

Extended Symmetry-Adapted Perturbation Theory (XSAPT): A Cubic-Scaling Platform for Computing Accurate Intermolecular Interaction Energies and *Ab Initio* Energy Decomposition Analysis



John M. Herbert , Montgomery Gray , Kuan-Yu Liu ,
and Kevin Carter-Fenk

Contents

1	Intermolecular Interactions the SAPT Way	139
1.1	Symmetry-Adapted Perturbation Theory	141
1.2	XSAPT Models	146
1.3	XSAPT Energy Decomposition	159
2	Performance Benchmarks	169
2.1	Accuracy	169
2.2	Cost	178
3	Illustrative Applications	181
3.1	Basic Intermolecular Physics	182
3.2	Comparison of Dispersion Models	188
3.3	Arene Stacking Interactions	191
3.4	Charge Transfer in Anion–Water Hydrogen Bonds	201
4	Summary	203
	References	204

Abstract This chapter provides an overview of some hybrid forms of symmetry-adapted perturbation theory (SAPT), developed over the past decade and known collectively as “extended” (X)SAPT. Two primary innovations are a self-consistent

J. M. Herbert (✉) and M. Gray

Department of Chemistry & Biochemistry, The Ohio State University, Columbus, OH, USA

e-mail: herbert@chemistry.ohio-state.edu

K.-Y. Liu

Q-Chem Inc., Pleasanton, CA, USA

K. Carter-Fenk

Department of Chemistry, University of Pittsburgh, Pittsburgh, PA, USA

charge embedding scheme to capture many-body polarization (the “XPol” procedure) and the use of low-cost dispersion models as replacements for SAPT’s perturbative description of dispersion. The latter modification reduces the formal complexity to $O(N^3)$ with system size. In conjunction with a many-body dispersion model (XSAPT+MBD) or empirical dispersion potentials fitted to *ab initio* data (XSAPT+*aiD*), the hybrid procedures achieve sub-kcal/mol accuracy with respect to high-level benchmarks. XSAPT is equipped with an energy decomposition that partitions the intermolecular interaction energy into components that include electrostatics, Pauli repulsion, dispersion, and induction, the latter of which can be further separated into polarization and charge transfer. As compared to energy decomposition analyses used in density functional theory, separation of the dispersion energy in XSAPT is less ambiguous, and the energy partition agrees well with accurate third-order SAPT benchmarks. Theoretical foundations of XSAPT are reviewed, and we provide a thorough discussion of its performance in terms of both accuracy and cost. Exemplary applications are presented that illustrate how XSAPT can be used to uncover the fundamental molecular physics of intermolecular interactions.

Keywords Dispersion · Energy decomposition analysis · Intermolecular forces · Noncovalent interactions · Symmetry-adapted perturbation theory

Abbreviations

ALMO	Absolutely localized molecular orbital
BSSE	Basis-set superposition error
CBS	Complete basis set
CCSD(T)	Coupled-cluster theory with single, double, and (perturbative) triple excitations
CP	Counterpoise
CT	Charge transfer
DFT	Density functional theory
EDA	Energy decomposition analysis
HF	Hartree–Fock
HOMO	Highest occupied molecular orbital
IE	Ionization energy
KS	Kohn–Sham
LRC	Long-range corrected
LUMO	Lowest unoccupied molecular orbital
MBD	Many-body dispersion
MO	Molecular orbital
MP2	Second-order Møller–Plesset perturbation theory
SAPT	Symmetry-adapted perturbation theory
SCF	Self-consistent field
vdW	van der Waals

XC	Exchange-correlation
XSAPT	Extended symmetry-adapted perturbation theory

1 Intermolecular Interactions the SAPT Way

With the nowadays widespread deployment of density functional theory (DFT) by nonexperts, it can easily seem as if *ab initio* description of the chemical bond is a solved problem. The treatment of nonbonded interactions, however, was certainly not a solved problem in the early days of DFT [1]. Since then, DFT’s ability to describe noncovalent interactions has been improved significantly [1–4], thanks to the development of both dispersion-corrected and dispersion-inclusive exchange-correlation (XC) functionals, which we have recently reviewed [1]. The availability of low-cost methods for noncovalent interactions has spurred more detailed scrutiny of their role in chemistry and in particular the important role of dispersion interactions in chemical bonding [5–7].

When it comes to *analyzing* intermolecular interactions, however, it is less straightforward to decide how DFT interaction energies can be decomposed into meaningful components, including dispersion (E_{disp}) but also electrostatics (E_{elst}), steric repulsion (E_{steric}), polarization (E_{pol}), and charge transfer (E_{CT}):

$$\Delta E_{\text{int}} = E_{\text{elst}} + E_{\text{steric}} + E_{\text{disp}} + \underbrace{E_{\text{pol}} + E_{\text{CT}}}_{E_{\text{ind}}} . \quad (1)$$

Here, ΔE_{int} is the intermolecular interaction energy computed via the supramolecular approach, meaning

$$\Delta E_{\text{int,AB}} = E_{\text{AB}} - E_{\text{A}} - E_{\text{B}} \quad (2)$$

for a dimer $\text{A} \cdots \text{B}$. In Eq. (1), we have defined “induction” (E_{ind}) to be the sum of polarization and charge transfer (CT),

$$E_{\text{ind}} = E_{\text{pol}} + E_{\text{CT}} . \quad (3)$$

This will aid in making contact between the present work (*vide infra*) and the partition in Eq. (1) that is often used in DFT-based energy decomposition analysis (EDA). A method to separate E_{pol} from E_{CT} is described in Sect. 1.3.2.

The decomposition of ΔE_{int} in Eq. (1) is not unique, insofar as there is no quantum-mechanical operator that represents dispersion (or polarization, etc.). As a result of this ambiguity, a variety of EDA schemes have been suggested for use with DFT [8–13], and there are a few EDAs for correlated wavefunction methods as well [14–16]. However, there is also no shortage of criticism over the fundamentally ill-defined nature of such analyses [17–20], with some going so far as to suggest that for “weak” interactions ($\lesssim 5$ kcal/mol), one should not even bother [20]. We agree

with the general sentiment that EDAs are often overinterpreted (“how many kJ/mol can dance on the head of a pin?”) [1], but a 5 kcal/mol threshold for abandoning any attempt at interpretation feels too conservative to us. After all, the energy of the single hydrogen bond in $(\text{H}_2\text{O})_2$ is 3.2 kcal/mol [21], and we *should* demand that a useful EDA can discern a hydrogen bond from some other type of intermolecular interaction.

The present chapter describes an alternative approach to EDA for intermolecular interactions that is not based on DFT but rather on many-body theory and aims to compute the interaction energy directly from first principles (not by energy difference), with quantitative accuracy. This approach, which is known as symmetry-adapted perturbation theory (SAPT) [10, 22–28], has a long history in quantum chemistry that is not reviewed here. Rather, the present work examines a relatively new family of hybrid methods that are built upon the SAPT framework but substitute alternative models for dispersion in place of perturbation theory. This is done in the interest of achieving quantitative accuracy at a cost that is cubic-scaling [$O(N^3)$] with system size, comparable to the cost of DFT. These methods are known collectively as “extended” (X)SAPT, originally built upon a combination of SAPT with the “XPol” self-consistent charge embedding procedure [29–31]. XSAPT has been developed in the Herbert group since 2010 [32–43] with the goal of providing *qualitative* insight, via a well-defined EDA that is built atop a method that delivers *quantitative* accuracy. Our philosophy is that physical insight should be independent of the ever-evolving milieu of density-functional approximations.

The remainder of this section outlines the SAPT formalism (Sect. 1.1), including definitions for its energy components. Modifications that form the basis of XSAPT are discussed in Sect. 1.2. The presentation is intended to be only moderately technical, an overview with enough detail so that a potential user can make informed choices with regard to the particular setup of an XSAPT calculation. Shorter and even less technical discussions of XSAPT can be found elsewhere [37, 42].

The accuracy and performance of XSAPT are discussed in Sect. 2. Accuracy is assessed (in Sect. 2.1) by comparing to benchmark *ab initio* results for standard noncovalent datasets [44], where interaction energies are available at the CCSD(T) level extrapolated to the complete basis-set (CBS) limit. Performance benchmarks for XSAPT (Sect. 2.2) are based on a newly rewritten version of the code that exists within the Q-Chem software package [45, 46], featuring improved single-node parallel efficiency that is documented here for the first time. Finally, illustrative applications are presented in Sect. 3. After making some elementary remarks about the distinction between electrostatics and Pauli repulsion, and the failure of multipole ideas at nonbonded close-contact distances (Sect. 3.1), we demonstrate that (X)SAPT provides a much cleaner definition of dispersion as compared to DFT-based EDA schemes (Sect. 3.2). We then describe how XSAPT calculations have been used to upend the conventional understanding of π – π interactions (Sect. 3.3) and to reinforce the conventional idea that CT plays an important role in ion–molecule hydrogen bonding (Sect. 3.4).

1.1 Symmetry-Adapted Perturbation Theory

Quite apart from DFT-based EDAs, the SAPT framework is older and has its basis in many-body perturbation theory [22]. It uses isolated-monomer wavefunctions as a zeroth-order starting point and then computes the intermolecular interaction using a form of perturbation theory. For a noncovalent dimer $A \cdots B$, this amounts to a partition of the dimer's Hamiltonian according to

$$\hat{H}_{AB} = \underbrace{\hat{F}_A + \hat{F}_B}_{\hat{H}_0} + \xi \underbrace{(\hat{W}_A + \hat{W}_B)}_{\hat{W}} + \lambda \hat{V}_{AB} = \hat{H}_0 + \xi \hat{W} + \lambda \hat{V}_{AB}. \quad (4)$$

Here, $\hat{F}_X$ is a Fock operator for the monomer ($X = A$ or B) and $\hat{W}_X$ is the corresponding Møller–Plesset fluctuation potential. The quantity $\hat{V}_{AB}$ contains the intermolecular Coulomb operators for both nuclei and electrons. In atomic units where $4\pi\epsilon_0 = 1$, the latter is

$$\begin{aligned} \hat{V}_{AB} = & \sum_{A \in A} \sum_{B \in B} \frac{Z_A Z_B}{R_{AB}} - \sum_{a \in A} \sum_{b \in B} \frac{Z_B}{R_{aB}} - \sum_{A \in A} \sum_{b \in B} \frac{Z_A}{R_{Ab}} \\ & + \sum_{a \in A} \sum_{b \in B} \frac{1}{R_{ab}}. \end{aligned} \quad (5)$$

Parameters ξ and λ in Eq. (4) serve to count orders in a double perturbation expansion, with the direct product $|\psi_0^A\rangle |\psi_0^B\rangle$ as a reference state.

A caveat is that $|\psi_0^A\rangle |\psi_0^B\rangle$ is not antisymmetric with respect to exchange of electrons between A and B , since $|\psi_0^A\rangle$ and $|\psi_0^B\rangle$ were computed in isolation. The “symmetry adaptation” in SAPT is really an *antisymmetry* adaptation in which $A \leftrightarrow B$ exchange is incorporated perturbatively, via an antisymmetrizer $\hat{\mathcal{A}}_{AB}$ [47]. In practice, this means that the perturbation is $\hat{V}_{AB} \hat{\mathcal{A}}_{AB}$, albeit with some additional simplifications for evaluating matrix elements of $\hat{\mathcal{A}}_{AB}$, most notably the “single-exchange” (or “S²”) approximation [48–50]. The presence of $\hat{\mathcal{A}}_{AB}$ in the perturbation gives rise to certain exchange terms in this antisymmetrized form of Rayleigh–Schrödinger perturbation theory, in addition to the usual terms that arise in the so-called polarization expansion [51]. The latter is the perturbative approximation to the $A \cdots B$ interaction energy that is based on $\hat{V}_{AB}$ alone, disregarding $\hat{\mathcal{A}}_{AB}$. That expansion is valid for the long-range forces when the isolated monomer densities $\rho_A(\mathbf{r})$ and $\rho_B(\mathbf{r})$ do not overlap and there is no need to enforce the Pauli principle for $A \leftrightarrow B$ exchange.

SAPT is a direct perturbative expansion for the $A \cdots B$ interaction energy, which we call $E_{\text{int},AB}$ (or simply E_{int}) in order to distinguish it from the supramolecular energy-difference approach. Whereas ΔE_{int} in Eq. (2) is computed by difference and is subject to basis-set superposition error (BSSE), the SAPT intermolecular interaction energy is not. In any EDA scheme that is based on a supramolecular calculation in a finite basis set, the BSSE is a spurious contribution that must be included

somewhere. Finally, the accuracy of SAPT is not beholden to a particular choice of density functional and can be systematically improved using higher orders in perturbation theory.

Application of symmetrized Rayleigh–Schrödinger perturbation theory to the perturbations $\hat{W}$ and $\hat{V}_{AB}\hat{\mathcal{A}}_{AB}$ affords an expression for E_{int} of the form

$$E_{\text{int}} = \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} \left(E_{\text{pol}}^{(nm)} + E_{\text{exch}}^{(nm)} \right), \quad (6)$$

whose terms are explicated elsewhere [22, 47, 52–54]. Superscripts n and m in Eq. (6) indicate orders in perturbation theory, for either the intermolecular perturbation ($\hat{V}_{AB}\hat{\mathcal{A}}_{AB}$) or for monomer electron correlation ($\hat{W}$), respectively. The terms $E_{\text{pol}}^{(nm)}$ constitute the polarization expansion (valid at long range) but each is accompanied by an exchange contribution $E_{\text{exch}}^{(nm)}$, arising from the antisymmetrizer.

The simplest form of SAPT sets $m=0$ and thus does not include monomer electron correlation. Further limiting the expansion to $n \leq 2$ affords the simplest form of SAPT that includes all of the energy components identified in Eq. (1), albeit in different guises. Somewhat confusingly, this second-order (in $\hat{V}_{AB}\hat{\mathcal{A}}_{AB}$) approach is often called “SAPT0” [25], because it is zeroth order in $\hat{W}$. It is akin to a Hartree–Fock (HF) description of the monomers combined with dispersion resembling that of second-order Møller–Plesset perturbation theory (MP2). Regrouping the relevant terms in the polarization expansion, this approximation for E_{int} is

$$E_{\text{int}}^{\text{SAPT0}} = E_{\text{elst}}^{(10)} + E_{\text{exch}}^{(10)} + \underbrace{E_{\text{disp}}^{(20)} + E_{\text{exch-disp}}^{(20)}}_{E_{\text{disp}}^{(2)}} + \underbrace{E_{\text{ind}}^{(20)} + E_{\text{exch-ind}}^{(20)}}_{E_{\text{ind}}^{(2)}}. \quad (7)$$

Electrostatics and exchange repulsion appear at first order in $\hat{V}_{AB}\hat{\mathcal{A}}_{AB}$ [47], as $E_{\text{elst}}^{(10)}$ and $E_{\text{exch}}^{(10)}$, respectively. The latter is identified with Pauli or steric repulsion [43], for reasons that are discussed in Sect. 3.1. Induction and dispersion do not appear until second order, which is why the second-order approximation is the simplest reasonable form of Eq. (6). As indicated in Eq. (7), we combine the second-order dispersion and induction terms in the polarization expansion with their exchange analogues to define the total dispersion and induction energies at second order,

$$E_{\text{disp}}^{(2)} = E_{\text{disp}}^{(20)} + E_{\text{exch-disp}}^{(20)} \quad (8a)$$

$$E_{\text{ind}}^{(2)} = E_{\text{ind}}^{(20)} + E_{\text{exch-ind}}^{(20)}. \quad (8b)$$

The exchange components arise from $\hat{\mathcal{A}}_{AB}$ and reduce the attractiveness of the induction or dispersion terms from the polarization expansion, by eliminating

Pauli-forbidden contributions to that expansion. To make contact with the generic energy decomposition in Eq. (1), recall that the induction energy contains both polarization and CT (Eq. 3), for reasons that are discussed in Sect. 1.3.2 where we introduce a technique to separate these two contributions.

Second-order induction often lacks sufficient accuracy for hydrogen-bonded systems [55], so some higher-order effects are often packaged with SAPT0. Specifically, we can define “response” (or “relaxed”) versions of second-order induction,

$$E_{\text{ind,resp}}^{(2)} = E_{\text{ind,resp}}^{(20)} + E_{\text{exch-ind,resp}}^{(20)}, \quad (9)$$

obtained by solving coupled-perturbed HF equations [56]. Substituting $E_{\text{ind,resp}}^{(2)}$ in place of $E_{\text{ind}}^{(2)}$ in Eq. (7) affords the best possible treatment of induction at second order in perturbation theory, but even this may be inadequate to describe the significant induction effects associated with hydrogen bonding [57, 58]. A supramolecular HF calculation contains much of that effect as it formally sums induction to infinite order in perturbation theory, albeit at zeroth order in $\hat{W}$. As such, one may consider a correction for induction of the form

$$\delta E_{\text{HF}} = \Delta E_{\text{int}}^{\text{CP-HF}} - (E_{\text{elst}}^{(10)} + E_{\text{exch}}^{(10)} + E_{\text{ind,resp}}^{(20)} + E_{\text{exch-ind,resp}}^{(20)}) \quad (10)$$

where $\Delta E_{\text{int}}^{\text{CP-HF}}$ is the counterpoise (CP) corrected HF interaction energy, computed using the supramolecular approach (Eq. 2). The quantity defined in Eq. (10) is known as the “ δHF ” correction, and it is often added to SAPT0:

$$E_{\text{int}}^{\text{SAPT0}+\delta\text{HF}} = E_{\text{elst}}^{(10)} + E_{\text{exch}}^{(10)} + E_{\text{disp}}^{(2)} + \underbrace{E_{\text{ind,resp}}^{(20)} + E_{\text{exch-ind,resp}}^{(20)}}_{E_{\text{ind}}} + \delta E_{\text{HF}}. \quad (11)$$

Note that some authors consider δE_{HF} to be part of the definition of SAPT0 [28], but others do not [25]; in the present work, we take δE_{HF} to be an additional correction. Strictly speaking, the model defined in Eq. (11) does not simply add δE_{HF} to the model in Eq. (7), due to the replacement of the second-order induction amplitudes with their response analogues. However, it is common to refer to Eq. (11) as “SAPT0+ δHF ”, and we shall do so.

