Comprehensive Basis-Set Testing of Extended Symmetry-Adapted Perturbation Theory and Assessment of Mixed-Basis Combinations to Reduce Cost

Montgomery Gray and John M. Herbert*
Department of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210 USA
(Dated: December 25, 2021)

Hybrid or "extended" symmetry-adapted perturbation theory (XSAPT) replaces traditional SAPT's treatment of dispersion with better-performing alternatives, while at the same time extending two-body (dimer) SAPT to a many-body treatment of polarization using a self-consistent charge-embedding procedure. The present work presents a systematic study of how XSAPT interaction energies and energy components converge with respect to the choice of Gaussian basis set. Although errors can be reduced in a systematic way using correlation-consistent basis sets, similar performance at lower cost is obtained using Karlsruhe basis sets, and we introduce new versions with limited augmentation (diffuse functions) that are even more efficient. Pople-style basis sets, which are even more efficient, often afford good results if a large number of polarization functions are included. The dispersion models used in XSAPT afford much faster basis-set convergence as compared to the perturbative description of dispersion in conventional SAPT, meaning that "compromise" basis sets (such as jun-cc-pVDZ) are no longer required and benchmark-quality results can be obtained using basis sets of triple-ζ quality. The use of diffuse functions proves to be essential, especially for the description of hydrogen bonds. The "δ(Hartree-Fock)" correction that accounts for high-order induction can be performed in double-ζ basis sets without significant loss of accuracy, leading to a mixed-basis approach that offers 4× speedup over the existing (cubic-scaling) XSAPT approach.

1 Introduction

Noncovalent interactions are ubiquitous in nature and drive important chemical processes including crystal packing,[1–5] protein folding,[6–10] and host–guest binding in pharmaceuticals,[11–14] and are important to soft materials of interest in materials science applications.[15–17] Despite their prevalence, noncovalent forces are often misunderstood by chemists.[18–20] The framework of symmetry-adapted perturbation theory (SAPT)[20–26] offers an accurate and systematic *ab initio* approach to noncovalent interaction energies, including an energy decomposition into physically-meaningful components: electrostatics, Pauli repulsion, induction, and dispersion.[27–29] This decomposition is useful for obtaining physical insight that is backed by reliable calculations.[11,19,20,30–36] The SAPT energy decomposition is inherently better separable than methods based on supramolecular density functional theory (DFT)[37–39] and can be used to develop physically meaningful force fields.[1,14,40,41]

The most widely used variant of SAPT, known as SAPT0,[25,42] combines Hartree-Fock (HF) wave functions for the isolated monomers (as zeroth-order states) with second-order perturbation theory for the intermolecular Coulomb potentials.[22,25] When used in conjunction with Kohn-Sham (KS) density functionals with correct asymptotic behavior, intramolecular electron correlation can be incorporated by substituting the KS determinant for the HF one, in a method that we have called SAPT0(KS).[20,42] The use of asymptotically-correct functionals is crucial and the SAPT0 formalism should not be used with arbitrary density functionals.[42–44] With that caveat, both traditional SAPT0 and SAPT0(KS) afford semiquantitative results at $\mathcal{O}(N^5)$ cost.[25,42] The accuracy of both methods is limited by the accuracy of the dispersion interactions, which are not quantitative within the "uncoupled" second-order approximation that characterizes SAPT0,[45–47] and which is similar to second-order Møller-Plesset perturbation theory (MP2).[48,49]

To obtain quantitative accuracy, second-order dispersion must be replaced by better-performing alternatives. This has led to development of MP2 variants including MP2C[46] and MP2D,[47] as well as DFT-SAPT.[23,24] These methods are much more accurate than SAPT0 but retain that method's $\mathcal{O}(N^5)$ cost. For MP2C and DFT-SAPT, a density fitting approximation is required in order to obtain fifth-order scaling.[50,51]

The "extended" (X)SAPT approach follows a somewhat similar strategy,[20] replacing second-order dispersion with either *ab initio* dispersion potentials (XSAPT + *ai*D)[48,52–54] or else with a many-body dispersion model (XSAPT + MBD).[55,56] The latter approach is currently the best-performing variant of XSAPT,[20,55] and is the one used herein. Unlike any of the aforementioned methods, XSAPT incurs only $\mathcal{O}(N^3)$ cost, and it can furthermore be extended to clusters of molecules using a self-consistent charge embedding to capture nonadditive polarization effects.[54,56–59] For large systems, the monomer-based nature of XSAPT calculations makes this approach more affordable even than supramolecular DFT.[52,54,55]

Some limited basis-set testing of XSAPT methods has been reported in previous work,[42,54,55,60] but only for total interaction energies. Herein, we also examine convergence of individual energy components, which allows us to consider whether earlier tests may have benefited

*herbert@chemistry.ohio-state.edu

from error cancellation amongst energy components that might exhibit different convergence behavior. The basis-set dependence of traditional SAPT has also been carefully evaluated,[25] but the especially slow convergence of the dispersion energy in the traditional approach means that those tests are not directly applicable to XSAPT. Moreover, because XSAPT was designed for large systems, we want to test Pople and Karlsruhe basis sets that are commonly used for large systems, whereas small-molecule SAPT calculations have generally been limited to Dunning correlation-consistent basis sets. These tests allow us to consider whether hybrid basis-set combinations (in which different energy components are evaluated in different basis sets) might further improve the efficacy of XSAPT calculations. Such hybrid approaches take advantage of the inherent separability of the (X)SAPT interaction energy

2 Theory

The XSAPT formalism[48,57,58] and the XSAPT + MBD method[55,56] have been described in previous work, including a recent review.[20] For completeness, these methods are briefly summarized below.

2.1. SAPT0(KS). The starting point for XSAPT is second-order SAPT0(KS) for a dimer, for which the total interaction energy ($E_{\rm int}$) is expressed as[48,54]

$$E_{\rm int}^{\rm SAPT0} = E_{\rm elst}^{(1)} + E_{\rm exch}^{(1)} + E_{\rm ind}^{(2)} + E_{\rm exch\text{-}ind}^{(2)} + E_{\rm disp}^{(2)} + E_{\rm exch\text{-}disp}^{(2)} + \delta E_{\rm HF} \tag{1}$$

Superscripts indicate orders in intermolecular perturbation theory but we drop these henceforth, setting $E_{\rm elst} \equiv E_{\rm elst}^{(1)}$ (electrostatics) and $E_{\rm exch} \equiv E_{\rm exch}^{(1)}$ (exchange or Pauli repulsion). The total induction and dispersion energies are defined, respectively, as

$$E_{\rm ind} = E_{\rm ind}^{(2)} + E_{\rm ind\text{-}disp}^{(2)} + \delta E_{\rm HF} \tag{2}$$

and

$$E_{\rm disp} = E_{\rm disp}^{(2)} + E_{\rm exch\text{-}disp}^{(2)} \,. \tag{3}$$

The final term in eq. 1, which appears also in eq. 2, is a so-called "δHF" correction for higher-order induction effects,[20,22] defined as

$$\delta E_{\rm HF} = \Delta E_{\rm int}^{\rm HF} - \left(E_{\rm elst}^{(1)} + E_{\rm exch}^{(1)} + E_{\rm ind,resp}^{(2)} + E_{\rm exch\text{-}ind,resp}^{(2)}\right) . \tag{4}$$

Here, $\Delta E_{\rm int}^{\rm HF}$ is the counterpoise-corrected HF interaction energy. The two second-order quantities in eq. 4 are response ("resp") analogues of the second-order terms in eq. 2, and these require the solution of coupled-perturbed HF equations.[61] For SAPT0(KS) calculations, the first- and second-order SAPT terms in eq. 4 should be computed at the HF level even if the corresponding terms in eq. 1 are computed based on KS orbitals.[42] Notably, $\delta E_{\rm HF}$ is the only term in eq. 1 that requires self-consistent field (SCF) iterations in a dimer basis set, which becomes a bottleneck for large monomers.

The accuracy of SAPT0(KS) interaction energies depends critically on the use of asymptotically-correct exchange-correlation functionals.[20,42,43] Long-range corrected (LRC) functionals[62–65] offer a simple means to enforce this constraint but the range separation parameter must be tuned separately for each monomer, in order to obtain correct asymptotics.[42–44] To this end, we use the LRC-ωPBE functional[64] combined with the "global density-dependent" (GDD) tuning procedure.[42,66] As compared to tuning based on the ionization energy criterion,[43] GDD-tuned SAPT0(KS) and XSAPT + MBD results are essentially identical.[42]

2.2. XSAPT. The XSAPT approach uses a self-consistent charge embedding procedure based on the variational "explicit polarization" (XPol) formalism.[56–58,67] In this approach, SCF wave functions for the monomers are computed in the presence of wave function-derived atomic point charges. For this purpose we use Charge Model 5 (CM5),[68] whose efficient implementation for XSAPT calculations is described in Ref. 56. CM5 charges are based on the Hirshfeld atomic charge model procedure but introduce parameters in an effort to obtain more accurate dipole moments.

The XSAPT + MBD approach method starts from the charge-embedded XPol version of SAPT0(KS) but replaces the second-order dispersion energy (eq. 3) with a variant of the range-separated and self-consistently screened MBD (or "MBD@rsSCS") model developed by Tkatchenko and co-workers,[69,70] although the original approach must be modified at short range for use with SAPT.[55,56] For a dimer system, this completes the specification of the XSAPT + MBD method.

For a system composed of more than two monomers, the XSAPT interaction energy is[20]

$$E_{\rm int}^{\rm XSAPT} = \sum_{A,B>A} \left(E_{\rm elst}^{AB} + E_{\rm exch}^{AB} + E_{\rm disp}^{AB} + E_{\rm ind}^{AB} + \delta E_{\rm HF}^{AB}\right) + E_{\rm pol}^{\rm PW} + E_{\rm pol}^{\rm MB} \tag{5}$$

where the summand in parentheses is the SAPT0(KS) interaction energy for dimer AB, meaning eq. 1 (with MBD replacing second-order dispersion) but without charge embedding. The final two terms in eq. 5, $E_{\rm pol}^{\rm PW} + E_{\rm pol}^{\rm MB}$, are the pairwise and many-body polarization energies. The pairwise polarization energy is defined as

$$E_{\rm pol}^{\rm PW} = \sum_{A,B>A} \left[E_{AB}^{\rm XSAPT}(AB) - E_{AB}^{\rm SAPT}\right] \tag{6}$$

where the term in square brackets is the difference between the charge-embedded energy for dimer AB and the SAPT energy computed without charge embedding. The many-body polarization energy is

$$E_{\rm pol}^{\rm MB} = \sum_{A,B>A} \left[E_{AB}^{\rm XSAPT}(AB\cdots N) - E_{AB}^{\rm XSAPT}(AB) \right] \quad (7)$$

where $E_{AB}^{\rm XSAPT}(AB\cdots N)$ is the energy of dimer AB embedded in an environment of atomic charges corresponding to the entire supersystem $AB\cdots N$, whereas $E_{AB}^{\rm XSAPT}(AB)$ is the same dimer's energy when embedding charges are included only on monomers A and B. Thus $E_{AB}^{\rm XSAPT}(AB)$ is no different from a dimer XSAPT calculation performed on AB in isolation from the rest of the supersystem.

3 Computational Details

3.1. XSAPT Methods. All calculations were performed using Q-Chem v. 5.[71] Integral and shell-pair thresholds were set to 10^{-12} a.u., the SCF convergence threshold was set to 10^{-7} Ha, and the SAPT and XSAPT calculations use the "projected" basis set.[57,58] The latter is an alternative to the dimer basis set that is traditionally used to compute monomer wave functions for SAPT,[72] but in clusters with more than two monomers this would entail a pairwise SCF cost. In the projected-basis approach, each monomer's SCF wave function is computed in the monomer basis but then a pseudocanonicalized dimer basis is used for the subsequent pairwise SAPT calculations.[57,58]

Most calculations reported here are performed at the XSAPT + MBD level of theory[55] but some tests will be reported using XSAPT + aiD3,[48,54] where "aiD3" denotes the third-generation *ab initio* dispersion potential developed by Lao and Herbert.[54] The performance of XSAPT + aiD3 is marginally better than that of the second-generation version (XSAPT + aiD2),[53] with slightly better coverage of π-stacking in the training set used to parameterize aiD3. These methods outperform the first-generation version,[52] which we do not recommend.[53,54]

3.2. Density Functionals. For SAPT0(KS) and XSAPT calculations, the range separation parameter in the LRC-ωPBE functional must be tuned individually for each monomer.[42–44] We do this at the LRC-ωPBE/def2-TZVP level, then use that value in all subsequent calculations regardless of basis set. (The GDD-tuned values of ω, for each of the systems considered in this work, can be found in Tables S1–S7 of the Supporting information.) Previous work has shown that tuned values of ω are sensitive to the fraction of short-range exact exchange but for a given density functional, these values are rather *insensitive* to the choice of basis set.[73] The SG-1 quadrature grid[74] is used for LRC-ωPBE.

3.3. Basis Sets. A standard complement of Pople-style basis sets is tested, ranging in quality up to 6-311++G(3df,2pd).[75] In modern electronic structure theory these are often regarded as low-quality basis sets yet they are still widely used, in part because certain quantum chemistry programs (including Q-Chem) have been optimized to take advantage the use of *sp* functions in Pople basis sets (*i.e.*, s and p functions with a common exponent). If so, then Pople basis sets are considerably more efficient per unit basis function as compared to other alternatives. Karlsruhe (Ahlrichs) "def2" basis sets are also tested,[76,77] up to quadruple-ζ quality, as are the correlation-consistent basis sets cc-pVXZ[78] and aug-cc-pVXZ,[79] for X = D, T, and Q.

In addition to these well-established families of basis sets, we also tested "calendar" versions of the correlation-consistent basis sets,[80] in which diffuse functions are systematically removed starting from aug-cc-pVXZ. (These basis sets were added to Q-Chem as part of the present work.) The jul-cc-pVXZ basis sets (for X = D, T, or Q) consist of cc-pVXZ for hydrogen and aug-cc-pVXZ for other atoms, meaning that there are no diffuse functions on hydrogen. The jun-cc-pVXZ basis sets additionally remove the diffuse functions with highest angular moment from each non-hydrogen atom. We note that jun-cc-pVDZ has been suggested as a compromise basis set for SAPT0 calculations,[25,42] exploiting the slow basis-set convergence of the dispersion energy to limit the intrinsic overestimation of dispersion by second-order perturbation theory. This is a compromise because electrostatics is not generally converged at the double-ζ level,[42,54] but overall errors in SAPT0 interaction energies are worse in aug-cc-pVDZ and larger basis sets, as compared to jun-cc-pVDZ,[25] suggesting that it is more important to control the error in second-order dispersion rather than to converge the electrostatics.

We also modified the Karlsruhe basis sets to delete diffuse functions in an analogous manner. As compared to the correlation-consistent basis sets, the Karlsruhe basis sets contain fewer diffuse functions; for example, def2-SVPD contains a diffuse s function and a set of diffuse d functions for second-row atoms, but no diffuse p function, and for hydrogen there is a set of diffuse p functions but no diffuse s function.[77] As a first truncation, and in analogy to to jul-cc-pVXZ, we delete all of the diffuse functions on hydrogen, forming what is traditionally called a "heavy-augmented" ("ha") basis set. The basis set consisting of def2-SVP for hydrogen and def2-SVPD for other atoms will therefore be denoted def2-ha-SVP. As a second step, and in analogy to jun-cc-pVXZ, we delete the highest angular momentum set of diffuse functions on each non-hydrogen atom. For second-row atoms, this leaves only minimal augmentation ("ma") with a diffuse s function, so we refer to these basis sets as "def2-ma-", *e.g.*, def2-ma-SVP. This paradigm is sim-

ilar in spirit to partially-augmented Karlsruhe basis sets tested previously by Zheng *et al.*,[81] although the diffuse exponents differ from that work because Ref. 81 did not start from the standardized Karlsruhe diffuse exponents that were introduced in Ref. 77. The latter serve as the starting point for our def2-ha and def2-ma basis sets. Exponents and contraction coefficients for these new basis sets are provided in the Supporting Information.

3.4. Data Sets. We will use the S66 data set of noncovalent dimers[82] for high-throughput evaluation of a wide variety of basis sets. This database was developed to sample various types of noncovalent interactions and benchmark interaction energies were reported in Ref. 82 at the level of coupled-cluster theory with single, double, and perturbative triple excitations [CCSD(T)], extrapolated to the complete basis set (CBS) limit. Per the analysis in Ref. 82, the 66 dimers in this test set are divided into three subsets: hydrogen-bonded dimers, which are characterized by the condition $|E_{\rm elst}| \geq 2|E_{\rm disp}|$; dispersion-dominated dimers, for which $|E_{\rm disp}| \geq 2|E_{\rm elst}|$; and dimers of mixed-influence interactions, where neither of these conditions is met. It is useful to group the complexes in this way because the hydrogen-bonded subset (containing 23 dimers involving water, methanol, acetic acid, and other polar monomers) places stringent demands on the basis set as compared to the other S66 systems. The dispersion-bound subset contains 23 dimers involving monomers such as benzene, pyridine, and ethylene.

We will also consider three data sets containing ions:[83] AHB21, which consists of 21 anion–neutral hydrogen-bonded complexes with ions including F^-, Cl^-, N_3^-, and SH^-; CHB6, consisting of cation–molecule complexes of Na^+, Li^+, and K^+ with water and benzene; and finally IL16,[83] which is a set of 16 ion pairs representing constituent molecules or constituent moieties of ionic liquids.[84] Benchmark interaction energies at the CCSD(T)/CBS level are taken from Ref. 83, for all three of these ion-containing data sets.

