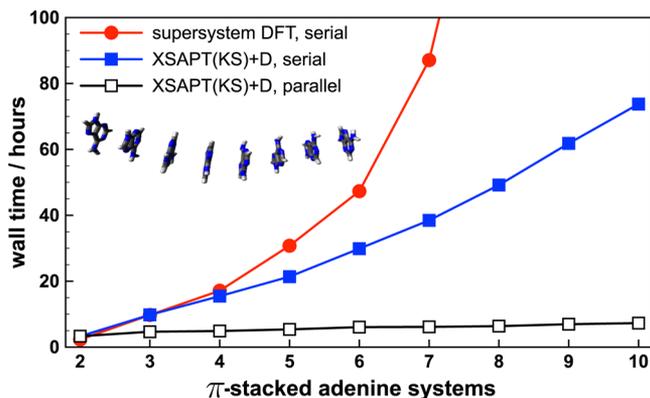
Figure 3. Timings for XSAPT(KS) and supersystem DFT calculations for π-stacked (adenine)$_n$ systems. All calculations use the LRC-ωPBE functional and the hpTZVPP basis set.

calculations on π-stacked (adenine)$_n$. Serial timings represent the total CPU time required, which scales as $O(n^2)$, whereas parallel timings represent the wall time required when the calculation is run in "embarrassingly parallel" mode [$n(n-1)/2$ processors for n monomers, so that all pairwise SAPT calculations can be performed simultaneously]. In the latter mode, wall time scales as $O(n)$ with a small prefactor. Even in serial, XSAPT(KS) is just as efficient as supersystem DFT for $n = 2$ monomers and is substantially more efficient for larger systems. In parallel, the wall time required for an XSAPT(KS) calculation on (adenine)$_{10}$ is only about twice as large as that required for (adenine)$_2$.

B. Validation of the D3 Dispersion Potential. To test the performance of the new D3 dispersion potential, we have used the D2 and D3 potentials to compute the stacking interaction between DNA base pairs in 10 different nucleobase tetramers,[132] as compared to the dispersion interaction ($=E_{disp}^{(2)} + E_{exch-disp}^{(2)}$) obtained from DFT-SAPT calculations. DFT-SAPT dispersion energies from ref 132 were multiplied by a factor of 1.1 as an approximate correction for basis-set incompleteness at the aTZ level, as suggested by Heßelmann.[124] The results, shown in Figure 4, indicate good agreement between E_{disp}(D3) and E_{disp}(DFT-SAPT), and the D3 dispersion potential is much closer to the benchmark as compared to D2 results.
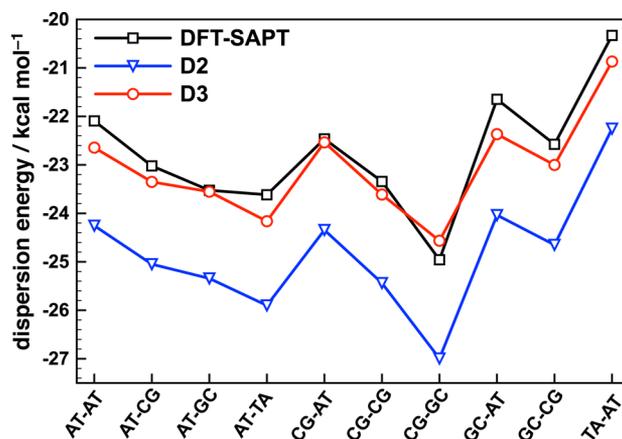
Figure 4. Comparison of E_{disp}(D2), E_{disp}(D3), and E_{disp}(DFT-SAPT) for the stacking interaction in nucleobase tetramers. The DFT-SAPT benchmarks come from ref 132, corrected by a factor of 1.1, as suggested in ref 124.

We note in passing that our D3 dispersion potential could be used to incorporate intramolecular correlation into the dispersion interaction at no additional cost. Specifically, the D3 dispersion potential could be combined with MP2 according to

$$E_{int}^{MP2+D3} = E_{int}^{MP2} - E_{disp}^{(2)} - E_{exch-disp}^{(2)} + E_{disp}(D3) \qquad (27)$$

in which we use MP2 to incorporate intramolecular correlation and then subtract out the second-order dispersion and replace it with the D3 empirical potential. This method is similar in spirit to the MP2(CCD) method.[106,148] We plan to investigate this "MP2+D3" approach in the future.

C. Basis Set Convergence. The $E_{elst}^{(1)}$ and $E_{exch}^{(1)}$ energy components are common to both SAPT(KS) and XSAPT(KS), and in the former case we can perform benchmark SAPT2+(3)/aTZ calculations for dimers, which we have done for the representative systems F$^-$(H$_2$O), (H$_2$O)$_2$, and both the T-shaped and parallel-displaced isomers of (C$_6$H$_6$)$_2$. SAPT(KS) results for the same systems, using either the dimer-centered or the "projected" (pseudocanonicalized monomer-centered) approach, were used to select an AO basis set. [The SAPT2+(3)/aTZ benchmarks employ the dimer-centered basis.] Mean errors in $E_{elst}^{(1)}$ and $E_{exch}^{(1)}$, with respect to the SAPT2+(3)/aTZ benchmarks, are provided in the Supporting Information for 21 different basis sets ranging from double-ζ to quadruple-ζ quality. A brief summary is presented here.

Using the dimer-centered approach, we find that the errors are quite small for both Dunning (aug-cc-pVXZ) and Ahlrichs (def2) basis sets, provided that diffuse functions are included. In that case, calculations of triple-ζ quality, or possibly even double-ζ quality, appear to be converged to the basis-set limit. Unfortunately, however, the dimer-centered construction is ill-defined for a many-body system, which is why we turn to the pseudocanonicalization approach.[51] The data in the Supporting Information show that diffuse functions are also essential in this approach, although in this case there remains significant discrepancy between augmented triple- and quadruple-ζ results. These differences are smaller when Ahlrichs basis sets are employed, in which case it matters little whether the diffuse functions are drawn from the Ahlrichs, Pople, or Dunning basis sets.

In consideration of both accuracy and computational efficiency, the Ahlrichs def2-TZVPP basis set will be the primary one used in this work, augmented with diffuse functions either from Pople's 6-311++G basis or from Dunning's aug-cc-pVTZ (aTZ) basis. We refer to these two choices as pTZVPP and aTZVPP, respectively. In some cases, we will omit the diffuse functions on hydrogen, to obtain "heavy-augmented" basis sets haTZVPP and hpTZVPP. The 6-31+G(3d,3pd) basis set will be used to compute the δE_{int}^{HF} corrections and the aTZ basis set to compute the δE_{MP2} corrections.

IV. ILLUSTRATIVE APPLICATIONS

The remainder of this paper is dedicated to illustrating the power and utility of XSAPT-based methods. Details regarding the benchmark calculations can be found in the Supporting Information.

A. Biologically-Relevant Dimers. Because the S22 data set[22] was used to fit both the D3 dispersion potential and the scaling factor for sd-XSAPT(KS), the S66 data set[36] will be used to evaluate the accuracy of these methods. S66 consists of CCSD(T)/CBS binding energies for 66 weakly bound dimers related to biomolecular structures, and we use the recently revised S66 binding energies.[133] Augmented triple-ζ basis sets are essential to obtain accurate results for individual energy components, and error statistics for XSAPT(KS)+D3 in a variety of triple-ζ basis sets are shown in Table 1. The best-performing basis sets are pTZVPP and hpTZVPP.

Table 1. Errors[a] in S66 Binding Energies for XSAPT(KS)+D3

	error/kcal mol^{-1}	
basis set[b]	MAE	max
hpTZVPP	0.27	1.11
pTZVPP	0.26	1.20
haTZVPP	0.30	1.23
aTZVPP	0.34	1.50
haTZ	0.44	1.81
aTZ	0.51	2.19

[a]With respect to CCSD(T)/CBS benchmarks.

Figure 5 shows S66 error statistics for a variety of methods that exhibit reasonably small MAEs. The new XSAPT(KS)+D3 method slightly outperforms the previous two generations (D1 and D2), and in particular reduces the errors in the π-stacked outliers. It is worth mentioning that the MP2C[42−44] and SAPT2+(3) methods,[9] which exhibit excellent performance for S66, are only formulated for dimers, and that the Coulomb-attenuated MP2 method[134−136] (att-MP2) contains a parameter that was optimized using this very data set. It also bears mention that all of the methods that outperform XSAPT(KS)+D3 exhibit at least fifth-order scaling with respect to the size of the *supersystem*, whereas XSAPT(KS)+D scales as $O(N^3)$ with respect to dimer size and $O(n^2)$ with respect to the number of monomers. XSAPT(KS)+D also affords an energy decomposition analysis that is discussed in section IV H.

