The Journal of Physical Chemistry Letters
Letter
pubs.acs.org/JPCL

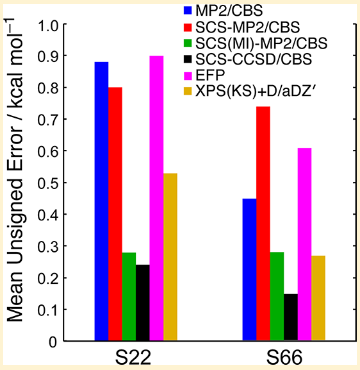
Accurate Intermolecular Interactions at Dramatically Reduced Cost: XPol+SAPT with Empirical Dispersion

Ka Un Lao and John M. Herbert*

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

Ⓢ *Supporting Information*

ABSTRACT: An efficient, monomer-based electronic structure method is introduced for computing noncovalent interactions in molecular and ionic clusters. It builds upon our explicit polarization plus symmetry-adapted perturbation theory approach, XPol+SAPT (XPS), but replaces the problematic and expensive sum-over-states dispersion terms with empirical potentials. This modification reduces the scaling from $O(N^5)$ to $O(N^3)$ with respect to monomer size and also facilitates the use of Kohn–Sham density functional theory (KS-DFT) as a low-cost means to capture intramolecular electron correlation. The accuracy of the resulting method [XPS(KS)+D], in conjunction with a double-ζ basis set, is superior to MP2-type methods extrapolated to the basis-set limit, with a mean unsigned error of 0.27 kcal/mol for the S66 data set. XPS(KS)+D yields accurate potential energy curves for a variety of challenging systems. As compared to traditional DFT-SAPT methods, it removes the limitation to dimers and extends SAPT-based methodology to many-body systems.



SECTION: Molecular Structure, Quantum Chemistry, and General Theory

Dispersion interactions are very important in biological systems, for example, in protein folding and in the structure of DNA.[1] In the latter case, the double helix is maintained both by dispersion-dominated π-stacking interactions within a strand and by hydrogen bonding between complementary strands. Although the H-bonding interactions are *dominated* by electrostatic effects, dispersion is still responsible for 20–30% of the base-pairing interaction,[2] and contrary to popular belief, experiments demonstrate that the stability of the double helix is mainly determined by base stacking rather than base pairing.[3] Therefore, in the theoretical description of noncovalent interactions, it is crucial to employ methods that furnish an accurate description of dispersion interactions.

The "gold standard" of electronic structure theory, CCSD(T), is accurate enough for this purpose but exhibits a cost that grows as $O(N^7)$ with respect to system size, N. This fact, in conjunction with the large basis sets that are required for accurate calculations of intermolecular interaction energies (e.g., to eliminate basis-set superposition error, BSSE) limits high-level ab initio calculations to small systems. Furthermore, even given accurate CCSD(T) results for the intermolecular interaction energy it is not easy to ascertain how much of this interaction is due to dispersion versus other effects such as exchange, electrostatics, or induction.

Symmetry-adapted perturbation theory (SAPT) is an alternative method to compute intermolecular interaction energies,[1,4,5] using a monomer-based formalism rather than a supersystem calculation. As such, the SAPT interaction energy is free of BSSE, by construction, and is also decomposable into a sum of physically meaningful contributions:[1]

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

Here, we have explicitly listed all terms up to second order in the intermolecular interaction, with subscripts that denote electrostatic (elst), exchange (exch), induction (ind), and dispersion (disp) contributions. These low-order terms neglect intramolecular electron correlation, and for high-accuracy SAPT calculations, it is essential to use a double-perturbation expansion that accounts for both inter- and intramolecular correlation.[1] Inclusion of intramolecular electron correlation, however, results in a SAPT method whose cost scales as $O(N^7)$, so there is no cost savings over CCSD(T), although the SAPT result can still be used to decompose the interaction energy.

In principle, Kohn–Sham density functional theory (KS-DFT) offers a low-cost means to describe intramolecular electron correlation, and has been introduced in the context of SAPT calculations as the so-called SAPT(KS) method.[6,7] This approach is disastrously unsuccessful, however, unless functionals with asymptotically corrected exchange-correlation potentials are employed, but even so, the dispersion energies computed with SAPT(KS) remain quite poor.[6−10] The solution to this dilemma is to replace the MP2-like sum-over-states dispersion formula used in traditional low-order SAPT with a formula involving frequency-dependent density susceptibilities, computed using KS-DFT. This method has variously been called DFT-SAPT[11] or SAPT(DFT).[12] Using density-fitting

Received: July 23, 2012
Accepted: October 21, 2012

techniques, it exhibits $O(N^5)$ scaling yet provides an accurate description of intermolecular interactions.[11,13]

SAPT methods have mainly been used to study dimers because calculation of nonadditive three-body interactions within the SAPT formalism requires computationally expensive triple excitations.[14,15] Many-body (nonpairwise-additive) effects are large in clusters of polar molecules but are dominated by induction (i.e., polarization).[16−19] In recognition of this, our group has recently introduced a low-cost, many-body generalization of the SAPT methodology that we call XPol+SAPT (XPS),[10,20] in which the variational explicit polarization (XPol) method[21] is used to generate one-body wave functions for subsequent pairwise SAPT calculations. The XPol procedure captures many-body polarization effects by means of a charge-embedded, monomer-based self-consistent field calculation whose cost scales as $O(n)$ with respect to the number of monomers, n. The subsequent second-order SAPT calculations scale as either $O(n^2)$ or $O(n^3)$ depending on the level of approximation,[10] but in any case these ${}_nC_2$ or ${}_nC_3$ SAPT calculations are embarrassingly parallelizable. With a suitable choice of basis set, dimer binding energies computed by XPS based on a Hartree−Fock description of the monomers [XPS(HF)] lie within 1 kcal/mol of high-level benchmarks.[10,20]