For hydrogen-bonded systems, the hybrid SAPT0+ δHF approach is sometimes necessary to obtain high accuracy, even when third-order induction terms (beyond SAPT0) are included [25, 57, 58]. However, mixing SAPT with an energy-difference approach has the potential to introduce BSSE into what is otherwise a BSSE-free formalism. As such, $\Delta E_{\text{int}}^{\text{CP-HF}}$ should always be evaluated using CP correction, as indicated by the notation. Even for self-consistent field (SCF) calculations, the CP correction is not negligible unless basis sets of at least quadruple- ζ quality are employed [59].

Substituting Eq. (10) into Eq. (11) affords

$$E_{\text{int}}^{\text{SAPT0}+\delta\text{HF}} = \Delta E_{\text{int}}^{\text{HF}} + E_{\text{disp}}^{(2)} . \quad (12)$$

This demonstrates that the SAPT0+ δ HF combination is nothing more than second-order dispersion added to the supramolecular HF interaction energy, with its infinite order but uncorrelated treatment of induction. (This is also a simple way to understand that dispersion is entirely an electron correlation effect, absent at the mean-field HF level.) For this reason, SAPT0+ δ HF and similar methods have sometimes been called “HF plus dispersion” [28, 60–62]. We avoid that terminology for SAPT0+ δ HF, however, because it is easily confused with very old approaches that augment a HF calculation with an empirical dispersion correction (e.g., of the form $-C_6/R^6$) [63–69], analogous to the manner in which contemporary DFT is typically augmented to obtain dispersion-corrected DFT+D methods [1, 2].

Along those lines, let us pause to discuss terminology. We will sometimes refer to the conventional, HF-based SAPT0 model as “SAPT0(HF),” in order to distinguish it from an analogous method that uses Kohn–Sham (KS) DFT for the monomers. We define SAPT0(KS) as second-order intermolecular perturbation theory based on a generalized description of the monomer wavefunctions [42, 55], meaning Eq. (7) with all SAPT terms evaluated using KS orbitals and one-particle energy levels. (There are important caveats concerning which XC functionals should be used, which are discussed in Sect. 1.2.2.) Ultimately, we want to eliminate $E_{\text{disp}}^{(2)}$ in favor of dispersion models that are both more accurate and less expensive (Sect. 1.2.4). Since SAPT0 implies second-order dispersion, we will refer to these hybrid models as “SAPT(KS),” without the zero. Explicitly,

$$E_{\text{int}}^{\text{SAPT(KS)}} = E_{\text{elst}}^{(10)} + E_{\text{exch}}^{(10)} + E_{\text{ind}}^{(20)} + E_{\text{exch-ind}}^{(20)} . \quad (13)$$

This is simply Eq. (7) without the dispersion or exchange-dispersion terms, but where the name suggests that KS-DFT is used to compute the SCF monomer wavefunctions. The SAPT(KS) model in Eq. (13) needs to be augmented with some form of dispersion, e.g., to obtain SAPT(KS)+D. It might or might not be augmented with a δ HF correction, which will be indicated explicitly if so. Even for SAPT(KS), we will always use HF orbitals to evaluate δE_{HF} because a supramolecular KS-DFT calculation mixes different energy components, whereas δ HF is intended as a correction for induction [55]. Thus,

$$\begin{aligned} E_{\text{int}}^{\text{SAPT(KS)}+\delta\text{HF}} &= E_{\text{elst}}^{(10),\text{KS}} + E_{\text{exch}}^{(10),\text{KS}} + E_{\text{ind,resp}}^{(20),\text{KS}} + E_{\text{exch-ind,resp}}^{(20),\text{KS}} \\ &\quad + \delta E_{\text{HF}}, \end{aligned} \quad (14)$$

where the “KS” superscripts indicate that these terms are evaluated using DFT orbitals and energy levels. This presents a potential mismatch between the KS response terms in Eq. (14) and their HF analogues that are subtracted in Eq. (10). We leave it to benchmark calculations to determine the extent to which this is a problem.

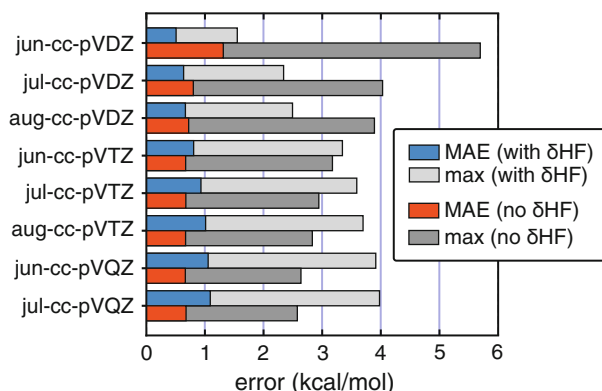


Fig. 1 Error statistics for SAPT0 (with and without the δ HF correction) in various basis sets, as applied to the S66 dataset and compared to CCSD(T)/CBS benchmarks. The “calendar” basis sets (jun- and jul-cc-pVXZ) include subsets of the diffuse functions from aug-cc-pVXZ. (Data are from Ref. [77])

Whatever the form of the monomer SCF wavefunctions, SAPT0 or SAPT0 (KS) is MP2-like in its treatment of dispersion. This lends some drawbacks because MP2 itself is not a benchmark-quality method for noncovalent interactions [26, 27, 38] and dispersion energies can be significantly overestimated [70], especially in large molecules [38, 71]. MP2 describes dispersion using uncoupled HF polarizabilities [72], but it is questionable whether any finite-order MP_n method is appropriate to describe dispersion in systems whose monomers are large and polarizable [71], e.g., due to extended conjugation. Exaggeration of dispersion interactions is exacerbated as the CBS limit is approached, as demonstrated in Fig. 1 using HF-based SAPT0 applied to the S66 dataset [73]. This is a standard database of 66 small dimers, containing ≤ 10 second-row atoms, for which CCSD(T)/CBS benchmarks are available for comparison.

Mean errors $\lesssim 1$ kcal/mol can be achieved for the S66 complexes using SAPT0+ δ HF in conjunction with basis sets up to jul-cc-pVQZ, but the outliers grow larger as the basis-set quality improves. These outliers correspond to dispersion-dominated systems such as the π -stacked uracil dimer and other π - π systems including uracil–benzene and uracil–pyridine, whose total interaction energies are relatively large by the standards of small, dispersion-dominated complexes. Interestingly, the importance of the δ HF correction diminishes in larger basis sets, and starting with jul-cc-pVTZ, both the MAE and the maximum error are smaller for SAPT0 (sans δE_{HF}) than for SAPT0+ δ HF. This behavior has not been recognized in previous comprehensive tests of wavefunction-based SAPT [26, 27], because in those studies the δ HF correction is always included with SAPT0.

The best accuracy is obtained for SAPT0+ δ HF/jun-cc-pVDZ, using a basis set that removes certain diffuse functions from aug-cc-pVDZ [74]. That basis, which is called aug-cc-pVDZ' in some older literature [75, 76], has been previously recommended for SAPT0 [26, 77]. However, the limited basis set amounts to a

form of error cancellation [1, 70], which works better in small dimers than it does in complexes with larger monomers [38]. For large-molecule vdW complexes, SAPT0 and other MP2-like methods fail badly [38, 42].

Terms beyond second order are required to obtain benchmark-quality interaction energies from the SAPT formalism. To this end, methods known as SAPT2+(3) and SAPT2+3 have been developed [25, 28], which include a subset of the possible third-order terms. For small molecules, SAPT2+(3)+ δ HF achieves sub-kcal/mol accuracy with respect to CCSD(T) [25–27, 57, 58], but both SAPT2+(3) and SAPT2+3 exhibit the same $O(N^7)$ scaling as CCSD(T). As such, these methods are prohibitively expensive for molecules with more than 10–20 non-hydrogen atoms. For larger systems, DFT-based EDAs have long been the only means available to analyze intermolecular interactions. The XSAPT family of methods was developed as a way around that bottleneck.

1.2 XSAPT Models

The most expensive terms in the SAPT0 energy expression are $E_{\text{disp}}^{(20)}$ and $E_{\text{exch-disp}}^{(20)}$, which scale as $O(N^4)$ and $O(N^5)$, respectively. Other terms including δE_{HF} scale no worse than $O(N^3)$, which is the same asymptotic scaling as DFT. Thus, what makes SAPT0 expensive is the same second-order dispersion that precludes it from being a benchmark-quality method [70–72]. Recognizing this, we have developed hybrid SAPT-based models in which $E_{\text{disp}}^{(2)}$ is replaced by a more accurate and less costly dispersion model [37, 42]. Several such models have been explored, leading to a family of methods that we call “extended” (X)SAPT [32–42]. In early work, the same approach was called “XPol + SAPT” [32–35], for reasons that will become clear in Sect. 1.2.3 when we discuss XPol-based charge embedding.

The formal scaling of XSAPT is $O(N^3)$ when evaluated in the atomic orbital basis [38, 78] and combined with density fitting [78]. Because it requires only monomer-based SCF calculations, XSAPT can sometimes be faster than supramolecular DFT [34, 37], although the δ HF correction and the many-body EDA do require calculations in line with the cost of DFT. Relative to SAPT0, however, accuracy is improved while cost is reduced [42].

1.2.1 Overview

XSAPT is based on a two-body energy decomposition of the form

$$E_{\text{int}}^{\text{XSAPT,2B}} = E_{\text{elst}} + E_{\text{exch}} + E_{\text{ind}} + \underbrace{E_{\text{disp}}}_{\text{model}}. \quad (15)$$

The superscript “2B” indicates that this is a two-body (dimer) EDA, which will be generalized to a many-body (cluster) EDA in Sect. 1.3.3. We equate the first three energy components in Eq. (15) with their SAPT0 analogues:

$$E_{\text{elst}} = E_{\text{elst}}^{(10)} \quad (16a)$$

$$E_{\text{exch}} = E_{\text{exch}}^{(10)} \quad (16b)$$

$$E_{\text{ind}} = \begin{cases} E_{\text{ind}}^{(20)} + E_{\text{exch-ind}}^{(20)} & \text{(simplest)} \\ E_{\text{ind,resp}}^{(20)} + E_{\text{exch-ind,resp}}^{(20)} & \text{(response version)} \\ E_{\text{ind,resp}}^{(20)} + E_{\text{exch-ind,resp}}^{(20)} + \delta E_{\text{HF}} & \text{(infinite-order induction)} \end{cases} \quad (16c)$$

As suggested in Eq. (16c), induction can be described using various levels of sophistication; these are listed in order of increasing accuracy and cost. By using SAPT0 energy components, electron correlation is excluded in the monomer calculations. This may be undesirable, especially for strong hydrogen bonds where monomer correlation effects can be significant [76]. Within conventional SAPT methodology, this requires terms that are first and second order in $\hat{W}$ (Eq. 4), an approach that has been called “SAPT2” and engenders $O(N^6)$ scaling, similar to CCSD but noniterative [28]. XSAPT instead uses KS-DFT for the monomers, as a low-cost means to incorporate monomer electron correlation effects. Importantly, the XC functional must have correct asymptotic behavior. This is accomplished using long-range corrected (LRC) density functionals [55, 79], as discussed in Sect. 1.2.2.

An additional consideration is how to represent the monomer wavefunctions in systems that contain more than two monomers: $A, B, C, \dots$. Our choice is to represent the SCF wavefunction on A using strictly those atomic orbital basis functions that are centered on A . Elsewhere, this has been called the “absolutely localized” molecular orbital (ALMO) basis [80], which is a starting point for a DFT-based EDA of the same name [13]. This choice is not entirely obvious from the point of view of conventional (dimer) SAPT calculations, where it is known as a monomer-centered basis, which is atypical because more accurate results are obtained using a dimer-centered basis [81]. The latter approach employs the combined basis functions of A and B to compute both $|\psi_0^A\rangle$ and $|\psi_0^B\rangle$, which allows for some intermolecular CT. For systems with more than two monomers, however, the appropriate definition for a dimer basis becomes ambiguous. For a large collection of monomers, this approach cannot be generalized without prohibitive cost.

As an alternative, XSAPT uses a monomer basis to compute the SCF wavefunctions, ensuring that the total SCF cost is $O(N)$ with respect to the number of monomers. As a means to capture some of the intermolecular CT that is lost in the

monomer-centered approach, XSAPT employs a “projected” (pseudocanonized) monomer basis to evaluate the dimer SAPT corrections [32, 33]. For the dimer AB , this means first obtaining density matrices $\mathbf{P}^A$ and $\mathbf{P}^B$ from monomer-centered SCF calculations and then building monomer Fock matrices $\mathbf{F}^A$ and $\mathbf{F}^B$ using the block-diagonal dimer density matrix $\mathbf{P}^A \oplus \mathbf{P}^B$. Each of these Fock matrices is then transformed into a pseudocanonical basis that diagonalizes the occupied–occupied and virtual–virtual blocks independently, as this does not alter the monomer densities that are invariant to such transformations. We omit the first-order (or “non-Brillouin singles”) correction, which is formally introduced into the perturbation expansion by this transformation [82]. As compared to a proper dimer basis, the use of this pseudocanonized basis results in acceptable errors in conventional pairwise SAPT0 calculations, e.g., a 10% error in the total interaction energy for $(\text{H}_2\text{O})_2$ [32]. For a system of more than two monomers, the aforementioned procedure is repeated for all pairs AB , and SAPT calculations are evaluated in a pairwise-additive manner as described in Sect. 1.3.3.

As indicated already in Eq. (16), XSAPT is not a single model but rather a family of methods [37, 42], or a platform for accurate calculation of intermolecular interactions that is also equipped with an energy decomposition scheme. By its nature, this requires some decisions on the user’s part and makes XSAPT less straightforward than a typical DFT-based EDA that simply requires the user to select an XC functional and basis set, push a button, and accept the results at face value. That final facet cannot be overemphasized, although it is overlooked by most users. XSAPT allows a user to test the impact of different approximations, with the goal of obtaining high-quality interaction energies and well-defined energy components, i.e., answers that are physically understandable. (In principle, the DFT-based approach could be tested with a variety of functionals, but few users do this, in our experience.) In our view, the added complexity of XSAPT is a virtue and not a drawback, and it is one that is shared by the more conventional and systematically improvable SAPT formalism [26, 27]. With conventional SAPT, however, the useful levels of approximation subsequent to SAPT0 come with prohibitive $O(N^7)$ cost that limits their application to very small dimers. XSAPT can be understood as an attempt to achieve the accuracy of higher-order SAPT while retaining the $O(N^3)$ complexity that is necessary to tackle large systems.

Our cynical view, honed from experience, is that any quantum chemistry method that requires physics-based decisions is too sophisticated for the average user. Nevertheless, we feel there is a need for alternatives to DFT-based EDA schemes. In our view, the latter are often used to reverse-engineer “explanations” that confirm preconceived expectations, with users shopping for an XC functional to achieve that result. This is a lousy approach in any case, but it is supremely dangerous in the context of EDAs, whose energy components do not correspond to observables and thus cannot be interrogated directly against experiment. Nevertheless, the well-defined nature of the XSAPT energy components facilitates comparison with SAPT2+(3) calculations, whose energy components may not be directly verifiable but whose total interaction energies are in good agreement with CCSD(T) results

[25–27, 57, 58], for reasons that are grounded in many-body theory [57, 83]. Such comparisons are discussed in Sect. 1.3.1.

1.2.2 Monomer SCF Calculations

We next consider how to compute individual monomer wavefunctions. Although SAPT’s double perturbation formalism is flexible enough to include monomer correlation from many-body theory, the simplest method that does so is SAPT2 [84], which exhibits $O(N^6)$ cost but does not afford benchmark-quality results for dispersion [25, 28]. As such, we restrict XSAPT to SCF wavefunctions for the monomers. These could be HF wavefunctions, or we could attempt to incorporate monomer electron correlation effects at low cost using KS-DFT, in which case the asymptotic behavior of the XC potential (v_{xc}) must be considered carefully. That discussion will occupy most of this section.