B. Potential Energy Curves. Comparison of the "sandwich" and "T-shaped" isomers of $(C_6H_6)_2$ represents a stringent test for theoretical models,[137] because the two are stabilized by very different types of interactions (dispersion competes with quadrupolar electrostatics). In Figure 6, the
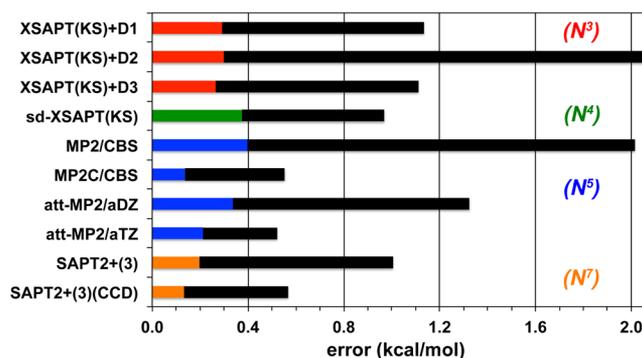
Figure 5. MAEs (colored bars) and maximum errors (in black) computed for S66 binding energies with respect to CCSD(T)/CBS benchmarks. These methods are color-coded according to how their cost scales with system size. The jun-cc-pVDZ basis set is used for XSAPT(KS)+D1 calculations and the hpTZVPP basis set for XSAPT(KS)+D2 and +D3 calculations. The 6-31G(d,2p) basis set is used for sd-XSAPT(KS) and aug-cc-pVTZ basis sets for SAPT2+(3) and SAPT2+(3)(CCD). The att-MP2 data were obtained from ref 134, and data for the other supermolecular methods were obtained from ref 36.
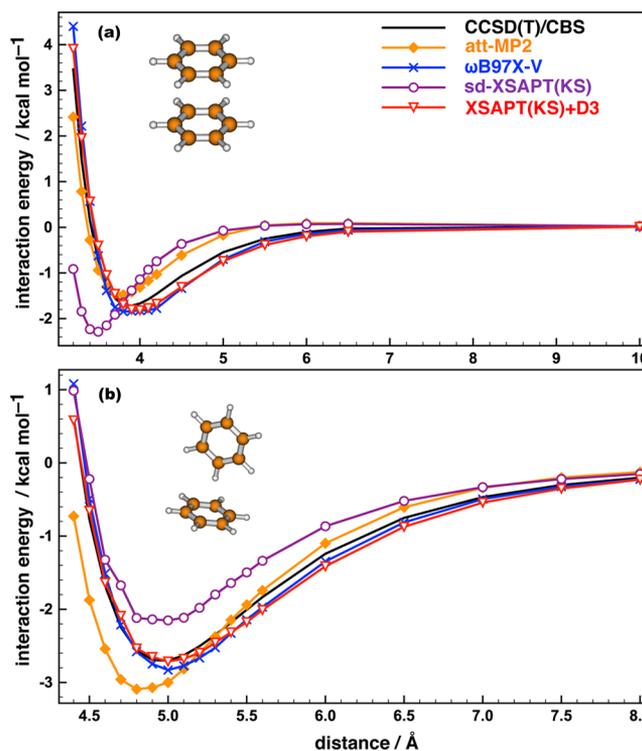
Figure 6. Potential energy curves for the (a) sandwich and (b) T-shaped isomers of $(C_6H_6)_2$. The distance coordinate is the center-to-center distance between the benzene rings. Benchmark CCSD(T)/CBS results are taken from ref 137. The aug-cc-pVTZ basis set is used for the att-MP2 and ωB97X-V calculations. The hpTZVPP basis set is used for the XSAPT(KS)+D3 calculations.

XSAPT(KS)+D3 method and the ωB97X-V density functional[19] reproduce CCSD(T)/CBS potential curves for both isomers nearly quantitatively. In contrast, the att-MP2 method shifts the minimum to shorter distances for both isomers. The sd-XSAPT(KS) method significantly shortens the van der Waals contact distance for the π-stacked isomer, while

242

dx.doi.org/10.1021/jp5098603 | J. Phys. Chem. A 2015, 119, 235−252
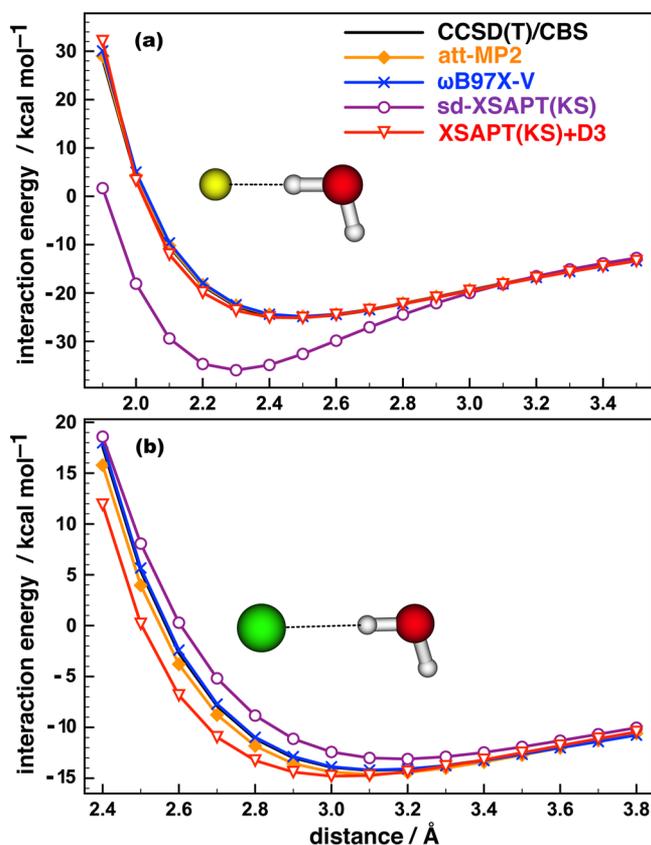
Figure 7. Potential energy curves for (a) $F^-(H_2O)$ and (b) $Cl^-(H_2O)$ at a fixed H_2O geometry. The distance coordinate is the halide−oxygen distance and the benchmark is CCSD(T)/CBS. The aug-cc-pVTZ basis set is used for the att-MP2 and ωB97X-V calculations. The hpTZVPP basis set is used for the XSAPT(KS)+D3 calculations.

significantly underestimating the binding energy for the T-shaped isomer.

In Figure 7, we plot the potential energy curves for $F^-(H_2O)$ and $Cl^-(H_2O)$, which are known to be challenging cases for SAPT.[52,105] CCSD(T)/CBS correlation energies were evaluated using a two-point (aTZ, aQZ) extrapolation[138] and then added to the HF/aQZ energy to obtain the CCSD(T)/CBS results. The XSAPT(KS)+D3, att-MP2, and ωB97X-V methods reproduce the CCSD(T)/CBS potential curves for $F^-(H_2O)$ nearly quantitatively, whereas sd-XSAPT(KS) significantly overestimates the binding energy. These methods perform even better for $Cl^-(H_2O)$, with only a slight underestimation of the binding energy on the part of sd-SAPT(KS).

The geometry of H_2O is held fixed in the calculation shown in Figure 7. Upon relaxation of the geometry, the $F^-(H_2O)$ binding energy increases from 25 to 32 kcal/mol at the minimum-energy structure, and the XSAPT(KS)+D3 method somewhat overestimates the binding energy. The relaxed geometry is somewhat problematic for fragment-based methods, as the proton affinity of F^- leads to an unusually long O−H bond length of 1.06 Å versus 0.96 Å when the H_2O geometry is optimized separately. [For $Cl^-(H_2O)$, the O−H bond length increases only to 0.98 Å upon relaxation.] We note that other fragment based methods, such as the effective fragment potential (EFP) method,[139] must also use rigid, EFP-optimized geometries; otherwise, large errors in noncovalent binding energies are obtained.[46]

C. Many-Body System: $(H_2O)_6$. We next consider a cluster of polar monomers that exhibits significant many-body polarization effects.[76,77] Eight low-lying structures of $(H_2O)_6$ are considered,[77] and their CCSD(T)/CBS binding energies are evaluated using a two-point extrapolation[138] of CCSD(T)-F12 correlation energies (cc-pVDZ-F12 and cc-pVTZ-F12 basis sets, using the corresponding near-complete auxiliary basis sets cc-pVDZ-F12-CABS and cc-pVTZ-F12-CABS).[140,141] The HF energy is evaluated using the cc-pVTZ-F12 basis set. For comparison, the MP2 correlation energy in the CBS limit was evaluated using a two-point (aTZ, aQZ) extrapolation[138] and then added to the HF/aQZ energy to obtain the MP2/CBS result.
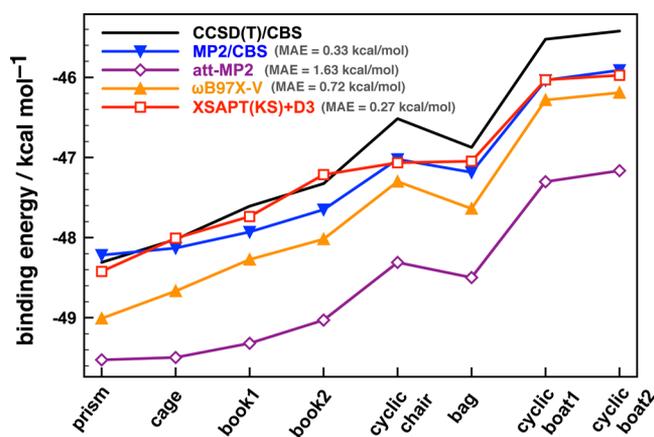
Figure 8. Binding energies for eight isomers of $(H_2O)_6$. The aTZ basis set is used for the att-MP2 and ωB97X-V calculations, whereas the hpTZVPP basis set is used for the XSAPT(KS)+D3 calculations. (MAEs for the whole data set, with respect to the CCSD(T)/CBS benchmarks, are also listed for each method.)

Figure 8 compares the binding energies for isomers of $(H_2O)_6$ computed using various methods, but sd-XSAPT(KS) results are not shown in Figure 8 because this method overestimates the binding energies by an average of 5.46 kcal/mol. This is consistent with the fact that sd-XSAPT(KS) overestimates the binding energy of water dimer already by 0.56 kcal/mol; hence, sd-XSAPT(KS) cannot be recommended for water clusters. The att-MP2,[134,135] ωB97X-V,[19] and MP2/CBS methods all afford accurate relative energies, as does XSAPT(KS)+D3 except for a slight overstabilization of the cyclic chair isomer, but the most accurate absolute binding energies are obtained using XSAPT(KS)+D3. Total binding energies predicted using XSAPT(KS)+D3 are more accurate than those obtained using ωB97X-V, which is one of the best DFT approaches for noncovalent interactions,[19] although it should be noted that *relative* energies are slightly better with ωB97X-V, which does not overstabilize the cyclic chair.

One factor that influences the accuracy of XSAPT(KS) results is the tuning of ω. Tuned values listed in the Supporting Information were obtained using a step size $\Delta\omega = 0.025$ a_0^{-1} to scan $\varepsilon_{HOMO}(\omega)$ and $-IP(\omega)$, and $\omega = 0.500$ a_0^{-1} is thus determined to be the optimal value for H_2O. However, if the step size is decreased to 0.005 a_0^{-1}, then the optimal value changes to 0.485 a_0^{-1}. For water dimer, the binding energy changes by only 0.07 kcal/mol between these two values, but errors accumulate as the number of fragments increases. For isomers of $(H_2O)_6$, the difference in binding energies between $\omega = 0.485$ and 0.500 a_0^{-1} is 0.68 kcal/mol on average but is

3.13 kcal/mol for the isomers of $(H_2O)_{20}$ that are discussed below. As such, we use the more finely tuned value ($\omega = 0.485$ a_0^{-1}) for water clusters.