As with SAPT(KS), XPS results obtained using KS orbitals are notably inferior to those obtained using HF orbitals. In particular, dispersion energies are vastly overestimated by XPS(KS),[10] which is an artifact of the sum-over-states dispersion formula in conjunction with HOMO/LUMO gaps that are significantly smaller than HF gaps.[6] Ironically, the dispersion and exchange-dispersion terms are not only the least accurate but also the most time-consuming to compute, scaling as the fourth and fifth powers, respectively, of monomer size, whereas other second-order terms are no worse than $O(N^3)$.[22] Recently, Hesselmann[22] introduced a method termed SAPT+D, in which these terms are replaced by empirical atom−atom potentials. In this work, we implement and test an analogous XPS(KS)+D method, which offers important advantages over alternative electronic structure methods for noncovalent interactions:

(1) Unlike XPS(HF), it incorporates intramolecular correlation, and in a relatively low-cost way.
(2) The expensive and inaccurate sum-over-states dispersion formulas are replaced by simple scalar potentials.
(3) Unlike SAPT+D, the method is applicable to any number of monomers.
(4) The cost of XPS(KS)+D calculations is $O(N^3)$ with respect to monomer size and no worse than $O(n^3)$ with respect to the number of monomers, and is thus amenable to large systems.

Several asymptotic correction (AC) schemes for the KS exchange-correlation potential, v_{xc}, have been used in SAPT(KS) and SAPT(DFT),[23,24] although a drawback of these techniques is that the corrected potential, $v_{\text{xc}}^{\text{AC}}$, is *not* the functional derivative of the energy. As an alternative to traditional AC schemes, we utilize long-range-corrected (LRC) density functionals[25,26] to obtain the correct asymptotic behavior. Specifically, we employ an *ansatz* that we have called LRC-ωPBEh,[26] which is based on the short-range ωPBE exchange functional.[27] Rather than using the empirically optimized LRC-ωPBEh parameters suggested in ref 26, however, we reoptimize the fraction of short-range HF exchange (C_{HF}) and then apply a system-specific tuning of the range separation parameter (ω), as suggested by Baer et al.,[28] in order to satisfy the condition

$$\varepsilon_{\text{HOMO}} = -\text{IP} \qquad (2)$$

where "IP" denotes the lowest ionization potential. For supersystems composed of well-defined monomers, as considered here, we assume that noncovalent interactions do not greatly alter the monomer IPs; hence we just need to determine ω separately for each monomer using eq 2. The value of ω appropriate for the supersystem is the one corresponding to the lowest monomer IP. Our tests reveal that this assumption is quite robust.

The empirical dispersion potential used in this work is the one suggested for SAPT+D:[22]

$$E_{\text{disp}}^{\text{SAPT+D}} = -s_\beta \sum_{i\in A} \sum_{j\in B} f_{\text{damp}}(r_{ij}) \frac{C_{ij}}{r_{ij}^{\beta}} \qquad (B\neq A) \qquad (3)$$

Here i and j represent nuclei located on different monomers and

$$f_{\text{damp}}(r_{ij}) = \text{erf}\left(\frac{\alpha r_{ij}}{R_i + R_j}\right) \qquad (4)$$

is a damping function. The latter differs from the damping function typically used in dispersion-corrected DFT.[29] The role of the dispersion correction is different in SAPT+D than it is in DFT+D,[22] and in the latter method, the dispersion correction should turn on only at large intermolecular separation because DFT models the short-range interactions. As such, $E_{\text{disp}}^{\text{DFT+D}}$ "is a model-dependent quantity with no real physical meaning",[30] and short-range damping is needed to avoid overcounting of interactions. In SAPT, however, the dispersion contribution to the energy is well-defined and should contribute at all intermolecular distances. However, short-range damping is still required to ensure that the empirical potential in eq 3 is finite when r_{ij} is small. The damping in eq 4 is much slower than that used in DFT+D.[22]

The parameters C_{ij} are defined in terms of atomic C_6 coefficients:

$$C_{ij} = (C_{6,i}C_{6,j})^{1/2} \qquad (5)$$

The atomic parameters $C_{6,i}$ and R_i used here are taken from those developed by Grimme.[31] The parameters $\alpha = 1.087$ and $\beta = 5.67$ are taken from SAPT+D,[22] where they were fit to reproduce benchmark intermolecular interaction energies. (The fact that dispersion is well-defined in SAPT suggests that these parameters may be transferable; results presented below will ultimately validate this choice.) Ultimately, β is an empirical parameter, but the fact that $\beta \neq 6$ can be understood based on the observation that higher-order multipole terms may contribute as much as 20−30% to the dispersion interaction in the middle-range region.[32] Moreover, $E_{\text{exch-disp}}^{(2)}$ also contributes to the SAPT dispersion energy, and this component varies exponentially with distance.

In this work, the parameter s_β in eq 3 was optimized for XPS(KS)+D using the S22A database,[33] which revises the energetics of the original S22 set of dimers.[34] For fixed α and β, the fit for s_β is a simple linear one that minimizes the absolute percent deviations, because this provides a more balanced fit for both weakly- and strongly-bound systems as compared to a least-squares fit.[22] It is possible that a global, nonlinear fit of all three parameters might improve the results, but this has not been attempted.

Optimal values of s_β were obtained for two different basis sets: aug-cc-pVDZ′ (abbreviated aDZ′)[1] and def2-TZVPP.[35] (Hereafter, we drop the "def2" on Ahlrichs basis sets.) The aDZ′ basis has been suggested for use in low-order SAPT0 and XPS(HF) calculations of dispersion-bound systems.[1,36,37] For XPS(HF) calculations on ion−water complexes, however, much better results are obtained using TZVP, which exhibits only slightly worse error statistics for S22A,[10] and we have since discovered that the use of TZVPP further improves the performance for these systems. As in previous work,[10,20] we use smooth ChElPG embedding charges[10] for the XPol calculations and "projected" (pseudocanonical dimer) basis sets for the SAPT calculations, along with a resolution-of-identity approximation combined with standard auxiliary basis sets. All calculations were performed using a locally-modified version of the Q-Chem program.[38]

The optimized values of s_β are 0.7267 and 0.8439 for aDZ′ and TZVPP, respectively. For these values of s_β, the optimal value of C_{HF} = 0.6. Selected error statistics for XPS-based methods, as applied to the S22A database, are listed in Table 1. The overall mean unsigned error (MUE) for XPS(KS) is larger than that for XPS(HF), as previously observed,[10] primarily due to overestimation (by as much as 5 kcal/mol) of binding energies for dispersion-dominated complexes. These errors are significantly reduced by replacing the MP2-like sum-over-states dispersion formula with the empirical potential in eq 3. The best results are obtained using XPS(KS)+D/aDZ′, for which the MUE is 0.5 kcal/mol or 9%.