4 Results and Discussion

4.1. Tests of Traditional SAPT0 Using S66. The basis-set convergence of traditional SAPT methods, including SAPT0 but also higher-order methods, has been reported previously for several small-molecule data sets.[25] Nevertheless we include our own SAPT0 convergence tests here, because they establish a baseline to which we can later compare the XSAPT methods, whose alternative descriptions of dispersion lead to accelerated basis-set convergence. Table 1 provides the mean absolute errors (MAEs) for the S66 database at SAPT0/cal-cc-pVXZ levels of theory, where X = D, T, or Q and cal = jun, jul, or aug. As previously reported,[25] the jun-cc-pVDZ basis set affords the smallest errors, with a MAE of 0.5 kcal/mol and a maximum error of 1.6 kcal/mol. We are unable to complete the SAPT0/aug-cc-pVQZ calculations due to memory limitations on our hardware, but previous results for other smaller dimers suggest that MAEs at the SAPT0/aug-cc-pVQZ level are only slightly larger than those at the SAPT0/aug-cc-pVTZ level, by $\lesssim$ 0.2 kcal/mol.[25]

Table 1: Error Statistics for SAPT0 Applied to the S66 Data Set.

Basis Set	Error (kcal/mol)	
	MAE	Max
jun-cc-pVDZ	0.51	1.55
jul-cc-pVDZ	0.64	2.34
aug-cc-pVDZ	0.67	2.49
jun-cc-pVTZ	0.81	3.34
jul-cc-pVTZ	0.93	3.56
aug-cc-pVTZ	1.01	3.70
jun-cc-pVQZ	1.05	3.91
jul-cc-pVQZ	1.09	3.98

SAPT0 errors increase both with increasing cardinality of the basis set (double-, triple-, or quadruple-ζ) and with increasing augmentation. The reason for this behavior is that the MP2-like second-order approximation that is used in SAPT0 tends to overestimate dispersion significantly, yet dispersion in particular converges very slowly to the CBS limit. More complete basis sets therefore afford increasingly poor dispersion energies and the use of jun-cc-pVDZ represents something of a "Pauling point",[85] balancing slow convergence against overestimation of the result. It is a remarkably robust compromise in small molecules,[25,42] although it may fare worse in larger systems.[48] For example, in the L7 set of large dispersion-bound complexes,[86] the MAE for SAPT0/jun-cc-pVDZ interaction energies is 4.8 kcal/mol, and the maximum error is 10.3 kcal/mol.[48]

SAPT0 dispersion energies for all of the S66 dimers are plotted in Fig. 1 using the jun-cc-pVXZ basis sets for X = D, T, and Q. These data make it clear that the triple-ζ dispersion energies are systematically much larger than double-ζ values, and quadruple-ζ dispersion energies are a bit larger still. As will become clear from the XSAPT data that are presented below, energy components other than dispersion are essentially converged at the triple-ζ level.

4.2. Survey of Basis Sets for XSAPT + MBD Using S66. We next consider the performance of the hybrid XSAPT + MBD method as applied to the S66 dimers. These systems are small, with the largest being pentane dimer, and thereby facilitate high-throughput testing. XSAPT + MBD errors for each of the S66 dimers are plotted in Fig. 2 across a wide range of Pople, Karls-

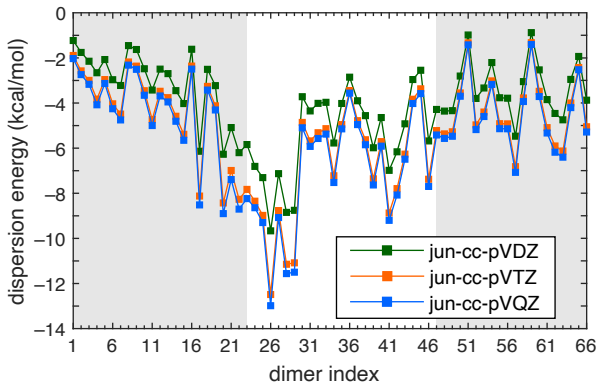

Fig. 1: Dispersion energies for the S66 dimers computed at the SAPT0/jun-cc-pVXZ level. Shaded regions delineate the three subsets of S66: hydrogen-bonded dimers (1–23), dispersion-dominated complexes (24–46), and dimers with mixed-influence interactions (47–66).

ruhe, and Dunning basis sets, and these various basis set families are analyzed separately in the subsections that follow. A statistical survey of the results is presented in Table 2, broken down into the three subsets that were described in Section 3.4, namely, hydrogen-bonded complexes, dispersion-bound complexes, and dimers with mixed-influence interactions. As with the SAPT0 assessment in Section 4.1, error is measured relative to CCSD(T)/CBS benchmarks.[82]

4.2.1. Karlsruhe Basis Sets. Errors for XSAPT + MBD using the Karlsruhe "def2" basis sets are plotted in Fig. 2a, where the double-, triple-, and quadruple-ζ basis sets have been grouped together by color. Clearly, the double-ζ errors are much larger, exceeding 1 kcal/mol in many cases and typically 1–3 kcal/mol larger than what is obtained in more complete basis sets.

These data also clearly demonstrate the importance of including diffuse functions when calculating interaction energies. This can be seen most clearly from the double-ζ data but is true in triple- and quadruple-ζ basis sets as well, although the magnitude of the effect diminishes with the cardinality of the basis set. For the double-ζ basis sets, errors increase in a consistent way as the diffuse orbitals are trimmed, going from def2-SVPD (with a full complement of diffuse functions) to the "heavy-augmented" def2-ha-SVP basis set, to the "minimally-augmented" def2-ma-SVP basis, and finally to def2-SVP that contains no diffuse functions at all. The importance of diffuse basis functions is most significant in the hydrogen-bonded subset, where induction effects are important and where the aforementioned truncations of def2-SVPD increase the MAE from 1.2 kcal/mol (def2-SVPD) to 3.7 kcal/mol (def2-SVP), with an increase in the maximum error from 3.8 kcal/mol (def2-SVPD) to 8.6 kcal/mol (def2-SVP). Removal of the diffuse functions has a much smaller effect in systems that are dominated by dispersion.

Removal of the diffuse functions also has a smaller effect at the triple- and quadruple-ζ levels. Whereas in the double-ζ case the use of diffuse functions is are absolutely essential in order to obtain even semi-quantitative results, at the triple- or quadruple-ζ level it appears that minimal augmentation is sufficient. We conclude that diffuse functions added to double-ζ basis sets are partially compensating for the overall incompleteness of the monomer basis set, which becomes much less of an issue at the triple-ζ level. We will carefully document the importance of diffuse functions throughout this work, because in our experience many quantum chemistry users are extremely reluctant to include these functions, presumably for reasons of cost.

Examining the error statistics for the Karlsruhe basis sets in Table 2, it appears that absolute errors converge at the triple-ζ level. Difference between triple- and quadruple-ζ interaction energies are uniformly smaller than 1 kcal/mol and on average these differences are no more than 0.2–0.3 kcal/mol. The largest differences (approaching 1 kcal/mol) are for systems with very strong hydrogen bonds, such as acetic acid dimer, and in those cases the quadruple-ζ errors are actually larger than the triple-ζ errors, albeit by a very small amount.

4.2.2. Dunning Basis Sets. Errors for correlation-consistent basis sets and "calendar" variants thereof are provided in Fig. 2b, with statistics compiled in Table 2. Again we find that some augmentation with diffuse functions is necessary to obtain high-quality results for hydrogen-bonded systems, despite the fact that all of the monomers in S66 are charge-neutral. In the absence of any diffuse functions at all, even the cc-pVTZ basis set affords unacceptably large errors (up to 5.7 kcal/mol), specifically for the hydrogen-bonded complexes. These errors are substantially reduced, even at the double-ζ level, by minimal augmentation, *e.g.*, jun-cc-pVDZ or jun-cc-pVTZ, although some of the errors for hydrogen-bonded complexes remain larger than 1 kcal/mol even at the aug-cc-pVTZ level, where the average error is 0.5 kcal/mol but there are outliers up to 2.6 kcal/mol for hydrogen-bonded complexes. In terms of the diffuse functions that are required, the use of jun-cc-pVXZ (for X = D, T, or Q) affords very similar results to aug-cc-pVXZ, indicating that the full complement of diffuse functions is not required even for hydrogen-bonded complexes. This is an extremely useful observation because jun-cc-pVXZ is more scalable to large systems as compared to aug-cc-pVXZ, as the difference (for second-row atoms) is a set of diffuse f functions.

The takeaway from this analysis is that the accuracy obtained from these very large basis sets is comparable to what is possible using the somewhat more modest Karlsruhe basis sets. Unlike the case of SAPT0, once the slow convergence of the dispersion energy is taken out

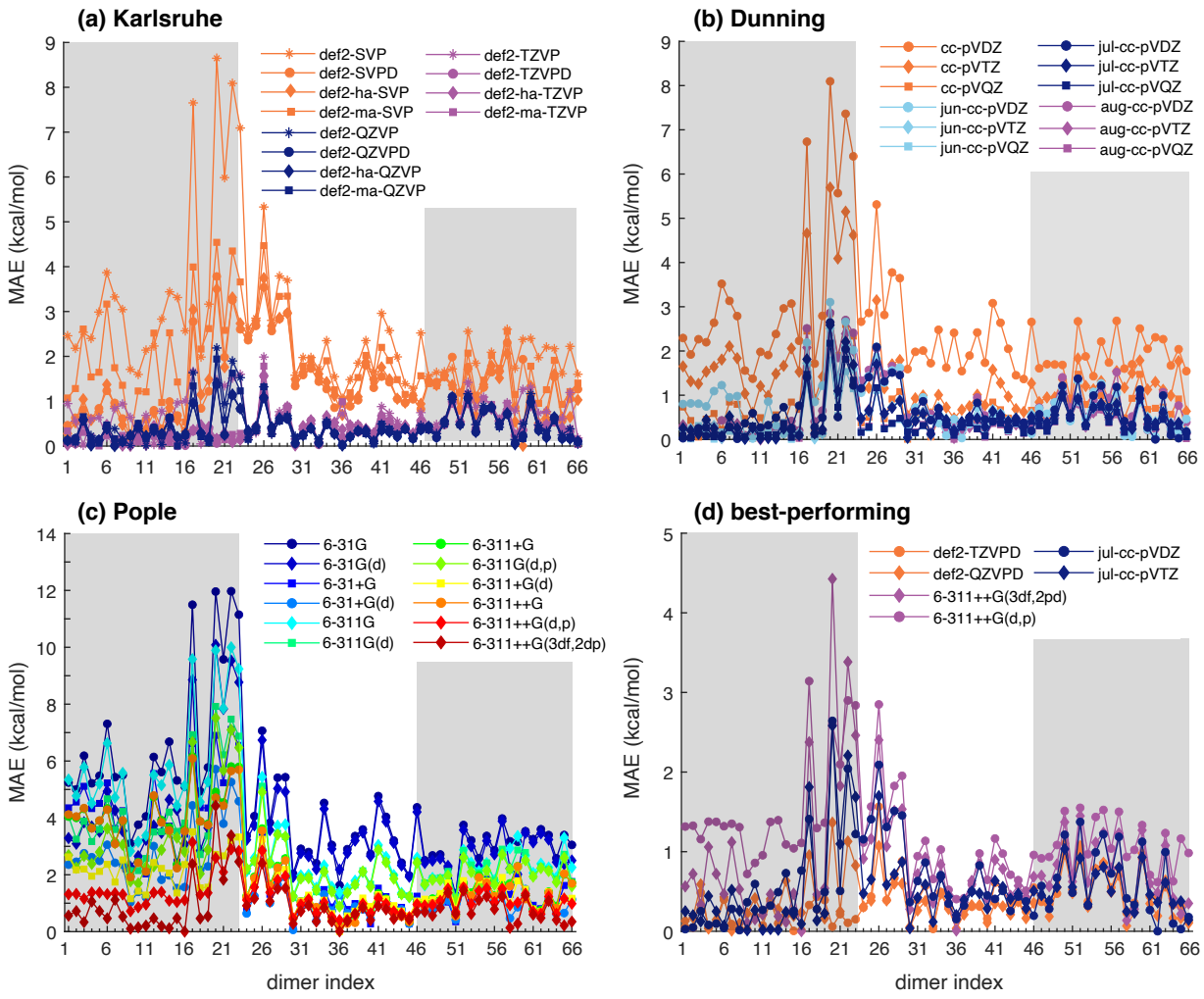

Fig. 2: Mean absolute errors (MAEs) in XSAPT + MBD total interaction energies for the S66 dimers, using (a) Karlsruhe basis sets, (b) Dunning correlation-consistent basis sets, (c) Pople basis sets, and (d) the best-performing basis sets from amongst these three categories. Indices along the horizontal axis refer to the ordering of the S66 dimers in Ref. 82 and the three regions delineated by shading indicate the hydrogen-bonded subset (dimers 1–23), the dispersion-dominated subset (24–46), and the subset of mixed-influence dimers (47–66). Errors are defined with respect to the CCSD(T)/CBS benchmarks from Ref. 82.

of the picture it is no longer the case that jun-cc-pVDZ affords the best results, and the accuracy is improved substantially in triple-ζ basis sets. The data provide no compelling reason to to push to quadruple-ζ basis sets, however.

4.2.3. Pople Basis Sets. XSAPT + MBD absolute errors using Pople basis sets are shown in Fig. 2c, color-coded across the visible spectrum based on the size of the basis set, with warmer colors (towards red) indicating larger basis sets. Errors are generally larger as compared to Karlsruhe or Dunning basis sets, although 6-311++G(d,p) and 6-311++G(3df,2dp) perform reasonable well, with MAEs of 1.2 and 0.8 kcal/mol, respectively, although the maximum errors (3.1 and 4.4 kcal/mol, respectively) are larger than those obtained using triple-ζ basis sets of the Karlsruhe or Dunning variety. Diffuse functions on hydrogen make very little difference, even for hydrogen-bonded systems, and 6-311+G(3df,2dp) affords results that are nearly indistinguishable from 6-311++G(3df,2dp).

These "(3df,2dp)" basis sets were originally developed for MP2 calculations[75] and provide relatively good performance for the hydrogen-bonded subset of S66, where large induction effects amplify the importance of polarization functions. For the other S66 dimers, however, the performance of these basis sets is nearly identical to that of several other Pople basis sets, notably 6-31+G(d). A set of diffuse functions is necessary even for the dispersion-bound complexes but otherwise the double-ζ basis sets perform just as well as the triple-ζ ones, for the dispersion-bound and mixed-influence dimers. Polarization functions matter little for dispersion-bound systems

Table 2: Error Statistics for XSAPT + MBD Applied to the S66 Data Set.

Method	Error (kcal/mol)							
	H-Bonded		Dispersion		Mixed		Total	
	MAE	Max	MAE	Max	MAE	Max	MAE	Max
def2-SVP	3.71	8.65	2.32	5.33	1.89	2.60	2.67	8.65
def2-ma-SVP	1.86	4.54	1.98	4.47	1.41	2.56	1.86	4.54
def2-ha-SVP	1.14	3.50	1.75	3.74	1.14	2.06	1.35	3.74
def2-SVPD	1.17	3.78	1.65	3.54	1.33	2.30	1.39	3.78
def2-TZVP	0.75	1.67	0.55	1.99	0.82	1.42	0.70	1.99
def2-ma-TZVP	0.19	0.36	0.49	1.78	0.60	1.17	0.42	1.78
def2-ha-TZVP	0.21	0.57	0.50	1.56	0.55	0.99	0.41	1.57
def2-TZVPD	0.17	0.48	0.43	1.57	0.54	0.99	0.37	1.57
def2-QZVP	0.53	2.19	0.36	1.32	0.55	1.17	0.48	2.19
def2-ma-QZVP	0.49	1.94	0.35	1.32	0.49	1.07	0.44	1.40
def2-ha-QZVP	0.40	1.40	0.35	1.08	0.49	1.09	0.41	1.40
def2-QZVPD	0.40	1.36	0.33	1.07	0.49	1.10	0.40	1.36
cc-pVDZ	3.34	8.10	2.42	5.31	1.90	2.68	2.58	8.10
jun-cc-pVDZ	0.98	3.10	0.69	2.09	0.65	1.32	0.78	3.10
jul-cc-pVDZ	0.58	2.64	0.73	2.09	0.62	1.37	0.64	2.64
aug-cc-pVDZ	0.75	2.86	0.77	2.06	0.74	1.52	0.75	2.86
cc-pVTZ	2.18	5.69	1.08	3.15	1.21	1.83	1.50	5.69
jun-cc-pVTZ	0.54	2.51	0.54	1.83	0.53	1.02	0.53	2.51
jul-cc-pVTZ	0.56	2.58	0.52	1.70	0.54	1.03	0.54	2.58
aug-cc-pVTZ	0.60	2.65	0.42	1.60	0.48	1.04	0.50	2.65
cc-pVQZ	0.87	2.84	0.52	2.08	0.78	1.47	0.72	2.84
jun-cc-pVQZ	0.38	1.80	0.32	1.11	0.47	1.00	0.39	1.80
jul-cc-pVQZ	0.44	2.17	0.32	1.17	0.46	0.99	0.40	2.18
aug-cc-pVQZ	0.49	2.41	0.30	1.14	0.44	0.97	0.41	2.41
6-31G(d)	4.74	10.08	3.41	6.74	2.73	3.88	3.67	10.08
6-31+G(d)	2.71	5.71	1.00	2.88	0.99	1.65	1.59	5.71
6-311G(d)	4.23	7.92	2.13	5.12	2.19	2.90	2.88	7.92
6-311G(d,p)	3.54	7.51	2.11	4.91	2.02	2.71	2.58	7.51
6-311+G(d)	2.20	3.51	1.15	3.25	1.31	1.91	1.56	3.51
6-311+G(3df,2pd)	1.04	4.50	0.84	2.47	0.82	1.53	0.90	4.50
6-311++G(d,p)	1.53	3.14	1.02	2.85	1.09	1.55	1.22	3.14
6-311++G(3df,2pd)	0.99	4.43	0.79	2.40	0.77	1.15	0.85	4.43

and affect the MAEs for that subset by $\lesssim 0.2$ kcal/mol.

Overall, even the best of the Pople basis sets afford larger errors than what is possible to achieve readily with either Karlsruhe or Dunning basis sets but can be competitive alternatives of only slightly lower quality, provided that both diffuse and polarization functions are present. The use of composite *sp* shells makes Pople basis sets more efficient, per unit basis function, as compared to other types of basis sets, assuming that one is using a quantum chemistry program that is written to exploit this. As such, 6-311+G(d,p) and 6-311+G(3df,2dp) may have a place in the pantheon of SAPT methods for large systems.