D. Larger Clusters: $(H_2O)_{20}$. Medium-sized water clusters have long attracted interest from the quantum chemistry community; for example, $(H_2O)_{16,17}$ are considered to be transition structures from "all-surface" to "internally solvated" arrangements of the hydrogen-bonding network,[142,143] and $(H_2O)_{20,24}$ are the building blocks of ice clathrates.[144] Here, we use ten low-energy isomers of $(H_2O)_{20}$, obtained using the TIP4P force field,[145] to benchmark the methods introduced above and, by comparison to $(H_2O)_6$ results, to understand whether errors increase with system size. MP2/CBS results for these ten isomers were estimated as described in section IV C.

To estimate the CCSD(T)/CBS binding energies, we use explicitly correlated CCSD(T) calculations reported recently using the generalized energy-based fragmentation (GEBF) method.[146] We take

$$E_{CBS}^{CCSD(T)} \approx E_{CBS}^{MP2} + \delta_{MP2\text{-}F12}^{CCSD(T)\text{-}F12a} \quad (28)$$

where the correction to the MP2/CBS result is equal to the difference between the CCSD(T)-F12a/aDZ and MP2-F12/aDZ binding energies that were reported in ref 146 using the GEBF approximation. Unlike the calculations reported in ref 146, which used the GEBF approximation at both the MP2 and CCSD(T) levels of theory, we evaluate the full MP2/CBS energy. This changes the binding energies by an average of 1.75 kcal/mol relative to the benchmarks reported in ref 146, and we believe that our CCSD(T)/CBS results for $(H_2O)_{20}$ are one of the most accurate binding benchmarks for large water clusters in the literature.

Table 2. MAEs^a for Ten Low-Energy Isomers of $(H_2O)_{20}$

method	MAE/kcal mol^{-1}	
	binding energy	relative energy
MP2/CBS	2.69	0.18
att-MP2/aTZ	3.96	0.37
ωB97X-V/aTZ	0.92	0.07
XSAPT+D3^b	5.42	0.97
XSAPT+D3+E_{3B}^{b}	0.57	0.69
XSAPT+D3^c	3.35	0.70
XSAPT+D3+E_{3B}^{c}	2.12	0.34

^{a}With respect to CCSD(T)/CBS benchmarks. ^{b}Using the hpTZVPP basis set. ^{c}Using the haTZVPP basis set.

Results for total binding energies are shown in Figure 9, with error statistics listed in Table 2. The correction to MP2/CBS in eq 28 is negative for these clusters, and CCSD(T)/CBS total binding energies are 2.7 kcal/mol larger, on average, than MP2/CBS binding energies. Unfortunately, XSAPT(KS)+D3 results for total binding energies are not significantly better than MP2/CBS results in this case and can be worse, depending on the basis set that is used.

In an attempt to understand this loss of accuracy relative to the $(H_2O)_6$ results, we investigated the neglected three-body induction couplings, E_{3B} in eq 16. For isomers of $(H_2O)_6$, $E_{3B} = 0.13$ kcal/mol (on average) at the XSAPT(KS)+D3/hpTZVPP level. However, $E_{3B} = 4.95$ kcal/mol (on average) for the $(H_2O)_{20}$ isomers considered here, and the XSAPT(KS)+D3 total binding energies are significantly improved by the addition of this three-body correction term.
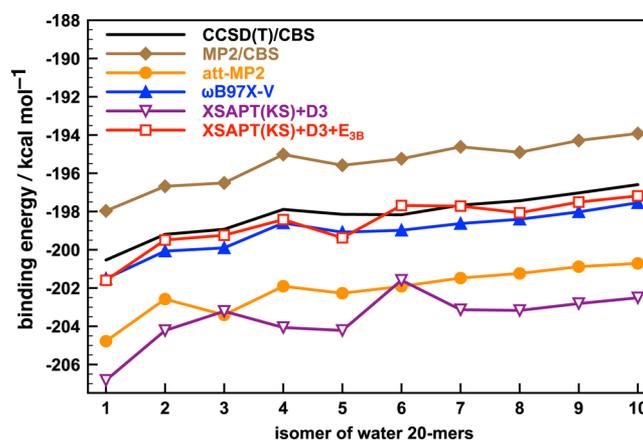
Figure 9. Binding energies for ten low-energy isomers of $(H_2O)_{20}$. The aTZ basis set is used for the att-MP2 and ωB97X-V calculations, and the hpTZVPP basis set is used for XSAPT methods. The isomers are numbered as in ref 146.

It is evident both from the plot in Figure 9 and from the error statistics in Table 2 that the ωB97X-V and MP2/CBS methods afford very good relative energies, as does att-MP2. XSAPT(KS)+D3+E_{3B} relative energies are somewhat worse, even if the absolute binding energies are slightly more accurate, at least when the hpTZVPP basis set is employed. The reasons for this are unclear, although in fitting the D3 potential we considered a variety of intermolecular distances but only one intermolecular orientation per monomer. Also unclear is why isomer **6** (Figure 9) poses such a problem for XSAPT, although this same isomer has been noted to pose problems for DFT methods as well.[146] It is the most highly coordinated of all $(H_2O)_{20}$ isomers considered here, so the problem may again be orientational dependence, although Wang et al.[146] suggest that both basis set and electron correlation effects must be considered to obtain an accurate relative energy for this isomer. Extension of XSAPT to larger basis sets, by means of an AO rather than an MO implementation, is currently underway in our group and may help in this capacity.

E. Halide–Water Clusters. Halide–water clusters are difficult cases for popular DFT methods.[19,46] Table 3 shows binding-energy errors, with respect to CCSD(T)/CBS benchmarks, for various methods applied to the minimum-energy structures of $X^-(H_2O)_{n=1-6}$, for X = F, Cl. (Benchmarks were obtained as described for water hexamer in section IV C.)

Table 3. MAEs^a for Binding Energies of $X^-(H_2O)_{n=1-6}$

method	MAE/kcal mol^{-1}	
	X = F	X = Cl
MP2/CBS	1.01	0.23
att-MP2/aTZ	0.22	1.02
ωB97X-V/aTZ	0.20	0.43
XSAPT(KS)+D3^b	3.32	1.72
XSAPT(KS)+D3+E_{3B}^{b}	1.64	0.98
XSAPT(KS)+D3+E_{3B}-δMP2^{b,c}	0.70	0.59
XSAPT(KS)+D3^d	3.67	1.85
XSAPT(KS)+D3+E_{3B}^{d}	1.73	0.94
XSAPT(KS)+D3+E_{3B}-δMP2^{c,d}	0.98	0.55

^{a}With respect to CCSD(T)/CBS benchmarks. ^{b}Using the hpTZVPP basis set for XSAPT. ^{c}Using the aug-cc-pVTZ basis for δMP2. ^{d}Using the haTZVPP basis set for XSAPT.

The MP2/CBS method cannot be considered a benchmark, sub-kcal/mol level of theory for $F^-(H_2O)_n$, with a MAE of 1.0 kcal/mol and a maximum error of 1.6 kcal/mol. Errors are smaller for $Cl^-(H_2O)_n$ (MAE = 0.2 kcal/mol, maximum = 0.3 kcal/mol). Interestingly, the att-MP2 method performs in the opposite way: highly accurate results for $F^-(H_2O)_n$, but a MAE of 1.0 kcal/mol for $Cl^-(H_2O)_n$. The ωB97X-V functional performs well for both, as shown previously.[19]

For XSAPT calculations of $X^-(H_2O)_n$, we find that the three-body induction couplings are important, reducing the errors by approximately a factor of 2, although the MAE for $F^-(H_2O)_n$ remains >1 kcal/mol. Parker et al.[148] have proposed a "δMP2" correction,

$$\delta E_{MP2} = E_{int}^{MP2} - E_{int}^{SAPT2} \qquad (29)$$

to account for missing terms such as high-order coupling between induction and dispersion. This correction, which we find is especially important in ionic systems,[106] is equal to the difference between the counterpoise-corrected MP2 binding energy for the dimer and the SAPT2 binding energy. In XSAPT calculations, we apply the δE_{MP2} correction in a pairwise way, for dimers that include X^-. This further reduces the errors, especially for X = F. [The SAPT2 part of δE_{MP2} also makes this correction $O(N^6)$ with respect to dimer size, although this can be reduced to $O(N^5)$ using density-fitting techniques.[113]] The hpTZVPP basis set works better than the haTZVPP basis set, which may be an overpolarization problem[147] in the latter case, due to the larger number of diffuse functions in haTZVPP. In any case, MAEs of <1 kcal/mol in total binding energies are achievable for the difficult case of $X^-(H_2O)_n$ clusters, if both E_{3B} and δE_{MP2} are included.

We next examine relative energies for ten isomers of $F^-(H_2O)_{10}$ that were considered in ref 46. The following scheme is used to obtain CCSD(T)/CBS benchmarks:

$$E_{CBS}^{CCSD(T)} \approx E_{CBS}^{MP2} + \delta_{MP2-F12}^{CCSD(T)-F12} \qquad (30)$$

Here, the correction to the MP2/CBS result is equal to the difference between CCSD(T)-F12/aDZ and MP2-F12/aDZ energies. Binding energies for a variety of XSAPT and DFT approaches are shown in Figure 10, and error statistics are listed in Table 4. Among supersystem methods, ωB97X-V and att-MP2 perform the best.