A SAPT0(KS) formalism, using wavefunction-based perturbation theory expressions for the energy components but orbitals from KS-DFT, was first attempted in the early 2000s with disappointing results [85–87]. This was ultimately attributed to incorrect asymptotic behavior in standard XC functionals [87–89]. For an electron that is well separated from its parent molecule, the asymptotic behavior of the XC potential ought to be

$$v_{xc}(r) \sim -\frac{1}{r} + \Delta_{\infty} \quad (17)$$

for large r [90, 91], where

$$\Delta_{\infty} = \text{IE} + \epsilon_{\text{HOMO}}. \quad (18)$$

Here, “IE” is the molecule’s first ionization energy, computed using a ΔSCF approach, and $\epsilon_{\text{HOMO}} < 0$ is the energy level of the highest occupied molecular orbital (HOMO). Proper asymptotic behavior (for exact KS theory) is $v_{xc}(r) \sim -1/r$, which is recovered from Eq. (17) provided that the IE theorem of DFT is satisfied [92–95], meaning that

$$\epsilon_{\text{HOMO}} = -\text{IE}. \quad (19)$$

Improper asymptotic behavior has numerous undesirable consequences for KS-DFT and in the early days of its application to molecular problems, *ad hoc* potentials were sometimes grafted onto the potential obtained from the XC energy functional, to achieve correct asymptotic behavior by construction [91, 96, 97]. For an energy functional $E_{xc}[\rho(\mathbf{r})]$, the XC potential is

$$v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})}, \quad (20)$$

but the grafting procedure spoils this relationship between energy and potential, leading to “stray” potentials that cannot be derived from any energy functional [98]. Nevertheless, the grafting procedure is still used in the SAPT(DFT) approach [23, 24, 28], in which frequency-dependent density susceptibilities for the monomers, computed from KS-DFT calculations, are used to replace second-order dispersion in the SAPT0 energy formula (Eq. 7).

A simple way to satisfy Eq. (17) is to use LRC functionals [99–101]. These start from a semilocal exchange functional E_x^{GGA} , obtained within the generalized gradient approximation (GGA), then replace the Coulomb potential with a shorter-range attenuated potential. This short-range exchange functional ($E_x^{\text{GGA,SR}}$) is then combined with long-range HF exchange ($E_x^{\text{HF,LR}}$) to afford an XC functional

$$E_{xc}^{\text{LRC}} = E_c + (1 - \alpha)E_x^{\text{GGA,SR}} + \alpha E_x^{\text{HF,SR}} + E_x^{\text{HF,LR}}. \quad (21)$$

Here, E_c is a correlation functional and α is any fraction of exact exchange that is included at short range ($E_x^{\text{HF,SR}}$). Functionals that employ range separation of this sort are generally known as range-separated hybrids, but only those that use 100% exact exchange at long range are designated as LRC functionals [102–104]. Of the standard GGAs, meta-GGAs, and hybrid functionals in common use, only LRC functionals afford $v_{xc}(r) \sim -1/r$. For XSAPT calculations, we typically use the LRC- ω PBE functional [103].

That choice alone does not guarantee that the IE condition in Eq. (19) is satisfied, and adjusting the range-separation parameter (ω) in order to satisfy that condition has been suggested as a nonempirical means to obtain a functional with correct asymptotics [99]. Alternatively, a global density-dependent (GDD) tuning procedure can be used [55], in which the length scale for range separation ($1/\omega$) is set by considering the distance between the outermost electron in a molecule and the region of localized orbitals [105]. The value of ω that is determined in this way is designed to be a good approximation to that obtained via “IE tuning,” to satisfy Eq. (19) [105, 106], yet the GDD procedure is automatable and sidesteps certain issues with IE tuning in extended systems [55, 106].

The GDD procedure is especially convenient for SAPT(KS) calculations because ω must be tuned individually for each monomer. The monomer-centric nature of SAPT means that tuned LRC-DFT maintains size consistency because dimer DFT calculations are not required. For any DFT-based EDA that requires both monomer and supramolecular calculations, the “optimal” or IE-tuning procedure generally violates size consistency [107].

Unfortunately, the admonition to maintain correct asymptotic behavior of $v_{xc}(r)$ is not always heeded. When orbitals from a semilocal or a global hybrid functional are inserted into wavefunction-based SAPT0 formulas, the too-small energy denominators (resulting from HOMO/LUMO gaps that are much smaller than those in HF

theory) cause $E_{\text{disp}}^{(2)}$ to increase unrealistically [55, 79]. This is the reason why SAPT0 (KS) was abandoned soon after the first numerical experiments in the early 2000s; it was not resuscitated until the use of optimally tuned LRC functionals became commonplace [79]. Meanwhile the SAPT0(B3LYP) method, meaning Eq. (7) with B3LYP orbitals, was the lone example of the SAPT formalism to be considered in a survey of contemporary EDA methods for protein–drug interactions [108]. From that unfair comparison, the authors concluded that DFT-based EDAs offered a better balance of properties [108], without considering the fact that SAPT0(B3LYP) exacerbates the already too-large dispersion energies predicted by SAPT0 (HF) calculations.

Indeed, direct comparison of SAPT0(KS) methods with different XC functionals demonstrates that SAPT0(B3LYP) is considerably *less* accurate than SAPT0(HF), despite the incorporation of monomer electron correlation [55]. This is illustrated in Table 1 using the S66 dataset. For the dispersion-bound subset of S66, SAPT0 (B3LYP) exhibits a maximum error of 4.5 kcal/mol versus only 1.5 kcal/mol for conventional SAPT0(HF), and the B3LYP-based approach is also less accurate for hydrogen-bonded systems, where monomer correlation is known to be important [76]. Errors for the SAPT0(BLYP) approach are larger still, as the lack of any exact exchange whatsoever leads to worse asymptotic behavior as compared to B3LYP and further exaggeration of dispersion energies as a result. In contrast, the SAPT0 (LRC- ω_{GDD} PBE) approach reduces the maximum error for the dispersion-bound complexes to 2.0 kcal/mol, although this method remains less accurate than SAPT0 (HF) for each of the standard subsets of S66 [55].

Two sets of XSAPT calculations in Table 1 use electrostatic embedding charges that are discussed in Sect. 1.2.3 and dispersion models that are discussed in Sect. 1.2.4. Once those aspects of the formalism have been described, we will present a more complete compendium of accuracy assessments, in Sect. 2.1.

Table 1 Error statistics for SAPT0(KS) methods and XSAPT methods (the latter of which include charge embedding), as applied to the S66 dataset and two subsets^a

Method	Error (kcal/mol) ^b					
	H-bonded		Disp.-bound		All S66	
	MAE ^c	Max ^d	MAE ^c	Max ^d	MAE ^c	Max ^d
SAPT0(HF)+ δ HF	0.4	0.8	0.6	1.5	0.5	1.5
SAPT0(BLYP)+ δ HF	0.8	1.5	2.6	6.1	1.5	6.1
SAPT0(B3LYP)+ δ HF	0.6	2.1	1.7	4.5	1.1	4.5
SAPT0(LRC- ω_{GDD} PBE)+ δ HF	1.0	2.3	1.0	2.0	0.8	2.7
XSAPT ^e + <i>ai</i> D3+ δ HF	0.2	0.4	0.4	1.0	0.4	1.0
XSAPT ^e +MBD+ δ HF	0.2	0.8	0.5	1.1	0.4	1.1

^a Data are from Ref. [55]. Basis set is jun-cc-pVDZ for SAPT0(KS) and def2-TZVPPD for XSAPT

^b With respect to CCSD(T)/CBS benchmarks from Ref. [73]

^c Mean absolute error

^d Maximum absolute error

^e LRC- ω_{GDD} PBE and CM5 embedding charges

1.2.3 XPol: Self-Consistent Charge Embedding

XSAPT was originally intended as a method for molecular clusters and liquids, in which many-body polarization effects would be treated self-consistently (via the XPol procedure) while resorting to a pairwise approximation for other energy components, hence, XPol + SAPT [32, 33]. That design reflects the reality that nonadditive effects are generally small, except for polarization [109]. For example, three-body (trimer) dispersion effects contribute no more than 5–6% (and typically < 2%) of the total interaction energy in small clusters composed of small molecules [110, 111]. In crystalline benzene, three-body effects contribute 0.8–0.9 kcal/mol to the lattice energy (about 7%) [112–114], with three-body dispersion as the largest contribution [114]. In contrast, three-body polarization ranges from -9 to -13 kcal/mol in the stable structures of $(\text{H}_2\text{O})_6$ [115]. In water trimers extracted from a liquid simulation, the total three-body energies range from -6 kcal/mol to $+1$ kcal/mol [116].

These polarization effects are not negligible, and to describe them at a cost that remains tractable for sizable clusters, we use the XPol method of Gao and coworkers [29–31]. XPol starts from single-monomer SCF calculations, following which a set of atomic point charges is computed from those SCF wavefunctions and subsequent monomer SCF calculations are performed in the presence of these embedding charges. This entire procedure is then iterated to self-consistency, with an outer loop over monomers and an inner loop consisting of SCF iterations on a given monomer, in the presence of point charges representing the other $N - 1$ monomers.

In more detail, and considering closed-shell fragments for convenience, the XPol energy is [33]

$$E_{\text{XPol}} = \sum_{A=1}^N \left[2 \sum_n^{\text{occ}} (\mathbf{c}_n^A)^\dagger \left(\mathbf{h}^A + \mathbf{J}^A - \frac{1}{2} \mathbf{K}^A \right) \mathbf{c}_n^A + E_{\text{nuc}}^A \right] + E_{\text{embed}}, \quad (22)$$

where the term in square brackets is the Hartree–Fock energy for monomer A expressed in terms of ALMOs $\{\mathbf{c}_n^A\}$, meaning MOs that have support only from Gaussian basis functions centered on monomer A . The quantity E_{embed} in Eq. (22) is the sum of electrostatic embedding energies from the wavefunction-derived point charges.

Self-consistent variation of E_{XPol} is accomplished via monomer Fock matrices $\mathbf{F}^A$ whose matrix elements in the atomic orbital basis $(\mu, \nu, \dots)$ are

$$F_{\mu\nu}^A = f_{\mu\nu}^A - \frac{1}{2} \sum_{B \neq A} \sum_{b \in B} q_b (\Phi_b)_{\mu\nu} + \sum_{a \in A} \frac{\partial E_{\text{embed}}}{\partial q_a} \frac{\partial q_a}{\partial P_{\mu\nu}^A}. \quad (23)$$

Here, $\mathbf{f}^A$ is the Fock matrix for isolated monomer A and

$$(\Phi_b)_{\mu\nu} = \left\langle \mu \left| \frac{1}{\|\mathbf{r} - \mathbf{R}_b\|} \right| \nu \right\rangle \quad (24)$$

is the electrostatic potential generated by the function pair $\mu\nu$ at the point $\mathbf{R}_b$, the location of nucleus $b \in B$. The integral that is implicit in Eq. (24) is over the electron coordinate $\mathbf{r}$. The second term in Eq. (23), involving $(\Phi_b)_{\mu\nu}$, comes from the electrostatic interaction between $\rho_A(\mathbf{r})$ and the embedding charges $\{q_b\}$ on monomers $B \neq A$ [117]. The final term in Eq. (23) includes the response of the embedding charges to changes in the fragment wavefunctions ($\partial q_a / \partial P_{\mu\nu}^A$) and ensures that the converged XPol energy is variational [30].

The embedding potential E_{embed} is a straightforward sum of interactions between point charges $q_a \in A$ and monomer densities $\rho_B(\mathbf{r})$, along with interactions between the charges and the nuclei, and $\partial E_{\text{embed}} / \partial q_a$ is easy to evaluate [33]. In contrast, the derivative $\partial q_a / \partial P_{\mu\nu}^A$ depends on how the embedding charges $\{q_a\}$ are determined from the SCF wavefunction for monomer A . Mulliken charges were used in the original XPol method [29–31], but we find them to be unstable when extended basis sets are employed, as are Löwdin charges [32].

A physically appealing choice is to use “ChElPG” charges derived from the molecular electrostatic potential [118]; these are the atomic charges that best reproduce the SCF electrostatic potential near the vdW surface of the molecule. The requisite derivatives $\partial q_a / \partial P_{\mu\nu}^A$ are complicated but they have been implemented [33, 119]. In recent work, however, we have taken the q_a from “Charge Model 5” (CM5) [120], which is an empirical modification of the Hirshfeld scheme [121] that is designed to reproduce molecular dipole moments. This avoids some significant computational overhead associated with evaluating $\partial q_a / \partial P_{\mu\nu}^A$ for ChElPG charges, as illustrated by timing data in Fig. 2 for a C₆₀–buckycatcher complex where the time to compute the derivatives $\partial q_a / \partial P_{\mu\nu}^A$ is reduced from 16.7 h (ChElPG) to 2.0 h (CM5) [41]. The former value represents about 35% of the total cost of the XSAPT calculation, whereas the cost to differentiate the CM5 charges is almost negligible, resembling the cost to evaluate E_{xc} on the DFT quadrature grid.

Regardless of which charges are used, the XPol SCF time greatly exceeds the time required to evaluate the SAPT corrections. This is predicated on replacing $O(N^5)$ dispersion in SAPT0 with a low-cost dispersion model, and those models are discussed next.

1.2.4 Dispersion Models

Perhaps the most important innovation in XSAPT is its hybrid dispersion model, replacing $E_{\text{disp}}^{(2)}$ in Eq. (7). This, too, has several variants that can be selected, including empirical dispersion potentials [36, 37, 42] and a version of the many-body dispersion (MBD) model [40–42] that was originally developed for DFT by

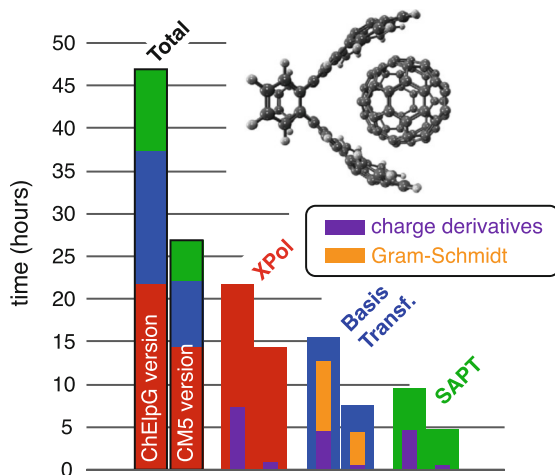


Fig. 2 Timing data for XSAPT+aiD3/hp-TZVPP calculations on a $C_{60}@C_{60}H_{28}$ complex that is shown (4592 total basis functions), running on 28 processors (Dell Intel Xeon E5-2680 v4). The total time is broken down into three color-coded steps: the XPol dual SCF procedure (in red), pseudocanonicalization to prepare the dimer basis (in blue), and the SAPT steps (in green). Charge derivatives $\partial q_a / P_{\mu\nu}^A$ are required in all three steps, and the time to evaluate them is indicated in purple. Improvements in Gram–Schmidt orthogonalization (timings in orange) reflect better parallelization beginning with Q-Chem v. 6.0. (Reprinted with permission from Ref. [41]; copyright 2019 American Institute of Physics)

Tkatchenko and coworkers [4, 122–124]. Both approaches are described in this section.

Empirical dispersion potentials take the usual atomic pairwise form [42],

$$E_{\text{disp}}^{\text{aiD}} = - \sum_{a \in A} \sum_{b \in B} \left[f_6(R_{ab}) \frac{C_{6,ab}}{R_{ab}^6} + f_8(R_{ab}) \frac{C_{8,ab}}{R_{ab}^8} \right], \quad (25)$$

written here for a dimer $A \cdots B$. The quantities f_6 and f_8 are damping functions that are needed to prevent divergence as $R \rightarrow 0$ [125]. Pairwise dispersion coefficients $C_{6,ab}$ and $C_{8,ab}$ are obtained from the combination rule

$$C_{n,ab} = (C_{n,a} C_{n,b})^{1/2} \quad (26)$$

for $n=6$ or 8 . In the context of XSAPT, we refer to Eq. (25) as an *ab initio* dispersion potential (*aiD*), because the atomic parameters $\{C_{6,a}\}$ and $\{C_{8,a}\}$ were obtained by fitting to *ab initio* dispersion energies [34–37], primarily from SAPT2+(3) calculations.

Notably, Eq. (25) has the same form as the dispersion corrections used in DFT+D [1, 2]. Several other groups have explored the idea of replacing $E_{\text{disp}}^{(2)}$ in SAPT0 with

an empirical model [126–128], which has sometimes been called “SAPT0+D” [127, 128], but we prefer to reserve the term SAPT0 to mean second-order dispersion and will instead call these models “SAPT(KS)+D”; see the discussion that accompanies Eq. (13). In terms of the physics, Grimme’s DFT+D3 and DFT+D4 dispersion corrections [129, 130] have been adapted for use with SAPT(HF) by refitting the damping parameters, leaving the C_6 and C_8 parameters unmodified [127, 128]. Unlike earlier dispersion corrections by Grimme [131], beginning with D3 the atomic dispersion coefficients are computed from DFT polarizabilities and should resemble genuine dispersion when the monomers are well separated [129]. However, these corrections *do not* resemble genuine dispersion at vdW contact distances, because the damping is significant in that range. As a result, the energy components in these methods are not reliable even if the total interaction energies may be accurate [128].