4.2.4. Summary. Several of the best-performing basis sets from each of the categories discussed above are compared in Fig. 2d. It is clear that even the best-performing Pople basis sets are outperformed by both Dunning and Karlsruhe basis sets, with absolute errors that are larger by up to 2 kcal/mol. All of the largest discrepancies are found amongst the hydrogen-bonded subset of S66, hence the problem is likely the inadequate description of induction energies. In contrast, the best of the Dunning and Karlsruhe basis sets are nearly identical in their performance, with MAEs < 0.5 kcal/mol. The Karlsruhe basis sets achieve this level of accuracy with fewer basis functions and for that reason we will focus on the Karlsruhe basis sets in much of the rest of this work.

4.3. Energy Component Analysis. It is clear from the results in Section 4.2 that the largest variations amongst basis sets occur for hydrogen-bonded complexes,

suggesting that the induction energy is more sensitive to the choice of basis set as compared to other energy components. To investigate this further, we next examine how individual energy components converge with respect to basis set. We do this first for the S66 systems that were examined above (in Section 4.3.1) and then for the S22 data set[87] (Section 4.3.2). For the latter, high-level SAPT benchmarks are available[88] so we can examine not just convergence but also errors in each energy component.

4.3.1. Averaged Components for S66. Individual energy components for the S66 systems are plotted in Fig. S1, where the calculations were performed at the XSAPT + MBD level in each of the six best-performing basis sets that were identified in Section 4.2. These data will not be discussed in detail because the variations between basis sets are mostly quite small. This does demonstrate that the basis sets that we previously identified as the best-performing ones achieve this status not through any kind of error cancellation but rather because they offer a converged (or nearly converged) description of each energy component.

In an attempt to distill the S66 data into a form that can be used for quick side-by-side comparison of different basis sets, we will instead examine average values of $E_{\rm elst}$, $E_{\rm exch}$, $E_{\rm ind}$, and $E_{\rm disp}$ across the entire S66 data set. These averages are not physically meaningful because they do not exemplify any one system; nevertheless, inspection of how the average changes from one basis set to the next provides a simple means to gauge convergence as a function of basis set. This analysis will underscore the fact that not all of the energy components converge in the same way and that some components place different demands on the basis set than others.

As explained in Section 3.3, we have tested not only the traditional augmented Karlsruhe basis sets (def2-SVPD, etc.) but also heavy-augmented versions (def2-ha-SVP, etc.) in which diffuse functions are removed from the hydrogen atoms, and minimally-augmented versions (def2-ma-SVP, etc.) that further remove the diffuse functions having highest angular momentum on the other atoms. Finally, the def2-SVP, def2-TZVP, and def2-QZVP basis sets contain no diffuse functions whatsoever.

Figure 3 presents the basis-set dependence for the errors in total interaction energies along with S66-averaged values of each energy component: $E_{\rm elst}$, $E_{\rm exch}$, $E_{\rm ind}$, and $E_{\rm disp}$. The average dispersion energies are considered in Fig. 3e and in view of the results presented in Section 4.3, it is not surprising that all of the basis sets examined afford values within 0.3 kcal/mol of the def2-QZVPD result. As such, we will not consider the dispersion energies in any more detail.

Considering the electrostatic energies (Fig. 3b), the def2-QZVPD basis set affords an S66-averaged value $\langle E_{\rm elst}\rangle = -6.7$ kcal/mol and all of the basis sets except def2-SVP come within ± 0.2 kcal/mol of this value, including the rather compact def2-ma-SVP basis set. In contrast to this, when the cardinality of the basis set is increased to triple- or quadruple-ζ level, augmentation does not appear to be necessary in order to obtain a converged result for electrostatics; the def2-TZVP and def2-QZVP values of $\langle E_{\rm elst}\rangle$ are both within 0.05 kcal/mol of the def2-QZVPD value.

S66-averaged exchange energies are given in Fig. 3c and these prove to be more sensitive to the choice of basis set. The def2-SVP basis is clearly inadequate and differs by about 4 kcal/mol (or 44%) from the def2-QZVPD value of $\langle E_{\rm exch}\rangle$. This should be contrasted with the behavior of $\langle E_{\rm elst}\rangle$, where the entire collection of Karlsruhe basis sets spans a range of less than 1.5 kcal/mol. Also in contrast to the electrostatics case, for exchange it does not seem to be possible to obtain a converged result simply by adding diffuse functions to a double-ζ basis set. However, a triple-ζ basis (with or without diffuse functions) does appear to be adequate: the def2-TZVP result differs by just 0.6 kcal/mol (8%) from the def2-QZVPD result and the def2-TZVPD value differs by just 0.1 kcal/mol (1%). What is similar to the case of electrostatics, and perhaps surprising, is that triple- and quadruple-ζ calculations of the exchange energy do not appear to benefit at all from diffuse basis functions. This is reflected in the overall error statistics in the interaction energies (Fig. 3a).

The quantity $E_{\rm elst} + E_{\rm exch}$ is often grouped together as "electrostatics plus finite size",[36,60] or equivalently as the electrostatic interaction between antisymmetrized monomer wave functions.[34] This is convenient because the sum of these two energy components is often more comparable in magnitude to the remaining energy components ($E_{\rm ind}$ and $E_{\rm disp}$), where electrostatics and exchange individually are often much larger in magnitude but opposite in sign, at least where equilibrium geometries are concerned. The average basis-set dependence of $E_{\rm elst} + E_{\rm exch}$ is presented in Fig. 3f. On the ~ 1 kcal/mol energy scale it is clear that a triple-ζ basis set is required to obtain converged results, but that only minimal augmentation is needed in that case. The def2-SVP result is qualitatively wrong but even with a full complement of diffuse functions (*i.e.*, def2-SVPD), the results are clearly inferior to def2-TZVP. Convergence is reached at def2-ma-TZVP.

Average induction energies are presented in Fig. 3d. Whereas the def2-SVP value of $E_{\rm ind}$ is underestimated by 2.0 kcal/mol (17%), all of the other values lie within 0.2 kcal/mol (8%) of the def2-QZVPD result. Remarkably, a double-ζ basis set with diffuse functions affords values of $E_{\rm ind}$ that are close to def2-QZVPD values, and in a triple-ζ basis set the diffuse functions are not needed to obtain a value that is negligibly different from the def2-QZVPD value. This may suggest a tractable computational scheme for large systems, in which the induction energy is computed in a smaller basis set as compared to electrostatics or exchange. First, however, we will more rigorously examine this conclusion in Section 4.5, using

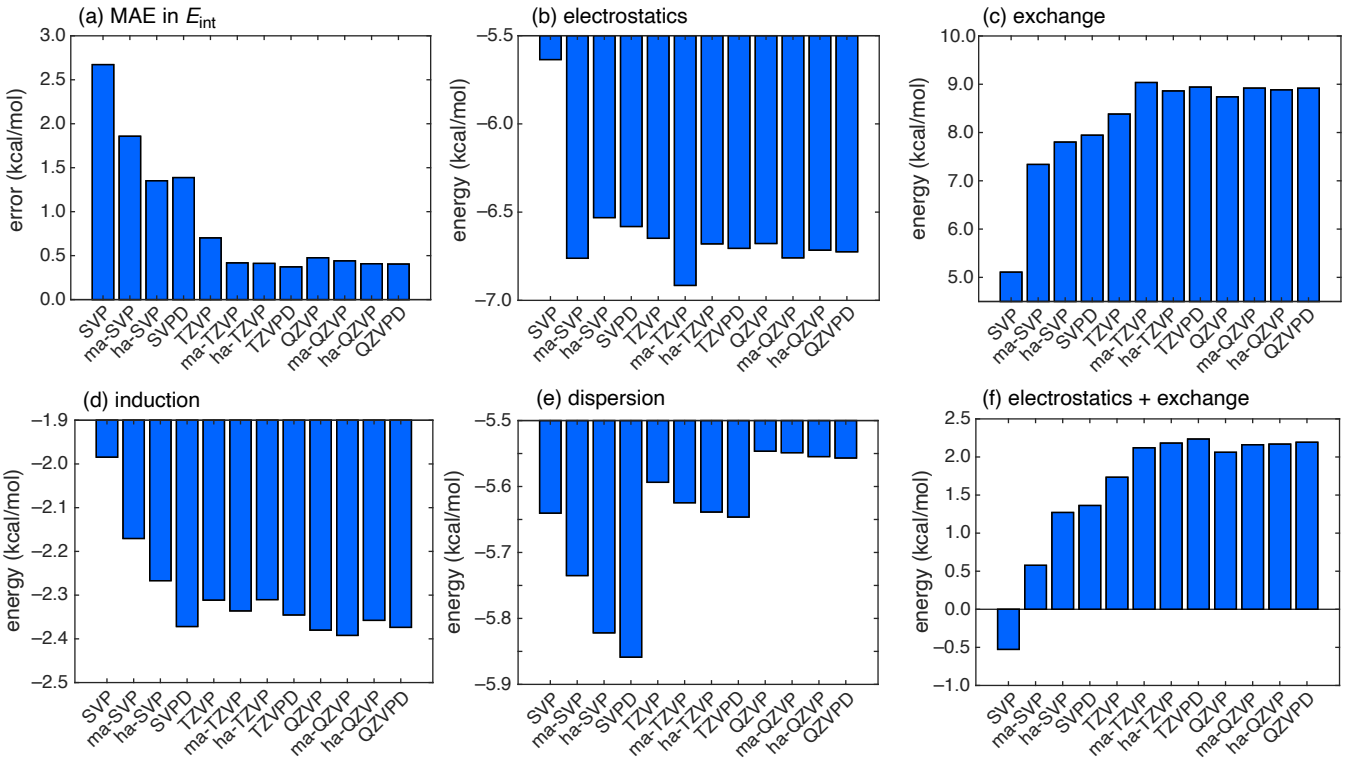

Fig. 3: (a) MAEs in $E_{\rm int}$, along with S66-averaged values of (b) $E_{\rm elst}$, (c) $E_{\rm exch}$, (d) $E_{\rm ind}$, (e) $E_{\rm disp}$, and (f) $E_{\rm elst} + E_{\rm exch}$, each computed at the XSAPT + MBD level in various Karlsruhe basis sets. For brevity, "def2" is omitted from the basis set names.

ion–water clusters that exhibit much larger induction energies as compared to the S66 systems.

Looking back at the errors in total interaction energies (Fig. 3a), we see that none of the double-ζ basis sets achieves a MAE below 1 kcal/mol. Based on the analysis presented in this section, we ascribe this failure primarily to inadequate description of $E_{\rm elst}$ and $E_{\rm exch}$, since $E_{\rm ind}$ and $E_{\rm disp}$ are reasonably well described in double-ζ basis sets so long as diffuse functions are included. The def2-TZVP basis set affords a mean accuracy < 1 kcal/mol despite the absence of diffuse functions, without benefiting significantly from error cancellation. With minimal augmentation (def2-ma-TZVP), the MAE drops below 0.5 kcal/mol and the average deviation in each energy component is $< 3\%$ of the def2-QZVPD value. The def2-ma-TZVP basis set therefore appears to afford converged XSAPT + MBD results, at least for the S66 dimers.

4.3.2. Benchmark Components for S22. In this section, we examine the basis-set behavior of individual energy components for the S22 data data,[87] comparing XSAPT + MBD (in various basis sets) to energy components computed at the SAPT2+(3)/aug-cc-pVTZ level.[88] The latter approach affords an overall MAE of 0.5 kcal/mol for the S22 interaction energies, with respect to CCSD(T)/CBS benchmarks, which can be separated into an MAE of 0.8 kcal/mol for the electrostatically-dominated subset of the S22 dimers and 0.2 kcal/mol for the remaining systems.[25] We therefore take the SAPT2+(3)/aug-cc-pVTZ results to be benchmarks for each individual energy component.

Figure 4 plots the error in each energy component, for each S22 dimer, in several basis sets. We restrict out attention to the best-performing Pople and Karlsruhe basis sets that were identified in Section 4.2, omitting the Dunning basis sets since the S66 results suggest that comparable, high-quality XSAPT + MBD energetics can be obtained more efficiently using def2-TZVPD and def2-QZVPD. Examining the electrostatic energies in Fig. 4a, we notice that 6-311++G(d,p) exhibits errors $\gtrsim 1.0$ kcal/mol for several of the electrostatically-dominated systems (dimers 1–7 in Fig. 4), with the largest errors for $(H_2O)_2$ and $(HCO_2H)_2$ (dimers 2 and 3, respectively). With additional polarization functions, to obtain 6-311++G(3df,2pd), the electrostatic errors for these same complexes are comparable to those obtained using def2-TZVPD or def2-QZVPD. More surprising is that 6-311++G(3df,2pd) outperforms both Karlsruhe basis sets for the induction energies (Fig. 4c). That said, errors in the exchange energies (Fig. 4b) are significantly larger for both Pople basis sets than they are for either Karlsruhe basis set. This is especially true for the formic acid dimer (system #3) and for the four π-stacked dimers, systems 11–15, representing (respectively) the dimers of benzene, pyrazine,

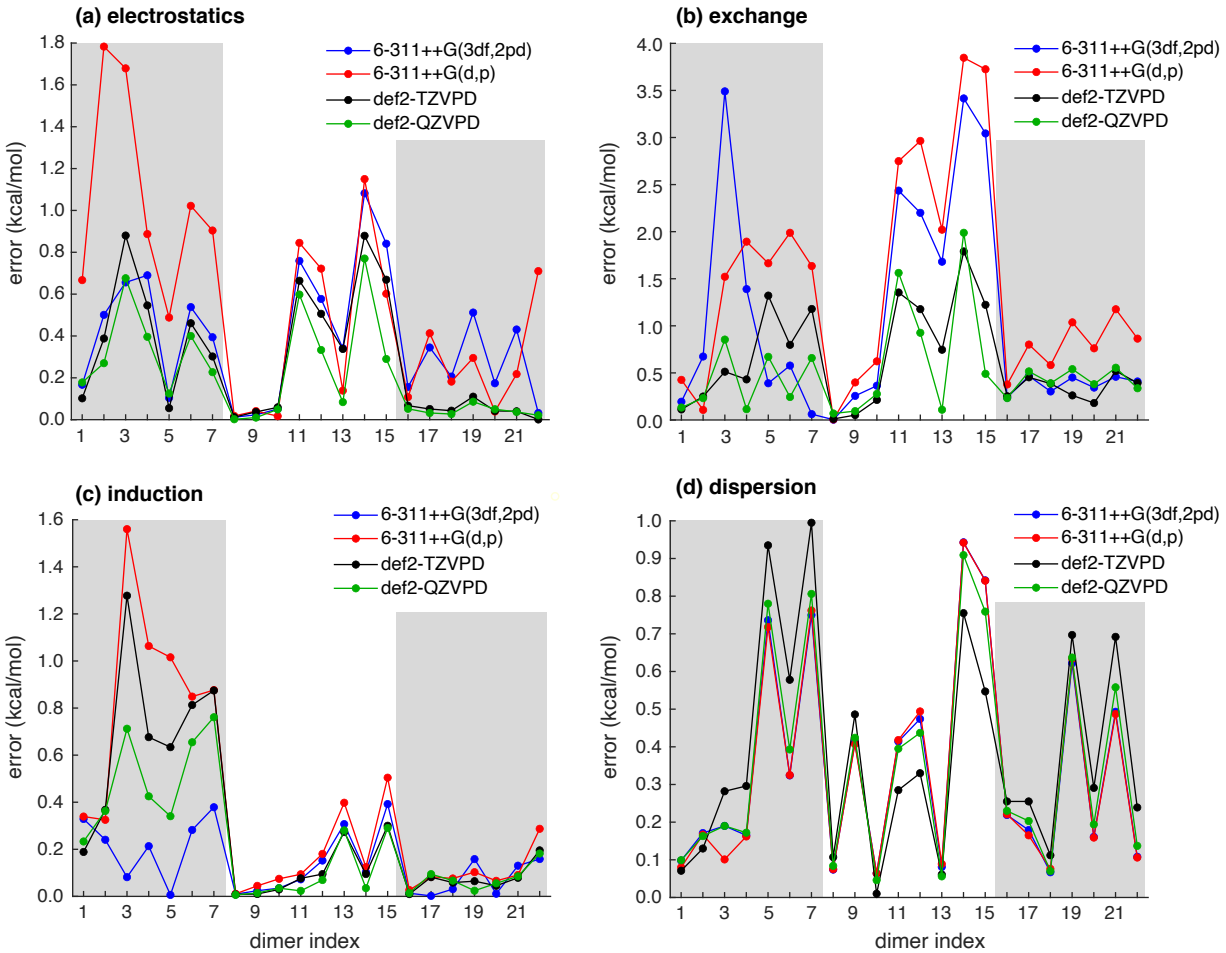

Fig. 4: MAEs in energy components for the S22 dimers, computed at the XSAPT + MBD level in the basis sets that are shown and compared to benchmark energy components computed at the SAPT2+(3)/aug-cc-pVTZ level. Shaded regions delineate the electrostatics-dominated dimers (1–7), the dispersion-dominated dimers (8–15), and the dimers with mixed-influence interactions (16–22). Note that the energy scale is different in each panel.

and uracil, and the heterodimers indole$\cdots$benzene and adenine$\cdots$thymine. Consistent with the analysis in Section 4.3.1, the dispersion energy is rather insensitive to the choice of basis set.

Overall, the picture that emerges from these data are that the somewhat larger errors for hydrogen-bonded complexes that were documented for S66 dimers when the 6-311++G(3df,2pd) basis set is used (MAE of 1.0 kcal/mol and maximum error of 4.4 kcal/mol, whereas the maximum error for def2-TZVPD is 0.5 kcal/mol) result primarily from inadequate description of exchange rather than electrostatics or induction. The def2-TZVPD basis set offers reasonably consistent performance for each energy component, with errors that are $\lesssim$ 1.5 kcal/mol in each.