In the case of XSAPT calculations, we find that the E_{3B} contribution is 5−6 kcal/mol, much larger than its contribution in neutral systems, which makes sense given that E_{3B} is an induction correction. The δMP2 term contributes 1.6 kcal/mol on average, but in contrast to its effect in the smaller halide−water clusters, here the δMP2 term has a deleterious effect on the accuracy of total binding energies. This discrepancy may arise from the manner in which we obtain the CCSD(T)/CBS benchmarks. Specifically, we used the CCSD(T)-F12 results to extrapolate directly to CBS limit in small halide−water clusters but the additive scheme in eq 30 is used for $F^-(H_2O)_{10}$. Although this additive scheme seems to work well in neutral systems, it is not well tested for anionic systems and may not be appropriate in such cases. We note that the deviation between XSAPT results and these putative CCSD(T)/CBS benchmarks is typically comparable to, or smaller than, the ∼2 kcal/mol magnitude of the $\delta_{MP2-F12}^{CCSD(T)-F12}$ correction in eq 30 although XSAPT(KS)+D3+E_{3B}/hpTZVPP offers the best binding energies among all methods considered in Table 4.
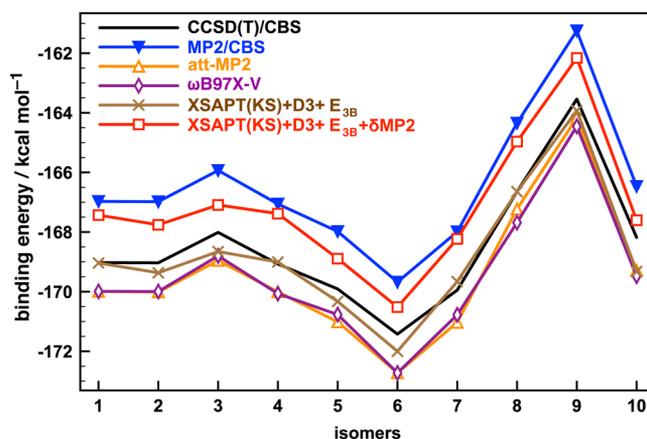
Figure 10. Binding energies for ten isomers of $F^-(H_2O)_{10}$. The att-MP2 and ωB97X-V calculations use the aTZ basis set, and XSAPT calculations use the hpTZVPP basis set, except that the aTZ basis set was used to evaluate the δMP2 correction.

Table 4. MAEs[a] in Binding Energies and Relative Energies for Ten Isomers of $F^-(H_2O)_{10}$

	MAE/kcal mol^{-1}	
method	binding energy	relative energy
MP2/CBS	2.02	0.16
att-MP2/aTZ	0.96	0.17
ωB97X-V/aTZ	0.99	0.13
ωB97X-D/aTZ	4.81	0.55
LC-VV10/aTZ	7.02	0.27
M06-2X/aTZ	10.18	0.45
M06-2X-D3(zero)[b]/aTZ	11.53	0.47
B2PLYP/aTZ	4.32	0.32
B2PLYP-D3(zero)[b]/aTZ	6.90	0.17
XSAPT(KS)+D3+E_{3B}[c]	0.38	0.41
XSAPT(KS)+D3+E_{3B}-δMP2[c,d]	1.28	0.41
XSAPT(KS)+D3+E_{3B}[e]	0.72	0.33
XSAPT(KS)+D3+E_{3B}-δMP2[d,e]	2.31	0.32

[a]With respect to CCSD(T)/CBS benchmarks. [b]Using the "zero-damping" function of ref 40. [c]Using the hpTZVPP basis set for XSAPT. [d]Using the aTZ basis set for δMP2. [e]Using the haTZVPP basis set for XSAPT.

Even if the deviation persists and is indeed a problem with XSAPT, we note that the ωB97X-D, LC-VV10, M06-2X, M06-2X-D3(zero), B2PLYP, and B2PLYP-D3(zero) supersystem methods all afford errors that are unacceptably large, and cannot be recommended for binding energies of halide−water clusters. (Although accurate relative energies are sufficient for structure determination, accurate *total* binding energies are needed to compute, e.g., the binding affinity of a drug molecule to a protein.) XSAPT works better with the hpTZVPP basis set than with haTZVPP, which may again be an overpolarization effect. Such effects are well-known in QM/MM calculations, where a simple solution is a Gaussian "blurring" of point charges nearby to the QM region.[147] In future work, we plan to test a Gaussian-blurred version of the XPol procedure.

F. CH_4 in a Dodecahedral $(H_2O)_{20}$ Cage. The isolated $CH_4@(H_2O)_{20}$ gas-phase cluster has been used as a model system to study the interaction between methane and clathrate hydrates.[150−152] A recent quantum Monte Carlo (QMC) benchmark affords a binding energy of −5.3 ± 0.5 kcal/mol,[149] for $CH_4@(H_2O)_{20} \rightarrow CH_4 + (H_2O)_{20}$, whereas various other

245

dx.doi.org/10.1021/jp5098603 | J. Phys. Chem. A 2015, 119, 235−252

Table 5. Binding Energy of CH_4 to $(H_2O)_{20}$

method	binding energy/kcal mol^{-1}
QMC[a]	−5.3 ± 0.5
MP2/CBS	−5.04
MP2C-F12/aTZ[a]	−4.60
att-MP2/aTZ	−4.01
ωB97X-V/aTZ	−6.29
ωB97X-D/aTZ	−6.39
LC-VV10/aTZ	−1.17
M06-2X/aTZ	−6.11
M06-2X-D3(zero)/aTZ	−7.32
B2PLYP/aTZ	−1.09
B2PLYP-D3(zero)/aTZ	−6.04
DFT-SAPT/aTZ[a]	−3.88
XSAPT(KS)+D3[b]	−3.48

[a]From ref 149. [b]Using the hpTZVPP basis set.

electronic structure methods predict binding energies ranging from −4 to −7 kcal/mol (Table 5). DFT methods generally overestimate the binding energy by about 1 kcal/mol whereas MP2 and MP2C afford accurate binding energies. It is therefore curious that the double-hybrid B2PLYP functional requires an empirical dispersion correction to get anywhere close to the benchmark binding energy. The nonlocal LC-VV10 functional also severely underestimates the binding energy despite its very good performance for S66. The att-MP2, DFT-SAPT, and XSAPT(KS)+D3 methods each underestimate the binding energy somewhat. It should be noted that the DFT-SAPT calculation requires the use of $(H_2O)_{20}$ as one monomer unit, whereas in XSAPT the monomers are CH_4 and H_2O.

G. Anti-Cancer Drug Intercalated into DNA. Predicting accurate noncovalent interaction between biomolecules and drug candidates (i.e., protein−ligand interactions) is a crucial component in drug discovery and design, where docking[153,154] (with force fields or empirical scoring functions) and *ab initio* screening[155,156] (with low-level quantum-chemical methods) are theoretical mainstays. Here, we consider intercalation of the anticancer agent ellipticine[157] into DNA, which involves insertion between two Watson−Crick CG base pairs, linked by their respective phosphate sugar puckers as depicted in Figure 11. The structure depicted in the figure consists of 157 atoms, and a benchmark binding energy is available from QMC calculations.[158]

The sd-XSAPT(KS) method overestimates the binding energy by about 10 kcal/mol with respect to this benchmark. The sd-XSAPT(KS) method ignores the exchange−dispersion component and scales the dispersion component by a factor optimized against CCSD(T)/CBS benchmarks for small dimers; this method may therefore overestimate binding energies in systems with very large monomers. Dispersion-corrected DFT is also known to overestimate binding energies in such systems[159] and can be improved in such cases by a three-body interatomic dispersion energy ($E_{\mathrm{disp}}^{(3)}$) based on the Axilrod−Teller−Muto three-body dispersion formula.[159] For ellipticine intercalated into DNA, we obtain $E_{\mathrm{disp}}^{(3)}$ = 8.90 kcal/mol, and the corrected sd-XSAPT(KS)+$E_{\mathrm{disp}}^{(3)}$ binding energy is −34.4 kcal/mol, which lies within the statistical error bars of the QMC benchmark. The PBE+MBD* method,[30] where "MBD*" is a many-body dispersion correction, yields a binding energy of −35.4 kcal/mol,[158] which is also within the QMC error bars. PBE+MBD* and sd-XSAPT(KS)+$E_{\mathrm{disp}}^{(3)}$ are the methods that come closest to the QMC result so far.

QMC: −33.6 ± 0.9 kcal/mol
sd-XSAPT(KS): −43.3 kcal/mol
sd-XSAPT(KS)+$E_{\mathrm{disp}}^{(3)}$: −34.4 kcal/mol
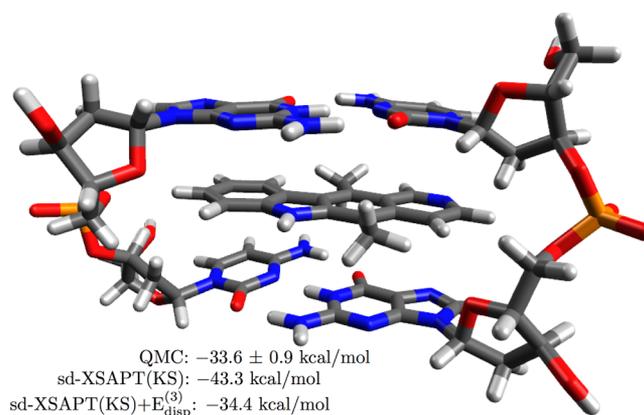
Figure 11. Ellipticine molecule intercalated into a GC:GC segment of DNA. Binding energies computed with various methods are shown. XSAPT calculations used three fragments: neutral ellipticine and two single-stranded GC complexes, each with a −1 charge.

H. Energy Decomposition. Table 6 shows a statistical summary of the energy components for the dimers in the S22 and S66 data sets, as computed by XSAPT methods and also by EFP.[139] The benchmark is SAPT2+(3)/aTZ, results of which were reported for S22 in ref 54 and are reported here for S66, for the first time.

For S22, MAEs for the individual energy components calculated using XSAPT(KS)+D1 are large. As suggested above, this method shows good results for total binding energies but only due to favorable error cancellation. In contrast, the XSAPT(KS)+D2 and+D3 methods afford very good results for individual energy components.

In contrast to the $X^-(H_2O)_n$ systems, for which the hpTZVPP basis set works slightly better than haTZVPP (possibly owing to overpolarization), here the haTZVPP basis set affords slightly better results for induction energies. We note that charge-transfer interactions show up in the induction energy within the SAPT formalism, but only if the basis set is diffuse enough so that basis functions centered on monomer *A* extend significantly over monomer *B*. This may be why the more diffuse haTZVPP performs better here for induction energies. In neutral systems, the overpolarization caused by the diffuse basis functions is not large, and it is better to use a large basis set to capture charge transfer, whereas overpolarization is more significant in anionic systems, and the more diffuse basis leads to larger errors in binding energies.[106]

For the S66 data set, XSAPT(KS)+D3 affords errors of <5% for the electrostatic, exchange, and dispersion components, and <10% for the induction component. In short, XSAPT(KS)+D3 is reliable for energy decomposition analysis. EFP, another fragment-based method, affords errors of 1−2 kcal/mol in each of the energy components.