The same is true of our first-generation *aiD1* model [34], whose parameters were fit to obtain total interaction energies in agreement with CCSD(T)/CBS benchmarks. It was subsequently discovered that SAPT(KS) + *aiD1* benefits significantly from error cancellation, such that the energy *components* are not in good agreement with SAPT2+(3) benchmarks [36], and the same is true for the SAPT(HF)+D3 and SAPT(HF)+D4 models [127, 128]. These methods yield good interaction energies at the expense of significant error cancellation between the dispersion and exchange-repulsion terms [128]. Because energy decomposition is one of the primary motivations to use the SAPT framework, we believe it is crucial to achieve high accuracy not only for E_{int} but also for its components. For this reason, we do not recommend the first-generation *aiD1* model

In subsequent generations of the model, parameters were fit so that the model dispersion energy $E_{\text{disp}}^{\text{aiD}}$ reproduces *ab initio* dispersion energies obtained from third-order calculations such as SAPT2+(3) [36, 37]. The most recent *aiD3* version [37] uses an improved training set and is the recommend version. Unlike *aiD1*, these models offer consistent accuracy for E_{int} and each of its individual components, indicating that the total does not benefit from error cancellation so the energy components ought to be reliable [36, 37]. As shown in Fig. 3 using the S66 dataset, the accuracy of XSAPT+*aiD3* (which includes charge embedding) is competitive with a variety of post-SCF methods including SAPT2+(3), with a maximum error of 1.1 kcal/mol and a mean absolute error (MAE) less than 0.3 kcal/mol.

Nevertheless, the performance of the atomic pairwise dispersion approximation in Eq. (25) breaks down for large molecules [38], degrading the performance of XSAPT+*aiD* and any similar SAPT+D model when the monomers are large. This can be seen in benchmark tests on the L7 dataset [132], which consists of larger complexes with up to 73 second-row atoms (Fig. 4). Errors obtained using XSAPT+*aiD3* are much larger for L7 (up to 5 kcal/mol) as compared to those for the small S66 dimers, and a variety of dispersion-inclusive and dispersion-corrected DFT approaches are significantly more accurate. At the other end of the spectrum, MP2/CBS significantly overestimates dispersion energies for dimers involving highly conjugated molecules, such as coronene and circumcoronene in L7. The

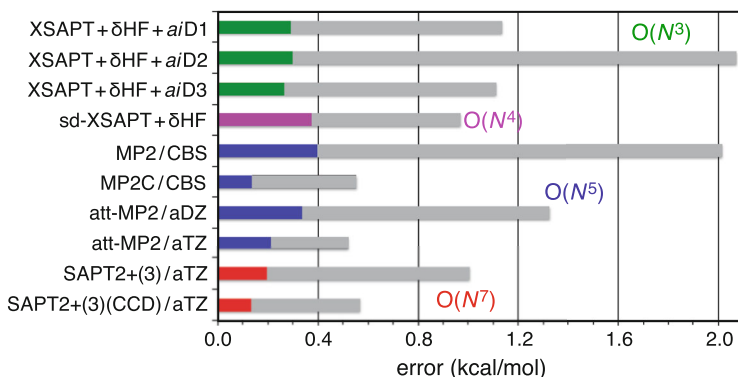


Fig. 3 Error statistics for total interaction energies in the S66 dataset, as compared to CCSD(T)/CBS benchmarks, for various post-SCF correlated methods. Colored bars indicate MAEs and gray bars are maximum absolute errors. The colors themselves suggest the formal scaling with system size, which is indicated at the right. XSAPT calculations employ ChEIPG embedding charges with LRC- ω_{IE} PBE, meaning a IE-tuned version of LRC- ω PBE, and use the following basis sets: jun-cc-pVDZ for XSAPT + *ai*D1, hpTZVPP for XSAPT+*ai*D2 and +*ai*D3, and 6-31G(d,2p) for scaled-dispersion (sd-)XSAPT. The MP2-based methods used augmented Dunning basis sets. (Adapted from Ref. [37])

only MP2-based method with reasonable accuracy for this dataset is “attenuated” (att-)MP2 [133], which removes the long-range portion of the Coulomb operator from the post-SCF calculation and whose success for noncovalent interactions likely rests on error cancellation.

The issue for XSAPT+*ai*D in large complexes is dispersion nonadditivity arising from the presence of polarizable centers that screen the $C_{6,ab}$ (and higher-order) dispersion coefficients within a molecule, relative to what would be predicted based on isolated-atom C_6 coefficients using the combination rule in Eq. (26). Dobson has called this “type-B” dispersion nonadditivity [134] and to account for it, Grimme’s D3 and D4 corrections incorporate three-body (triatomic) dispersion corrections of the Axilrod–Teller–Muto (ATM) “triple-dipole” form [3]. This is a different usage of “many-body” as compared to the discussion in Sect. 1.2.3, where the “bodies” were monomers. Here they are atoms, reflecting the atom–atom pairwise starting point in Eq. (25).

The particular version of the ATM correction that is used in Grimme’s models is [1, 129]

$$E_{\text{3B-ATM}}^{(\text{Grimme})} = - \sum_a \sum_{b>a} \sum_{c>b} \frac{g_{abc} C_{9,abc} f_{\text{3B}}(\bar{R}_{abc})}{R_{ab}^3 R_{ac}^3 R_{bc}^3} \quad (27)$$

where g_{abc} is a geometric factor and $f_{\text{3B}}(\bar{R}_{abc})$ is a damping function that depends on the mutual three-body distance, $\bar{R}_{abc}$ [38]. The requisite C_9 coefficients can be estimated from the C_6 coefficients [129],

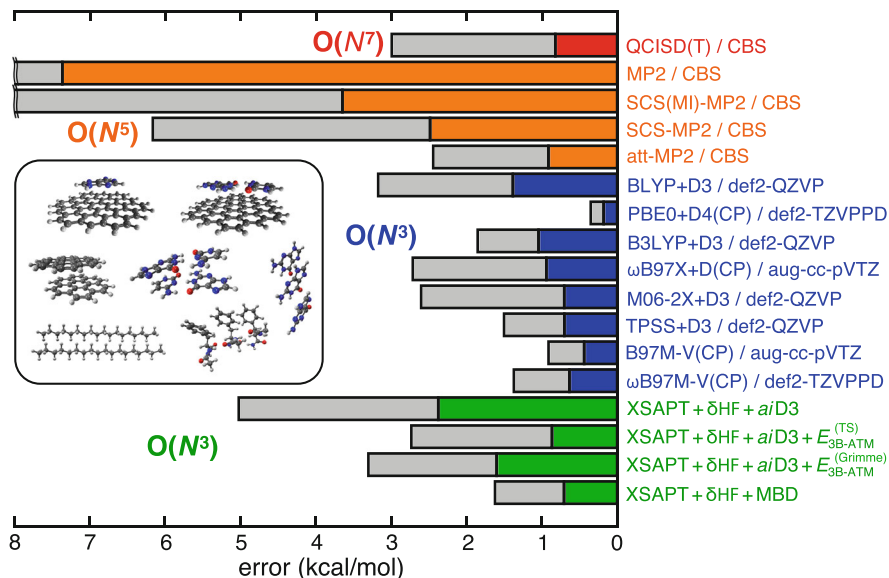


Fig. 4 MAEs (colored bars) and maximum absolute errors (gray bars) for the L7 dataset, assessed against CCSD(T)/CBS benchmarks. (The L7 complexes are shown in the inset.) Triple- ζ DFT results are CP-corrected and all XSAPT calculations use the def2-TZVPPD basis set with LRC- ω GDDPBE. Each method is color-coded according to its formal asymptotic scaling with system size. Maximum errors for MP2/CBS and SCS(MI)-MP2/CBS are both 10 kcal/mol [38]. (Adapted with permission from Ref. [42]; copyright 2021 American Chemical Society)

$$C_{9,abc} = (C_{6,ab} C_{6,bc} C_{6,ac})^{1/2}. \quad (28)$$

A version with a modified form of damping has been used by Tkatchenko and Scheffler [38, 135–137]. When either this form ($E_{3B-ATM}^{(TS)}$) or else $E_{3B-ATM}^{(Grimme)}$ in Eq. (27) is combined with XSAPT+ $aiD3$, results for the L7 complexes are greatly improved as documented in Fig. 4.

Nevertheless, the significance of nonadditive dispersion for large vdW complexes prompted a search for a less *ad hoc* way to describe this phenomenon. To that end, we have explored the use of XSAPT in conjunction with the MBD model developed for DFT by Tkatchenko and coworkers [4, 122–124]. It builds upon an earlier dispersion model developed by Tkatchenko and Scheffler [135], which is sometimes called “vdW-TS” in the condensed matter DFT literature [138]. The theory behind both MBD and its predecessor has been reviewed recently [1]. Briefly, the vdW-TS method determines an *in situ* (or “effective”) value for $C_{6,a}$ by means of a volume-scaling relationship,

$$C_{6,a}^{\text{eff}} = \left(\frac{\langle r^3 \rangle_a}{\langle r^3 \rangle_a^{\text{free}}} \right) C_{6,a}^{\text{free}}, \quad (29)$$

where $C_{6,a}^{\text{free}}$ is the isolated atom C_6 coefficient, which can be determined using time-dependent DFT calculations [139]. The scaling ratio in Eq. (29), which measures how the volume of atom a is compressed by its molecular environment, is determined by Hirshfeld partition [1, 140–142]. Pairwise coefficients $C_{6,ab}$ can be determined using a combination rule, exploiting a volume scaling analogous to Eq. (29) for the atoms-in-molecules polarizabilities $\bar{\alpha}_a^{\text{eff}}$ [1]. It should also be noted that the relationship in Eq. (29) is simply a model, based on a long-presumed relationship between polarizability and volume [143]. Recent work suggests that

$$\bar{\alpha}_a^{\text{TS}} = \left(\frac{\langle r^4 \rangle_a}{\langle r^4 \rangle_a^{\text{free}}} \right) \bar{\alpha}_a^{\text{free}} \quad (30)$$

may be the more fundamental relationship, leading to improvements in the vdW-TS method [144]. Thus, the model may be improvable.

In the MBD extension of the vdW-TS method, imaginary-frequency atom-in-molecule polarizabilities, consistent with the dispersion coefficients $C_{6,a}$ in Eq. (29), are determined according to

$$\bar{\alpha}_a^{\text{TS}}(i\omega) = \frac{\bar{\alpha}_a^{\text{free}}}{1 + (\omega/\omega_a)^2} \quad (31)$$

where the characteristic excitation frequency ω_a can be obtained from $C_{6,a}$ [1, 135]. These atomic polarizabilities could be used to compute pairwise dispersion coefficients using the Casimir–Polder relation [3, 28, 145, 146],

$$C_{6,ab} = \frac{3\hbar}{\pi} \int_0^\infty \bar{\alpha}_a(i\omega) \bar{\alpha}_b(i\omega) d\omega. \quad (32)$$

Instead, many-body dispersion effects are introduced via mutual (self-consistent) dipole screening of the atomic polarizabilities $\bar{\alpha}_a^{\text{TS}}$, which are mapped onto a $3N_{\text{atoms}}$ -dimensional quantum harmonic oscillator Hamiltonian [122, 123]. Eigenmodes of that Hamiltonian provide information about collective (multiatom) contributions to dispersion [147, 148].

This approach was originally developed as an *ad hoc*, density-dependent dispersion correction for DFT [122–124], using range separation to make sure that the model accounts for long-range dispersion only (and to avoid short-range polarization catastrophes [149]), while the XC functional is tasked with the description of short-range correlation effects including dispersion. For this reason, the model described

above has been called “MBD@rsSCS” [4], meaning “range-separated self-consistent screening,” although we will simply call it MBD.

For use with SAPT, we require a version of MBD that describes dispersion at all length scales. This will ultimately be combined with an XC functional like PBE that is largely free of dispersion (even in the short range) [1], made asymptotically correct in LRC- ω PBE form. To that end, “effectively screened” C_8 coefficients were added to the original MBD model [40], which is a dipole polarizability (C_6) model. This effectively screened dipole–quadrupole (esDQ) dispersion model augments the original MBD, and the complete dispersion model to be used with SAPT is [40, 41]

$$E_{\text{disp}}^{\text{MBD+esDQ}} = E_{\text{disp}}^{\text{MBD}} - s_8 \sum_a \sum_{b>a} f_8(R_{ab}) \frac{C_{8,ab}}{R_{ab}^8} . \quad (33)$$

Coefficients $C_{8,ab}$ can be obtained from the $\{C_{6,a}\}$ computed for the MBD model [40].

The SAPT(KS)+MBD method consists in adding the dispersion model of Eq. (33) to the nondispersion SAPT(KS) terms in Eq. (13). XSAPT+MBD augments this with XPol charge embedding and uses LRC- ω_{GDD} PBE to compute the monomer wavefunctions. Results in Fig. 5 for the S66 dimers demonstrate that the accuracy of XSAPT+MBD is very similar to that of XSAPT+*ai*D3, without any three-body dispersion corrections, since type-B dispersion nonadditivity is insignificant for these small complexes. (See Table 1 for the corresponding numerical data.) For the L7 complexes, however, XSAPT+MBD is appreciably more accurate than XSAPT+*ai*D3 even when a nonadditive dispersion correction is included in the latter (Fig. 4). At present, XSAPT+MBD is the most accurate version of XSAPT, at least for charge-neutral complexes. (Performance for ions is discussed in Sect. 2.1.4.) Additional accuracy benchmarks are provided in Sect. 2.1, but we first introduce a few more aspects of the XSAPT methodology.

1.3 XSAPT Energy Decomposition

Having established the XSAPT formalism, we next discuss its inherent EDA scheme, focusing first on the accuracy of the energy components (Sect. 1.3.1). Then, we describe how the SAPT induction energy can be separated into polarization and CT contributions (Sect. 1.3.2). Finally, we discuss how many-body effects manifest in the EDA for systems with more than two monomers (Sect. 1.3.3).

1.3.1 Comparison to High-Level SAPT

Because SAPT2+(3) affords total interaction energies in good agreement with CCSD(T) benchmarks [25–27, 57, 58], we take its energy components as

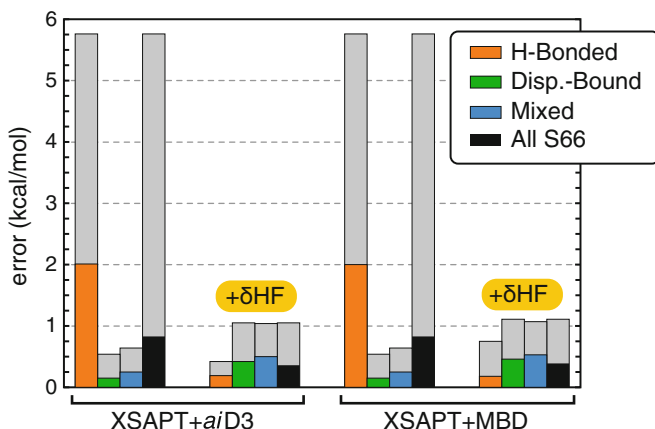


Fig. 5 Performance of XSAPT/def2-TZVPPD versus CCSD(T)/CBS benchmarks for the S66 dataset, using CM5 embedding charges. Results with and without the δ HF correction are shown separately. Colored bars indicate mean absolute errors and gray bars indicate maximum errors. (Data are from Ref. [55] and the figure is reprinted with permission from Ref. [42]; copyright 2021 American Chemical Society)

benchmarks for other EDAs [1, 77]. Figure 6 plots those benchmarks against SAPT (HF)+MBD, SAPT(KS)+MBD, and XSAPT+MBD energy components, using the S22 dataset of small-molecule dimers [150, 151]. These three models allow us to compare the effects of monomer electron correlation [SAPT(HF) versus SAPT(KS)] and charge embedding [SAPT(KS) versus XSAPT]. Results are partitioned into three standard subsets defined by the ratio of E_{elst} to E_{disp} [152], as indicated in Fig. 6. Electrostatics-dominated complexes, defined as cases where $|E_{\text{elst}}| \geq 2|E_{\text{disp}}|$, correspond to hydrogen bonding in this dataset.

Broadly speaking, there is reasonable agreement among these methods (and with the benchmarks) for E_{elst} and E_{exch} , although the SAPT(HF) method slightly exaggerates electrostatic interactions for two of the hydrogen-bonded complexes where those interactions are largest, yet still affords sub-kcal/mol accuracy (MAE = 0.7 kcal/mol). However, that method significantly underestimates exchange repulsion in many of the hydrogen-bonded dimers, and the MAE is 3.6 kcal/mol for that component with a maximum error of 6.1 kcal/mol [77]. The two methods that use LRC- ω_{GDP} PBE wavefunctions for the monomers are much more accurate for these two energy components.

Induction energies are rather small except for the hydrogen-bonded dimers. Here, XSAPT's use of charge embedding clearly leads to better accuracy as compared to the SAPT(KS)+MBD method that includes monomer correlation but not self-consistent embedding (Fig. 6c). Considering only the hydrogen-bonded dimers, the MAE for XSAPT+MBD is 0.2 kcal/mol with respect to SAPT2+(3) benchmarks versus 1.5 kcal/mol for SAPT(KS)+MBD and 2.0 kcal/mol for SAPT(HF)+MBD. This is notable, given that all three methods include the δ HF correction, yet there is still value in self-consistent embedding.