4.4. Benchmarks Containing Ions. Systems containing ions tend to exhibit much larger interaction energies as compared to cases where the monomers are charge-neutral, and this may place different demands on basis sets. We next consider results for the IL16, AHB21, and CHB6 data sets of ion-containing dimers that were described in Section 3.4.[83] Despite some promising preliminary results using XSAPT + MBD to describe ion–molecule interactions,[36,56,60] we have identified other cases where performance of the MBD model is erratic for ion-containing systems, perhaps due to issues relating to the non-uniqueness of the Hirshfeld partition in such cases, for which iterative schemes have been suggested[89–91] and implemented with supramolecular DFT-based versions of the MBD model.[92–94] (Alternative means of obtaining polarizabilties for use in the MBD model have also been shown to be superior to the original MBD@rsSCS approach.[95]) In view of these difficulties, we consider that a proper MBD model for ion-containing systems is still under development. For the present purposes, however, we have already demonstrated that the MBD dispersion energy converges rapidly with basis set, because the density converges rapidly, and do not expect

this to be different for an ionic MBD model. We will therefore focus on the other energy components in this section, using the XSAPT + *ai*D3 method.[54] In this approach, $E_{\rm disp}$ is modeled using atom–atom C_6/R^6 and C_8/R^8 potentials and does not depend on the density at all, but the other three energy components are computed in the same manner as in XSAPT + MBD, so the conclusions should be transferable.

XSAPT + *ai*D3 error statistics for the IL16, AHB21, and CHB6 data sets are compiled in Table 3. For these calculations, monomer SCF wave functions are computed using the monomer basis set, in order to preclude any charge delocalization at the SCF level, although the dimer basis is still used for the $\delta E_{\rm HF}$ calculations because $\Delta E_{\rm int}^{\rm HF}$ in eq. 4 needs to be counterpoise-corrected. Because the IL16 and AHB21 data sets contain anions, only basis sets containing diffuse functions are considered in these tests.

The first column of Table 3 gives the error statistics for the IL16 database of cation–anion pairs, and these errors are also plotted in Fig. 5a for the Dunning and Karlsruhe basis sets. Notably, 6-311++G(3df,2pd) proves to be on par with the other triple-ζ basis sets in systems containing either cations (CHBH) or anions (AHB21) with an MAEs of 2.95 and 1.74 kcal/mol respectively. However, when cations and anions are in the system (IL16), the MAE increases substantially to 8.31 kcal/mol

Perhaps serendipitously, the best-performing basis set for IL16 is jun-cc-pVTZ with a MAE of 0.8 kcal/mol although the def2-ha-SVP and def2-TZVPPD basis sets afford comparable MAEs, and def2-TZVPPD affords a smaller maximum error. (The def2-TZVPD basis set is nearly as good as def2-TZVPPD.) In contrast to the behavior for S66, here we see no systematic improvement in accuracy for larger basis sets, although neither do the errors get significantly larger. For example, the XSAPT + *ai*D3/aug-cc-pVXZ methods with X = D, T, and Q afford MAEs of 1.3, 1.7, and 1.7 kcal/mol, respectively, and maximum errors of 3.5, 3.8, and 3.7 kcal/mol. Similar trends hold for the Karlsruhe sequence of basis sets (def2-SVPD, def2-TZVPD, and def2-QZVPD), although the errors are slightly but systematically smaller in the Karlsruhe basis sets as compared to the Dunning basis sets. As noted in a previous SAPT study,[83] the most challenging systems in the ion-containing data sets that are examined here require a "δMP2" correction for higher-order induction (and not just a δHF correction), in order to achieve sub-kcal/mol accuracy with respect to CCSD(T)/CBS benchmarks. As such, errors of 1.7 kcal/mol (MAE) or 3.7 kcal/mol (maximum) obtained at the XSAPT + *ai*D3/aug-cc-pVQZ level likely reflect the inherent accuracy of this particular XSAPT-based method as applied to these very challenging ion-pair systems. Some error cancelation is then responsible for the somewhat better performance of the smaller basis sets mentioned above.

Figures 5b–d show the energy components (excluding dispersion) for the IL16 data set. The electrostatics term (Fig. 5b) is especially flat, with variations of no more than 1 kcal/mol across a wide range of Karlsruhe and Dunning basis sets. Excluding the double-ζ basis sets, because we know that $E_{\rm elst}$ is not yet converged in double-ζ basis sets even for the S66 dimers, the variations in $E_{\rm elst}$ are $\lesssim 0.1$ kcal/mol. Exchange energies (Fig. 5c) also vary by only ~ 1 kcal/mol if double-ζ basis sets are excluded from the comparison. In contrast, induction energies (Fig. 5d) span a range of ~ 4 kcal/mol in the various triple- and quadruple-ζ basis sets. This analysis points to induction as the energy component wherein the overall errors in the interaction energies reside, which is consistent with the need for a δMP2 correction to achieve sub-kcal/mol accuracy.

In contrast to IL16, overall errors in the interaction energies for the AHB21 and CHB6 data sets follow more discernible and systematic trends; see Table 3. The smallest MAEs (1.2–1.5 kcal/mol) are obtained using the aug-cc-pVQZ and def2-QZVPD basis sets. Trimming the diffuse functions at the triple-ζ level has a noticeably detrimental effect on the anion–neutral dimers in the AHB21 data set. This effect, which was not seen in dimers of neutral molecules, is less evident at the quadruple-ζ level. In general the basis-set demands for these ion-containing dimers are more severe than they are for the S22 and S66 dimers.

4.5. Halide–Water Clusters. For the S66 data set we found that induction energies exhibit remarkably little basis set dependence (see Fig. 3d), which might simply reflect that the induction energies for these small, charge-neutral dimers are not especially large. That the ionic test systems considered in Section 4 .4 exhibit somewhat more stringent basis-set requirements comports with this hypothesis. To test this further, we next consider some larger anion–water clusters whose induction energies are more significant.

4.5.1. Aqueous Chloride Ion. We first consider $Cl^-(H_2O)_n$ clusters extracted from a molecular dynamics simulation of Cl^-(aq),[96] and which have previously been used to explore the energetics of ion hydration.[60] A set of 51 clusters containing an average of $n = 28$ water molecules (representing two solvation shells) was taken from previous work,[60,96] and we will examine basis-set effects on the energy components across this trajectory. These calculations are meant to investigate what (if any) cancellation one can expect across a set of similar geometries for the same system and for the purposes of these calculations each $Cl^-(H_2O)_n$ snapshot is treated as a dimer, with Cl^- as one monomer and $(H_2O)_n$ as the other. As a representative example of a high-throughput application, we examine some very efficient basis sets such as 6-31+G(d) and 6-311+G(d,p) in addition to higher-quality ones such as def2-TZVPD and some intermediate choices as well. Calculations were performed

Table 3: Error Statistics for XSAPT + *ai*D3 Applied to Ion-Containing Data Sets.[a]

Basis Set	Error (kcal/mol)								
	IL16			AHB21			CHB6		
	MAE		Max	MAE		Max	MAE		Max
jun-cc-pVDZ	2.53	(2.3%)	5.75	3.12	(12.9%)	13.12	1.46	(5.0%)	3.11
jul-cc-pVDZ	1.09	(1.0%)	3.18	2.26	(11.0%)	7.69	2.05	(6.3%)	6.19
aug-cc-pVDZ	1.29	(1.2%)	3.54	1.86	(8.0%)	7.36	2.01	(6.2%)	6.50
jun-cc-pVTZ	**0.80**	**(0.8%)**	**3.07**	1.76	(8.4%)	6.44	2.09	(6.4%)	7.21
jul-cc-pVTZ	1.28	(1.2%)	3.46	1.48	(7.0%)	5.69	2.32	(7.1%)	7.57
aug-cc-pVTZ	1.71	(1.6%)	3.83	1.27	(5.8%)	5.70	2.09	(6.3%)	7.06
jun-cc-pVQZ	1.17	(1.1%)	2.79	1.43	(6.9%)	6.18	1.53	(4.8%)	4.40
jul-cc-pVQZ	1.44	(1.3%)	3.38	1.28	(6.2%)	7.80	1.69	(5.1%)	5.53
aug-cc-pVQZ	1.72	(1.6%)	3.69	**1.23**	**(5.5%)**	**8.18**	**1.31**	**(3.8%)**	**4.74**
def2-ma-SVP	3.11	(2.8%)	5.60	3.70	(16.7%)	11.93	1.83	(6.7%)	3.43
def2-ha-SVP	1.03	(1.0%)	2.60	2.86	(13.5%)	8.38	1.20	(4.3%)	2.56
def2-SVPD	1.23	(1.1%)	2.60	2.49	(11.3%)	8.14	1.52	(5.3%)	2.89
def2-ma-TZVP	1.95	(1.8%)	3.78	2.61	(11.2%)	8.57	1.74	(5.9%)	3.86
def2-ha-TZVP	1.24	(1.1%)	2.30	2.73	(12.3%)	8.24	1.28	(4.1%)	3.68
def2-TZVPD	1.08	(1.0%)	2.29	2.54	(11.1%)	8.36	1.26	(4.0%)	3.65
def2-TZVPPD	**0.85**	(0.8%)	**2.16**	1.89	(8.9%)	6.09	1.63	(5.0%)	5.78
def2-ma-QZVP	1.44	(1.3%)	3.44	**1.39**	**(7.0%)**	**3.20**	1.76	(5.3%)	3.20
def2-ha-QZVP	1.23	(1.1%)	2.60	1.51	(7.6%)	3.03	1.43	(4.4%)	4.97
def2-QZVPD	1.24	(1.1%)	2.63	**1.46**	**(7.3%)**	**3.06**	**1.20**	**(3.8%)**	**3.96**
def2-QZVPPD	1.24	(1.1%)	2.63	**1.46**	**(7.3%)**	**3.06**	**1.21**	**(3.8%)**	**3.96**
6-311+G(3df,2pd)	8.33	(8.1%)	18.12	1.76	(8.2%)	8.28	2.98	(8.8%)	10.38
6-311++G(3df,2pd)	8.32	(8.1%)	18.09	1.74	(8.1%)	8.09	2.95	(8.7%)	10.26

[a]Relative to CCSD(T)/CBS benchmarks from Ref. 83. Boldface values indicate the best-performing basis sets.

at the XSAPT + *ai*D3 level and δE_{HF} is *not* included in the induction energy for these calculations, in the interest of computational expedience. We expect the convergence behavior of the second-order induction terms in eq. 2 to be similar to that of δE_{HF}. The first half of the data set is sufficient to exhibit both the minimum and maximum value of each energy component so for clarity only those data are shown, in Fig. 6.

While there are certainly systematic errors in particular basis sets, we observe that the overall step-to-step fluctuations (both in E_{int} and in its components) are very similar, across basis sets ranging in quality from 6-31+G(d) to def2-TZVPD. Systematic errors are different in various energy components and this leads to some error cancellation, such that E_{int} computed at the XSAPT + *ai*D3/6-31+G(d) level differs from the corresponding def2-TZVPD value by only 1.9 ± 0.8 kcal/mol. This is just 2% of the ensemble-averaged interaction energy, $\langle E_{\mathrm{int}} \rangle = -87.0$ kcal/mol. Error statistics for E_{int}, with respect to XSAPT + *ai*D3/def2-TZVPD values that represent our best estimates for this system, are listed in Table 4 along with timing data for various basis sets.

Examining the individual energy components in Fig. 6, we note that the relatively low-quality 6-31+G(d) and 6-311+G(d,p) basis sets both overestimate E_{elst} (making it too attractive) but underestimate E_{ind} (making it insufficiently attractive). Partial cancellation of these errors is responsible for the reasonable interaction energies that are obtained in either basis set. Errors in both E_{elst} and E_{ind} are larger for 6-31+G(d), which therefore relies more heavily on error cancellation. In contrast, the 6-311+G(3df,2pd) basis set affords individual energy components that are fully converged with respect to def2-TZVPD values, so the accuracy of this approach does not

Table 4: Timing Data and Error Statistics for XSAPT + *ai*D3 Calculations on $Cl^-(H_2O)_n$.[a]

Basis Set	E_{int} Error (kcal/mol)[b]		Time (sec)[c]	
	MAE[d]	Max	CPU	wall
6-31+G(d)	1.9 ± 0.8	3.8	1954	152
6-311+G(d,p)	1.3 ± 0.5	2.4	4912	402
6-311+G(2df,2p)	0.1 ± 0.1	0.4	14935	1143
6-311+G(3df,2pd)	0.3 ± 0.1	0.6	32916	2501
def2-SVPD	1.2 ± 0.2	1.5	22492	2144
def2-TZVPD	0.0	0.0	73063	6326

[a]Data represent averages over 51 snapshots with $\langle n \rangle = 28.4 \pm 2.4$ water molecules treated as a single fragment. [b]With respect to the def2-TZVPD value. [c]Calculations run on 14 processors. [d]Uncertainty represents one standard deviation.

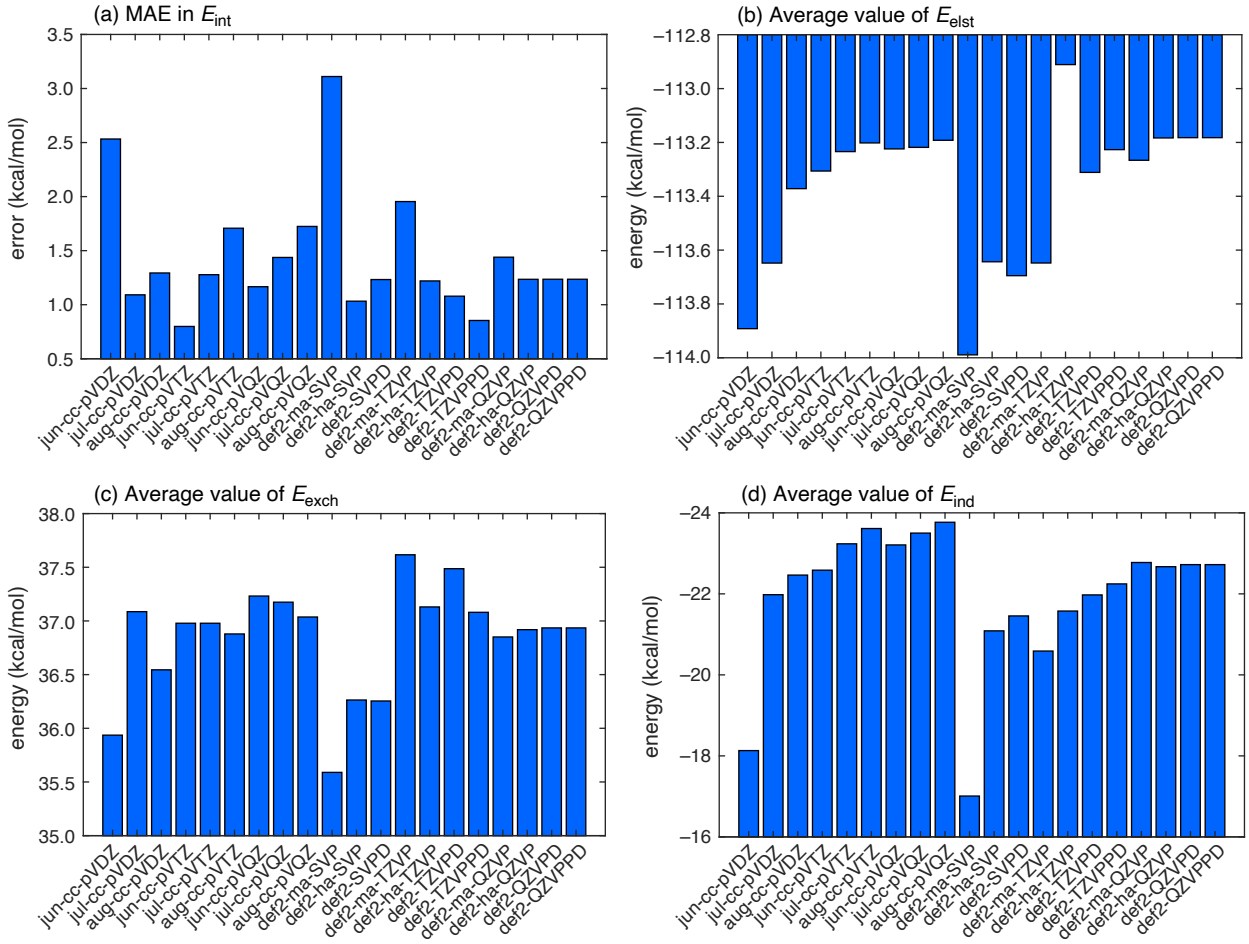

Fig. 5: XSAPT + aiD3 results for the IL16 data set: (a) error in the total interaction energy, and values of (b) $E_{\rm elst}$, (c) $E_{\rm exch}$, and (d) $E_{\rm ind}$ in various basis sets.

rely on error cancellation, and in fact the maximum deviation with respect to the def2-TZVPD interaction energies is only 0.4 kcal/mol when this high-quality Pople basis set is used. Interestingly, the exchange energies appear to be converged in any triple-ζ basis set but also in def2-SVPD, although they are somewhat overestimated when 6-31+G(d) is used.

Timing data in Table 4 demonstrate that the use of 6-311+G(3df,2pd) affords a 2× speedup relative to def2-TZVPD. Interestingly, the latter actually comprises *fewer* basis functions. For $Cl^-(H_2O)_{28}$, which is the average size of the clusters examined in this section, the use of def2-TZVPD involves 1,670 basis functions versus 1,923 for 6-311+G(3df,2pd). The smaller timings for the larger calculation provides an unambiguous indication of the much better efficiency of Pople basis sets, provided that the integrals engine can exploit *sp* shells.

If one removes some of the polarization functions from the highest-quality Pople basis set, to obtain 6-311+G(2df,2p), the speedup relative to def2-TZVPD increases to 5× while preserving the accuracy of $E_{\rm int}$, at the expense of introducing a small amount of error cancellation between $E_{\rm elst}$ and $E_{\rm ind}$. This basis is superior to def2-SVPD in both accuracy and efficiency. Nevertheless, it is worth noting that the error in the 6-31+G(d) calculations appears to be quite systematic while affording a speedup of 37× relative to def2-TZVPD. The more affordable Pople basis sets therefore offer good insight as to how the energy components change as a function of geometry in large systems.[60]

4.5.2. Basis-Set Convergence for $F^-(H_2O)_6$. In a similar vein, we next examine energy components for $F^-(H_2O)_6$, for which CCSD(T)/CBS benchmarks[97] allow us to characterize absolute errors in the interaction energies. In contrast to the Cl^-(aq) system, where each $Cl^-(H_2O)_n$ cluster was treated as a dimer, here we exploit the many-body power of XSAPT to describe $F^-(H_2O)_6$ as a system of 7 fragments.