Table 1. Error Statistics, with Respect to CCSD(T)/CBS[a] Benchmarks, for the S22A Data Set of Ref 33

		error/kcal mol^{-1}		% error	
method	data set[b]	MUE[c]	max[d]	MUE[c]	max[d]
XPS(HF)/aDZ′	all	0.60	1.67	16.7	79.1
	H-bond	0.60	1.05	6.5	25.5
	disp.	0.85	1.67	32.1	79.1
	mixed	0.30	0.55	9.3	17.6
XPS(HF)/TZVPP	all	0.53	1.94	12.4	49.8
	H-bond	0.40	0.81	3.6	8.4
	disp.	0.78	1.94	22.0	49.8
	mixed	0.37	0.78	10.1	17.2
XPS(KS)/aDZ′	all	1.03	4.43	23.3	81.0
	H-bond	0.53	1.43	5.2	18.6
	disp.	1.79	4.43	46.3	81.0
	mixed	0.66	1.38	15.1	30.9
XPS(KS)/TZVPP	all	1.04	4.95	22.2	69.0
	H-bond	0.30	0.70	2.1	4.3
	disp.	1.80	4.95	38.2	69.0
	mixed	0.90	1.82	24.0	38.8
XPS(KS)+D/aDZ′	all	0.53	1.16	9.4	24.4
	H-bond	0.73	1.16	5.9	9.9
	disp.	0.38	1.01	9.8	24.4
	mixed	0.52	0.99	12.3	21.8
XPS(KS)+D/TZVPP	all	0.61	1.48	10.9	33.0
	H-bond	0.62	1.22	4.1	7.3
	disp.	0.63	1.48	13.0	33.0
	mixed	0.59	1.28	15.5	23.2

[a] CBS = complete basis set. [b] Statistics listed separately for the entire S22A data set, along with subsets consisting of strongly H-bonded dimers, dispersion-dominated dimers, and dimers whose interactions are of mixed influence, as classified in ref 33. [c] Mean unsigned error. [d] Maximum unsigned error.

Note that XPS(KS) outperforms XPS(HF) for the strongly H-bonded subset of S22A, where the intermolecular interactions are dominated by electrostatics and induction. We attribute this to the effects of intramolecular electron correlation because SAPT(KS) has previously been shown to provide accurate values for components of the intermolecular interaction other than dispersion, provided that v_{xc} is asymptotically correct.[7,39] Our results demonstrate that LRC functionals afford an alternative means to enforce correct asymptotic behavior of v_{xc} without abandoning the relationship $v_{\mathrm{xc}}(\mathbf{r}) = \delta E_{\mathrm{xc}}/\delta\rho(\mathbf{r})$.

The performance of CBS-extrapolated MP2 and CCSD methods, along with spin-component-scaled (SCS) versions thereof,[1] was evaluated previously for the S22A data set.[33] The relevant MUEs are:

- 0.88 kcal/mol (MP2/CBS)
- 0.80 kcal/mol (SCS-MP2/CBS)
- 0.28 kcal/mol [SCS(MI)-MP2/CBS]
- 0.24 kcal/mol (SCS-CCSD/CBS)

Our XPS(HF) and XPS(KS)+D results, obtained using a double-ζ basis set, are better than the MP2/CBS and SCS-MP2/CBS results. The XPS calculations may benefit from some cancellation of errors because the double-ζ basis sets used here are far from complete, but on the other hand BSSE is absent from XPS by construction. As such, one could argue that large basis sets are not an essential requirement for XPS calculations.

The XPS(KS) method will not be considered further because of known problems with the SAPT(KS) treatment of dispersion, evident from results in Table 1. The XPS(KS)+D method contains parameters that were fit to the S22A database, so it is possible that S22A error statistics are overly optimistic. For a blind test of this approach, we turn to the S66 database,[40] for which error statistics are summarized in Table 2.

Results for S66 obtained using XPS(HF) and XPS(KS)+D, with the two basis sets considered here, are comparable to or better than CBS-extrapolated results obtained with the MP2,

Table 2. Error Statistics, with Respect to CCSD(T)/CBS Benchmarks, for the S66 Data Set of Ref 40

	error/kcal mol^{-1}		% error	
method	MUE	max	MUE	max
XPS(HF)/aDZ′	0.54	1.91	15.3	53.9
XPS(HF)/TZVPP	0.39	1.11	9.9	29.0
XPS(KS)+D/aDZ′	0.27	0.94	7.0	53.9
XPS(KS)+D/TZVPP	0.46	1.38	11.9	60.9
MP2/CBS[a]	0.45			40
SCS-MP2/CBS[a]	0.74			79
SCS(MI)-MP2/CBS[a]	0.28			54
MP2.5/CBS[a]	0.12			16
CCSD/CBS[a]	0.62			73
SCS-CCSD/CBS[a]	0.15			6
SCS(MI)-CCSD/CBS[a]	0.06			6
EFP[b]	0.61			

[a] Results taken from ref 40. [b] Results for the effective fragment potential (EFP) from ref 41.

SCS-MP2, and CCSD methods. Particularly impressive is the XPS(KS)+D/aDZ′ method, whose MUE is just 0.27 kcal/mol and whose largest unsigned error is <1 kcal/mol. This method does exhibit a 54% error for the benzene-ethene dimer (which is underbound by 0.78 kcal/mol), the maximum percentage error that we observe for S66 with this approach. If we eliminate benzene-ethene and also pyridine-ethene, which is underbound by 0.66 kcal/mol (35%), then the MUE is reduced to 6% and the maximum unsigned error is reduced to 20% for XPS(KS)+D/aDZ′.

The high accuracy of XPS(KS)+D for dimers is primarily a result of the empirical dispersion correction, rather than the XPol procedure, as evident from a comparison to SAPT(KS)+D for dimers. Refitting of s_β for use with SAPT(KS)+D/aDZ′ barely changes this parameter (<0.01), and the MUE for the S66 database is 0.32 kcal/mol as compared with 0.27 kcal/mol for XPS(KS)+D/aDZ′.