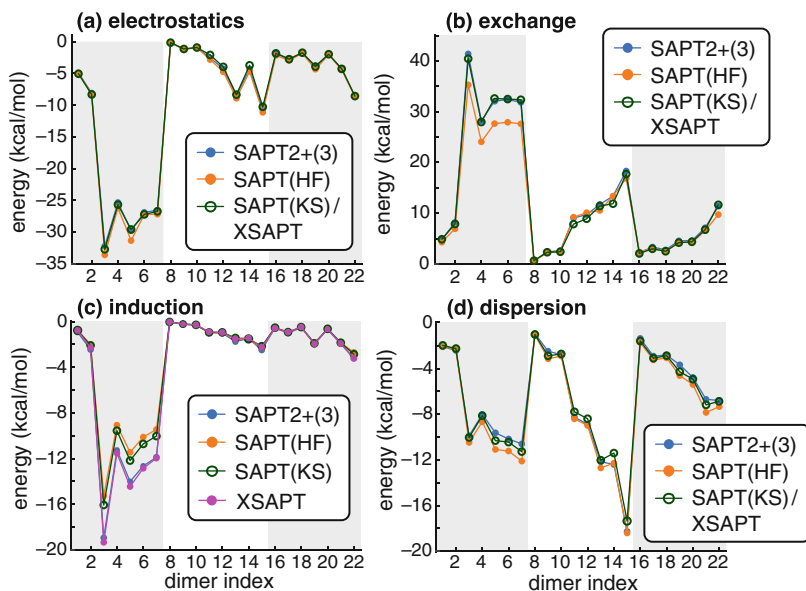


Fig. 6 Comparison of energy components for the S22 dimers: (a) E_{elst} , (b) E_{exch} , (c) E_{ind} , and (d) E_{disp} . SAPT(HF)+MBD, SAPT(KS)+MBD, and XSAPT+MBD calculations were performed using the aug-cc-pVQZ basis set and include the δHF correction, with LRC- ω_{GDD} PBE used for the latter two methods and CM5 charge embedding for XSAPT. The SAPT(KS)+MBD and XSAPT+MBD results are indistinguishable (and thus plotted together), except for E_{ind} in (c). Alternating shaded regions delineate the three standard subsets of S22: hydrogen-bonded dimers on the left, dispersion-bound dimers in the middle, and dimers with mixed influence interactions on the right. (Reproduced with permission from Ref. [77]; copyright 2022 American Chemical Society)

Finally, MBD dispersion energies are in very good agreement with the benchmarks regardless of which orbitals are used for the monomer wavefunctions (Fig. 6d). All together, this means that XSAPT+MBD reproduces all of the energy component accurately. Notably, results in Fig. 6 are converged with respect to basis set (aug-cc-pVQZ), so this accuracy does not rest on error cancellation as is the case for a method like SAPT0(HF). Basis-set convergence is systematic, as shown for E_{int} and several components in Fig. 7 using Karlsruhe basis sets. Other energy components exhibit similar or better convergence behavior, and the minimally augmented def2-ma-TZVP basis set affords converged results for all components [77]. This basis set, which was introduced in Ref. 77 as a proper subset of def2-TZVPD, differs from other minimally augmented extensions of def2-TZVP [153].

1.3.2 Separating Polarization and Charge Transfer

For chemists, the “induction energy” in the XSAPT-EDA of Eq. (15) may be unfamiliar; missing are the more familiar contributions from polarization and

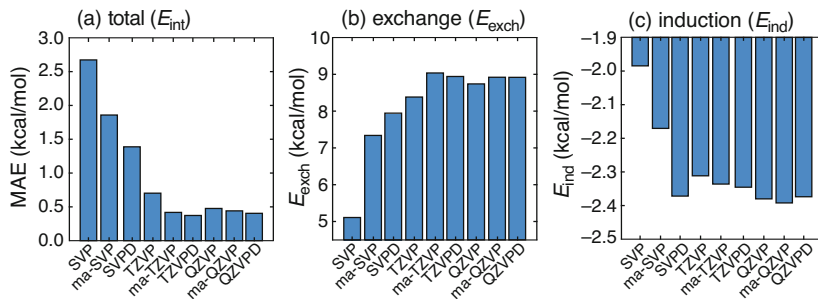


Fig. 7 Basis-set convergence of (a) E_{int} , (b) E_{exch} , and (c) E_{ind} for the S66 dimers, computed at the XSAPT+MBD+ δ HF level of theory. Data in (a) are MAEs with respect to CCSD(T)/CBS interaction energies. Data in (b) and (c) illustrate convergence of the energy components. All basis sets are Karlsruhe “def2.” (Data are from Ref. [77])

intermolecular CT. The latter has proven challenging to define uniquely [154–157], a debate that has played out recently in the context of halogen bonding [157–165]. Some authors, citing the Hellman–Feynman theorem [166], insist that all intermolecular forces—including polarization, CT, and even dispersion—are fundamentally electrostatic in origin [166–170]. This is not precisely *wrong*, insofar as they are all governed by a Coulombic Hamiltonian, but it ignores the idea of emergent complexity in many-body systems [171].

Within SAPT-based methods, both polarization and CT are contained within the induction energy as indicated in Eq. (3). Figure 8 illustrates the inherent difficulty in separating the two. Fundamentally, polarization is orbital relaxation in response to a perturbing influence, from a nearby nonbonded partner in this case. Mathematically, orbital relaxation can be cast as a unitary transformation from the isolated monomer MOs to the relaxed orbitals obtained in the presence of the partner, and such a transformation can always be written in terms of singly substituted Slater determinants expressed in the original, unrelaxed basis [172]. In a compact basis set, where the atomic orbital basis functions centered on monomer A have little overlap with those centered on B, one can readily identify the A-centered excitations that relax A’s MOs in the presence of B, distinguishing them from A $\rightarrow$ B excitations. That dichotomy could provide a basis for separating E_{pol} from E_{CT} . In larger basis sets, however, functions centered on one monomer extend significantly over the other monomer, and the distinction between orbital relaxation and CT is obscured.

As a result, orbital-based definitions of CT are notoriously unstable with respect to expansion of the atom-centered Gaussian basis set [173–175]. Within SAPT, it has been suggested to define E_{CT} as the difference between $E_{\text{ind}}^{(2)}$ computed in the monomer- and dimer-centered basis sets [176]; this has been called the second-order CT energy, $E_{\text{CT}}^{(2)}$:

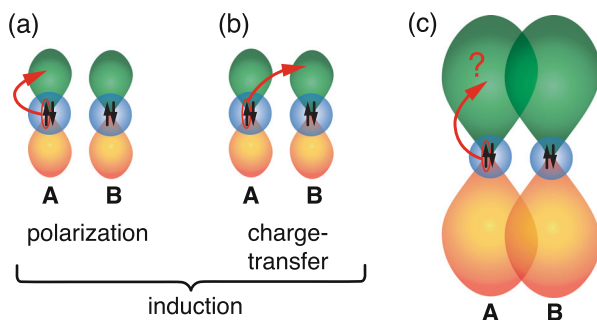


Fig. 8 Cartoon depiction of single excitations that contribute to induction. For compact atom-centered basis sets, these can be easily grouped into excitations that contribute to either (a) polarization (in this case, orbital relaxation of monomer A due to the presence of B) or else (b) CT (in this case, $A \rightarrow B$). For larger basis sets, however, the distinction is muddled, as in (c). (Adapted with permission from Ref. [173]; copyright 2016 American Chemical Society)

$$E_{\text{CT}}^{(2)} = E_{\text{ind}}^{(2)}[\text{dimer basis}] - E_{\text{ind}}^{(2)}[\text{monomer basis}]. \quad (34)$$

However, $E_{\text{CT}}^{(2)} \rightarrow 0$ in the CBS limit, where monomer and dimer basis sets should afford the same wavefunctions. A numerical demonstration is provided in Fig. 9 using the prototypical Lewis acid/base system $\text{H}_3\text{N} \cdots \text{BH}_3$ (borazane) and extending previously reported data [173] to basis sets up to aug-cc-pV7Z. Alternatively, one may isolate E_{CT} via a regularization procedure that removes the Coulomb driving force that attracts electrons on A to monomer B [174].

A simpler procedure is based on constrained (c)DFT [159, 173, 175], a general technique in which the KS equation is solved in the presence of a constraint that forces a stipulated group of atoms to integrate to a user-specified number of electrons [177]. (There is some variability in how the constraint is implemented and our approach is documented elsewhere [178, 179].) Using cDFT, one can define a CT-free reference state in which the $A \cdots B$ complex is assembled in space at the geometry of the dimer, and its KS-DFT energy is computed under the constraint that both monomers contain an integer number of electrons. Thus, orbitals are polarized but cannot transfer any (net) charge. Relaxing this constraint defines E_{CT} (as the $\text{cDFT} \rightarrow \text{DFT}$ relaxation energy), which is subtracted from a SAPT-based calculation of E_{ind} to define E_{pol} [159, 173, 175]. For $\text{H}_3\text{N} \cdots \text{BH}_3$, Fig. 9 shows that this definition is exquisitely stable with respect to basis set, with a very minor change in E_{CT} between double- and triple- ζ basis sets but no change thereafter. This is consistent with the fact that DFT densities tend to be *almost* converged in double- ζ basis sets and well converged in triple- ζ basis sets.

It is worth noting that E_{CT} as defined by the second-generation ALMO-EDA2 method [13] does appear to have a well-defined CBS limit, as a result of a new definition of E_{CT} as compared to the first-generation ALMO-EDA1 [180]. Even for the second-generation EDA, however, convergence is slow and not complete until

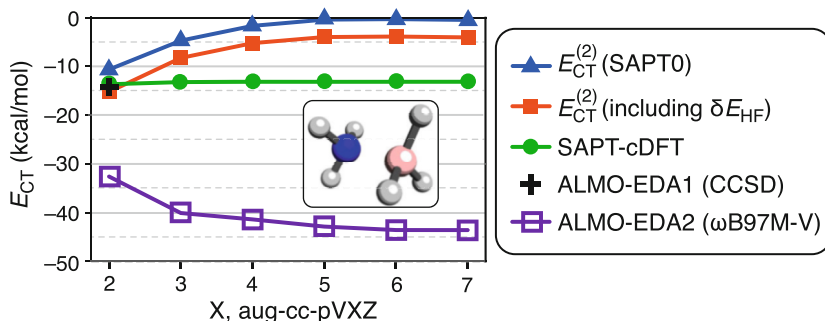


Fig. 9 Various definitions of the CT energy applied to $H_3N \cdots BH_3$ using aug-cc-pVXZ basis sets. (Angular momentum functions beyond g were removed from aug-cc-pV5Z and larger basis sets.) The quantity $E_{CT}^{(2)}$ is defined in Eq. (34) and might be evaluated with or without the δE_{HF} correction

aug-cc-pV6Z, with an 11 kcal/mol difference between double- and sextuple- ζ . That is less variability, however, as compared to ALMO-EDA1, where aug-cc-pV6Z results for the same borazane complex do not appear to have converged [173]. The SAPT-cDFT value of E_{CT} is considerably smaller than the ALMO-EDA2 value, which has been observed previously and attributed to a failure by cDFT to fully inhibit CT [156]. An alternative explanation is that the decision to use a pseudocanonicalized dimer basis for SAPT (Sect. 1.2.1) puts some intermolecular CT effects into the monomer wavefunctions [32], classified as electrostatics. In contrast, ALMO-EDA uses the eponymous, fragment-localized ALMO basis, leading to larger CT energies. This hypothesis is at least consistent with the fact that Eq. (34) affords smaller CT energies as compared to ALMO-EDA2, even in double- ζ basis sets where $|E_{CT}^{(2)}|$ is largest.

The SAPT-cDFT definition of E_{CT} affords what we believe to be a correct periodic trend for complexes $M^+ \cdots (C_6H_6)$ involving alkali cations (Fig. 10a). Specifically, we expect the magnitude of E_{CT} to decrease rapidly with ion–molecule distance, due to exponential decay of the monomer wavefunctions. For the optimized geometries that are used in Fig. 10a, that distance is largest for K^+ and smallest for Li^+ , and the SAPT-cDFT scheme predicts that $|E_{CT}(Li^+)| > |E_{CT}(Na^+)| > |E_{CT}(K^+)|$. The same cannot be said of several other EDA schemes. For this reason, we judge the cDFT-based approach to be reliable. For $M^+ \cdots (H_2O)$ complexes, the same periodic trend is not observed (using either SAPT-cDFT or the ALMO-EDA2 scheme), as $|E_{CT}(K^+)|$ is slightly larger than $|E_{CT}(Na^+)|$, although $|E_{CT}(Li^+)|$ remains larger than both. However, the magnitude of the CT energies is smaller than it is for $M^+ \cdots (C_6H_6)$ so something close to the anticipated group trend is still observed.

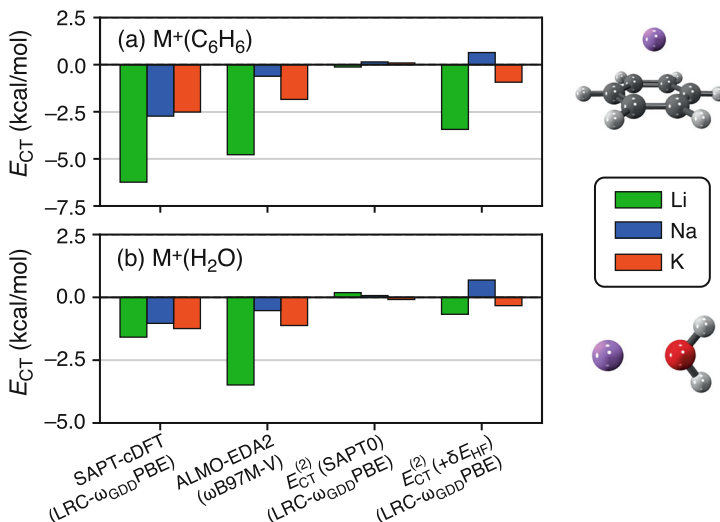


Fig. 10 CT energies for complexes of alkali cations with (a) C_6H_6 and (b) H_2O , computed according to various schemes. Geometries were optimized at the MP2/def2-SVPD level, and the ion–molecule distance increases in the order $Li^+ < Na^+ < K^+$. EDA calculations were performed using aug-cc-pVTZ for $M^+(C_6H_6)$ and aug-cc-pVQZ for $M^+(H_2O)$

1.3.3 Many-Body Systems

The XSAPT methodology was originally envisaged as “XPol + SAPT” [32–34], using a combination of XPol to describe many-body polarization in a cluster or molecular liquid and pairwise-additive SAPT0 for the remaining components of the intermolecular interaction energy [32]. Alternative dispersion models quickly became the focus [34], and there have been many applications to dimer systems. The methodology is efficient enough that a relatively large ion–water cluster, $X^-(H_2O)_n$, can be handled by treating $(H_2O)_n$ as a single monomer [77, 179], and results up to $X^-(H_2O)_{70}$ are used in Sect. 2.2.2 to establish how the cost scales with system size. Nevertheless, the methodology is formulated for many-body systems and that formalism is described next.

By definition, the XSAPT energy for a collection of N monomers is given by a pairwise sum [33, 36, 37],

$$E_{\text{int}}^{\text{XSAPT,MB}} = \sum_A \sum_{B>A}^N E_{\text{int,AB}}^{\text{XSAPT,2B}}(A, B, \dots, N), \quad (35)$$

where $E_{\text{int,AB}}^{\text{XSAPT,2B}}(A, B, \dots, N)$ is the two-body XSAPT-EDA in Eq. (15), applied to the dimer AB . The additional notation suggests that this quantity is a function of all N monomers, because the dimer SAPT calculations employ XPol monomer wavefunctions that are computed in the presence of embedding charges for the

remaining $N - 2$ monomers. As such, even the pairwise energies are many-body properties.

It is useful to rearrange Eq. (35) into the form of an EDA. This will contain some pairwise-additive parts but also a many-body contribution. To identify and isolate the latter, we need to define pairwise SAPT energy components. Unlike XSAPT components, these are based on isolated monomer wavefunctions with no embedding charges, so the total SAPT energy components are strictly pairwise-additive. Let us define these to be

$$E_{\text{elst}}^{\text{SAPT}} = \sum_A^N \sum_{B>A}^N E_{\text{elst},AB}^{\text{SAPT}}, \quad (36a)$$

$$E_{\text{exch}}^{\text{SAPT}} = \sum_A^N \sum_{B>A}^N E_{\text{exch},AB}^{\text{SAPT}}, \quad (36b)$$

etc., with analogous definitions for $E_{\text{ind}}^{\text{SAPT}}$ and $E_{\text{disp}}^{\text{SAPT}}$. For the latter, we might still use a hybrid dispersion model; the “SAPT” superscripts in Eq. (36) only signify that there is no charge embedding. Lastly, let us define

$$\delta E_{\text{HF}}^{\text{total}} = \sum_A^N \sum_{B>A}^N \delta E_{\text{HF},AB}. \quad (37)$$

For generality, the following EDA expressions will include the dimer corrections $\delta E_{\text{HF},AB}$ but these might be omitted for nonpolar systems, in the interest of cost, or evaluated using a smaller basis set as compared to the monomer wavefunctions and SAPT corrections [77].