Figure 7a shows the errors in $E_{\rm int}$ using XSAPT + MBD in various Karlsruhe basis sets. Errors are unacceptably large in the absence of diffuse functions (even using def2-QZVP), which is not surprising, but

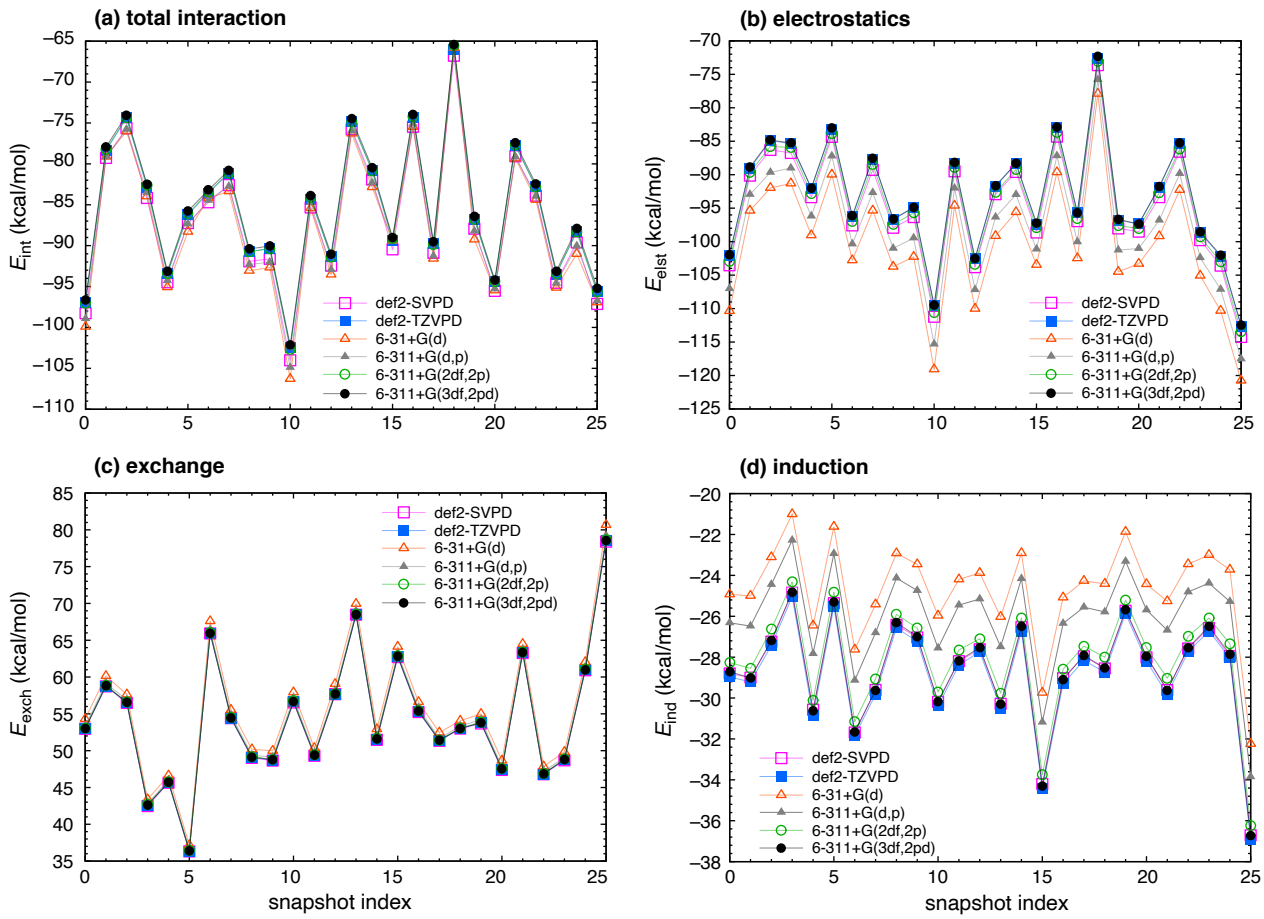

Fig. 6: XSAPT + aiD3 calculations along a Cl^-(aq) trajectory with two solvation shells of explicit water: (a) $E_{\rm int}$, (b) $E_{\rm elst}$, (c) $E_{\rm exch}$, and (d) $E_{\rm ind}$. Note that the energy scales are different in each panel. The induction energy in (d) contains the second-order terms only but omits $\delta E_{\rm HF}$ in eq. 2.

minimal augmentation is sufficient to reduce the errors to $\leq$ 5 kcal/mol versus a CCSD(T)/CBS reference energy $E_{\rm int} = -116.6$ kcal/mol,[97] even at the double-ζ level. In quadruple-ζ basis sets with diffuse functions, the XSAPT + MBD errors are all ≤ 1.3 kcal/mol.

Examining the basis-set convergence of the $F^-(H_2O)_6$ energy components, we find that electrostatic energies (Fig. 7b) span a wide range but if the minimally-augmented basis sets are excluded then all of the values of $E_{\rm elst}$ lie between -175 and -172 kcal/mol. At the triple-ζ level it appears that diffuse functions can safely be deleted from the hydrogen atoms, as the def2-ha-TZVP and def2-TZVP values of $E_{\rm elst}$ differ by only 0.2 kcal/mol. That said, the def2-TZVPD value ($E_{\rm elst} = -174.4$ kcal/mol) differs by 2 kcal/mol from the def2-QZVPD value ($E_{\rm elst} = -172.2$ kcal/mol), and in fact that def2-SVPD value is much closer: $E_{\rm elst} = -174.6$ kcal/mol. As we observed for S66, a double-ζ basis set with a full complement of diffuse functions can stand in for a triple-ζ basis set for electrostatics calculations.

Exchange energies (Fig. 7c) span a similarly wide range and here the diffuse functions are important, to prevent severe underestimation of $E_{\rm exch}$ caused by the tails of the anion's wave function. That said, the def2-SVPD value ($E_{\rm exch} = 144.6$ kcal/mol) is only about 1 kcal/mol different from the def2-QZVPD value ($E_{\rm exch} = 143.8$ kcal/mol). Often, the sum $E_{\rm elst} + E_{\rm exch}$ is easier to interpret than either of these energy components alone,[36,60] in part because the sum tends to be more comparable to values of $|E_{\rm ind}|$ and $|E_{\rm disp}|$, whereas $|E_{\rm elst}|$ and $E_{\rm exch}$ are individually much larger. Data for $E_{\rm elst} + E_{\rm exch}$ in $F^-(H_2O)_6$ are presented in Fig. 7f. Using heavy-augmented basis sets, this sum appears to converge at the triple-ζ level, where the difference with respect to def2-QZVPD is only 0.8 kcal/mol, which once again suggests that diffuse functions on hydrogen are not essential at the triple-ζ level. With a full complement of diffuse functions, the def2-SVPD basis set can be substituted for the triple-ζ one with $\lesssim$ 1 kcal/mol loss in accuracy.

Induction energies for $F^-(H_2O)_6$ are presented in Fig. 7d and the def2-QZVPD value is $E_{\rm ind} = -50.8$ kcal/mol. Double-ζ basis sets grossly under-

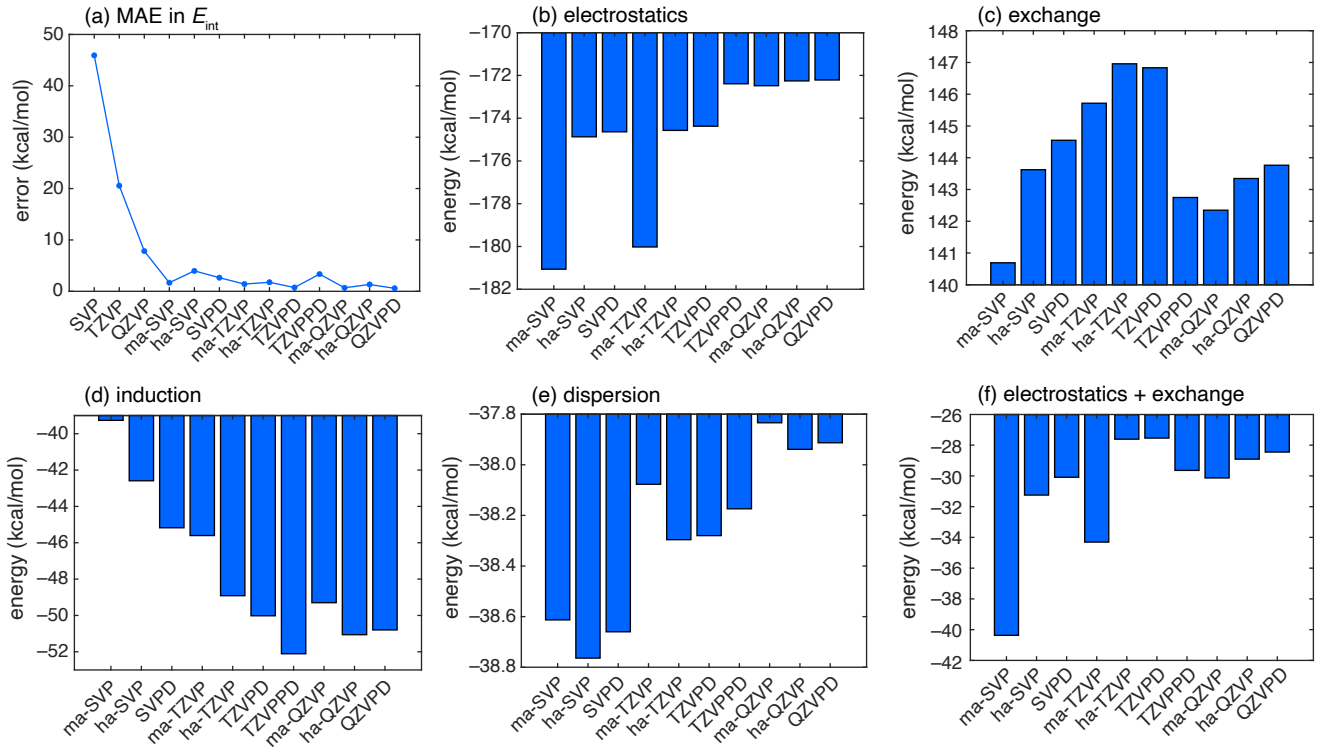

Fig. 7: Basis-set dependence of XSAPT + MBD for $F^-(H_2O)_6$ using Karlsruhe basis sets: (a) Errors in $E_{\rm int}$ versus the CCSD(T)/CBS benchmark ($E_{\rm int} = -116.6$ kcal/mol),[97] and convergence of (b) $E_{\rm elst}$, (c) $E_{\rm exch}$, (d) $E_{\rm ind}$, (e) $E_{\rm disp}$, and (f) $E_{\rm elst} + E_{\rm exch}$ as a function of basis set. For brevity, the "def2" prefix is omitted from the basis set names. Note also that the basis-set selection (and the ordering) is different in (a) as compared to the other panels, in order to highlight the significant errors obtained in the absence of diffuse functions.

Table 5: Error Statistics for SAPT0(HF) + MBD Applied to the S66 Data Set.

Method	Error (kcal/mol)							
	H-Bonded		Dispersion		Mixed		Total	
	MAE	Max	MAE	Max	MAE	Max	MAE	Max
def2-SVP	5.24	12.53	2.77	6.73	2.39	3.31	3.52	12.53
def2-SVPD	3.11	8.70	2.28	5.43	1.92	2.95	2.46	8.70
def2-TZVP	2.45	5.38	1.12	3.71	1.44	2.33	1.61	5.38
def2-TZVPD	1.73	4.85	1.06	3.36	1.21	2.07	1.34	4.85
def2-QZVP	2.01	5.88	1.04	3.24	1.24	2.17	1.44	5.88
def2-QZVPD	1.91	5.72	1.04	3.12	1.20	2.09	1.39	5.72
cc-pVDZ	4.86	11.84	2.81	6.73	2.46	3.46	3.42	11.84
jun-cc-pVDZ	3.52	8.93	1.64	4.68	1.61	2.48	2.29	8.93
jul-cc-pVDZ	2.02	5.77	1.06	3.40	1.19	2.07	1.44	5.77
aug-cc-pVDZ	2.63	7.22	1.68	4.43	1.61	2.41	1.99	7.22
cc-pVTZ	3.43	8.38	1.69	4.76	1.82	2.80	2.34	8.38
jun-cc-pVTZ	2.28	6.28	1.27	3.97	1.29	2.06	1.63	6.28
jul-cc-pVTZ	2.29	6.29	1.28	3.86	1.31	2.07	1.64	6.29
aug-cc-pVTZ	2.30	6.41	1.23	3.80	1.26	2.07	1.61	6.41
cc-pVQZ	2.32	6.09	1.17	3.72	1.42	2.43	1.64	6.09
jun-cc-pVQZ	1.98	5.65	1.06	3.35	1.20	2.08	1.43	6.65
jul-cc-pVQZ	2.48	7.29	1.61	4.44	1.49	2.31	1.87	7.29
aug-cc-pVQZ	1.98	5.83	1.07	3.39	1.20	2.04	1.43	5.83
6-311+G(3df,2pd)	2.92	8.40	1.56	4.49	1.53	2.41	2.02	8.40
6-311++G(3df,2pd)	2.90	8.37	1.59	4.52	1.58	2.46	2.04	8.37

estimate this value, even with a full complement of diffuse functions; the def2-SVPD value is $E_{\rm ind} = -45.2$ kcal/mol, which represents 11% error with respect to the def2-QZVPD value. In contrast, the def2-TZVPD value ($E_{\rm ind} = -50.0$ kcal/mol) represents an underestimate of 0.8 kcal/mol or 2%. To obtain converged results, triple-ζ basis sets with diffuse functions are required and here the hydrogen atom diffuse functions contribute 1.0 kcal/mol, comparing def2-ha-TZVP to def2-TZVPD.

It is also abundantly clear that the dispersion energy is not very sensitive to basis set, and changes by only 0.7 kcal/mol between the def2-ma-SVP and def2-QZVPD basis sets. This is consistent with earlier results for other data sets where the dispersion energy is nearly constant across a wide range of basis sets, suggesting that MBD gives both accurate dispersion energies and ones that converge quickly.

4.6. Performance of Different SAPT Methods. In this section, we compared the performance of three different SAPT variants: XSAPT, SAPT0(HF), and SAPT0(KS), using either MBD or else *ai*D3 to replace second-order dispersion (eq. 3) in each case. The purpose of these tests are to examine the importance of the self-consistent XPol charge embedding that is used in XSAPT + MBD, thus the two SAPT0 variants do not include such an embedding. The SAPT0(HF) + MBD method consists of a standard HF-based SAPT0 calculation but with a modified dispersion term, while SAPT0(KS) + MBD uses GDD-tuned LRC-ωPBE in place of HF theory for the monomer wave functions and is therefore similar to XSAPT + MBD except that the self-consistent embedding is omitted.

4.6.1. Total Interaction Energies. Error statistics for the SAPT0(HF) + MBD and SAPT0(KS) + MBD methods, as applied to the S66 dimers data set, are provided in Tables 5 and 6, respectively, which should be compared to the statistics for XSAPT + MBD that are listed in Table 2. SAPT0(HF) + MBD is clearly less accurate than either of the approaches based on LRC-ωPBE for the monomers, with MAEs that are consistently 1.0–1.5 kcal/mol larger. (This is analyzed in terms of energy components in Section 4.6.2.) Statistically speaking, the performance of SAPT0(KS) + MBD and XSAPT + MBD is about the same for these dimers, which by definition do not experience many-body polarization. This demonstrates that the charge-embedding procedure does not improve the quality of the zeroth-order reference state small dimers in any way that is statistically meaningful.

Error statistics for the ion-containing data sets are provided in Tables S8, S9 and S10, computed with the +*ai*D3 versions of XSAPT, SAPT0(HF), and SAPT0(KS). Here, the trend is different and XSAPT + *ai*D3 outperforms both of the other methods by 1–2 kcal/mol for the IL16 data set and by a lesser amount for the AHB21 data set. The fact that XSAPT + *ai*D3 exhibits statistically-significant error reduction as compared to SAPT0(KS) + *ai*D3 suggests that induction effects are large enough in these systems so that the charge-embedded XPol determinant is a better reference state for perturbation theory as compared to the use of isolated-monomer wave functions.

In contrast to the behavior for IL16 and AHB21, where XSAPT + *ai*D3 appears to offer clear improvement over the other two +*ai*D3 methods, results from all three approaches are much more similar for the CHB6 data set; see Table S10. This data set consists of $M^+(H_2O)$ and $M^+(C_6H_6)$ complexes with M = Li, Na, or K. Specifically for the cations, the tuning procedure affords rather large values of $\omega_{\rm GDD}$ (Table S4): $\omega_{\rm GDD} = 1.30\ a_0^{-1}$, $0.84\ a_0^{-1}$, and $0.59\ a_0^{-1}$ for Li^+, Na^+, and K^+, respectively. The values for Li^+ and Na^+ are much larger than the tuned values for small, charge-neutral monomers or anions, and even the value obtained for K^+ is somewhat larger than what we observe for other systems. The range separation in LRC-ωPBE occurs at a distance of $\approx \omega_{\rm GDD}^{-1}$, and that value corresponds to 0.41 Å for Li^+, 0.63 Å for Na^+, and 0.90 Å for K^+. In each case, this is smaller than the van der Waals radius of the cation itself,[98] meaning that for the purpose of intermolecular interactions computations, the XSAPT + *ai*D3 calculations are using HF theory for the exchange functional. This explains the close similarity with SAPT0(HF) + *ai*D3.

4.6.2. Energy Components. In order to further investigate these three variants of SAPT and to understand the origins of the differences between them, we wish to examine individual energy components. For this we turn to the S22 data set,[87] for which benchmark energy components are available,[88] computed at the SAPT2+(3)/aug-cc-pVTZ level. Figure 8 compares benchmarks against the SAPT0(HF) + MBD, SAPT0(KS) + MBD, and XSAPT + MBD methods and Table 7 lists the error statistics in each energy component. These data indicate that the $E_{\rm exch}$ and $E_{\rm ind}$, for the electrostatically-dominated subset of S22, are responsible for the diminished accuracy of SAPT0(HF) + MBD relative to other methods; these two components exhibit MAEs of 3.6 and 2.0 kcal/mol, respectively, for the subset that includes the hydrogen-bonded dimers.