For the eight $(H_2O)_6$ isomers discussed in section IV C, energy components have been quantified by Chen and Li[77] at the MP2/a5Z-h level, using a localized molecular orbital energy decomposition analysis. Many-body effects in $(H_2O)_6$ are dominated by polarization interactions, whereas the other energy components are strictly or nearly pairwise additive.[77] Figure 12 compares the many-body polarization and total many-body energies for these $(H_2O)_6$ isomers, as reported by Chen and Li, to XSAPT(KS)+D3 results. The latter method inherently assumes that the many-body part of the interaction arises exclusively from polarization. The many-body polar-

Table 6. MAEs (kcal/mol) and Percent Errors (in Parentheses) for Individual Energy Components of the S22 and S66 Data Sets

method	energy components[a]								binding energy[b]	
	electrostatic		exchange		induction		dispersion			
					S22					
XSAPT(KS)+D1[c]	0.55	(11.21)	3.00	(22.90)	1.96	(60.98)	1.55	(20.72)	0.52	(9.34)
XSAPT(KS)+D2[d]	0.19	(2.80)	0.45	(4.10)	0.14	(9.46)	0.39	(5.70)	0.74	(9.92)
XSAPT(KS)+D3[e]	0.20	(3.04)	0.44	(4.04)	0.22	(10.80)	0.12	(3.19)	0.45	(7.47)
EFP	1.77	(32.66)	2.07	(14.87)	1.81	(51.53)	0.95	(14.20)	1.79	(27.19)
					S66					
XSAPT(KS)+D3[e]	0.20	(3.92)	0.31	(4.42)	0.18	(9.60)	0.23	(4.50)	0.27	(7.14)

[a]Errors with respect to SAPT2+(3)/aTZ energy components. [b]Error with respect to CCSD(T)/CBS binding energy. [c]Using LRC-ωPBEh/jun-cc-pVDZ with 60% short-range HF exchange. [d]Using LRC-ωPBE/haTZVPP. [e]Using LRC-ωPBE/hpTZVPP.
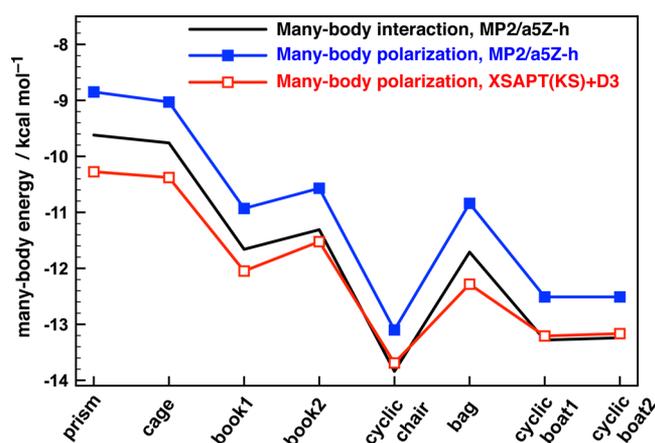
Figure 12. Many-body interactions for isomers of $(H_2O)_6$. MP2/a5Z-h results are taken from ref 77, and the hpTZVPP basis is used for XSAPT(KS)+D3 calculations.

ization energies using XSAPT(KS)+D3 are consistently overestimated as compared to the many-body polarization energies evaluated at the MP2/a5Z-h level but are much closer to the total many-body interaction energies computed at the MP2/a5Z-h level.

V. SUMMARY

Two new XSAPT-based methods based on a modified dispersion interaction, XSAPT(KS)+D3 and sd-XSAPT(KS), are reported in this article. It has been demonstrated that XSAPT(KS)+D3 is very successful in predicting binding energies for a wide range of challenging systems ranging from benzene dimer to large water— and halide—water clusters. The sd-XSAPT(KS) method performs well for large, dispersion-bound systems, such as a ligand—DNA intercalation complex considered here and (based on preliminary calculations) the L7 database[160] of large organic dimers. However, this method performs less well for water clusters, where the double-ζ basis set that is used in fitting the dispersion scaling parameter cannot adequately describe electrostatic and induction interactions.

Based on this survey of applications, it appears that the $O(n)$—when run in "embarrassingly parallel" mode—XSAPT family of methods, and especially XSAPT(KS)+D3, should routinely be used to explore noncovalent interactions in large assemblies of molecules. The many-body XSAPT energy decomposition can be used to understand the meaning of such interactions.

There is still room for progress with XSAPT, including the formulation and implementation of analytic energy gradients for geometry optimizations and simulations, and the combination of XSAPT and TDDFT response theory for dispersion energy along the lines of SAPT(DFT). Extension of XSAPT to include intramolecular correlation based on either a Møller—Plesset[5,56] or coupled-cluster formalism[57−59] is possible, as are improvements to the empirical dispersion potential and the charge embedding scheme. Several of these lines of development are currently underway in our group.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Energy components, various benchmark binding energies, tuned ω values, and D3 dispersion parameters. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*J. M. Herbert. E-mail: herbert@chemistry.ohio-state.edu.

Notes

The authors declare no competing financial interest.

Biographies

Ka Un Lao was born in Macau and received his B.S. and M.S. degrees in chemistry from National Tsing Hua University in Hsinchu, Taiwan, where he worked with Prof. Chin-Hui Yu. He is currently a graduate student under Prof. John Herbert at The Ohio State University, where his research focuses on intermolecular interactions using fragment-based methods and symmetry-adapted perturbation theory.

John M. Herbert received a Ph.D. in chemistry from the University of Wisconsin—Madison in 2003. Following postdoctoral work at The Ohio State University and at the University of California—Berkeley, he joined The Ohio State University as an Assistant Professor in 2006 and was promoted to Associate Professor in 2011 and then to Professor in 2014. His research interests include intermolecular interactions, solvation, excited states, and electronic processes in macromolecules and aqueous systems. His research group is a major contributor to the Q-Chem electronic structure software.

■ REFERENCES

(1) Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **2006**, *34*, 564−574.

(2) Guvench, O.; MacKerell, A. D. Computational evaluation of protein-small molecule binding. *Curr. Opin. Struct. Biol.* **2009**, *19*, 56−61.

(3) Durdagi, S.; Zhao, C.; Cuervo, J. E.; Noskov, S. Y. Atomistic models for free energy evaluation of drug binding to membrane proteins. *Curr. Med. Chem.* **2011**, *18*, 2601−2611.

(4) Woodley, S. M.; Catlow, R. Crystal structure prediction from first principles. *Nat. Mater.* **2008**, *7*, 937−946.

(5) Jeziorski, B.; Moszynski, R.; Ratkiewicz, A.; Rybak, S.; Szalewicz, K.; Williams, H. L. SAPT: A program for many-body symmetry-adapted perturbation theory calculations of intermolecular interaction energies. In *Methods and Techniques in Computational Chemistry: METECC-94*; Clementi, E., Ed.; STEF: Cagliari, 1993; Vol. B, Chapter 3, pp 79−129.

(6) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes. *Chem. Rev.* **1994**, *94*, 1887−1930.

(7) Szalewicz, K.; Patkowski, K.; Jeziorski, B. Intermolecular interactions via perturbation theory: From diatoms to biomolecules. In *Intermolecular Forces and Clusters II*; Wales, D. J., Ed.; Vol. *116* of Structure and Bonding; Springer-Verlag: Berlin, 2005; pp 43−117.

(8) Szalewicz, K. Symmetry-adapted perturbation theory of intermolecular forces. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 254−272.

(9) Hohenstein, E. G.; Sherrill, C. D. Wavefunction methods for noncovalent interactions. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 304−326.

(10) Jansen, G. Symmetry-adapted perturbation theory based on density functional theory for noncovalent interactions. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 127−144.

(11) Perdew, J. P. Climbing the ladder of density functional approximations. *MRS Bull.* **2013**, *38*, 743−750.

(12) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215−241.

(13) Grimme, S. Semiempirical GGA-type density functional constructed with long-range dispersion correction. *J. Comput. Chem.* **2006**, *27*, 1787−1799.

(14) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate *ab initio* parameterization of density functional dispersion correction (DFT-D) for the 94 elements H−Pu. *J. Chem. Phys.* **2010**, *132*, 154104:1−19.

(15) Grimme, S. Density functional theory with London dispersion corrections. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 211−228.

(16) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom−atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615−6620.

(17) Grimme, S. Semiempirical hybrid density functional with perturbative second-order correction. *J. Chem. Phys.* **2006**, *124*, 034108:1−16.

(18) Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals density functional: The simpler the better. *J. Chem. Phys.* **2010**, *133*, 244103:1−9.

(19) Mardirossian, N.; Head-Gordon, M. ωB97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9904−9924.

(20) Howard, J. C.; Tschumper, G. S. Wavefunction methods for the accurate characterization of water clusters. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2014**, *4*, 199−224.

(21) Del Ben, M.; Schönherr, M.; Hutter, J.; VandeVondele, J. Bulk liquid water at ambient temperature and pressure from MP2 theory. *J. Phys. Chem. Lett.* **2013**, *4*, 3753−3759. Erratum: Del Ben, M.; Schönherr, M.; Hutter, J.; VandeVondele, J. *J. Phys. Chem. Lett.* **2014**, *5*, 3066−3067.

(22) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985−1993.

(23) Cremer, D. Møller−Plesset perturbation theory: From small molecule methods to methods for thousands of atoms. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 509−530.

(24) Szabo, A.; Ostlund, N. S. The correlation energy in the random phase approximation: Intermolecular forces between closed-shell systems. *J. Chem. Phys.* **1977**, *67*, 4351−4360.

(25) Tkatchenko, A.; DiStasio, R. A., Jr.; Head-Gordon, M.; Scheffler, M. Dispersion-corrected Møller−Plesset second-order perturbation theory. *J. Chem. Phys.* **2009**, *131*, 094106:1−7.