Using the notation introduced above, a many-body generalization of the two-body XSAPT-EDA in Eq. (15) can be expressed as [36, 37]

$$E_{\text{int}}^{\text{XSAPT,MB}} = E_{\text{elst}}^{\text{SAPT}} + E_{\text{exch}}^{\text{SAPT}} + E_{\text{disp}}^{\text{SAPT}} + E_{\text{ind}}^{\text{SAPT}} + \delta E_{\text{HF}}^{\text{total}} + \Delta E_{\text{ind}}^{\text{pairwise}} + E_{\text{int}}^{\text{MB}} \quad (38)$$

where

$$\Delta E_{\text{ind}}^{\text{pairwise}} = \sum_A^N \sum_{B>A}^N [E_{\text{int},AB}^{\text{XSAPT}}(AB) - E_{\text{int},AB}^{\text{SAPT}}] \quad (39)$$

is a pairwise induction correction. Equation (38) is a trivial reformulation of Eq. (35), since the terms $-E_{\text{int},AB}^{\text{SAPT}}$ within the summation in Eq. (39) exactly cancel $E_{\text{elst}}^{\text{SAPT}} + E_{\text{exch}}^{\text{SAPT}} + E_{\text{disp}}^{\text{SAPT}} + E_{\text{ind}}^{\text{SAPT}}$. The many-body contribution in Eq. (38) can be expressed explicitly as

$$E_{\text{int}}^{\text{MB}} = \sum_A^N \sum_{B>A}^N \left[E_{\text{int},AB}^{\text{XSAPT},2\text{B}}(AB \cdots N) - E_{\text{int},AB}^{\text{XSAPT},2\text{B}}(AB) \right]. \quad (40)$$

The first term in this summand represents the AB energy computed using embedding charges for the full N -body system, whereas the second term is a self-consistently polarized dimer computed using only charges on the dimer AB . In other words, $E_{\text{int},AB}^{\text{XSAPT},2\text{B}}(AB)$ is a dimer XSAPT calculation that uses charge embedding, to be contrasted with a dimer SAPT calculation $E_{\text{int},AB}^{\text{SAPT}}$, which does not. Energy components of the latter are what appear in Eq. (36).

In practical terms, the energy decomposition in Eq. (38) requires three calculations per dimer if the δHF correction is included. One is $\delta E_{\text{HF},AB}$ in Eq. (37), while the other two are XSAPT(KS) with embedding charges and a SAPT(KS) calculation without them. It can be argued that there is some double-counting of induction effects if the pairwise $\delta E_{\text{HF},AB}$ corrections are included, as some higher-order induction is partially (though approximately) captured by the self-consistent XPol procedure [36, 37]. In the formalism developed here, the δHF corrections are held apart as a separate term, and accuracy benchmarks both with and without this correction are presented in Sect. 2.1. Note that the $\delta E_{\text{HF},AB}$ corrections are decoupled from the (X)SAPT-SCF equations, so there is no difficulty in using different basis sets to evaluate them. Previous work has shown that def2-SVPD affords similar accuracy to def2-QZVPD for the δHF correction, at a small fraction of the cost [77].

Rewriting Eq. (38) using Eqs. (36) and (39), we obtain

$$\begin{aligned} E_{\text{int}}^{\text{XSAPT},\text{MB}} &= \sum_A^N \sum_{B>A}^N \left(E_{\text{elst},AB}^{\text{SAPT}} + E_{\text{exch},AB}^{\text{SAPT}} + E_{\text{disp},AB}^{\text{SAPT}} + E_{\text{ind},AB}^{\text{SAPT}} + \delta E_{\text{HF},AB} \right) \\ &\quad + \Delta E_{\text{ind}}^{\text{pairwise}} + E_{\text{int}}^{\text{MB}}. \end{aligned} \quad (41)$$

Assuming that we use HF wavefunctions for the monomers, the definition of δE_{HF} (Eq. 10) greatly simplifies the term in parentheses in Eq. (41), leaving

$$E_{\text{int}}^{\text{XSAPT}(\text{HF}),\text{MB}} = \underbrace{\sum_A^N \sum_{B>A}^N \left(\Delta E_{\text{int},AB}^{\text{HF}} + E_{\text{disp},AB}^{\text{SAPT}} \right)}_{\text{HF+D}} + \Delta E_{\text{ind}}^{\text{pairwise}} + E_{\text{int}}^{\text{MB}}. \quad (42)$$

Here, $\Delta E_{\text{int},AB}^{\text{HF}}$ is the HF interaction energy for the dimer AB computed in the supramolecular way. Equation (42) reveals that the many-body XSAPT (HF) interaction energy can be interpreted as a form of pairwise HF+D plus additional induction, both pairwise and nonadditive.

In the case of XSAPT(KS), intramonomer correlation changes how the monomers polarize. We can write an expression similar to Eq. (42) by considering KS orbitals in Eq. (41), expressing the result as

$$E_{\text{int}}^{\text{XSAPT(KS),MB}} = \sum_A \sum_{B>A}^N \left(\Delta E_{\text{int},AB}^{\text{dlKS}} + E_{\text{disp},AB}^{\text{SAPT}} \right) + \Delta E_{\text{ind}}^{\text{pairwise}} + E_{\text{int}}^{\text{MB}} \quad (43)$$

where

$$\Delta E_{\text{int}}^{\text{dlKS}} = \Delta E_{\text{int}}^{\text{CP-HF}} + \underbrace{E_{\text{elst}}^{\text{KS}} + E_{\text{exch}}^{\text{KS}} + E_{\text{ind}}^{\text{KS}} - E_{\text{elst}}^{\text{HF}} - E_{\text{exch}}^{\text{HF}} - E_{\text{ind}}^{\text{HF}}}_{\Delta E_{\text{KS}}} . \quad (44)$$

The term ΔE_{KS} that is defined in Eq. (44) represents the difference brought about in the HF interaction energy by including electron correlation in the zeroth-order wavefunctions used for subsequent intermolecular energy corrections. This implies that the quantity $\Delta E_{\text{int}}^{\text{dlKS}}$ defined in Eq. (44) can be interpreted as a “dispersionless” KS interaction energy, which is added to SAPT(KS) dispersion and XPol polarization in Eq. (43).

The many-body EDA in Eq. (38) has been discussed in prior work [36, 37] and was used in a recent benchmark study of small-molecule trimers [181], where nonadditive dispersion and induction effects computed using XSAPT+MBD were compared against those obtained from CCSD(T) calculations and from SAPT(DFT) augmented with three-body corrections taken from wavefunction theory. Consistent with the discussion in Sect. 1.2.3, these nonadditive dispersion effects are quite small, as shown in Fig. 11a. Even so, they are better described by XSAPT+MBD than by SAPT(DFT), based on comparison to CCSD(T) benchmarks.

As a final note, we observe that second-order perturbation theory is sufficient to connect three monomers. Three-body induction couplings for XPol + SAPT0 have been derived and implemented [33] but only tested in a few cases [33, 37, 181]. Recent three-body benchmarks suggest that the XPol treatment affords an accurate treatment of nonadditive polarization for small-molecule trimers, which is degraded by the inclusion of the three-body induction couplings [181]. This is demonstrated in Fig. 11b, where XSAPT nonadditive induction energies computed *without* the three-body induction couplings are in reasonable agreement with SAPT (DFT) results whose three-body corrections are taken from wavefunction-based SAPT [181]. In contrast, XSAPT induction energies that include the three-body induction couplings are much larger. It is suggested that this overestimation likely results from neglect of the complementary nonadditive exchange-repulsion interactions [181], which have not been implemented. As such, these three-body induction couplings are not considered further, and they were not included in the many-body EDA scheme described above. See Ref. [37] for a version of the XSAPT-EDA that does include those couplings.

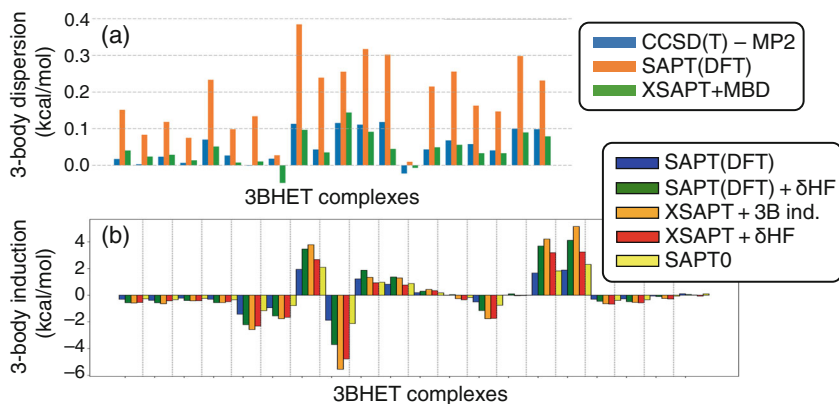


Fig. 11 Three-body results for a small-molecule dataset of trimers (“3BHET”), including (a) dispersion energies and (b) induction energies. XSAPT+MBD and SAPT(DFT) calculations employ the aug-cc-pVTZ basis set, whereas CCSD(T) and MP2 results are extrapolated to the CBS limit. The dataset consists of trimers of molecules ranging in size up to phenol and bromobenzene, along with some ions (NH_4^+ and Cl^-). (Adapted with permission from Ref. [181]; copyright 2023 the PCCP Owner Societies)

2 Performance Benchmarks

Throughout Sect. 1.2, we introduced data documenting the accuracy of various approximations, as a way to motivate the historical development of XSAPT. With the complete formalism now assembled, additional accuracy benchmarks are provided in Sect. 2.1. In Sect. 2.2, we document the cost and effective computational scaling with system size.

2.1 Accuracy

To provide a realistic overview of a family of XSAPT methods, we divide this discussion into several parts. Sect. 2.1.1 compares different XC functionals used in SAPT0(KS) and SAPT(KS) methods. This choice has important consequences for accuracy, and some poor choices that should not be emulated can nevertheless be found in the literature. Basis-set demands are discussed in Sect. 2.1.2, and because they are relatively forgiving (by *ab initio* standards), calculations can be performed in systems composed of large monomers. This capability is illustrated with examples in Sect. 2.1.3. Finally, Sect. 2.1.4 provides an overview of XSAPT for ions, which is one place where there is certainly room for future improvement.

2.1.1 Flavors of (X)SAPT

We first compare different versions of SAPT0(KS), meaning second-order perturbation theory with various SCF methods to generate the monomer wavefunctions. We use error statistics versus CCSD(T)/CBS interaction energies, for the S66 dataset [73], as a means to evaluate different approaches. Figure 12 summarizes results for four different SCF functionals and two basis sets.

Results in Fig. 12a use the jun-cc-pVDZ basis set that is found to be a good compromise for SAPT0(HF) [26, 27, 77]. This basis is a proper subset of aug-cc-pVDZ, eliminating diffuse functions on hydrogen along with the highest angular momentum set of diffuse functions on other atoms [74]. Its success for SAPT0 (HF) calculations (and for MP2) rests on error cancellation [27, 75], as is evident from the fact that both the MAEs and the maximum errors increase when aug-cc-pVTZ is used instead (Fig. 12b). Nevertheless, MAEs are no larger than 1.1 kcal/mol for each of the traditional subsets of S66 including hydrogen-bonded dimers, dispersion-dominated complexes, and dimers with mixed influence interactions. Interestingly, opposite basis-set behavior is obtained with and without the δ HF correction. At the SAPT0(HF)/jun-cc-pVDZ level, the errors for hydrogen-bonded

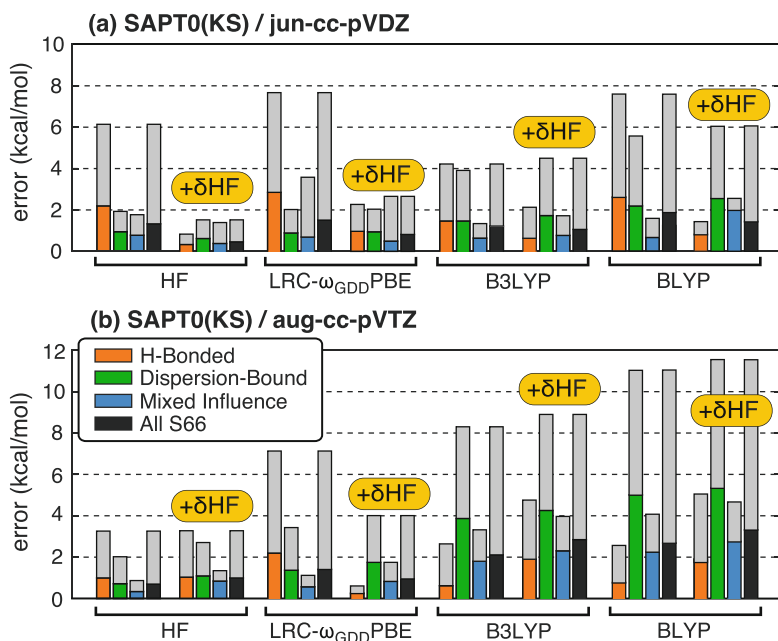


Fig. 12 Performance of second-order perturbation theory for interaction energies in the S66 dataset and its subsets, using either (a) the jun-cc-pVDZ basis set or (b) the aug-cc-pVTZ basis set. SAPT0 (KS) results are compared using various SCF methods, as indicated, either with or without the δ HF correction. Colored bars indicate MAEs with respect to CCSD(T)/CBS benchmarks and gray bars are maximum absolute errors. (Data are from Ref. [55])

dimers are unacceptably large (up to 6 kcal/mol) when this correction is omitted, yet for SAPT0(HF)/aug-cc-pVTZ calculations, the δHF correction has little statistical effect. This is a curious result that went unnoticed in comprehensive benchmarking of various levels of SAPT approximations [26, 27], because the δHF correction was always included at the SAPT0 level. This observation points to an additional source of error cancellation.

Interestingly, error statistics for S66 are not improved, and are actually slightly worsened, by the use of LRC- ω_{GDD} PBE for the monomers. The δHF correction remains essential, and results using jun-cc-pVDZ are reasonable when it is included. Differences with respect of HF-based SAPT0 may lie in the imbalance between $E_{\text{ind}}^{(2)}$ based on KS orbitals and the infinite-order induction correction that is contained in δE_{HF} , which subtracts out second-order corrections based on HF theory (Eq. 10). This is done to avoid contaminating E_{ind} with other energy components that are present in a supramolecular KS-DFT calculation [55].

While SAPT0(LRC- ω_{GDD} PBE)/jun-cc-pVDZ errors are perhaps acceptable, dispersion-dominated outliers are larger when the aug-cc-pVTZ basis set is used; see Fig. 12b. For second-order SAPT0 calculations, there seems to be no reason to recommend using a SCF method other than HF theory. That is especially true when XC functionals lacking correct asymptotic behavior are inserted into the SAPT0 formalism, as demonstrated by results using BLYP and B3LYP in Fig. 12.

In a review of different EDAs targeted at drug discovery [108], the SAPT0 (B3LYP) method was included as the sole representative of the entire SAPT formalism. The present comparison demonstrates that this approach is a poor substitute even for conventional HF-based SAPT0, even if the δHF correction were included. In the larger aug-cc-pVTZ basis set, the reason is obvious: significant overestimation of dispersion-dominated interaction energies. This is the net result of a second-order perturbation theory formalism that overestimates dispersion, combined with a functional that lacks correct asymptotic behavior and affords smaller HOMO/LUMO gap as compared to HF theory [182]. Within a perturbative formulation of E_{disp} , smaller gaps manifest as dispersion energies that are too attractive. GGA functionals narrow the gaps even further as compared to global hybrids like B3LYP, with correspondingly worse results for SAPT0(KS) calculations, as documented in Fig. 12. Functionals that lack correct asymptotic behavior should not be used in SAPT0(KS) calculations.

These problems are alleviated through the use of MBD. Figure 13 juxtaposes SAPT0(HF) and SAPT0(LRC- ω_{GDD} PBE) results from Fig. 12b with the corresponding results when MBD replaces $E_{\text{disp}}^{(2)}$. The SAPT(LRC- ω_{GDD} PBE) +MBD method affords very good results in a high-quality basis set (aug-cc-pVTZ), demonstrating that its favorable accuracy does not rely on a limited basis set for error cancellation. That said, note that SAPT(HF)+MBD results are somewhat worse than SAPT0(HF) when the δHF correction is included in both. The damping function that defines the MBD + esDQ model in Eq. (33) was only fitted for use with LRC- ω PBE, which may result in slightly degraded performance for HF monomers.