For the DFT-based second-order methods, most of the energy components exhibit sub-kcal/mol MAEs, with the lone exception being $E_{\rm ind}$ for the electrostatically-dominated dimers described at the SAPT0(KS) + MBD level, for which the MAE is 1.5 kcal/mol. That is reduced to 0.2 kcal/mol using XSAPT + MBD, demonstrating the importance of charge embedding even in a dimer system, at least where hydrogen bonds are concerned. For the dimers that are not dominated by electrostatics, charge embedding makes little difference and the error statistics for both SAPT0(KS) + MBD and

Table 6: Error Statistics for SAPT0(KS) + MBD Applied to the S66 Data Set.

Method	Error (kcal/mol)							
	H-Bonded		Dispersion		Mixed		Total	
	MAE	Max	MAE	Max	MAE	Max	MAE	Max
def2-SVP	3.56	7.38	2.39	5.37	1.84	2.52	2.60	7.38
def2-SVPD	0.67	1.92	1.66	3.47	1.19	2.25	1.17	3.47
def2-TZVP	0.40	0.95	0.58	2.02	0.79	1.35	0.58	2.02
def2-TZVPD	0.50	1.62	0.43	1.51	0.48	0.95	0.47	1.62
def2-QZVP	0.22	0.77	0.38	1.30	0.53	1.11	0.37	1.30
def2-QZVPD	0.43	0.98	0.32	1.00	0.46	1.10	0.40	1.10
cc-pVDZ	3.13	6.90	2.45	5.37	1.89	2.61	2.52	6.90
jun-cc-pVDZ	0.81	1.96	0.72	2.06	0.67	1.32	0.74	2.06
jul-cc-pVDZ	0.34	0.92	0.31	1.10	0.40	0.91	0.35	1.10
aug-cc-pVDZ	0.19	0.81	0.77	2.02	0.68	1.45	0.54	2.02
cc-pVTZ	1.74	3.55	1.17	3.20	1.19	1.73	1.38	3.55
jun-cc-pVTZ	0.17	0.60	0.55	1.79	0.49	1.00	0.40	1.79
jul-cc-pVTZ	0.16	0.60	0.53	1.65	0.49	1.01	0.39	1.65
aug-cc-pVTZ	0.21	0.67	0.42	1.53	0.41	1.00	0.34	1.53
cc-pVQZ	0.36	0.66	0.59	2.10	0.76	1.40	0.56	2.10
jun-cc-pVQZ	0.39	1.01	0.31	1.04	0.42	0.95	0.37	1.04
jul-cc-pVQZ	0.31	1.01	0.76	2.06	0.62	1.27	0.56	2.06
aug-cc-pVQZ	0.42	0.87	0.28	1.05	0.38	0.95	0.36	1.05
6-311+G(3df,2pd)	0.56	2.05	0.86	2.41	0.79	1.41	0.73	2.41
6-311++G(3df,2pd)	0.50	1.95	0.81	2.34	0.74	1.40	0.68	2.34

Table 7: Error Statistics for Energy Components in the S22 Data Set.[a]

Method	Subset	Error (kcal/mol)							
		Electrostatics		Exchange		Induction		Dispersion	
		MAE	Max	MAE	Max	MAE	Max	MAE	Max
SAPT0(HF)+MBD	Elst.	0.72	1.57	3.59	6.10	2.00	3.68	0.77	1.51
SAPT0(HF)+MBD	Disp.	0.26	0.48	0.45	1.45	0.08	0.20	0.24	0.67
SAPT0(HF)+MBD	Mixed	0.23	0.50	0.51	1.45	0.12	0.53	0.60	1.15
SAPT0(HF)+MBD	All	0.40	1.57	1.47	6.10	0.70	3.68	0.53	1.51
SAPT0(KS)+MBD	Elst.	0.27	0.52	0.38	0.95	1.53	2.88	0.31	0.68
SAPT0(KS)+MBD	Disp.	0.25	0.59	0.60	1.50	0.10	0.33	0.44	1.00
SAPT0(KS)+MBD	Mixed	0.04	0.08	0.29	0.38	0.10	0.36	0.25	0.61
SAPT0(KS)+MBD	All	0.19	0.59	0.45	1.50	0.62	2.88	0.35	1.00
XSAPT+MBD	Elst.	0.27	0.52	0.38	0.95	0.24	0.44	0.31	0.68
XSAPT+MBD	Disp.	0.25	0.59	0.60	1.50	0.08	0.25	0.44	1.00
XSAPT+MBD	Mixed	0.04	0.08	0.29	0.38	0.04	0.08	0.25	0.61
XSAPT+MBD	All	0.19	0.59	0.45	1.50	0.12	0.44	0.35	1.00

[a]Error is relative to SAPT2+(3)/aug-cc-pVTZ benchmarks from Ref. 88.

XSAPT + MBD are very similar, for each energy component. Notably, error statistics for the induction energy evaluated using SAPT0(KS) + MBD are more similar to the SAPT0(HF) + MBD errors than they are to the XSAPT + MBD errors. Each of these methods uses the same δE_{HF} correction (evaluated in each case using HF theory and not DFT), which again supports the important role that charge embedding plays even dimers, at least where hydrogen bonds are involved. Swapping DFT for HF orbitals reduces the MAE for the hydrogen-bonded subset from 2.0 to 1.5 kcal/mol, then adding charge embedding reduces it even more, to 0.2 kcal/mol.

4.7. Hybrid Basis-Set Calculations. We now consider the possibility of a hybrid method in which different energy components are computed with different basis sets, exploiting the separability of the SAPT decomposition. A simple-to-implement version of such a procedure is to focus on the δE_{HF} correction, which requires a dimer HF calculation. This is the only part of an XSAPT calculation that requires an iterative SCF calculation on something larger than a monomer, and it is usually the computational bottleneck step when large basis sets are employed. That δE_{HF} is defined by difference (eq. 4) also suggests that this term might converge more rapidly than other energy components in eq. 1, similar to the way that a correction $\delta = E_{\mathrm{CCSD(T)}} - E_{\mathrm{MP2}}$ converges more

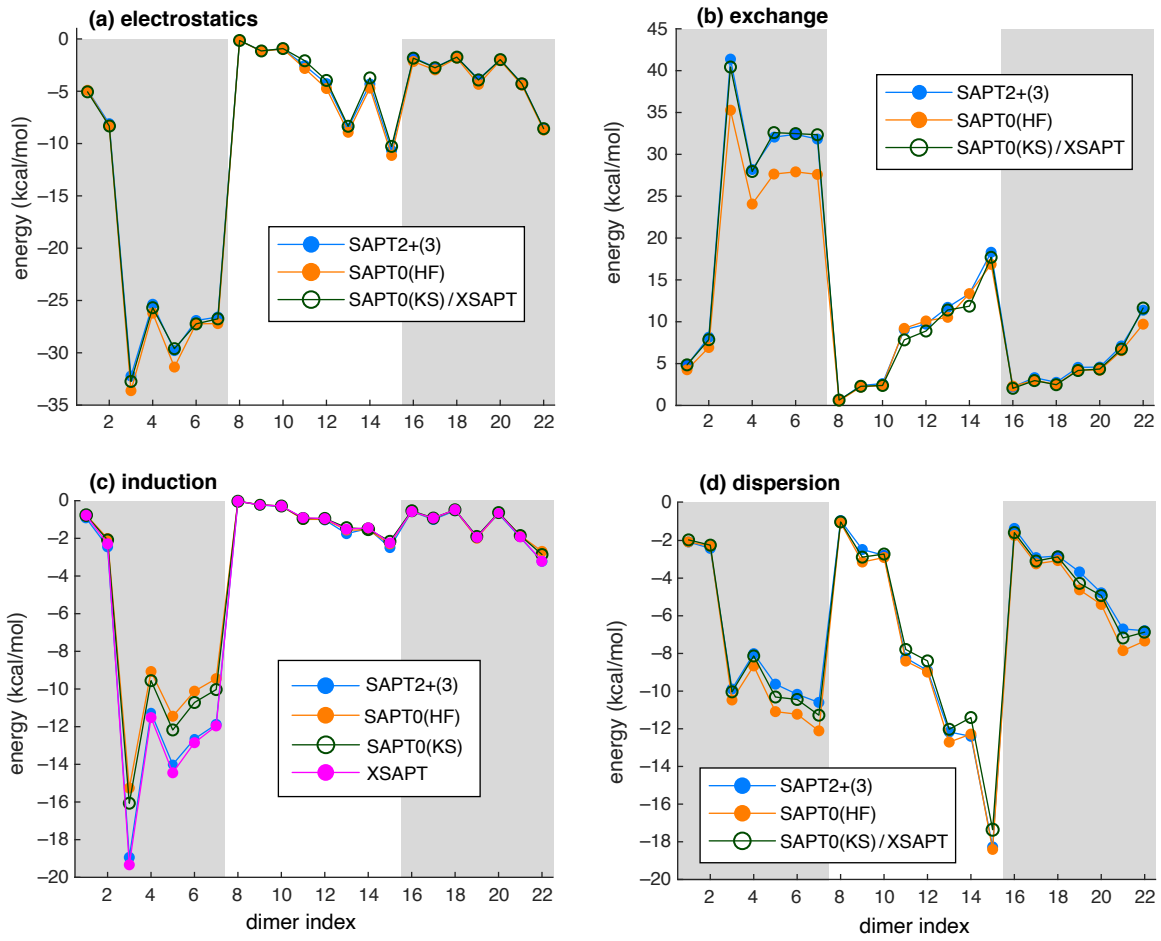

Fig. 8: Comparison of energy components of the S22 set of dimers: (a) E_{elst}, (b) E_{exch}, (c) E_{ind}, and E_{disp}. Benchmark SAPT2+(3) calculations were performed using the aug-cc-pVTZ data set and are taken from Ref. 88, whereas SAPT0(HF) + MBD, SAPT0(KS) + MBD, and XSAPT + MBD calculations were performed using the aug-cc-pVQZ basis set, and the latter two methods are indistinguishable except for E_{ind} in (c). Shaded regions delineate the electrostatics-dominated dimers (1–7), the dispersion-dominated dimers (8–15), and the dimers with mixed-influence interactions (16–22). Note that the energy scale is different in each panel.

rapidly than either $E_{\mathrm{CCSD(T)}}$ or E_{MP2}.[99]

To test this, we repeated the XSAPT + MBD calculations on the S66 dimers, using the def2-QZVPD basis set to evaluate all of the terms in eq. 1 except for δE_{HF}, for which we use a smaller basis set. The smaller basis set is used both for the supramolecular HF calculation (for $\Delta E_{\mathrm{int}}^{\mathrm{HF}}$) and to solve the coupled-perturbed equations that are used to obtain $E_{\mathrm{ind,resp}}^{(2)} + E_{\mathrm{exch\text{-}ind,resp}}^{(2)}$ in eq. 4. Table 8 shows the error statistics for this hybrid scheme, averaged over the S66 data set and using various choices for the smaller (δE_{HF}) basis set, and Fig. S2 shows the individual errors across S66 for a selection of basis sets. Even when δE_{HF} is evaluated with a basis set as small as 6-31G*, the errors (with respect to the full def2-QZVPD result) are uniformly < 1 kcal/mol. Average errors with respect to the CCSD(T)/CBS benchmarks, using def2-QZVPD for all the terms except δE_{HF}, are all ≈ 0.4 kcal/mol (see Table 8), using a variety of small basis sets to evaluate δE_{HF}.

Timing data for these hybrid calculations are shown in Fig. 9 using the largest of the S66 systems, uracil dimer. These data demonstrate the dramatic speedups that are afforded by this simple hybrid procedure. If δE_{HF} is evaluated using def2-ma-SVPD or 6-31+G(d), with all other parts of the calculation performed using def2-QZVPD, then the total cost of the calculation is reduced by 3.8× when either 6-31+G(d) or def2-ma-SVPD is substituted for def2-QZVPD to evaluate δE_{HF}. This substitution introduces an error of < 0.1 kcal/mol. The speedup can be expected to improve as system size increases, since the cost of the δE_{HF} correction scales with dimer size whereas other parts of the XSAPT procedure scale only with monomer size, meaning that the δE_{HF} calculation will constitute an increasing fraction of the total cost as the system grows larger. This will be evident in the timing data presented below for a much larger supramolec-

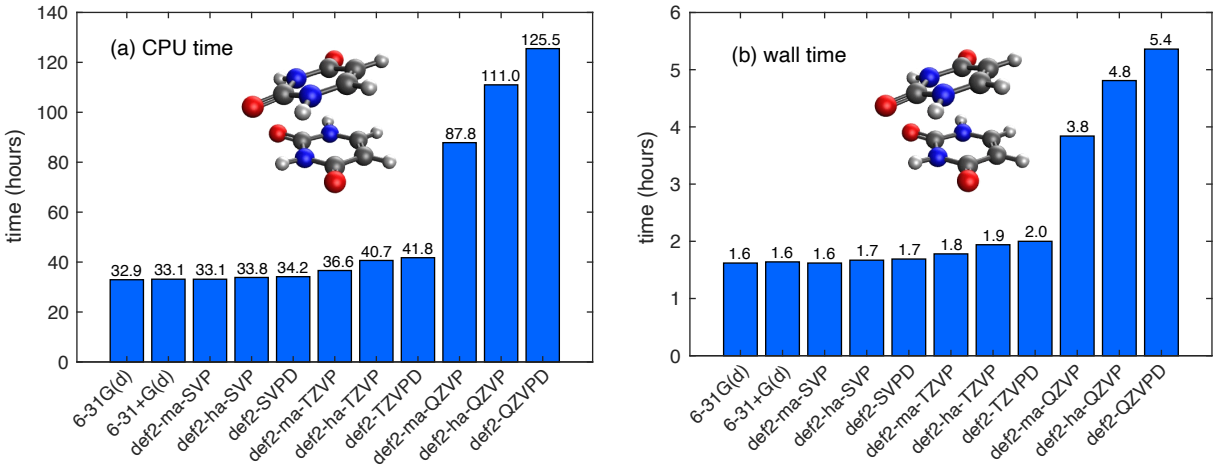

Fig. 9: (a) Total CPU time and (b) wall time for a XSAPT + MBD/def2-QZVPD calculation of uracil dimer (shown), using a hybrid approach in which the $\delta E_{\rm HF}$ correction is evaluated using a smaller basis set (as indicated). All calculations were performed on a single 28-core node and the time (in hours) is indicated at the top of each bar.

Table 8: Accuracy of XSAPT + MBD/def2-QZVPD for the S66 Data Set, Using a Smaller Basis for $\delta E_{\rm HF}$.

Basis Set for $\delta E_{\rm HF}$	MAE (kcal/mol) vs. CCSD(T)/CBS[a]	MAE (kcal/mol) vs. XSAPT + MBD/def2-QZVPD[b]
6-31G(d)	0.36	0.12
6-31+G(d)	0.37	0.06
def2-ma-SVP	0.36	0.11
def2-ha-SVP	0.38	0.07
def2-SVPD	0.38	0.07
def2-ma-TZVP	0.39	0.03
def2-ha-TZVP	0.40	0.02
def2-TZVPD	0.40	0.02
def2-ma-QZVP	0.40	0.00
def2-ha-QZVP	0.40	0.00

[a]MAE with respect to the CCSD(T)/CBS benchmark value of $E_{\rm int}$. [b]MAE with respect to conventional XSAPT + MBD/def2-QZVPD.

ular complex.

4.8. Results for Large Complexes. To complement the small dimer and cluster systems examined above, we next present two examples of XSAPT + MBD calculations on much larger systems: coronene dimer, which is one of the dispersion-bound systems in the L7 data set,[86] and a DNA–ellipticine intercalation complex[100] that has become something of a standard test system in noncovalent quantum chemistry.[20,54,56,101–103] The latter system contains 157 atoms including the ligand, two hydrogen-bonded base pairs, and the sugar/phosphate backbone. In the calculations presented below, the ellipticine ligand (33 atoms) is treated as one monomer and the entire double-stranded DNA model as a second monomer (with 124 atoms and −2 charge).

These two systems are interesting in part because there is a discrepancy in both cases between various benchmark interaction energies available in the literature. For the DNA intercalation complex, a recent CCSD(T)/CBS benchmark ($E_{\rm int}$ = 38.0±2.2 kcal/mol[103]) is inconsistent with an older quantum Monte Carlo (QMC) estimate ($E_{\rm int}$ = −33.6 ± 0.9 kcal/mol[102]), even after accounting for the reported uncertainties in both values. For coronene dimer, the QMC benchmark ($E_{\rm int}$ = −18.1 ± 0.8 kcal/mol[104]) and the CCSD(T)/CBS benchmark ($E_{\rm int}$ = −20.93 ± 0.44 kcal/mol[103]) remain 1.6 kcal/mol apart after accounting for the uncertainties, although the smaller uncertainty in the CCSD(T)/CBS value in this case means that the two values of $E_{\rm int}$ are not so far apart as they are for the DNA intercalation complex. A few other examples where QMC calculations are inconsistent with CCSD(T)/CBS values have also been reported recently,[104,105] and it is unclear to us which values are more reliable. QMC calculations are subject to an unknown fixed-node error that is not reflected in the statistical uncertainty, while the CCSD(T)/CBS benchmarks are computed using the domain-localized pair natural orbital (DLPNO) approximation.[106–108] For noncovalent interactions, this approximation is especially sensitive to numerical thresholds,[109] although this has been taken into consideration in the benchmark calculations cited above. Accuracy of the DLPNO approximation is also sensitive to system size.[110]

For these large systems, the $\delta E_{\rm HF}$ correction is the most time-consuming part of an XSAPT calculation and the mixed-basis approach introduced above stands to yield considerable savings. We therefore benchmark this approach for these larger systems, using a variety of small

Table 9: Accuracy of the Mixed-Basis δE_{HF} Approach for Large Complexes.[a]

Basis Set	(coronene)$_2$		DNA/ellipticine	
for δE_{HF}	δE_{HF}	error[b]	δE_{HF}	error[c]
6-31+G(d)	−1.82	0.24	−2.90	0.27
6-311+G(3df,2pd)	−2.04	0.03	−3.15	0.01
6-311++G(3df,2pd)	−2.04	0.03	−3.15	0.01
def2-SVP	−1.75	0.31	−2.63	0.53
def2-ma-SVP	−1.79	0.30	−2.70	0.47
def2-ha-SVP	−1.77	0.29	−2.87	0.29
def2-SVPD	−1.77	0.30	−2.87	0.29
def2-TZVP	−2.04	0.03	−3.24	0.07
def2-ma-TZVP	−2.05	0.02	−3.25	0.09
def2-ha-TZVP	−2.06	0.01	−3.28	0.12
def2-TZVPD	−2.06	0.01	−3.16	0.00
jun-cc-pVDZ	−1.99	0.08	−3.07	0.05
jul-cc-pVDZ	−2.00	0.06	−3.21	0.04
aug-cc-pVDZ	−2.00	0.07	−3.22	0.10
jun-cc-pVTZ	−2.06	0.01	−3.17	0.004
jul-cc-pVTZ	−2.06	0.00	—[d]	—[d]

[a]All values are in kcal/mol. [b]With respect to the def2-ma-QZVP value, $\delta E_{\mathrm{HF}} = -2.07$ kcal/mol. [c]With respect to the def2-TZVPD value, $\delta E_{\mathrm{HF}} = -3.16$ kcal/mol. [d]Omitted for reasons of cost.