For clusters of polar molecules, however, the use of XPol monomer wave functions has a significant effect. Figure 1 shows binding energies for certain[10] $(H_2O)_n$ clusters up to n = 20, as compared with MP2/CBS benchmarks.[55−59] Not surprisingly, XPS-based methods are significantly more accurate than the corresponding pairwise-additive SAPT methods, which neglect many-body polarization. Furthermore, XPS(KS)+D is more accurate than the XPS(HF) method that neglects the monomer electron correlation and uses the MP2-like sum-over-states dispersion formula. The accuracy of the XPS and SAPT binding energies would likely improve if the cluster geometries were self-consistently optimized at these same levels of theory,[10] whereas MP2 geometries are used in Figure 1.

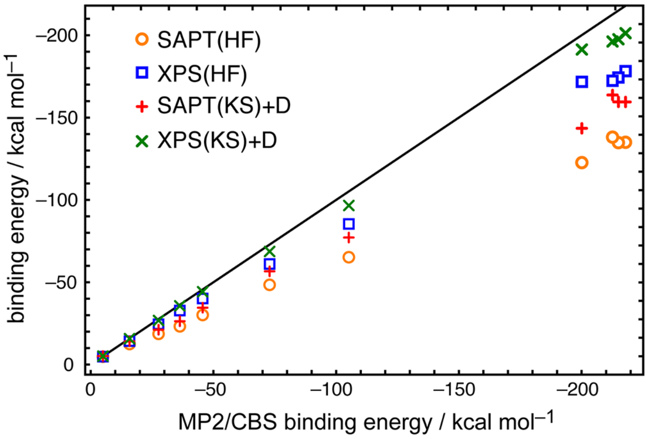


Figure 1. Binding energies for some $(H_2O)_n$ clusters, n = 2−20, computed at MP2 geometries. Binding energies on the vertical axis were computed with various XPS and SAPT methods using the aDZ′ basis, whereas the horizontal axis represents MP2/CBS benchmarks. Along the solid line, the XPS or SAPT result would be equal to the benchmark value.

The error statistics presented in Tables 1 and 2 demonstrate the excellent performance of XPS(KS)+D at equilibrium geometries, but it is also important to understand how this method performs across a range of intermolecular distances. To test this, we have chosen several challenging systems: Ar⋯Ne, formic acid dimer, benzene dimer, and $X^-(H_2O)$ with X = F, Cl. In the case of Ar⋯Ne, several DFT approaches thought to be accurate for noncovalent interactions predict qualitatively *incorrect* potential energy curves (PECs) for this system, even in cases where the binding energy evaluated at the minimum-energy geometry is accurate.[42] (Examples can be found in the Supporting Information.) The binding energy of the formic acid dimer has a large contribution from intramolecular electron correlation[37] and among the S22A molecules exhibits the maximum error at the SAPT0/aDZ′ level. Benzene dimer also represents a stringent test of dispersion interactions, and we consider both the "sandwich" isomer, which is dispersion-dominated, as well as the T-shaped isomer, where quadrupolar electrostatic interactions are important as well.[33] Finally, the $X^-(H_2O)$ complexes exhibit much larger binding energies than any of the S22A dimers.

One-dimensional PECs for these systems are depicted in Figures 2−5. The XPS(KS)+D/aDZ′ method yields PECs that are quite comparable to benchmark results for all of these difficult cases except $F^-(H_2O)$. For the charge-neutral systems, binding energy errors evaluated at equilibrium geometries range from 0.02 kcal/mol for Ar⋯Ne to 0.77 kcal/mol for $(HCO_2H)_2$. For $Cl^-(H_2O)$ and $F^-(H_2O)$, the XPS(KS)+D/aDZ′ method affords larger binding energy errors: 2.07 and 2.94 kcal/mol, respectively, at the equilibrium geometries. Similar binding energies are obtained at the XPS(HF)/aDZ′

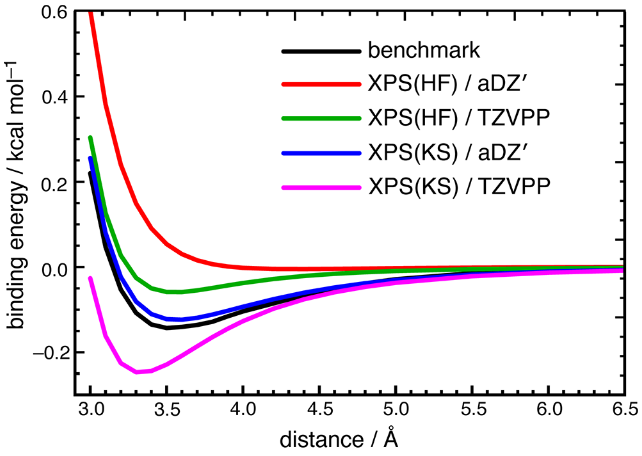


Figure 2. Binding energy curves for Ar⋯Ne. Benchmark results were computed using a dispersionless density functional (dlDF) augmented with a dispersion correction (the dlDF+D_{as}/aTZ method of ref 42).

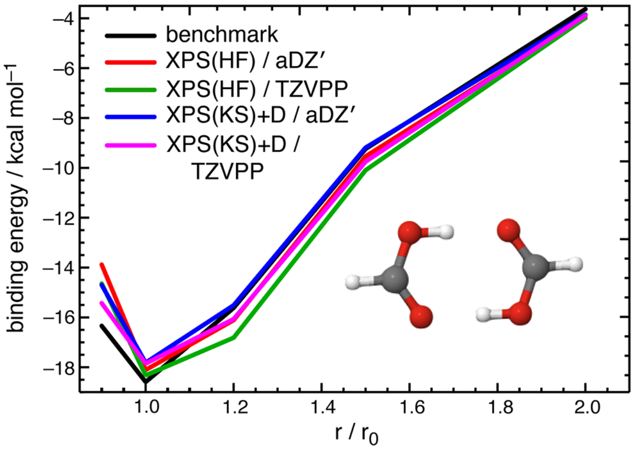


Figure 3. Binding energy curves for formic acid dimer. Benchmark CCSD(T)/CBS results are taken from ref 43. The horizontal axis is the scaled equilibrium distance between the centers of mass of the two monomers.