(26) van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H. State of the art in counterpoise theory. *Chem. Rev.* **1994**, *94*, 1873−1885.

(27) Hopkins, B. W.; Tschumper, G. S. Ab initio studies of π···π interactions: The effects of quadruple excitations. *J. Phys. Chem. A* **2004**, *108*, 2941−2948.

(28) Hobza, P. Calculations on noncovalent interactions and databases of benchmark interaction energies. *Acc. Chem. Res.* **2012**, *45*, 663−672.

(29) Dubecký, M.; Jurečka, P.; Derian, R.; Hobza, P.; Otyepka, M.; Mitas, L. Quantum Monte Carlo methods describe noncovalent interactions with subchemical accuracy. *J. Chem. Theory Comput.* **2013**, *9*, 4287−4292.

248

dx.doi.org/10.1021/jp5098603 | *J. Phys. Chem. A* 2015, 119, 235−252

(30) Ambrosetti, A.; Alfè, D.; DiStasio, R. A., Jr.; Tkatchenko, A. Hard numbers for large molecules: Toward exact energetics for supramolecular systems. *J. Phys. Chem. Lett.* **2014**, *5*, 849−855.

(31) Řezáč, J.; Hobza, P. Describing noncovalent interactions beyond the common approximations: How accurate is the "gold standard", CCSD(T) at the complete basis set limit? *J. Chem. Theory Comput.* **2013**, *9*, 2151−2155.

(32) Šimová, L.; Řezáč, J.; Hobza, P. Convergence of the interaction energies in noncovalent complexes in the coupled-cluster methods up to full configuration interaction. *J. Chem. Theory Comput.* **2013**, *9*, 3420−3428.

(33) Smith, D. G. A.; Jankowski, P.; Slawik, M.; Witek, H. A.; Patkowski, K. Basis set convergence of the post-CCSD(T) contribution to noncovalent interaction energies. *J. Chem. Theory Comput.* **2014**, *10*, 3140−3150.

(34) Marshall, M. S.; Burns, L. A.; Sherrill, C. D. Basis set convergence of the coupled-cluster correction, $\delta_{MP2}^{CCSD(T)}$: Best practices for benchmarking non-covalent interactions and the attendant revision of the S22, NBC10, and HSG databases. *J. Chem. Phys.* **2011**, *135*, 194102:1−10.

(35) Yoo, S.; Aprà, E.; Zeng, X. C.; Xantheas, S. S. High-level ab initio electronic structure calculations of water clusters $(H_2O)_{16}$ and $(H_2O)_{17}$: A new global minimum for $(H_2O)_{16}$. *J. Phys. Chem. Lett.* **2010**, *1*, 3122−3127.

(36) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427−2438. Erratum: Řezáč, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2014**, *10*, 1359−1360.

(37) Goerigk, L.; Kruse, H.; Grimme, S. Benchmarking density functional methods against the S66 and S66 × 8 datasets for non-covalent interactions. *ChemPhysChem* **2011**, *12*, 3421−3433.

(38) Vydrov, O. A.; Van Voorhis, T. Benchmark assessment of the accuracy of several van der Waals density functionals. *J. Chem. Theory Comput.* **2012**, *8*, 1929−1934.

(39) Lin, Y.-S.; Li, G.-D.; Mao, S.-P.; Chai, J.-D. Long-range corrected hybrid density functionals with improved dispersion corrections. *J. Chem. Theory Comput.* **2013**, *9*, 263−272.

(40) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.

(41) Pitoňňák, M.; Řezáč, J.; Hobza, P. Spin-component scaled coupled-clusters singles and doubles optimized towards calculation of noncovalent interactions. *Phys. Chem. Chem. Phys.* **2010**, *12*, 9611−9614.

(42) Heßelmann, A. Improved supermolecular second order Møller−Plesset intermolecular interaction energies using time-dependent density functional response theory. *J. Chem. Phys.* **2008**, *128*, 144112:1−9.

(43) Pitoňňák, M.; Heßelmann, A. Accurate intermolecular interaction energies from a combination of MP2 and TDDFT response theory. *J. Chem. Theory Comput.* **2010**, *6*, 168−178.

(44) Huang, Y.; Shao, Y.; Beran, G. J. O. Accelerating MP2C dispersion corrections for dimers and molecular crystals. *J. Chem. Phys.* **2013**, *138*, 224112:1−8.

(45) Huang, Y.; Goldey, M.; Head-Gordon, M.; Beran, G. J. O. Achieving high-accuracy intermolecular interactions by combining Coulomb-attenuated second-order Møller−Plesset perturbation theory with coupled Kohn−Sham dispersion. *J. Chem. Theory Comput.* **2014**, *10*, 2054−2063.

(46) Lao, K. U.; Herbert, J. M. An improved treatment of empirical dispersion and a many-body energy decomposition scheme for the explicit polarization plus symmetry-adapted perturbation theory (XSAPT) method. *J. Chem. Phys.* **2013**, *139*, 034107:1−16. Erratum: Lao, K. U.; Herbert, J. M. *J. Chem. Phys.* **2014**, *140*, 119901.

(47) Chen, Z.; Li, Y.; Chen, E.; Hall, D. L.; Darke, P. L.; Culberson, C.; Shafer, J. A.; Kuo, L. C. Crystal structure at 1.9-Å resolution of human immunodeficiency virus (HIV) II protease complexed with L-735,524, an orally bioavailable inhibitor of the HIV proteases. *J. Biol. Chem.* **1994**, *269*, 26344−26348.

(48) Ucisik, M. N.; Dashti, D. S.; Faver, J. C.; Merz, K. M., Jr. Pairwise additivity of energy components in protein-ligand binding: The HIV II protease-indinavir case. *J. Chem. Phys.* **2011**, *135*, 085101:1−12.

(49) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation methods: A route to accurate calculations on large systems. *Chem. Rev.* **2012**, *112*, 632−672.

(50) Jacobson, L. D.; Richard, R. M.; Lao, K. U.; Herbert, J. M. Efficient monomer-based quantum chemistry methods for molecular and ionic clusters. *Annu. Rep. Comput. Chem.* **2013**, *9*, 25−56.

(51) Jacobson, L. D.; Herbert, J. M. An efficient, fragment-based electronic structure method for molecular systems: Self-consistent polarization with perturbative two-body exchange and dispersion. *J. Chem. Phys.* **2011**, *134*, 094118:1−17.

(52) Lao, K. U.; Herbert, J. M. Accurate intermolecular interactions at dramatically reduced cost: XPol+SAPT with empirical dispersion. *J. Phys. Chem. Lett.* **2012**, *3*, 3241−3248.

(53) Herbert, J. M.; Jacobson, L. D.; Lao, K. U.; Rohrdanz, M. A. Rapid computation of intermolecular interactions in molecular and ionic clusters: Self-consistent polarization plus symmetry-adapted perturbation theory. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7679−7699.

(54) Lao, K. U.; Herbert, J. M. Symmetry-adapted perturbation theory with Kohn-Sham orbitals using non-empirically tuned, long-range-corrected density functionals. *J. Chem. Phys.* **2014**, *140*, 044108:1−8.

(55) Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. The variational explicit polarization potential and analytical first derivative of energy: Towards a next generation force field. *J. Chem. Phys.* **2008**, *128*, 234108:1−9.

(56) Rybak, S.; Jeziorski, B.; Szalewicz, K. Many-body symmetry-adapted perturbation theory of intermolecular interactions. H_2O and HF dimers. *J. Chem. Phys.* **1991**, *95*, 6576−6601.

(57) Korona, T. First-order exchange energy of intermolecular interactions from coupled cluster density matrices and their cumulants. *J. Chem. Phys.* **2008**, *128*, 224104:1−14.

(58) Korona, T. Second-order exchange-induction energy of intermolecular interactions from coupled cluster density matrices and their cumulants. *Phys. Chem. Chem. Phys.* **2009**, *10*, 6509−6519.

(59) Korona, T. Exchange-dispersion energy: A formulation in terms of monomer properties and coupled cluster treatment of intra-monomer correlation. *J. Chem. Theory Comput.* **2009**, *5*, 2663−2678.

(60) Korona, T. A coupled cluster treatment of intramonomer electron correlation within symmetry-adapted perturbation theory: Benchmark calculations and a comparison with a density-functional theory description. *Mol. Phys.* **2013**, *111*, 3705−3715.

(61) Williams, H. L.; Chabalowski, C. F. Using Kohn-Sham orbitals in symmetry-adapted perturbation theory to investigate intermolecular interactions. *J. Phys. Chem. A* **2001**, *105*, 646−659.

(62) Jansen, G.; Heßelmann, A. Comment on "Using Kohn−Sham orbitals in symmetry-adapted perturbation theory to investigate intermolecular interactions". *J. Phys. Chem. A* **2001**, *105*, 11156−11157.

(63) Misquitta, A. J.; Szalewicz, K. Intermolecular forces from asymptotically corrected density functional description of monomers. *Chem. Phys. Lett.* **2002**, *357*, 301−306.

(64) Tozer, D. J.; Handy, N. C. Improving virtual Kohn−Sham orbitals and eigenvalues: Application to excitation energies and static polarizabilities. *J. Chem. Phys.* **1998**, *109*, 10180−10189.

(65) Perdew, J. P.; Burke, K. Comparison shopping for a gradient-corrected density functional. *Int. J. Quantum Chem.* **1996**, *57*, 309−319.

(66) Heßelmann, A.; Jansen, G. First-order intermolecular interaction energies from Kohn-Sham orbitals. *Chem. Phys. Lett.* **2002**, *357*, 464−470.

(67) Heßelmann, A.; Jansen, G. Intermolecular induction and exchange-induction energies from coupled-perturbed Kohn-Sham density functional theory. *Chem. Phys. Lett.* **2002**, *362*, 319−325.

(68) Misquitta, A. J.; Jeziorski, B.; Szalewicz, K. Dispersion energy from density-function theory description of monomers. *Phys. Rev. Lett.* **2003**, *91*, 033201:1−4.

(69) Heßelmann, A.; Jansen, G. Intermolecular dispersion energies from time-dependent density functional theory. *Chem. Phys. Lett.* **2003**, *367*, 778−784.