Table 10: XSAPT + MBD Results for the DNA–Ellipticine Intercalation Complex.

Basis Set[a]	E_{int} (kcal/mol)		Time[c]
	sans δE_{HF}	with δE_{HF}[b]	(hours)
6-311+G(3df,2pd)	−43.2	−46.1	603.6
6-311++G(3df,2pd)	−42.6	−45.5	621.0
def2-SVP	−61.3	−64.2	56.6
def2-ma-SVP	−55.7	−58.6	69.3
def2-ha-SVP	−54.1	−57.0	121.4
def2-SVPD	−54.1	−57.0	143.6
def2-TZVP	−42.9	−45.8	199.4
def2-ma-TZVP	−41.1	−44.0	267.7
def2-ha-TZVP	−40.5	−43.4	586.2
def2-TZVPD	−40.4	−43.3	636.7
jun-cc-pVDZ	−43.8	−46.7	133.8
jul-cc-pVDZ	−43.3	−46.2	232.3
aug-cc-pVDZ	−43.2	−46.1	322.1
jun-cc-pVTZ	−40.3	−43.2	995.1
CCSD(T)/CBS[d]	-38.6 ± 2.2		—
QMC[e]	-33.6 ± 0.9		—

[a]Calculations performed at XSAPT + MBD level except where indicated, using the listed basis set for all parts of the calculation except δE_{HF}. [b]The δE_{HF} calculation is performed using 6-31+G(d). [c]Aggregate time on 28 processors including the time to evaluate δE_{HF}, which is 41.2 hours in each case. [d]DLPNO-CCSD(T)/CBS value from Ref. 103. [e]Fixed-node diffusion Monte Carlo value, from Ref. 102.

basis sets to evaluate δE_{HF}; see Table 9. With only one exception, errors in δE_{HF} are ≤ 0.3 kcal/mol for each of the smaller basis sets that we test, as measured with respect to the value computed in the largest basis set that is practical in either case. The one slightly larger error (of 0.5 kcal/mol) occurs when def2-SVP is used to evaluate δE_{HF} for the DNA–ellipticine complex. Given the −2 charge on the DNA backbone, this would seem to be an inappropriate choice and is included here mostly for completeness, although it is interesting that the δE_{HF} error remains rather small even in the absence of any diffuse functions at all. Based on these tests, we recommend 6-31+G(d) for the δE_{HF} correction, as the error is < 0.3 kcal/mol but this basis is much more efficient than any of the others that are listed in Table 9.

We now turn to total interaction energies for these large complexes, using 6-31+G(d) to evaluate δE_{HF}. (Although we have large-basis benchmarks for δE_{HF}, this allows us to present timing data for realistic applications using the mixed-basis procedure.) Interaction energies and timing data for DNA–ellipticine are presented in Table 10, using a variety of basis sets for the other energy components besides δE_{HF}.

Only the highest-quality basis sets that we can afford for this system, def2-TZVPD and jun-cc-pVTZ, afford interaction energies that begin to approach the CCSD(T)/CBS benchmark, although even these XSAPT + MBD values are overbound by ≈ 1.7 kcal/mol in the absence of the δE_{HF} correction. When that correction is included ($\delta E_{\mathrm{HF}} = -2.9$ kcal/mol), the XSAPT + MBD interaction energy moves even further away from the CCSD(T)/CBS value, and further still from the QMC reference value, which is less strongly bound. Nevertheless, these XSAPT + MBD values with triple-ζ basis sets and including δE_{HF} represent the highest-level XSAPT methods that have been applied to this system, and this set of calculations appears to reach a consensus value of $E_{\mathrm{int}} \approx -43.3$ kcal/mol for XSAPT + MBD in the large-basis limit. The origins of the discrepancy with respect to either the DLPNO-CCSD(T)/CBS or the QMC benchmark remains a topic for further investigation.

Analogous XSAPT + MBD results for coronene dimer are presented in Table 11. For the Karlsruhe basis sets, the difference between minimal and full augmentation is mostly insignificant (equal to 1.0 kcal/mol at the double-ζ level but much smaller in larger basis sets), whereas the difference between jun-cc-pVXZ and aug-cc-pVXZ (X = D or T) is completely insignificant. It is worth noting that jun-cc-pVXZ retains some higher-angular momentum diffuse functions beyond minimal augmentation and these appear to be adequate to afford an interaction energy that is converged with respect to inclusion of further diffuse functions. The calculations appear to converge to a value $|E_{\mathrm{int}}| = 21$–22 kcal/mol in the absence of the δE_{HF} correction, or $|E_{\mathrm{int}}| = 23$–24 kcal/mol when δE_{HF} is included. The latter value represents what is in principle the most complete version of XSAPT + MBD that

Table 11: XSAPT + MBD Results for Coronene Dimer.

Basis Set[a]	E_{int} (kcal/mol)		Time[c]
	sans δE_{HF}	with δE_{HF}[b]	(hours)
6-311+G(3df,2pd)	−25.8	−27.7	81.6
6-311++G(3df,2pd)	−25.6	−27.5	82.6
def2-ma-SVP	−31.1	−32.9	12.0
def2-ha-SVP	−32.2	−33.9	19.2
def2-SVPD	−32.1	−33.9	22.7
def2-ma-TZVP	−20.5	−22.3	50.4
def2-ha-TZVP	−20.7	−22.5	89.3
def2-TZVPD	−20.8	−22.6	106.8
def2-ma-QZVP	−21.4	−23.3	462.1
def2-ha-QZVP	−21.3	−23.2	750.3
def2-QZVPD	−21.1	−22.9	850.8
jun-cc-pVDZ	−27.2	−29.0	21.1
jul-cc-pVDZ	−26.8	−28.6	37.2
aug-cc-pVDZ	−27.2	−29.0	45.5
jun-cc-pVTZ	−22.4	−24.2	146.8
jul-cc-pVTZ	−22.3	−24.2	316.9
aug-cc-pVTZ	−22.2	−24.0	415.1
CCSD(T)/CBS[d]	-20.93 ± 0.44		—
QMC[e]	-18.1 ± 0.8		—

[a]Calculations performed at XSAPT + MBD level except where indicated, using the listed basis set for all parts of the calculation except δE_{HF}. [b]The δE_{HF} calculation is performed using 6-31+G(d). [c]Aggregate time on 40 processors including the time to evaluate δE_{HF}, which is 7.3 hours in each case. [d]DLPNO-CCSD(T)/CBS value from Ref. 103. [e]Fixed-node diffusion Monte Carlo value, from Ref. 104.

we have applied to this system, yet remains somewhat overbound with respect to the DLPNO-CCSD(T)/CBS benchmark ($|E_{\text{int}}| = 21$ kcal/mol[103]). As in the DNA–ellipticine case, the QMC benchmark is somewhat less strongly bound ($|E_{\text{int}}| = 18$ kcal/mol[104]).

For both of the large systems examined in this section, inclusion of the δE_{HF} correction worsens the agreement with the benchmarks. Reasons for this are unclear. In small-molecule benchmarks, δE_{HF} has its most important effects on hydrogen-bonded systems whereas this correction is negligible for dispersion-dominated systems.[42] Although neither $(\text{coronene})_2$ nor the DNA intercalation complex involves hydrogen bonding, the correction is 2–3 kcal/mol (and therefore not negligible) in these large systems. It is worth noting the hybrid nature of the δE_{HF} correction in XSAPT: $E^{(2)}_{\text{ind}} + E^{(2)}_{\text{ind-disp}}$ in eq. 2 is computed using DFT (tuned LRC-ωPBE functional), whereas δE_{HF} (eq. 4) is computed using HF theory. Although this mixed procedure has been carefully examined in small systems,[42] it is possible that those benchmarks misrepresent the behavior in larger systems. This remains a topic for further investigation.

Examining the timing data for these large supramolecular complexes (Tables 10 and 11), it is clear that the cost can be reduced substantially by trimming the diffuse functions. For coronene dimer, the change from def2-QZVPD to def2-ma-QZVP is accompanied by a 0.3 kcal/mol (or 1%) increase in the interaction energy and a 2× speedup. The change from def2-TZVPD to def2-ma-TZVP also changes the interaction energy by 0.3 kcal/mol and also results in a 2× speedup. The "sweet spots" for this system would appear to be def2-TZVPD or jun-cc-pVTZ, for which the interaction energies are within 1 kcal/mol of the def2-QZVPD result but the cost is less (in either case) than that associated with def2-ma-QZVP. As applied to the DNA–ellipticine complex, these same two basis sets appear to afford converged results, to the extent that we can tell without prohibitively expensive quadruple-ζ calculations.

Examining lower-cost alternatives to def2-TZVPD or jun-cc-pVTZ, it seems that def2-SVPD is wholly inadequate, even if it was able to function as a stand-in for triple-ζ basis sets in some small-molecule complexes. The jun-cc-pVDZ basis is better (and 7× faster than jun-cc-pVTZ) but overestimates the XSAPT + MBD/jun-cc-pVTZ interaction energy by 3.5 kcal/mol (or 9%) for the DNA complex and by 4.8 kcal/mol (21%) for the coronene dimer.

5 Conclusions

In this comprehensive assessment of basis-set behavior for the XSAPT, SAPT0, and SAPT0(KS) methods, we have demonstrated that the energy components E_{elst}, E_{exch}, E_{ind}, and E_{disp} converge in different ways with respect to the underlying Gaussian basis set. The XSAPT approach[20,54] replaces conventional second-order dispersion with a model that is both cheaper and more accurate: either MBD,[55,56] or else aiD3.[48,54] The former is a first-principles, density-dependent description of dispersion,[69,70] but one that substantially mitigates the basis-set dependence of E_{disp}, which for XSAPT + MBD is converged (or nearly converged) already in double-ζ basis sets. In XSAPT + aiD3, the classical dispersion potential has no basis-set dependence at all. In contrast, for conventional SAPT or MP2 calculations, dispersion exhibits the slowest convergence with respect to basis set amongst all the energy components; this was a motivating factor in the early development of XSAPT.[52]

With the basis-set dependence of E_{disp} thus rendered manageable, it is no longer clear that a compromise (or "Pauling-point") basis set such as jun-cc-pVDZ, which is often used for conventional SAPT0,[25,42] is the best choice for hybrid XSAPT-type methods. Much of the present work has therefore been dedicated to finding basis sets that afford converged results for all four energy components. A general trend across all of the energy components is that diffuse functions are important, even in systems where the monomers are charge-neutral. This

is an important observation, given that (in our experience) many computational chemists are hesitant to include diffuse functions, presumably for reasons of cost. Data presented herein make it clear that this omission has deleterious consequences for accuracy. Notably, this is true for the perturbative components of XSAPT ($E_{\rm elst}$, $E_{\rm exch}$, and $E_{\rm ind}$) that have identical forms as compared to the SAPT0 method, hence our results underscore the need for diffuse functions in conventional SAPT calculations as well.

As compared to electrostatics or dispersion, the exchange and induction energies are more sensitive to the choice of basis set and these are not converged unless triple-ζ basis sets are used. Dunning's correlation-consistent basis sets can certainly be used for this purpose, but insofar as hybrid XSAPT methods have not been designed with the intention to extrapolate to the complete-basis limit, the Dunning basis sets are more expensive than other alternatives that afford similar accuracy. Karlsruhe "def2" basis sets work well in this capacity, although we find that Pople basis sets with additional polarization functions, such as 6-311+G(3df,2pd), can also provide high-quality results, comparable to def2-TZVPD in many cases. The use of Pople basis sets is sometimes maligned in modern electronic structure theory but they are very efficient in certain quantum chemistry codes that can take advantage of *sp* shells. As applied to $Cl^-(H_2O)_{28}$, for example (with Cl^- as one monomer and the water cluster as the other), XSAPT calculations with 6-311+G(3df,2pd) afford a 2× speedup over def2-TZVPD despite the fact that the Pople basis set has 253 more functions! The 6-311+G(2df,2p) basis set affords a 5× speedup over def2-TZVPD, and both of these Pople basis sets afford interaction energies within 0.5 kcal/mol of the def2-TZVPD value.

Further speedups can be obtained by exploiting separability of the SAPT energy decomposition to evaluate different energy components at different levels of theory. A preliminary version of this idea, which is trivial to implement, is to use a smaller basis set to evaluate $\delta E_{\rm HF}$, the correction for induction effects beyond second-order perturbation theory. We find that a minimally-augmented double-ζ basis, such as 6-31+G(d) or def2-ma-SVP (the latter of which is introduced as part of the present work) is sufficient to evaluate $\delta E_{\rm HF}$, which is defined by energy difference and thus converges rapidly. For quadruple-ζ calculations, this mixed-basis procedure affords speedups of 3–4×.

That said, for dimers of charge-neutral complexes there is little reason to push XSAPT to quadruple-ζ basis sets and we find that a good catch-all basis set is def2-TZVPD, although 6-311++G(3df,2pd) also works very well for charge-neutral monomers but exhibits much larger errors for systems containing ions. For ionic systems, where induction effects are very large, best results are obtained using def2-ma-QZVP for anions and def2-QZVPD for cations, although def2-TZVPPD affords results that are only slightly less accurate, with MAEs that are about 0.4 kcal/mol larger.

The conclusions outlined above are drawn from tests on small dimers where it is possible to use basis sets of quadruple-ζ quality. Tests on two larger systems, namely, a DNA intercalation complex with 157 atoms and the coronene dimer, $(C_{24}H_{12})_2$, appear to indicate that def2-TZVPD and jun-cc-pVTZ results are essentially converged. Unlike the results for smaller systems, such as the S22 and S66 dimers where XSAPT + MBD exhibits an accuracy of $\sim$ 1 kcal/mol with respect to CCSD(T)/CBS benchmarks, in these larger systems the XSAPT + MBD value of $|E_{\rm int}|$ is 2–3 kcal/mol larger (more strongly bound) as compared to the CCSD(T)/CBS result. The latter benchmark is itself is somewhat more strongly bound than alternative QMC benchmarks. Agreement with the benchmarks improves if the $\delta E_{\rm HF}$ correction is omitted, for reasons that are unclear but which warrant further study. Given that individual XSAPT + MBD energy components for the S22 dimers are in good agreement with SAPT2+(3)/aug-cc-pVTZ benchmarks, the origin of these discrepancies for larger systems remains an open question.

Acknowledgments

This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences under Award No. DE-SC0008550. Calculations were performed at the Ohio Supercomputer Center under project no. PAA-0003.[111] J.M.H. serves on the board of directors of Q-Chem, Inc.

[1] Szalewicz, K. Determination of structure and properties of molecular crystals from first principles. *Acc. Chem. Res.* **2014,** *47,* 3266–3274.

[2] Brandenburg, J. G.; Gerit, J.; Grimme, S. Dispersion corrected Hartree-Fock and density functional theory for organic crystal structure prediction. In *Prediction and Calculation of Crystal Structures*, Vol. 345; Ş. Atahan-Evrenk,; Aspuru-Guzik, A., Eds.; Springer: Heidelberg, 2014.

[3] Beran, G. J. O. Modeling polymorphic molecular crystals with electronic structure theory. *Chem. Rev.* **2016,** *116,* 5567–5613.

[4] Beran, G. J. O.; Heit, Y. N.; Hartman, J. D. Noncovalent interactions in molecular crystals. In *Non-Covalent Interactions in Quantum Chemistry and Physics*; de la Roza, A. O.; DiLabio, G. A., Eds.; Elsevier: Amster-

dam, 2017; Chapter 10, pages 303–331.

[5] Price, S. L.; Brandenburg, J. G. Molecular crystal structure prediction. In *Non-Covalent Interactions in Quantum Chemistry and Physics*; de la Roza, A. O.; DiLabio, G. A., Eds.; Elsevier: Amsterdam, 2017; Chapter 11, pages 333–363.

[6] Plevin, M.; Bryce, D. L.; Boisbouvier, J. Direct detection of CH/π interactions in proteins. *Nat. Chem.* **2010,** *2,* 466–471.

[7] McGaughey, G. B.; Gagné, M.; Rappé, A. K. π-stacking interactions: Alive and well in proteins. *J. Biol. Chem.* **1998,** *273,* 15458–15463.

[8] Brandl, M.; Weiss, M. S.; Jabs, A.; Sühnel, J.; Hilgenfeld, R. C-H$\cdots\pi$ interactions in proteins. *J. Mol. Biol.* **2001,** *307,* 357–377.

[9] He, X.; Fusti-Molnar, L.; Cui, G.; Merz, Jr., K. M. The importance of dispersion and electron correlation in ab initio protein folding. *J. Phys. Chem. B* **2009,** *113,* 5290–5300.

[10] Xu, Z.; Zhang, Q.; Shi, J.; Zhu, W. Underestimated noncovalent interactions in protein data bank. *J. Chem. Inf. Model.* **2019,** *59,* 3389–3399.