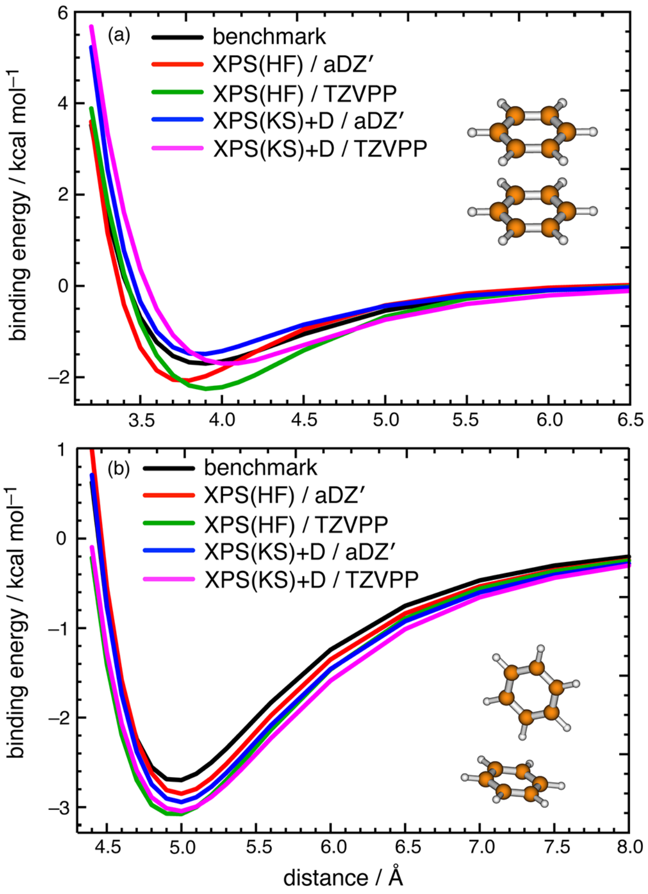


Figure 4. Binding energy curves for (a) the "sandwich" and (b) the "T-shaped" isomer of benzene dimer. The distance coordinate in either case is the center-to-center distance between the benzene rings. Benchmark CCSD(T)/CBS results are taken from ref 44.

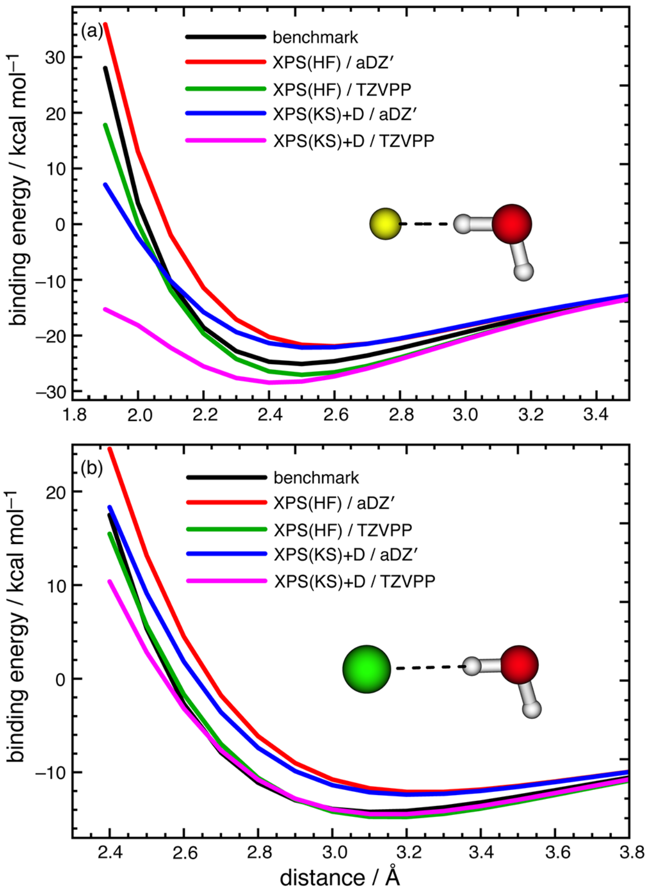


Figure 5. Binding energy curves for (a) F$^-$(H$_2$O) and (b) Cl$^-$(H$_2$O). The distance coordinate is the halide−oxygen distance, and the benchmarks are CCSD(T)/CBS.

Table 3. Error Statistics, with Respect to CCSD(T)/CBS Benchmarks, For a Set of Ionic H-bonded Dimers from Ref 45

	error/kcal mol^{-1}			% error	
method	MUE	RMSE	max	MUE	max
XPS(HF)/aDZ′	1.26	1.61	3.31	6.3	12.4
XPS(HF)/TZVPP	0.97	1.15	1.92	6.0	10.6
XPS(KS)/aDZ′	0.85	1.03	2.10	4.4	10.7
XPS(KS)/TZVPP	0.66	0.72	1.29	3.5	7.4
XPS(KS)+D/aDZ′	1.02	1.14	1.96	5.4	14.1
XPS(KS)+D/TZVPP	0.48	0.56	1.08	2.5	6.2
MP2/cc-pVTZ[a]		1.81			
BLYP+D3/QZVP[a]		0.59			

[a] Counterpoise-corrected results from ref 45.

level, whereas XPS(HF)/TZVPP binding energies are much more accurate, consistent with our previous results[10] suggesting that large basis sets are required for systems whose binding is dominated by electrostatics and induction.

To examine whether these larger errors are systemic to ions in general, we computed binding energies for a database[45] of 15 ionic H-bonded dimers, in which the ionic partner is either acetate, guanidinium, methylammonium, or imidazolium. Error statistics are listed in Table 3 and are somewhat larger than those seen for the neutral S22A and S66 data sets. Here the importance of monomer electron correlation and triple-ζ basis sets is clear: in the absence of either of those two factors, MUEs are $\gtrsim$1 kcal/mol and maximum errors are 2 to 3 kcal/mol. This is consistent with previous XPS results, indicating that the description of H-bonded and induction-bound systems generally improves in a systematic way with respect to basis-set quality and the treatment of electron correlation.[10] The XPS(KS)+D/TZVPP method delivers a root-mean-square error (RMSE) of 0.56 kcal/mol for the ionic H-bonded data set, which is significantly better than MP2/cc-pVTZ results and comparable to BLYP+D3/QZVP results but at significantly reduced cost.