(70) Heßelmann, A.; Jansen, G. The helium dimer potential from a combined density functional theory and symmetry-adapted perturbation theory approach using an exact exchange−correlation potential. *Phys. Chem. Chem. Phys.* **2003**, *5*, 5010−5014.

(71) Heßelmann, A.; Jansen, G.; Schütz, M. Density-functional theory symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies. *J. Chem. Phys.* **2005**, *122*, 014103:1−17.

(72) Misquitta, A. J.; Szalewicz, K. Symmetry-adapted perturbation-theory calculations of intermolecular forces employing density-functional description of monomers. *J. Chem. Phys.* **2005**, *122*, 214109:1−19.

(73) Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations. *J. Chem. Phys.* **2005**, *123*, 214103:1−14.

(74) Bukowski, R.; Podeszwa, R.; Szalewicz, K. Efficient calculation of coupled Kohn−Sham dynamic susceptibility functions and dispersion energies with density fitting. *Chem. Phys. Lett.* **2005**, *414*, 111−116.

(75) Podeszwa, R.; Bukowski, R.; Szalewicz, K. Density-fitting method in symmetry-adapted perturbation theory based on Kohn-Sham description of monomers. *J. Chem. Theory Comput.* **2006**, *2*, 400−412.

(76) Turki, N.; Milet, A.; Rahmouni, A.; Ouamerali, O.; Moszynski, R.; Kochanski, E.; Wormer, P. E. S. Theoretical study of the $OH^-(H_2O)_2$ system: Nature and importance of three-body interactions. *J. Chem. Phys.* **1998**, *109*, 7157−7168.

(77) Chen, Y.; Li, H. Intermolecular interaction in water hexamer. *J. Phys. Chem. A* **2010**, *114*, 11719−11724.

(78) Dahlke, E. E.; Truhlar, D. G. Electrostatically embedded many-body correlation energy, with applications to the calculation of accurate second-order Møller−Plesset perturbation theory energies for large water clusters. *J. Chem. Theory Comput.* **2007**, *3*, 1342−1348.

(79) McDaniel, J. G.; Schmidt, J. R. First-principles many-body force fields from the gas phase to liquid: A "universal" approach. *J. Phys. Chem. B* **2014**, *118*, 8042−8053.

(80) Lotrich, V. F.; Szalewicz, K. Symmetry-adapted perturbation theory of three-body nonadditivity of intermolecular interaction energy. *J. Chem. Phys.* **1997**, *106*, 9668−9687.

(81) Lotrich, V. F.; Szalewicz, K. Symmetry-adapted perturbation theory of three-body nonadditivity in Ar trimer. *J. Chem. Phys.* **1997**, *106*, 9688−9702.

(82) Lotrich, V. F.; Szalewicz, K. Perturbation theory of three-body exchange nonadditivity and application to helium trimer. *J. Chem. Phys.* **2000**, *112*, 112−121.

(83) Gao, J. Toward a molecular orbital derived empirical potential for liquid simulations. *J. Phys. Chem. B* **1997**, *101*, 657−663.

(84) Gao, J. A molecular-orbital derived polarization potential for liquid water. *J. Chem. Phys.* **1998**, *109*, 2346−2354.

(85) Xie, W.; Gao, J. Design of a next generation force field: The X-POL potential. *J. Chem. Theory Comput.* **2007**, *3*, 1890−1900.

(86) Gao, J.; Truhlar, D. G.; Wang, Y.; Mazack, M. J. M.; Löffler, P.; Provorse, M. R.; Rehak, P. Explicit polarization: A quantum mechanical framework for developing next generation force fields. *Acc. Chem. Res.* **2014**, *47*, 2837−2845.

(87) Breneman, C. M.; Wiberg, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* **1990**, *11*, 361−373.

(88) Holden, Z. C.; Richard, R. M.; Herbert, J. M. Periodic boundary conditions for QM/MM calculations: Ewald summation for extended Gaussian basis sets. *J. Chem. Phys.* **2013**, *139*, 244108:1−13.

(89) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment molecular orbital method: An approximate computational method for large molecules. *Chem. Phys. Lett.* **1999**, *313*, 701−706.

(90) Fedorov, D. G.; Kitaura, K. Extending the power of quantum chemistry to large systems with the fragment molecular orbital method. *J. Phys. Chem. A* **2007**, *111*, 6904−6914.

(91) Dahlke, E. E.; Truhlar, D. G. Electrostatically embedded many-body expansion for large systems, with applications to water clusters. *J. Chem. Theory Comput.* **2007**, *3*, 46−53.

(92) Richard, R. M.; Lao, K. U.; Herbert, J. M. Aiming for benchmark accuracy with the many-body expansion. *Acc. Chem. Res.* **2014**, *47*, 2828−2836.

(93) Hratchian, H. P.; Parandekar, P. V.; Raghavachari, K.; Frisch, M. J.; Vreven, T. QM:QM electronic embedding using Mulliken atomic charges: Energies and analytic gradients in an ONIOM framework. *J. Chem. Phys.* **2008**, *128*, 034107:1−11.

(94) Nagata, T.; Brorsen, K.; Fedorov, D. G.; Kitaura, K.; Gordon, M. S. Fully analytic energy gradient in the fragment molecular orbital method. *J. Chem. Phys.* **2011**, *134*, 124115:1−13.

(95) Fedorov, D. G.; Ishida, T.; Uebayasi, M.; Kitaura, K. The fragment molecular orbital method for geometry optimizations of polypeptides and proteins. *J. Phys. Chem. A* **2007**, *111*, 2722−2732.

(96) Hua, S.; Hua, W.; Li, S. An efficient implementation of the generalized energy-based fragmentation approach for general large molecules. *J. Phys. Chem. A* **2010**, *114*, 8126−8134.

(97) Li, W.; Hua, W.; Fang, T.; Li, S. The energy-based fragmentation approach for ab initio calculations of large systems. In *Computational Methods for Large Systems: Electronic Structure Approaches for Biotechnology and Nanotechnology*; Reimers, J. R., Ed.; Wiley: Hoboken, NJ, 2011; pp 227−258.

(98) Bottcher, C. J. F.; Belle, O. V.; Bordewijk, P.; Rip, A. *Theory of Electric Polarization*; Vol. 1 of Dielectrics in Static Fields; Elsevier: Amsterdam, 1973.

(99) Jacobson, L. D.; Williams, C. F.; Herbert, J. M. The static-exchange electron-water pseudopotential, in conjunction with a polarizable water model: A new Hamiltonian for hydrated-electron simulations. *J. Chem. Phys.* **2009**, *130*, 124115:1−18.

(100) Hirschfelder, J. O. Perturbation theory for exchange forces. I. *Chem. Phys. Lett.* **1967**, *1*, 325−329.

(101) Hirschfelder, J. O.; Meath, W. J. Nature of intermolecular forces. *Adv. Chem. Phys.* **1967**, *12*, 3−106.

(102) Jeziorski, B.; Bulski, M.; Piela, L. First-order perturbation treatment of the short-range repulsion in a system of many closed-shell atoms or molecules. *Int. J. Quantum Chem.* **1976**, *10*, 281−297.

(103) Schäffer, R.; Jansen, G. Intermolecular exchange-induction energies without overlap expansion. *Theor. Chem. Acc.* **2012**, *131*, 1235:1−10.

(104) Schäffer, R.; Jansen, G. Single-determinant-based symmetry-adapted perturbation theory without single-exchange approximation. *Mol. Phys.* **2013**, *111*, 2570−2584.

(105) Lao, K. U.; Herbert, J. M. Breakdown of the single-exchange approximation in third-order symmetry-adapted perturbation theory. *J. Phys. Chem. A* **2012**, *116*, 3042−3047.

(106) Lao, K. U.; Schäffer, R.; Jansen, G.; Herbert, J. M. Accurate description of intermolecular interactions in ionic systems using symmetry-adapted perturbation theory. Submitted for publication.

(107) Fedorov, D. G.; Nagata, T.; Kitaura, K. Exploring chemistry with the fragment molecular orbital method. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7562−7577.

(108) Moszynski, R.; Jeziorski, B.; Ratkiewicz, A.; Rybak, S. Many-body perturbation theory of electrostatic interactions between molecules: Comparison with full configuration interaction for four-electron dimers. *J. Chem. Phys.* **1993**, *99*, 8856.

(109) Moszyński, R.; Cybulski, S. M.; Chałasiński, G. Many-body theory of intermolecular induction interactions. *J. Chem. Phys.* **1994**, *100*, 4998−5010.

(110) Williams, H. L.; Mas, E. M.; Szalewicz, K.; Jeziorski, B. On the effectiveness of monomer-, dimer-, and bond-centered basis functions in calculations of intermolecular interaction energies. *J. Chem. Phys.* **1995**, *103*, 7374−7391.

(111) Wolinski, K.; Pulay, P. Second-order Møller-Plesset calculations with dual basis sets. *J. Chem. Phys.* **2003**, *118*, 9497−9503.

(112) Steele, R. P.; DiStasio, R. A., Jr.; Shao, Y.; Kong, J.; Head-Gordon, M. Dual-basis second-order Møller-Plesset perturbation theory: A reduced-cost reference for correlation calculations. *J. Chem. Phys.* **2006**, *125*, 074108:1−11.

(113) Hohenstein, E. G.; Sherrill, C. D. Density fitting of intramonomer correlation effects in symmetry adapted perturbation theory. *J. Chem. Phys.* **2010**, *133*, 014101:1−12.

(114) van Leeuwen, R.; Baerends, E. J. Exchange-correlation potential with correct asymptotic behavior. *Phys. Rev. A* **1994**, *49*, 2421−2431.

(115) Tozer, D. J.; Handy, N. C. Improving virtual Kohn-Sham orbitals and eigenvalues: Application to excitation energies and static polarizabilities. *J. Chem. Phys.* **1998**, *109*, 10180−10189.

(116) Gruning, M.; Gritsenko, O. V.; van Gisbergen, S. J. A.; Baerends, E. J. Shape corrections to exchange-correlation potentials by gradient-regulated seamless connection of model potentials for inner and outer region. *J. Chem. Phys.* **2001**, *114*, 652−660.

(117) Gaiduk, A. P.; Staroverov, V. N. How to tell when a model Kohn−Sham potential is not a functional derivative. *J. Chem. Phys.* **2009**, *131*, 044107:1−7.