[11] Parrish, R. M.; Sitkoff, D. F.; Cheney, D. L.; Sherrill, C. D. The surprising importance of peptide bond contacts in drug–protein interactions. *Chem. Eur. J.* **2017,** *23,* 7887–7890.

[12] Kumar, K.; Woo, S. M.; Siu, T.; Cortopassi, W. A.; Duarte, F.; Paton, R. S. Cation–π interactions in protein–ligand binding: Theory and data-mining reveal different roles for lysine and arginine. *Chem. Sci.* **2018,** *9,* 2655–2665.

[13] Sylvetsky, N. Toward simple, predictive understanding of protein-ligand interactions: Electronic structure calculations on Torpedo Californica Acetylcholinesterase join forces with the chemist's intuition. *Sci. Rep.* **2020,** *10,* 9219:1–12.

[14] Schriber, J. B.; Nascimento, D. R.; Koutsoukas, A.; Spronk, S. A.; Cheney, D. L.; Sherrill, C. D. CLIFF: A component-based, machine-learned, intermolecular force field. *J. Chem. Phys.* **2021,** *154,* 184110:1–17.

[15] Tkatchenko, A. Current understanding of van der Waals effects in realistic materials. *Adv. Funct. Mater.* **2015,** *25,* 2054–2061.

[16] Ravva, M. K.; Risko, C.; Brédas, J.-L. Noncovalent interactions in organic electronic materials. In *Non-Covalent Interactions in Quantum Chemistry and Physics*; de la Roza, A. O.; DiLabio, G. A., Eds.; Elsevier: Amsterdam, 2017; Chapter 9, pages 277–302.

[17] Cooper, V. R.; Lam, C. N.; Wang, Y.; Sumpter, B. G. Noncovalent interactions in nanotechnology. In *Non-Covalent Interactions in Quantum Chemistry and Physics*; de la Roza, A. O.; DiLabio, G. A., Eds.; Elsevier: Amsterdam, 2017; Chapter 14, pages 417–451.

[18] Wagner, J. P.; Schreiner, P. R. London dispersion in molecular chemistry—Reconsidering steric effects. *Angew. Chem. Int. Ed. Engl.* **2015,** *54,* 12274–1229.

[19] Herbert, J. M. Neat, simple, and wrong: Debunking electrostatic fallacies regarding noncovalent interactions. *J. Phys. Chem. A* **2021,** *125,* 7125–7137.

[20] Carter-Fenk, K.; Lao, K. U.; Herbert, J. M. Predicting and understanding non-covalent interactions using novel forms of symmetry-adapted perturbation theory. *Acc. Chem. Res.* **2021,** *54,* 3679–3690.

[21] Szalewicz, K.; Patkowski, K.; Jeziorski, B. Intermolecular interactions via perturbation theory: From diatoms to biomolecules. In *Intermolecular Forces and Clusters II*, Vol. 116; Wales, D. J., Ed.; Springer-Verlag: Berlin, 2005.

[22] Hohenstein, E. G.; Sherrill, C. D. Wavefunction methods for noncovalent interactions. *WIREs Comput. Mol. Sci.* **2012,** *2,* 304–326.

[23] Szalewicz, K. Symmetry-adapted perturbation theory of intermolecular forces. *WIREs Comput. Mol. Sci.* **2012,** *2,* 254–272.

[24] Jansen, G. Symmetry-adapted perturbation theory based on density functional theory for noncovalent interactions. *WIREs Comput. Mol. Sci.* **2014,** *4,* 127–144.

[25] 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.

[26] Patkowski, K. Recent developments in symmetry-adapted perturbation theory. *WIREs Comput. Mol. Sci.* **2020,** *10,* e1452:1–47.

[27] Stone, A. J. Physical basis of intermolecular interactions. In *Non-Covalent Interactions in Quantum Chemistry and Physics*; de la Roza, A. O.; DiLabio, G. A., Eds.; Elsevier: Amsterdam, 2017; Chapter 1, pages 3–26.

[28] Francisco, E.; Pendás, A. M. Energy partition analyses: Symmetry-adapted perturbation theory and other techniques. In *Non-Covalent Interactions in Quantum Chemistry and Physics*; de la Roza, A. O.; DiLabio, G. A., Eds.; Elsevier: Amsterdam, 2017; Chapter 2, pages 27–64.

[29] Misquitta, A. Intermolecular interactions. In *Handbook of Computational Chemistry*, 2nd ed.; Leszczynski, J.; Kaczmarek-Kedziera, A.; Puzyn, T.; Papadopoulos, M. G.; Reis, H.; Shukla, M. K., Eds.; Springer International Publishing: Switzerland, 2017; Chapter 8, pages 295–335.

[30] Hohenstein, E. G.; Duan, J.; Sherrill, D. C. Origin of the surprising enhancement of electrostatic energies by electron-donating substituents in substituted sandwich benzene dimers. *J. Am. Chem. Soc.* **2011,** *133,* 13244–13247.

[31] Sherrill, C. D. Energy component analysis of π interactions. *Acc. Chem. Res.* **2013,** *46,* 1020–1028.

[32] Parrish, R. M.; Sherrill, C. D. Quantum-mechanical evaluation of the π–π versus substituent–π interactions in π stacking: Direct evidence for the Wheeler–Houk picture. *J. Am. Chem. Soc.* **2014,** *136,* 17386–17389.

[33] Parrish, R. M.; Parker, T. M.; Sherrill, C. D. Chemical assignment of symmetry-adapted perturbation theory interaction energy components: The functional-group SAPT partition. *J. Chem. Theory Comput.* **2014,** *10,* 4417–4431.

[34] Carter-Fenk, K.; Herbert, J. M. Electrostatics does not dictate the slip-stacked arrangement of aromatic π–π interactions. *Chem. Sci.* **2020,** *11,* 6758–6765.

[35] Carter-Fenk, K.; Herbert, J. M. Reinterpreting π-stacking. *Phys. Chem. Chem. Phys.* **2020,** *22,* 24870–24886.

[36] Herbert, J. M.; Carter-Fenk, K. Electrostatics, charge transfer, and the nature of the halide–water hydrogen bond. *J. Phys. Chem. A* **2021,** *125,* 1243–1256.

[37] Shahbaz, M.; Szalewicz, K. Do semilocal density-functional approximations recover dispersion energies at small intermonomer separations? *Phys. Rev. Lett.* **2018,** *121,* 113402:1–6.

[38] Shahbaz, M.; Szalewicz, K. Evaluation of methods for obtaining dispersion energies used in density functional calculations of intermolecular interactions. *Theor. Chem. Acc.* **2019,** *138,* 25:1–17.

[39] Price, A. J. A.; Bryenton, K. R.; Johnson, E. R. Requirements for an accurate dispersion-corrected density functional. *J. Chem. Phys.* **2021,** *154,* 230902:1–12.

[40] Schmidt, J. R.; Yu, K.; McDaniel, J. G. Transferable next-generation force fields from simple liquids to complex materials. *Acc. Chem. Res.* **2015,** *48,* 548–556.

[41] Metz, M. P.; Szalewicz, K. Automatic generation of flexible-monomer intermolecular potential energy surfaces. *J. Chem. Theory Comput.* **2020,** *16,* 2317–2339.

[42] Gray, M.; Herbert, J. M. Simplified tuning of long-range corrected density functionals for symmetry-adapted perturbation theory. *J. Chem. Phys.* **2021,** *155,* 034103:1–8.

[43] 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.

[44] Hapka, M.; Rajchel, L.; Modrzejewski, M.; Chałasiǹski, G.; Szczęśniak, M. M. Tuned range-separated hybrid functionals in the symmetry-adapted perturbation theory. *J. Chem. Phys.* **2014,** *141,* 134120:1–10.

[45] 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.

[46] 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.

[47] Řezác, J.; Greenwell, C.; Beran, G. J. O. Accurate noncovalent interactions via dispersion-corrected second-order Møller–Plesset perturbation theory. *J. Chem. Theory Comput.* **2018,** *14,* 4711–4721.

[48] Lao, K. U.; Herbert, J. M. Atomic orbital implementation of extended symmetry-adapted perturbation theory (XSAPT) and benchmark calculations for large supramolecular complexes. *J. Chem. Theory Comput.* **2018,** *14,* 2955–2978.

[49] Nguyen, B.; Chen, G. P.; Agee, M. M.; Burow, A. M.; Tang, M.; Furche, F. Divergence of many-body perturbation theory for noncovalent interactions of large molecules. *J. Chem. Theory Comput.* **2020,** *16,* 2258–2273.

[50] 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.

[51] 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.

[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] 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: *J. Chem. Phys.* **140**, 119901 (2014).

[54] Lao, K. U.; Herbert, J. M. Accurate and efficient quantum chemistry calculations of noncovalent interactions in many-body systems: The XSAPT family of methods. *J. Phys. Chem. A* **2015,** *119,* 235–253.

[55] Carter-Fenk, K.; Lao, K. U.; Liu, K.-Y.; Herbert, J. M. Accurate and efficient *ab initio* calculations for supramolecular complexes: Symmetry-adapted perturbation theory with many-body dispersion. *J. Phys. Chem. Lett.* **2019,** *10,* 2706–2714.

[56] Liu, K.-Y.; Carter-Fenk, K.; Herbert, J. M. Self-consistent charge embedding at very low cost, with application to symmetry-adapted perturbation theory. *J. Chem. Phys.* **2019,** *151,* 031102:1–7.

[57] 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.

[58] 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.

[59] 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,* 25–58.

[60] Herbert, J. M.; Paul, S. K. Interaction energy analysis of monovalent inorganic anions in bulk water versus air/water interface. *Molecules* **2021,** *26,* 6719:1–20.

[61] Moszyński, R.; Cybulski, S. M.; Chałasiński, G. Many-body theory of intermolecular induction interactions. *J. Chem. Phys.* **1994,** *100,* 4998–5010.

[62] 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.

[63] Rohrdanz, M. A.; Herbert, J. M. Simultaneous benchmarking of ground- and excited-state properties with long-range-corrected density functional theory. *J. Chem. Phys.* **2008,** *129,* 034107:1–9.

[64] 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.

[65] Alam, B.; Morrison, A. F.; Herbert, J. M. Charge separation and charge transfer in the low-lying excited states of pentacene. *J. Phys. Chem. C* **2020,** *124,* 24653–24666.

[66] Modrzejewski, M.; Rajchel, Ł.; Chalasinski, G.; Szczesniak, M. M. Density-dependent onset of the long-range exchange: A key to donor–acceptor properties. *J. Phys. Chem. A* **2013,** *117,* 11580–11586.

[67] 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.

[68] Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge model 5: An extension of Hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases. *J. Chem.*

Theory Comput. **2012,** *8,* 527–541.

[69] Ambrosetti, A.; Reilly, A. M.; DiStasio Jr., R. A.; Tkatchenko, A. Long-range correlation energy calculated from coupled atomic response functions. *J. Chem. Phys.* **2014,** *140,* 18A508:1–14.

[70] Hermann, J.; DiStasio Jr., R. A.; Tkatchenko, A. First-principles models for van der Waals interactions in molecules and materials: Concepts, theory, and applications. *Chem. Rev.* **2017,** *117,* 4714–4758.

[71] Epifanovsky, E. *et al.* Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *J. Chem. Phys.* **2021,** *155,* 084801:1–59.

[72] 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.

[73] Carter-Fenk, K.; Mundy, C. J.; Herbert, J. M. Natural charge-transfer analysis: Eliminating spurious charge-transfer states in time-dependent density functional theory via diabatization, with application to projection-based embedding. *J. Chem. Theory Comput.* **2021,** *17,* 4195–4210.

[74] Gill, P. M. W.; Johnson, B. G.; Pople, J. A. A standard grid for density-functional calculations. *Chem. Phys. Lett.* **1993,** *209,* 506–512.

[75] Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-consistent molecular orbital methods 25. Supplementary functions for Gaussian basis sets. *J. Chem. Phys.* **1984,** *80,* 3265–3269.

[76] 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.

[77] Rappoport, D.; Furche, F. Property-optimized Gaussian basis sets for molecular response calculations. *J. Chem. Phys.* **2010,** *133,* 134105:1–11.

[78] Dunning, Jr., T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989,** *90,* 1007–1023.

[79] Kendall, R. A.; Dunning, Jr., T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992,** *96,* 6796–6806.

[80] Zheng, E. P. J.; Xu, X.; Leverentz, H. R.; Truhlar, D. G. Perspectives on basis sets beautiful: Seasonal plantings of diffuse basis functions. *J. Chem. Theory Comput.* **2011,** *7,* 3027–3034.

[81] Zheng, J.; Xu, X.; Truhlar, D. G. Minimally augmented Karlsruhe basis sets. *Theor. Chem. Acc.* **2011,** *128,* 295–305.

[82] 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. Erratum: *J. Chem. Theory Comput.* **10**, 1359–1360 (2014).

[83] Lao, K. U.; Schäffer, R.; Jansen, G.; Herbert, J. M. Accurate description of intermolecular interactions involving ions using symmetry-adapted perturbation theory. *J. Chem. Theory Comput.* **2015,** *11,* 2473–2486.

[84] Zahn, S.; MacFarlane, D. R.; Izgorodina, E. I. Assessment of Kohn–Sham density functional theory and Møller–Plesset perturbation theory for ionic liquids. *Phys. Chem. Chem. Phys.* **2013,** *15,* 13664–13675.

[85] Löwdin, P.-O. Twenty-five years of Sanibel symposia: A brief historic and scientific survey. *Int. J. Quantum Chem. Symp.* **1985,** *28,* 19–37.

[86] 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.

[87] 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.

[88] Flick, J. C.; Kosenkov, D.; Hohenstein, E. G.; Sherrill, C. D.; Slipchenko, L. V. Accurate prediction of noncovalent interaction energies with the effective fragment potential method: Comparison of energy components to symmetry-adapted perturbation theory for the S22 test set. *J. Chem. Theory Comput.* **2012,** *8,* 2835–2843. Erratum: *J. Chem. Theory Comput.* **10**, 4759–4760 (2014).

[89] Bultinck, P.; Ayers, P. W.; Fias, S.; Tiels, K.; Van Alsenoy, C. Uniqueness and basis set dependence of iterative Hirshfeld charges. *Chem. Phys. Lett.* **2007,** *444,* 205–208.

[90] Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. Critical analysis and extension of the Hirshfeld atoms in molecules. *J. Chem. Phys.* **2007,** *126,* 144111:1–9.

[91] Bultinck, P.; Cooper, D. L.; Van Neck, D. Comparison of the Hirshfeld-I and iterated stockholder atoms in molecules schemes. *Phys. Chem. Chem. Phys.* **2009,** *11,* 3424–3429.

[92] Bučko, T.; Lebègue, S.; Hafner, J.; Ángyán, J. G. Improved density dependent correction for the description of London dispersion forces. *J. Chem. Theory Comput.* **2013,** *9,* 4293–4299.

[93] Bučko, T.; Lebègue, S.; Ángyán, J. G.; Hafner, J. Extending the applicability of the Tkatchenko-Scheffler dispersion correction via iterative Hirshfeld partitioning. *J. Chem. Phys.* **2014,** *141,* 034114:1–17.

[94] Gould, T.; Lebègue, S.; Àngyán, J. G.; Bucčko, T. A fractionally ionic approach to polarizability and van der Waals many-body dispersion calculations. *J. Chem. Theory Comput.* **2016,** *12,* 5920–5930.

[95] Hermann, J.; Tkatchenko, A. Density functional model for van der Waals interactions: Unifying many-body atomic approaches with nonlocal functionals. *Phys. Rev. Lett.* **2020,** *124,* 146401:1–7.

[96] Paul, S. K.; Herbert, J. M. Probing interfacial effects on ionization energies: The surprising banality of anion–water hydrogen bonding at the air/water interface. *J. Am. Chem. Soc.* **2021,** *143,* 10189–10202.

[97] Lao, K. U.; Herbert, J. M. A simple correction for nonadditive dispersion within extended symmetry-adapted perturbation theory (XSAPT). *J. Chem. Theory Comput.* **2018,** *14,* 5128–5142.

[98] Shannon, R. D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Cryst. A* **1976,** *32,* 751–767.

[99] Marshall, M. S.; Burns, L. A.; Sherrill, C. D. Basis set convergence of the coupled-cluster correction, $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$: 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.

[100] 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.

[101] DiStasio, Jr., R. A.; von Lilienfeld, O. A.; Tkatchenko, A. Collective many-body van der Waals interactions in molecular systems. *Proc. Natl. Acad. Sci. USA* **2012,** *109,* 14791–14795.

[102] 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.

[103] Ballesteros, F.; Dunivan, S.; Lao, K. U. Coupled cluster benchmarks of large noncovalent complexes: The L7 dataset as well as DNA-ellipticine and buckycatcher-fullerene. *J. Chem. Phys.* **2021,** *154,* 154104:1–12.

[104] Al-Hamdani, Y. S.; Nagy, P. R.; Zen, A.; Barton, D.; Kállay, M.; Brandenburg, J. G.; Tkatchenko, A. Interactions between large molecules pose a puzzle for reference quantum mechanical methods. *Nat. Commun.* **2021,** *12,* 3927:1–12.

[105] Benali, A.; Shin, H.; Heinonen, O. Quantum Monte Carlo benchmarking of large noncovalent complexes in the L7 benchmark set. *J. Chem. Phys.* **2020,** *153,* 194113:1–7.

[106] Riplinger, C.; Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *J. Chem. Phys.* **2013,** *138,* 034106:1–18.

[107] Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *J. Chem. Phys.* **2013,** *139,* 134101:1–13.

[108] Sparta, M.; Neese, F. Chemical applications carried out by local pair natural orbital based coupled-cluster methods. *Chem. Soc. Rev.* **2014,** *43,* 5032–5041.

[109] Sorathia, K.; Tew, D. P. Basis set extrapolation in pair natural orbital theories. *J. Chem. Phys.* **2020,** *153,* 174112:1–8.

[110] Altun, A.; Ghosh, S.; Riplinger, C.; Neese, F.; Bistoni, G. Addressing the system-size dependence of the local approximation error in coupled-cluster calculations. *J. Phys. Chem. A* **2021,** *125,* 9932–9939.

[111] "Ohio Supercomputer Center", http://osc.edu/ark:/19495/f5s1ph73.

TOC Graphic