Returning to F$^-$(H$_2$O), we note that the PECs computed by XPS(KS)+D methods (Figure 5a) are qualitatively *incorrect* for intermolecular distances smaller than the equilibrium distance. Several factors may account for this. First, no ionic systems were used in fitting the empirical potential $E_{\text{disp}}^{\text{SAPT+D}}$. Second, this potential actually diverges for $r_{ij} \ll R_i + R_j$,[22] and we note that the van der Waals radius R_i for the anion is larger than that of the corresponding neutral atom, which may explain the fact that PECs for the neutral systems are accurate even at relatively short distances. Lastly, the intermolecular interaction in F$^-$(H$_2$O) is known to have substantial covalent character resulting from a low-energy FH···OH$^-$ diabatic state,[46] and as such this system may be especially problematic for monomer-based quantum chemistry.

As a final test of XPS(KS)+D, we have computed binding energies for nucleobase tetramers arranged in average B-DNA geometries,[47] for which MP2, SCS-MP2, and DFT-SAPT results are available.[48] The MP2 method is known to overestimate binding energies of π-stacked complexes;[44] SCS-MP2 is sometimes used to reduce this systematic error.[49,50] However, SCS-MP2 *underestimates* the binding energy of the π-stacked uracil dimer[51] by 3 kcal/mol as compared with the CCSD(T)/CBS benchmark of 9.7 kcal/mol.[33] The SCS(MI)-MP2 method,[52] in which the two SCS scaling parameters are optimized (in a basis-set-specific way) using the S22 database, reduces this error to 0.4 kcal/mol,[51] comparable to the MUE for the S66 database obtained at the SCS(MI)-MP2/CBS level (see Table 2).[40] The XPS(KS)+D/aDZ′ method exhibits comparable error statistics for S66.

Results for nucleobase tetramers are listed in Table 4. CCSD(T)/CBS benchmarks are only available for base pairs, but in that case comparison to DFT-SAPT suggests that the DFT-SAPT/aTZ results in Table 4 are systematically too small.[2,54] For example, the DFT-SAPT/aTZ binding energy[2] for the π-stacked AT dimer is underestimated by 1.4 kcal/mol, as compared with the CCSD(T)/CBS benchmark (11.66 kcal/mol), if the so-called δ(HF) correction[1] is used in the SAPT calculation. If it is not, then the binding energy is underestimated by 3 kcal/mol. Hence, in perusing the DFT-SAPT/aTZ results in Table 4, one should anticipate that binding energies are underestimated by $\gtrsim$1 kcal/mol. At the same time, the XPS(KS)+D/aDZ′ method *overestimates* the π-stacked AT binding energy by 0.8 kcal/mol. Therefore, the correct binding energies for the nucleobase tetramers probably lie between the DFT-SAPT/aTZ and XPS(KS)+D/aDZ′ results in Table 4.

Table 4. Binding Energies for Nucleobase Tetramers

	binding energy/kcal mol^{-1}				
			DFT-SAPT[a]		
system[b]	MP2/ aTZ[c,d]	SCS-MP2/ aTZ[c,d]	aTZ[d]	scaled disp.[e]	XPS(KS)+D/ aDZ′
AT-AT	−16.45	−11.10	−11.39	−12.41	−13.50
AT-CG	−14.77	−9.33	−9.82	−10.89	−11.57
AT-GC	−14.91	−9.34	−10.03	−11.12	−11.30
AT-TA	−12.22	−6.62	−7.33	−8.42	−8.99
CG-AT	−15.56	−10.33	−11.32	−12.36	−12.78
CG-CG	−13.55	−7.99	−8.57	−9.65	−9.59
CG-GC	−17.51	−11.83	−13.14	−14.30	−14.70
GC-AT	−15.10	−9.83	−9.97	−10.97	−11.79
GC-CG	−16.54	−11.15	−10.92	−11.97	−13.02
TA-AT	−14.82	−9.96	−10.54	−11.48	−12.09

[a] Includes the δ(HF) correction. [b] Notation WX-YZ means that WX and YZ are Watson−Crick pairs, treated as monomers. [c] Counterpoise corrected. [d] Values from ref 48. [e] Includes empirical scaling of the dispersion energy, as recommended in ref 53.

This conclusion is bolstered by DFT-SAPT results in which the dispersion energy is scaled by an empirical factor of 1.051, suggested in ref 53 as a means of accounting for the slow basis-set convergence of this term. Scaled DFT-SAPT/aTZ binding energies lie between the unscaled DFT-SAPT/aTZ results and the XPS(KS)+D/aDZ′ results, except for CG-CG where the scaled DFT-SAPT and XPS(KS)+D results are essentially identical. Together, the nucleobase calculations presented here place plausible bounds on the tetramer binding energies and lend support to the accuracy of the scaled DFT-SAPT/aTZ approach.[53]

As compared with DFT-SAPT, however, XPS(KS)+D has important advantages. First, it avoids the supermolecular HF calculation that is required to compute the δ(HF) correction.[1] Second, the cost of XPS(KS)+D scales as $O(N^3)$ with respect to dimer size, whereas DFT-SAPT or SAPT(DFT) scales as $O(N^5)$.[11,13] Finally, XPS(KS)+D can be extended to any number of monomer units, at $O(n^3)$ cost, whereas DFT-SAPT is restricted to dimers only. This restriction means that the nucleobase tetramer calculations in Table 4 must be computed using base *pairs* as the monomer units, which significantly adds to the cost of DFT-SAPT.