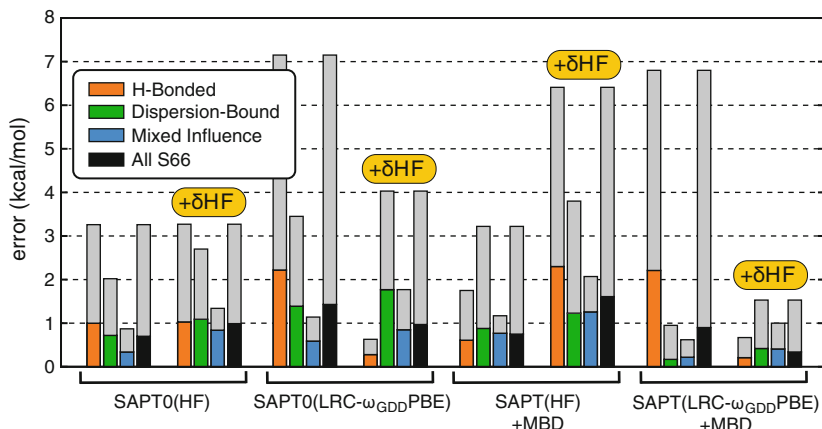


Fig. 13 MAEs (in color) and maximum absolute errors (in gray) for the S66 dimers, comparing SAPT(KS)-based interaction energies to CCSD(T)/CBS results. Methods labeled “SAPT0” use $E_{\text{disp}}^{(2)}$ (Eq. 8a), whereas the two SAPT(KS)+MBD methods on the right use the MBD model (Eq. 33). (All calculations use the aug-cc-pVTZ basis set and the data are from Ref. [77])

2.1.2 Basis Sets

One advantage of introducing dispersion models in place of perturbation theory is a reduction in basis-set demands, as *aiD* models are entirely independent of the basis set and MBD exhibits mild basis-set dependence [77]. Extensive benchmarking of the XSAPT+MBD energy components has been reported recently [77], and Fig. 14 summarizes how the choice of basis set impacts the accuracy of total interaction energies computed using SAPT(LRC- ω_{GDDPBE})+ δHF +MBD.

Errors incurred using double- ζ basis sets are unacceptably large except for jun-cc-pVDZ or aug-cc-pVDZ. This likely results from an imbalance in the convergence behavior of $E_{\text{disp}}^{\text{MBD}}$ as compared to other energy components; indeed, these errors are reduced dramatically in triple- ζ basis sets. The minimally augmented def2-ma-TZVP basis set [77], which is a proper subset of def2-TZVPD [183], is essentially converged with respect to its parent basis set but considerably more efficient, as documented below. Comparing def2-ma-TZVP and def2-QZVPD results in Fig. 14a, we see that differences in the MAEs are well within the intrinsic accuracy of the method itself, although there is some reduction in the outliers when the larger basis set is used. The same can be said for the comparison between jun-cc-pVTZ and aug-cc-pVQZ results in Fig. 14b. This suggests using minimally augmented triple- ζ basis sets for XSAPT+MBD calculations.

Of these, def2-ma-TZVP is by far the least expensive, adding only modest overhead upon def2-TZVP as documented in Fig. 15. In contrast, the upgrade from def2-ma-TZVP to def2-TZVPD improves the MAE for the S66 dimers by ≤ 0.1 kcal/mol but nearly doubles the cost for the π -stacked uracil dimer that is used

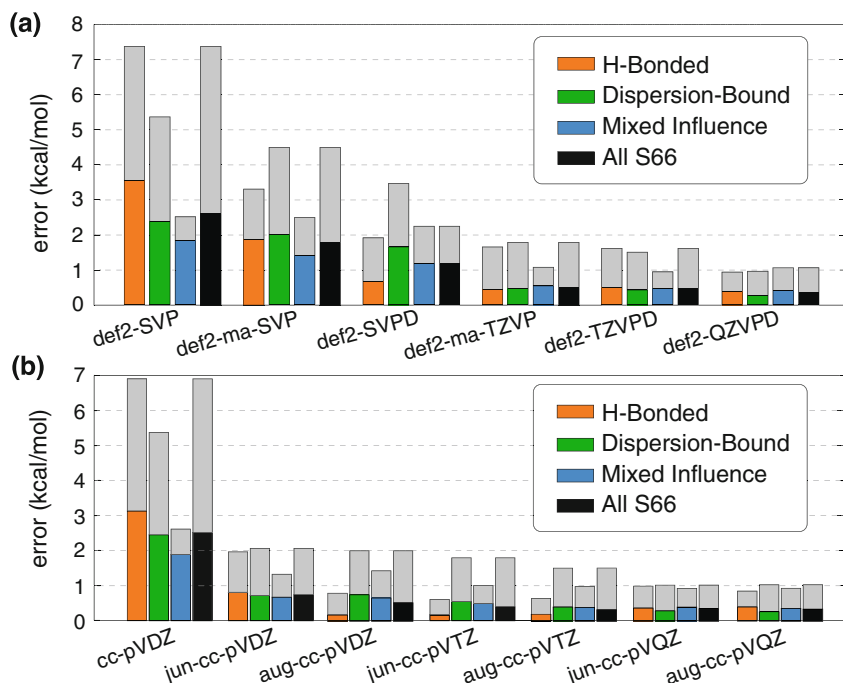


Fig. 14 Errors across the S66 dataset, for SAPT(LRC- ω_{GDDPBE})+ δHF +MBD calculations using (a) Karlsruhe and (b) Dunning basis sets. Colored bars represent MAEs and gray bars are maximum absolute errors, with respect to CCSD(T)/CBS benchmarks. (Data are from Ref. [77])

for timing benchmarks. The def2-QZVPD basis set improves the MAE by < 0.1 kcal/mol with respect to def2-TZVPD, although it does reduce some of the outliers, but increases the cost by nearly $5 \times$. These disparities are exacerbated in Dunning basis sets, where the full complement of diffuse functions in aug-cc-pVTZ reduces the MAE by < 0.1 kcal/mol, as compared to jun-cc-pVTZ results, at $1.4 \times$ the cost. Quadruple- ζ basis sets provide very limited improvements over aug-cc-pVTZ at prohibitive cost. We recommend def2-ma-TZVP unless one is pushing for absolute convergence to the CBS limit, in which case Dunning basis sets provide slightly more consistent results.

Convergence data for energy components were discussed above (see Fig. 7). With the exception of dispersion, all energy components exhibit similar convergence behavior and have nearly reached the CBS limit using def2-ma-TZVP [77]. The dispersion energy $E_{\text{disp}}^{\text{MBD}}$ converges even faster because MBD depends on the density but not directly on the wavefunction. This should be contrasted with the slow convergence of SAPT0 (Fig. 1), to an undesirable answer. Relatively uniform convergence of the energy components means that the XSAPT EDA does not benefit significantly from error cancellation, provided that basis sets of triple- ζ quality are used.

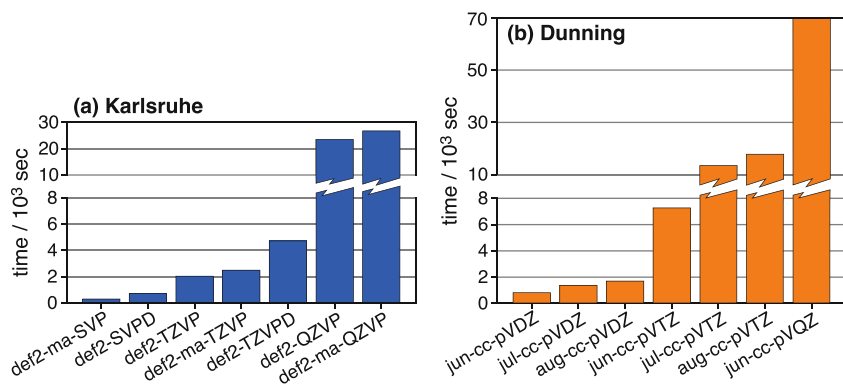


Fig. 15 Total CPU times for a SAPT(LRC- ω_{GDD} PBE) calculation on uracil dimer using (a) Karlsruhe basis sets or (b) Dunning basis sets. All calculations were performed using 12 processors of a Dual Intel Xeon 8268 node with 64 Gb of memory. Note the change in vertical scale beginning at 8000 s, in both panels

2.1.3 Large Supramolecular Complexes

Accuracy benchmarks in Sects. 2.1.1 and 2.1.2 employ the S66 dataset of small dimers in order to have decent statistics and good throughput on numerous variants. Where XSAPT+MBD truly shines, however, is in applications to large complexes. This was previewed in Fig. 4, demonstrating a MAE below 1 kcal/mol for the L7 dataset, which is comparable to the best available DFT methods for noncovalent interactions, while affording an EDA that is less ambiguous as discussed in Sect. 3.2.

Here, we consider some additional large-system benchmarks starting with another set of calculations for the L7 complexes, the results of which are shown in Fig. 16. Data for MP2 and SAPT0(HF) emphasize the failure of second-order perturbation theory for systems involving large, conjugated monomers. In contrast, XSAPT+*ai*D3 and (especially) XSAPT+MBD perform extremely well for these systems. This is true even in comparison to state-of-the-art semiempirical methods designed for noncovalent interactions, exemplified in Fig. 16 by the HF-3c [184] and PBEh-3c [185] methods. These are designed for geometry optimization of large molecules, and they complement XSAPT by providing an affordable means to obtain geometries. Reliable interaction energies can then be computed using XSAPT-based methods.

Figure 17 presents E_{int} benchmarks for a set of large host–guest complexes common to the S12L and S30L datasets [186, 187]. Unlike other benchmarks discussed in this chapter, which come primarily from CCSD(T)/CBS calculations, these are obtained from experimental association energies (ΔG°), back-corrected to gas-phase interaction energies using theoretical estimates of the vibrational entropy and the change in solvation energy upon complexation. These benchmarks come with estimated uncertainties of ± 3 kcal/mol, with the solvation correction as the largest source of error [186, 187].

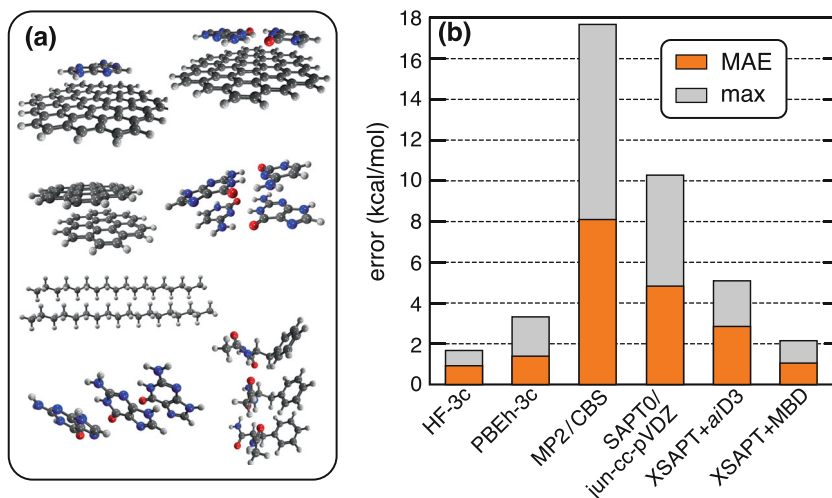


Fig. 16 (a) Dimers in the L7 dataset and (b) error statistics for L7 versus CCSD(T)/CBS benchmarks. XSAPT+*aiD3*/def2-TZVPP and XSAPT+MBD/def2-TZVPPD calculations use LRC- ω GDD PBE and ChElPG embedding charges. None of these calculations employs the δ HF correction. (Data are from Ref. [38])

MP2-based methods are effectively useless for these systems and even B97M-V, which is usually an outstanding semilocal functional for noncovalent interactions [188], exhibits a MAE of 6 kcal/mol and a maximum error of 11 kcal/mol. Notably, these are CP-corrected B97M-V/aug-cc-pVTZ calculations that should lie near the CBS limit [59]. In complexes of this size, BSSE effects can reach 50 kcal/mol or more in double- ζ basis sets [1, 59].

As with the large-monomer L7 complexes, the XSAPT+*aiD3* method requires three-body (triatomic) dispersion corrections to achieve reasonable accuracy for this S12L/S30L dataset, whereas those effects are inherent in XSAPT+MBD. The latter method exhibits accuracy comparable to B97M-V(CP)/aug-cc-pVTZ, in a more modest triple- ζ basis set and without the need for CP correction.

2.1.4 Performance for Ions

Early versions of XSAPT used ChElPG embedding charges for the XPol procedure [32, 33]. These are designed to approximate the molecular electrostatic potential and seem fit to purpose but engender significant computational overhead (Fig. 2). In contrast, CM5 charges [120] are designed only to reproduce the molecular dipole moment but can be evaluated much more efficiently using DFT quadrature grids [41]. Although implemented for reasons of efficiency, CM5 embedding charges also turn out to improve the accuracy, especially for complexes involving ions and ion

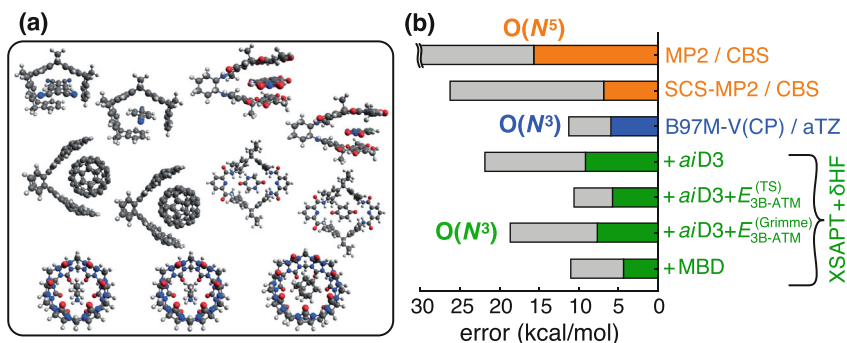


Fig. 17 (a) The S12L/S30L dataset of host–guest complexes. (b) Mean absolute errors (colored bars) and maximum absolute errors (gray bars) for various methods, compared to back-corrected experimental benchmarks with estimated uncertainties of ± 3 kcal/mol. XSAPT+ δ HF calculations use ChEIPG embedding charges and the def2-TZVPPD basis set. (Adapted with permission from Ref. [42]; copyright 2021 American Chemical Society)

pairs, for which ChEIPG embedding charges afford errors up to 15 kcal/mol that are reduced to < 4 kcal/mol using CM5 charges [41].

Table 2 provides an accuracy assessment for XSAPT+ δ HF+aiD3 and XSAPT+ δ HF+MBD calculations on ion–molecule complexes, using the AHB21 and CHB6 datasets [189] that consist of anions and cations, respectively. For AHB21, the benchmarks range up to $|E_{\text{int}}| = 66$ kcal/mol, and ten of the 21 dimers have absolute interaction energies larger than 20 kcal/mol. (For comparison, $|E_{\text{int}}| < 20$ kcal/mol for all of the S66 dimers.) For CHB6, the benchmarks range from -20 kcal/mol to -39 kcal/mol. As such, these are more challenging tests and errors are correspondingly larger than what we have seen for neutral complexes.

For the anion–molecule dataset, MAEs with respect to CCSD(T)/CBS benchmarks are $\lesssim 3$ kcal/mol when the def2-ma-TZVP basis set is used, although maximum errors are about 8.5 kcal/mol. Independent of basis set, the largest error is incurred for the very challenging $\text{Cl}^- (\text{HCl})$ complex consisting of a proton that is shared equally between two Cl^- ions yet must be assigned to one monomer or the other for the purpose of (X)SAPT calculations. For the XSAPT+ δ HF+MBD calculations, this outlier is noticeably reduced by additional diffuse functions so in that case there is a reason to prefer def2-ha-TZVP over def2-ma-TZVP, which is not surprising for a calculation involving an anion.

The CHB6 dataset consists of $\text{M}^+ (\text{C}_6\text{H}_6)$ and $\text{M}^+ (\text{H}_2\text{O})$ complexes for alkali metal cations: $\text{M} = \text{Li}, \text{Na}, \text{and K}$. As compared to the anions, XSAPT+ δ HF+aiD3 calculations result in smaller errors for these cationic complexes, with MAEs below 2 kcal/mol and maximum errors < 4 kcal/mol. However, the MBD model affords excessively large dispersion energies for the two complexes involving K^+ , e.g., $E_{\text{disp}}^{\text{MBD+esDQ}} = -84$ kcal/mol for $\text{K}^+ (\text{C}_6\text{H}_6)$, as compared to $E_{\text{disp}}^{\text{aiD3}} = -4$ kcal/mol for the same complex. These very large MBD dispersion energies manifest as significant

Table 2 Error statistics for the AHB21 and CHB6 datasets of ionic complexes using XSAPT with CM5 embedding charges

Basis set	Error (kcal/mol) ^a							
	XSAPT+ δ HF+ <i>ai</i> D3 ^b				XSAPT+ δ HF+MBD			
	AHB21		CHB6		AHB21		CHB6	
	MAE	Max	MAE	Max	MAE	Max	MAE ^c	Max ^c
def2-ma-SVP	3.7	11.9	1.8	3.4	3.2	8.3	10.6	19.0
def2-ha-SVP	2.9	8.4	1.2	2.6	2.4	6.1	11.8	21.0
def2-SVPD	2.5	8.1	1.5	2.9	2.8	12.1	12.4	22.0
def2-ma-TZVP	2.6	8.6	1.7	3.9	3.2	8.5	11.6	23.0
def2-ha-TZVP	2.7	8.2	1.3	3.7	1.8	5.0	10.9	23.0
def2-TZVPD	2.5	8.4	1.3	3.6	1.7	4.8	10.9	23.0

^a With respect to CCSD(T)/CBS benchmarks from Ref. [189]^b Data from Ref. [55]^c K⁺(H₂O) and K⁺(C₆H₆) removed from the dataset

overbinding of the K⁺ complexes by XSAPT+ δ HF+MBD, with errors of 23 kcal/mol for K⁺(H₂O) and 80 kcal/mol for K⁺(C₆H₆), using def2-ma-TZVP. As such, these two complexes have been removed from the error statistics in Table 2 in order to provide a more representative perspective of the other cation–molecule dimers. Even so, errors remain considerably larger than they are for XSAPT+ δ HF+*ai*D3 applied to the CHB6 dataset.