(118) Iikura, H.; Tsuneda, T.; Yanai, T.; Hirao, K. A long-range correction scheme for generalized-gradient-approximation exchange functionals. *J. Chem. Phys.* **2001**, *115*, 3540−3544.

(119) Henderson, T. M.; Janesko, B. G.; Scuseria, G. E. Generalized gradient approximation model exchange holes for range-separated hybrids. *J. Chem. Phys.* **2008**, *128*, 194105:1−9.

(120) Livshits, E.; Baer, R. A well-tempered density functional theory of electrons in molecules. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2932−2941.

(121) Rohrdanz, M. A.; Martins, K. M.; Herbert, J. M. A long-range-corrected density functional that performs well for both ground-state properties and time-dependent density functional theory excitation energies, including charge-transfer excited states. *J. Chem. Phys.* **2009**, *130*, 054112:1−8.

(122) Stein, T.; Kronik, L.; Baer, R. Prediction of charge-transfer excitations in coumarin-based dyes using a range-separated functional tuned from first principles. *J. Chem. Phys.* **2009**, *131*, 244119:1−5.

(123) Baer, R.; Livshits, E.; Salzner, U. Tuned range-separated hybrids in density functional theory. *Annu. Rev. Phys. Chem.* **2010**, *61*, 85−109.

(124) Heßelmann, A. Comparison of intermolecular interaction energies from SAPT and DFT including empirical dispersion contributions. *J. Phys. Chem. A* **2011**, *115*, 11321−11330.

(125) Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comput. Chem.* **2004**, *25*, 1463−1473.

(126) Podeszwa, R.; Pernal, K.; Patkowski, K.; Szalewicz, K. Extension of the Hartree−Fock plus dispersion method by first-order correlation effects. *J. Phys. Chem. Lett.* **2010**, *1*, 550−555.

(127) Tang, K. T.; Toennies, J. P. An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients. *J. Chem. Phys.* **1984**, *80*, 3726−3741.

(128) Williams, H. L.; Szalewicz, K.; Moszynski, R.; Jeziorski, B. Dispersion energy in the coupled pair approximation with noniterative inclusion of single and triple excitations. *J. Chem. Phys.* **1995**, *103*, 4586−4599.

(129) Parrish, R. M.; Hohenstein, E. G.; Sherrill, C. D. Tractability gains in symmetry-adapted perturbation theory including coupled double excitations: CCD+ST(CCD) dispersion with natural orbital truncations. *J. Chem. Phys.* **2013**, *139*, 174102:1−15.

(130) Lange, A. W.; Herbert, J. M. Both intra- and interstrand charge-transfer excited states in B-DNA are present at energies comparable to, or just above, the $^1\pi\pi^*$ excitonic bright states. *J. Am. Chem. Soc.* **2009**, *131*, 3913−3922.

(131) Maurer, S. A.; Beer, M.; Lambrecht, D. S.; Ochsenfeld, C. Linear-scaling symmetry-adapted perturbation theory with scaled dispersion. *J. Chem. Phys.* **2013**, *139*, 184104:1−7.

(132) Fiethen, A.; Jansen, G.; Hesselmann, A.; Schütz, M. Stacking energies for average B-DNA structures from the combined density functional theory and symmetry-adapted perturbation theory approach. *J. Am. Chem. Soc.* **2008**, *130*, 1802−1803.

(133) DiLabio, G. A.; Johnson, E. R. Application of 25 density functionals to dispersion-bound homomolecular dimers. *Phys. Chem. Chem. Phys.* **2013**, *15*, 12821−12828.

(134) Goldey, M.; Dutoi, A.; Head-Gordon, M. Attenuated second-order Møller−Plesset perturbation theory: performance within the aug-cc-pVTZ basis. *Phys. Chem. Chem. Phys.* **2013**, *15*, 15869−15875.

(135) Goldey, M.; Head-Gordon, M. Attenuating away the errors in inter- and intramolecular interactions from second-order Møller−Plesset calculations in the small aug-cc-pVDZ basis set. *J. Phys. Chem. Lett.* **2012**, *3*, 3592−3598.

(136) Huang, Y.; Goldey, M.; Head-Gordon, M.; Beran, G. J. O. Achieving high-accuracy intermolecular interactions by combining Coulomb-attenuated second-order Møller-Plesset perturbation theory with coupled Kohn-Sham dispersion. *J. Phys. Chem. A* **2014**, *10*, 2054−2063.

(137) Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. An assessment of theoretical methods for nonbonded interactions: Comparison to complete basis set limit coupled-cluster potential energy curves for benzene dimer, the methane dimer, benzene-methane, and benzene-H_2S. *J. Phys. Chem. A* **2009**, *113*, 10146−10159.

(138) Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; Wiley: New York, 2000.

(139) Gordon, M. S.; Smith, Q. A.; Xu, P.; Slipchenko, L. V. Accurate first principles model potentials for intermolecular interactions. *Annu. Rev. Phys. Chem.* **2013**, *64*, 553−578.

(140) Valeev, E. Improving on the resolution of the identity in linear R12 ab initio theories. *Chem. Phys. Lett.* **2004**, *395*, 190−195.

(141) Bischoff, F. A.; Wolfsegger, S.; Tew, D. P.; Klopper, W. Assessment of basis sets for F12 explicitly-correlated molecular electronic-structure methods. *Mol. Phys.* **2009**, *107*, 963−975.

(142) Hartke, B. Size-dependent transition from all-surface to interior-molecule structures in pure neutral water clusters. *Phys. Chem. Chem. Phys.* **2003**, *5*, 275−284.

(143) Lagutschenkov, A.; Fanourgakis, G. S.; Niedner-Schatteburg, G.; Xantheas, S. S. The spectroscopic signature of the "all-surface" to "internally-solvated" structural transition in water clusters in the $n = 17−21$ size regime. *J. Chem. Phys.* **2005**, *122*, 194310:1−9.

(144) Yoo, S.; Xantheas, S. S. Structures, energetics, and spectroscopic fingerprints of water clusters $n = 2−24$. In *Handbook of Computational Chemistry*; Leszczynski, J., Ed.; Springer Science + Business Media: Berlin, 2012; Chapter 21, pp 761−792.

(145) Kazimirski, J. K.; Buch, V. Search for low energy structures of water clusters $(H_2O)_n$, $n = 20−22$, 48, 123, and 293. *J. Phys. Chem. A* **2003**, *107*, 9762−9775.

(146) Wang, K.; Li, W.; Li, S. Generalized energy-based fragmentation CCSD(T)-F12a method and application to the relative energies of water clusters $(H_2O)_{20}$. *J. Chem. Theory Comput.* **2014**, *10*, 1546−1553.

(147) Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodošček, M.; Brooks, B. R. Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method. *J. Chem. Phys.* **2002**, *117*, 10534−10547.

(148) Parker, T. M.; Burns, L. A.; Parrish, R. M.; Ryno, A. G.; Sherrill, C. D. Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies. *J. Chem. Phys.* **2014**, *140*, 094106:1−16.

(149) Deible, M. J.; Tuguldur, O.; Jordan, K. D. Theoretical study of the binding energy of a methane molecule in a $(H_2O)_{20}$ dodecahedral cage. *J. Phys. Chem. B* **2014**, *118*, 8257−8263.

(150) Kumar, P.; Sathyamurthy, M. Theoretical studies of host−guest interaction in gas hydrates. *J. Phys. Chem. A* **2011**, *115*, 14276−14281.

(151) Ramya, K. R.; Venkatnathan, A. Stability and reactivity of methane clathrate hydrates: Insights from density functional theory. *J. Phys. Chem. A* **2012**, *116*, 7742−7745.

(152) Liu, Y.; Zhao, J.; Li, F.; Chen, Z. Appropriate description of intermolecular interactions in the methane hydrates: An assessment of DFT methods. *J. Comput. Chem.* **2013**, *34*, 121−131.

(153) Weill, N.; Therrien, E.; Campagna-Slater, V.; Moitessier, N. Methods for docking small molecules to macromolecules: A user's perspective. 1. The theory. *Curr. Pharm. Design* **2014**, *20*, 3338−3359.

(154) Campagna-Slater, V.; Therrien, E.; Weill, N.; Moitessier, N. Methods for docking small molecules to macromolecules: A user's perspective. 2. Applications. *Curr. Pharm. Design* **2014**, *20*, 3360−3372.

(155) von Lilienfeld, O. A.; Tuckerman, M. E. Alchemical variations of intermolecular energies according to molecular grand-canonical ensemble density functional theory. *J. Chem. Theory Comput.* **2007**, *3*, 1083−1090.

(156) Bryce, R. A.; Hillier, I. H. Quantum chemical approaches: Semiempirical molecular orbital and hybrid quantum mechanical/molecular mechanical techniques. *Curr. Pharm. Design* **2014**, *20*, 3293−3302.

(157) Stiborova, M.; Sejbal, J.; Borek-Dohalska, L.; Aimova, D.; Poljakova, J.; Forsterova, K.; Rupertova, M.; Wiesner, J.; Hudecek, J.; Wiessler, M.; Frei, E. The anticancer drug ellipticine forms covalent DNA adducts, mediated by human cytochromes P450, through metabolism to 13-hydroxyellipticine and ellipticine N^2-oxide. *Cancer Res.* **2004**, *64*, 8374−8380.

(158) Benali, A.; Shulenburger, L.; Romero, N. A.; Kim, J.; von Lilienfeld, O. A. Application of diffusion Monte Carlo to materials dominated by van der Waals interactions. *J. Chem. Theory Comput.* **2014**, *10*, 3417−3422.

(159) von Lilienfeld, O. A.; Tkatchenko, A. Two- and three-body interatomic dispersion energy contributions to binding in molecules and solids. *J. Chem. Phys.* **2010**, *132*, 234109:1−11.

(160) Sedlak, R.; Janowski, T.; Pitoňák, M.; Řezáč, J.; Pulay, P.; Hobza, P. Accuracy of quantum chemical methods for large noncovalent complexes. *J. Chem. Theory Comput.* **2013**, *9*, 3364−3374.

252

dx.doi.org/10.1021/jp5098603 | *J. Phys. Chem. A* 2015, 119, 235−252