Figure 6 shows timings for XPol, XPS(KS)+D, and supersystem DFT calculations on (adenine)$_n$ systems, arranged as in single-stranded B-DNA. For the XPS calculations, we plot both serial timings (representing the total CPU time required) as well as timings when run in "embarrassingly parallel" mode (representing wall time, assuming that $_nC_2$ processors are available). For these particular systems, the cost of any one dimer SAPT calculation (with or without the dispersion and exchange-dispersion terms) is negligible in comparison to the time required for the XPol procedure, which in these calculations is performed entirely on a single processor. As such, XPS(KS) and XPS(KS)+D timings are very similar. More importantly, the cost of such calculations scales as $O(n)$, with a very small prefactor. The wall time required for an XPS(KS)+D calculation on (adenine)$_{10}$ is only about four times larger than what is required for (adenine)$_2$!

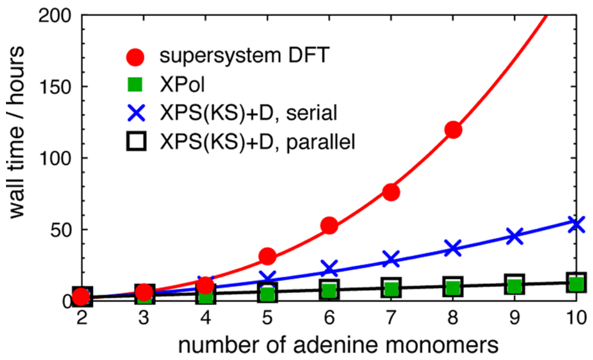


Figure 6. Timings for XPol, XPS(KS)+D, and supersystem DFT calculations for π-stacked (adenine)$_n$ systems. All calculations use the LRC-ωPBEh functional and the aDZ′ basis set.

In summary, the expensive and problematic sum-over-states dispersion terms in XPS(KS) have been replaced by empirical atom−atom potentials. Using a "tuned" LRC functional[26,28] and modest basis sets, the method exhibits an accuracy that is comparable to or better than MP2, SCS-MP2, SCS(MI)-MP2, and CCSD results extrapolated to the complete basis limit, and we have used this approach (in conjunction with existing DFT-SAPT calculations) to obtain the best estimates to date of the binding energies for nucleobase tetramers. The cost of XPS(KS)+D calculations grows cubically with respect to monomer size, just like traditional DFT, but no worse than cubically with respect to the number of monomer units. This makes XPS(KS)+D a promising method for use in fragment-based drug design, although the performance for ions suggests that extensions to larger basis sets and a more extensive parametrization of the dispersion potential may be in order. Such work is currently underway in our group.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Optimized range separation parameters for each monomer along with additional PECs for Ar···Ne. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

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

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

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

■ REFERENCES

(1) Hohenstein, E. G.; Sherrill, C. D. Wavefunction Methods for Noncovalent Interactions. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 304−326.

(2) Hesselmann, A.; Jansen, G.; Schütz, M. Interaction Energy Contributions of H-Bonded and Stacked Structures of the AT and GC Base Pairs from the Combined Density Functional Theory and Intermolecular Perturbation Theory Approach. *J. Am. Chem. Soc.* **2006**, *128*, 11730−11731.

(3) Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D. Base-Stacking and Base-Pairing Contributions to Thermal Stability of the DNA Double Helix. *Nucleic Acids Res.* **2006**, *34*, 564−574.

(4) 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.; Structure and Bonding *116*; Springer-Verlag: Berlin, 2005; pp 43−117.

(5) Szalewicz, K. Symmetry-Adapted Perturbation Theory of Intermolecular Forces. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 254−272.

(6) 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.

(7) Misquitta, A. J.; Szalewicz, K. Intermolecular Forces from Asymptotically Corrected Density Functional Description of Monomers. *Chem. Phys. Lett.* **2002**, *357*, 301−306.

(8) Heßelmann, A.; Jansen, G. First-Order Intermolecular Interaction Energies from Kohn-Sham Orbitals. *Chem. Phys. Lett.* **2002**, *351*, 464−470.

(9) 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.

(10) 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.

(11) 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.

(12) 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.

(13) Podeszwa, R.; Cencek, W.; Szalewicz, K. Efficient Calculations of Dispersion Energies for Nanoscale Systems from Coupled Density Response Functions. *J. Chem. Theory Comput.* **2012**, *8*, 1963−1969.

(14) Lotrich, V. F.; Szalewicz, K. Symmetry-Adapted Perturbation Theory of Three-Body Nonadditivity of Intermolecular Interaction Energy. *J. Chem. Phys.* **1997**, *106*, 9668−9687.

(15) Lotrich, V. F.; Szalewicz, K. Perturbation Theory of Three-Body Exchange Nonadditivity and Application to Helium Trimer. *J. Chem. Phys.* **2000**, *112*, 112−121.

(16) 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.

(17) 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.

(18) 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.

(19) Chen, Y.; Li, H. Intermolecular Interaction in Water Hexamer. *J. Phys. Chem. A* **2010**, *114*, 11719−11724.

(20) 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.

(21) 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.

(22) Hesselmann, A. Comparison of Intermolecular Interaction Energies from SAPT and DFT Including Empirical Dispersion Contributions. *J. Phys. Chem. A* **2011**, *115*, 11321−11330.

(23) 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.

(24) 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.

(25) Lange, A. W.; Rohrdanz, M. A.; Herbert, J. M. Charge-Transfer Excited States in a π-Stacked Adenine Dimer, as Predicted Using Long-Range-Corrected Time-Dependent Density Functional Theory. *J. Phys. Chem. B* **2008**, *112*, 6304−6308.

(26) 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.

(27) 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.

(28) Baer, R.; Livshits, E.; Salzner, U. Tuned Range-Separated Hybrids in Density Functional Theory. *Annu. Rev. Phys. Chem.* **2010**, *61*, 85−109.

(29) Grimme, S. Density Functional Theory with London Dispersion Corrections. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 211−228.

(30) Grimme, S. Accurate Description of van der Waals Complexes by Density Functional Theory Including Empirical Corrections. *J. Comput. Chem.* **2004**, *25*, 1463−1473.

(31) Grimme, S. Semiempirical GGA-Type Density Functional Constructed with Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787−1799.

(32) Adamovic, I.; Gordon, M. S. Dynamic Polarizability, Dispersion Coefficient C_6 and Dispersion Energy in the Effective Fragment Potential Method. *Mol. Phys.* **2005**, *103*, 379−387.

(33) Takatani, T.; Hohenstein, E. G.; Malagoli, M.; Marshall, M. S.; Sherrill, C. D. Basis Set Consistent Revision of the S22 Test Set of Noncovalent Interaction Energies. *J. Chem. Phys.* **2010**, *132* (144104), 1−5.

(34) 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.

(35) Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.

(36) Hohenstein, E. G.; Sherrill, C. D. Density Fitting and Cholesky Decomposition Approximations in Symmetry-Adapted Perturbation Theory: Implementation and Application to Probe the Nature of π-π Interactions in Linear Acenes. *J. Chem. Phys.* **2010**, *132* (184111), 1−10.

(37) Hohenstein, E. G.; Sherrill, C. D. Density Fitting of Intramonomer Correlation Effects in Aymmetry Adapted Perturbation Theory. *J. Chem. Phys.* **2010**, *133* (014101), 1−12.

(38) Shao, Y.; Fusti-Molnar, L.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; et al. Advances in Methods and Algorithms in a Modern Quantum Chemistry Program Package. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172−3191.

(39) 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.

(40) Řezác, 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.

(41) Flick, J. C.; Kosenkov, D.; Hohenstein, E. G.; Sherrill, C. D.; Slipchenko, L. V. Accurate Prediction of Non-Covalent Interaction Energies with the Effective Fragment PotentialMethod: Comparison of Energy Components to Symmetry-Adapted Perturbation Theory for the S22 TestSet. *J. Chem. Theory Comput.* **2012**, *8*, 2835−2843.

(42) Pernal, K.; Podeszwa, R.; Patkowski, K.; Szalewicz, K. Dispersionless Density Functional Theory. *Phys. Rev. Lett.* **2009**, *103* (263201), 1−4.

(43) Grafova, L.; Pitonak, M.; Řezác, J.; Hobza, P. Comparative Study of Selected Wave Function and Density Functional Methods for Noncovalent Interaction Energy Calculations Using the Extended S22 Data Set. *J. Chem. Theory Comput* **2010**, *6*, 2365−2376.

(44) 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 the Benzene Dimer, the Methane Dimer, Benzene−Methane, and Benzene−H_2S. *J. Phys. Chem. A* **2009**, *113*, 10146−10159.

(45) Řezác, J.; Hobza, P. Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *J. Chem. Theory Comput.* **2012**, *8*, 141−151.

(46) Herbert, J. M.; Head-Gordon, M. Charge Penetration and the Origin of Large O−H Vibrational Red-Shifts in Hydrated-Electron Clusters, $(H_2O)_n^-$. *J. Am. Chem. Soc.* **2006**, *128*, 13932−13939.

(47) Olson, W. K.; Bansal, M.; Burley, S. K.; Dickerson, R. E.; Gerstein, M.; Harvey, S. C.; Heinemann, U.; Lu, X. J.; Neidle, S.; Shakked, Z.; et al. A Standard Reference Frame for the Description of Nucleic Acid Base-Pair Geometry. *J. Mol. Biol.* **2001**, *313*, 229−237.

(48) 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.

(49) Grimme, S. Improved Second-Order Møller−Plesset Perturbation Theory by Separate Scaling of Parallel- and Antiparallel-Spin Pair Correlation Energies. *J. Chem. Phys.* **2003**, *118*, 9095−9102.

(50) Antony, J.; Grimme, S. Is Spin-Component Scaled Second-Order Møller-Plesset Perturbation Theory an Appropriate Method for the Study of Noncovalent Interactions in Molecules? *J. Phys. Chem. A* **2007**, *111*, 4862−4868.

(51) Pitonak, M.; Riley, K. E.; Neogrady, P.; Hobza, P. Highly Accurate CCSD(T) and DFT−SAPT Stabilization Energies of H-Bonded and Stacked Structures of the Uracil Dimer. *ChemPhysChem* **2008**, *9*, 1636−1644.

(52) DiStasio, R. A., Jr.; Head-Gordon, M. Optimized Spin-Component Scaled Second-Order Møller-Plesset Perturbation Theory for Intermolecular Interaction Energies. *Mol. Phys.* **2007**, *105*, 1073−1083.

(53) Řezác, J.; Hobza, P. Extrapolation and Scaling of the DFT-SAPT Interaction Energies Toward the Basis Set Limit. *J. Chem. Theory Comput.* **2011**, *7*, 685−689.

(54) Sedlak, R.; Jurečka, P.; Hobza, P. Density Functional Theory-Symmetry Adapted Perturbation Treatment Energy Decomposition of Nucleic Acid Base Pairs Taken from DNA Crystal Geometry. *J. Chem. Phys.* **2007**, *127* (075104), 1−3.

(55) Xantheas, S. S. On the Importance of the Fragment Relaxation Energy Terms in the Estimationof the Basis Set Superposition Error Correction to the Intermolecular Interaction Energy. *J. Chem. Phys.* **1996**, *104*, 8821−8824.

(56) Xantheas, S. S.; Burnham, C. J.; Harrison., R. J. Development of Transferable Interaction Models for Water. II. Accurate Energetics of the First Few Water Clusters from First Principles. *J. Chem. Phys.* **2002**, *116*, 1493−1499.

(57) Xantheas, S. S.; Aprà, E. The Binding Energies of the D_2d and S_4 Water Octamer Isomers: High-Level Electronic Structure and Empirical Potential Results. *J. Chem. Phys.* **2004**, *120*, 823−828.

(58) Fanourgakis, G. S.; Aprà, E.; Xantheas, S. S. High-Level *ab Initio* Calculations for the Four Low-Lying Families of Minima of $(H_2O)_{20}$. I. Estimates of MP2/CBS Binding Energies and Comparison with Empirical Potentials. *J. Chem. Phys.* **2004**, *121*, 2655−2663.

(59) Bulusu, S.; Yoo, S.; Aprà, E.; Xantheas, S. S.; Zeng, X. C. Lowest-Energy Structures of Water Clusters $(H_2O)_{11}$ and $(H_2O)_{13}$. *J. Phys. Chem. A* **2006**, *110*, 11781−11784.