The source of these errors remains a topic for future work and development, but we can speculate. It has been demonstrated that MBD is sometimes vulnerable to polarization catastrophes for short-range interactions [149], as one might anticipate in strong interactions with ions, although the reason why cations are more problematic is unclear. Especially for ions [190], the model is sensitive to the input polarizabilities [149], which are controlled by the oscillator frequency ω_a in Eq. (31). Alternative models for obtaining these polarizabilities have been suggested [191], which are available in the libMBD library that is interfaced to Q-Chem [192], but these have not yet been adapted for use with XSAPT. At a minimum, iterative determination of Hirshfeld volumes [193–195], to assign excess charge to appropriate atoms in a molecular density, is probably necessary to make MBD generally applicable to ions [190, 194, 196] and the ordinary Hirshfeld partition is known to produce unrealistically large polarizabilities for cations [196]. However, it is unclear that this should improve the description of complexes involving monatomic ions, such as those in the CHB6 dataset. A more rigorous treatment of dipole–quadrupole (and higher-order) polarizability contributions to MBD, as compared to the MBD + esDQ model in Eq. (33), might improve the results for strong, short-range interactions. Clearly, there remains room for improvement to the performance of XSAPT-based methods for ions.

2.2 Cost

In this section, we document the cost for XSAPT calculations and how it scales with system size. All calculations were performed using the Q-Chem program, which remains the only platform in which XSAPT is implemented [45, 46]. Timing benchmarks were performed using a pre-release copy of Q-Chem v. 6.3, in order to document improved multithreading for conventional SAPT0 calculations. The resolution-of-identity (RI) technique leads to considerable reduction in the cost of this $O(N^5)$ method [76, 78, 197, 198] and the SAPT0 timings reported in Sect. 2.2.1 use this technique. Additional parallelization improvements were made in Q-Chem v. 6.0, as documented in Fig. 2, so that multithreaded performance should be significantly improved as compared to Q-Chem v. 5.4. Further improvements in parallel efficiency since that release are documented in Sect. 2.2.1.

These timing benchmarks use an SCF convergence threshold of $10^{-5} E_h$ in the electronic energy gradient $\|\mathbf{FP} - \mathbf{PF}\|$ [199] and a screening threshold of 10^{-12} a.u. (The latter is appropriate for medium-size systems with diffuse basis functions [200].) Timing benchmarks are run on dedicated nodes.

2.2.1 Parallel Performance

Figure 18 documents the parallel efficiency of the new RI-SAPT0 code, running on a single compute node using OpenMP parallelization. Even for the smallest basis set that we consider (cc-pVDZ), the scaling across an entire 48-processor node is reasonably good. The parallel efficiency exceeds 70% in most cases (Fig. 18b) and is considerably improved relative to the old code. Memory usage is also improved, enabling cc-pVQZ calculations (1345 basis functions) that were not previously feasible on our hardware.

We next examine parallel efficiencies for XSAPT+*aiD3*, which has a much different cost profile. For these tests, we have extended upon data for $\text{Cl}^-(\text{H}_2\text{O})_n$ clusters that were published elsewhere [77], adding new basis sets and updated timings. These $\text{Cl}^-(\text{H}_2\text{O})_n$ clusters were extracted from a molecular dynamics simulation of $\text{Cl}^-(\text{aq})$ [201], with $\langle n \rangle = 28.4 \pm 2.4$ water molecules per snapshot, and the data presented in Table 3 are averages across 51 snapshots, treating $(\text{H}_2\text{O})_n$ as a single monomer. For Pople basis sets, parallel efficiencies approach or exceed 80% on 26 processors, although they are lower (46–67%) for Karlsruhe basis sets. The difference likely lies in the use of compound (“sp”) shells for Pople basis sets, which Q-Chem’s integrals code handles efficiently [200].

Although Pople basis sets have fallen out of favor for modern electronic structure calculations (for good reason in certain cases [202–205]), they can yield high-quality results for DFT calculations if sufficient polarization functions are included, beyond the usual 6-31G(d) or 6-311G(d,p) [200]. In the present context, Pople basis sets can reproduce induction energies obtained in much larger Karlsruhe or Dunning basis sets, to an acceptable degree, if additional polarization functions are included

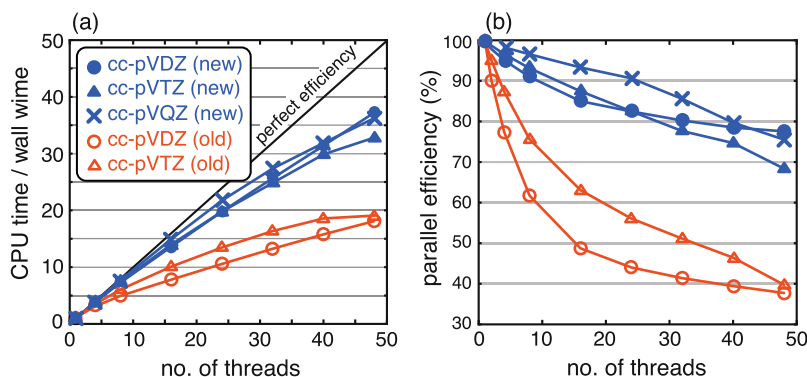


Fig. 18 Strong-scaling data for multitreaded RI-SAPT0 calculations on a guanine–cytosine base pair, $(\text{C}_5\text{H}_5\text{N}_5\text{O}) \cdots (\text{C}_4\text{H}_5\text{N}_3\text{O})$, using either the original SAPT0 code (Q-Chem versions up to v. 6.2) or the new code (pre-release Q-Chem v. 6.3). **(a)** Speedup measured by the ratio of CPU time to wall time ($t_{\text{CPU}}/t_{\text{wall}}$). **(b)** Parallel efficiency, defined as $t_{\text{CPU}}/t_{\text{wall}}/n_{\text{threads}}$ and expressed as a percentage. All calculations were run on a single Dual Intel Xeon 8268 node with 48 processors and 192 Gb memory

Table 3 Error statistics and timing data for XSAPT+*aiD3* calculations on $\text{Cl}^- (\text{H}_2\text{O})_n^{\text{a}}$

Basis set	E_{int} Error (kcal/mol) ^b		Time (sec) ^c		Parallel efficiency ^e
	MAE ^d	Max	CPU	Wall	
6-31+G(d)	1.9 ± 0.8	3.8	1226	57	82%
6-311+G(d,p)	1.3 ± 0.5	2.4	2834	140	78%
6-311+G(2df,2p)	0.1 ± 0.1	0.4	7813	355	85%
6-311+G(3df,2pd)	0.3 ± 0.1	0.6	16,564	761	84%
def2-ma-SVP	3.1 ± 0.7	4.4	2678	154	67%
def2-SVPD	1.2 ± 0.2	1.5	11,063	928	46%
def2-ma-TZVP	1.3 ± 0.3	1.9	9982	586	66%
def2-TZVPD	—	—	34,812	2371	56%

^a Data represent averages over 51 snapshots with $\langle n \rangle = 28.4 \pm 2.4$. The $(\text{H}_2\text{O})_n$ cluster is treated as a single monomer

^b With respect to the def2-TZVPD value

^c Using 26 processors (Dell Intel CPU Max 9470 HBM)

^d Uncertainty represents one standard deviation

^e $\text{PE} = t_{\text{CPU}}/t_{\text{wall}}/n_{\text{proc}}$

[77]. For example, XSAPT+ δHF +MBD induction energies computed using the 6-311++G(3df,2pd) basis set exhibit slightly smaller errors, as compared to SAPT2+(3)/aug-cc-pVTZ benchmarks [151], than when the def2-QZVPD basis set is used [77]. Errors are small in both cases, but the Pople basis set is dramatically more efficient.

This is a rather general observation [77, 200]. To see this, we take XSAPT+MBD/def2-TZVPD results as converged and examine interaction energies for $\text{Cl}^- (\text{H}_2\text{O})_n$

computed in other basis sets (Table 3). The data indicate that 6-311+G(2df,2p) affords negligible differences as compared to def2-TZVPD but is $4.5 \times$ less expensive, when measured in aggregate CPU time, and $6.7 \times$ faster in wall time. For rough estimates of ~ 1 kcal/mol accuracy, even 6-311+G(d,p) can be used and is more than an order of magnitude faster than def2-TZVPD.

2.2.2 Scaling with System Size

To ascertain the effective scaling of cost versus system size, we will present timing data for homologous systems starting with $\text{Cl}^- (\text{H}_2\text{O})_n$ clusters of increasing size, treating $(\text{H}_2\text{O})_n$ as a single monomer. Figure 19 presents timing data for XSAPT+MBD calculations. The quantity that is plotted is the total CPU time aggregated across all 20 processors that were used, not the time to completion. The latter is much shorter, as may be inferred from the reasonably good parallel efficiencies that are documented in Table 3, but it is total CPU time that captures the true cost of the calculation [206]. Fits of the total time t to the function $t = an^\beta$ reveal that the scaling is effectively cubic, for both the XPol-SCF procedure (which contributes 80% of the total cost, for the largest systems examined) and for the post-SCF SAPT and MBD corrections (comprising the remaining 20% of the cost).

For completeness, we also present timings for RI-SAPT0(HF)/jun-cc-pVDZ calculations on (guanine) $_n$ stacks extracted from an idealized structure of B-DNA. Timing data are presented in Fig. 20, for calculations in which (guanine) $_{n-1}$ is treated as a single monomer. For the post-SCF steps, we obtain an effective scaling of $O(N^{4.2})$ that is better than the theoretical result. However, the SCF time constitutes a significant fraction of the overall time for the system sizes considered (up to $n = 6$), resulting in an effective scaling of $O(N^{3.1})$ for the entire calculation. For the same

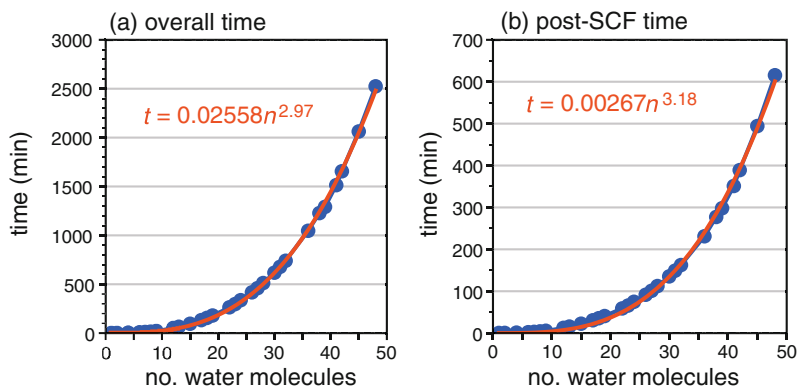


Fig. 19 Timing data for XSAPT+MBD/def2-ma-TZVP calculations on $\text{Cl}^- (\text{H}_2\text{O})_n$ clusters of increasing size, treating $(\text{H}_2\text{O})_n$ as a single monomer and examining (a) the total compute time and (b) the post-SCF time. Fits to the function $t = an^\beta$ are indicated. Calculations were performed on 20 processors of a single Dual Intel Xeon 8268 node and the δHF correction is not included

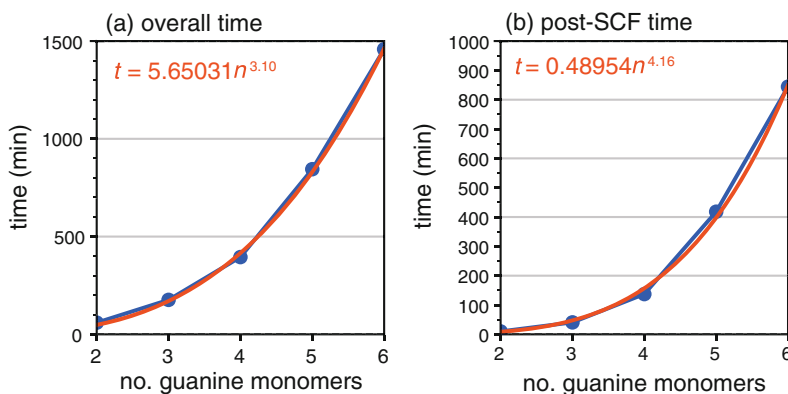


Fig. 20 Timing data for RI-SAPT0/jun-cc-pVDZ calculations on (guanine)_n, including (a) the overall time, aggregated across all processors, and (b) the post-SCF time. The system is treated as a dimer of guanine and (guanine)_{n-1}. Calculations were performed using 40 processors of a single Dual Intel Xeon 8268 node

systems and basis set, XSAPT+MBD calculations afford effective scaling of $O(N^{2.6})$ for both the overall time and the post-SCF time, demonstrating that this approach does actually reduce the effective scaling, relatively to RI-SAPT0, even when super-theoretical scaling is obtained. Moreover, the memory footprint of XSAPT+MBD is much smaller.

3 Illustrative Applications

This section explores the use of XSAPT+MBD as a tool for understanding the physics of intermolecular interactions, taking advantage of the dimer EDA in Eq. (1) and relying on the sub-kcal/mol accuracy that was established in Sect. 2.1 and elsewhere. We first establish some basics about short-range repulsion, which originates in the Pauli principle and not with Coulomb interactions, as discussed in Sect. 3.1. Then, the MBD dispersion model is compared directly to dispersion energies obtained from a popular DFT-based EDA scheme [13], in Sect. 3.2. That discussion highlights the fact that $E_{\text{disp}}^{\text{MBD}+\text{esDQ}}$ in Eq. (33) agrees very well with SAPT2+(3) dispersion energies, even though it was not parameterized using these data (unlike the *aiD* models). That discussion also highlights ambiguities in the definition of E_{disp} in the DFT-based EDA. Having established the veracity of the XSAPT+MBD energy decomposition, we apply it to analyze π - π interactions in Sect. 3.3. Finally, hydrogen bonding in $\text{F}^-(\text{H}_2\text{O})$ is examined in Sect. 3.4, using the SAPT-cDFT decomposition to isolate E_{CT} and thereby elucidate the role of CT in the anion–water hydrogen bond.

3.1 Basic Intermolecular Physics

Chemistry textbooks are replete with half-truths and outright falsehoods regarding the role of multipolar interactions in short-range molecular physics, some of which are addressed here. Additional examples can be found in Ref. [43].

3.1.1 Electrostatics Beyond Multipoles

In the interest of pedagogy, let us introduce the electrostatic (or Coulomb) potential

$$\tilde{\varphi}(\mathbf{R}) = \int \frac{\tilde{\rho}(\mathbf{r})}{\|\mathbf{r} - \mathbf{R}\|} d\mathbf{r} \quad (45)$$

for a continuous charge distribution $\tilde{\rho}(\mathbf{r})$. We use $\tilde{\rho}(\mathbf{r})$ to indicate a molecular charge distribution that includes both nuclei and electrons. If $\rho(\mathbf{r})$ is the charge distribution for the electrons only, which is positive-valued by convention, then

$$\tilde{\rho}(\mathbf{r}) = -\rho(\mathbf{r}) + \sum_a^{\text{nuclei}} Z_a \delta(\mathbf{r} - \mathbf{R}_a). \quad (46)$$

Given charge densities $\tilde{\rho}_A(\mathbf{r})$ and $\tilde{\rho}_B(\mathbf{r})$ for isolated molecules A and B, along with the corresponding electrostatic potentials $\tilde{\varphi}_A(\mathbf{r})$ and $\tilde{\varphi}_B(\mathbf{r})$, the total electrostatic interaction in the SAPT scheme (Eq. 16a) is equivalent to

$$E_{\text{elst}} = \int \tilde{\varphi}_A(\mathbf{R}) \tilde{\rho}_B(\mathbf{R}) d\mathbf{R} = \int \tilde{\varphi}_B(\mathbf{R}) \tilde{\rho}_A(\mathbf{R}) d\mathbf{R}. \quad (47)$$

For interpretative purposes, we might approximate one or both of the electrostatic potentials using a multipole expansion. For example,

$$\tilde{\varphi}_A(R) = \frac{Q_A}{R} + \frac{\boldsymbol{\mu}_A \cdot \mathbf{R}}{R^3} + \frac{\mathbf{R}^\dagger \boldsymbol{\Theta}_A \mathbf{R}}{R^5} + \cdots \quad (48)$$

where Q_A , $\boldsymbol{\mu}_A$, and $\boldsymbol{\Theta}_A$ are the total charge, the dipole moment vector, and the quadrupole moment tensor of molecule A, respectively. One could make a multipole expansion for the density $\tilde{\rho}_B(\mathbf{R})$ and then recast Eq. (47) into a form involving multipole–multipole interactions [207]. In doing so, it is important to remember that Eq. (48) is an asymptotic expansion that is valid only for a test charge that is located at a distance R that is large compared to the details of the molecular charge distribution, $\tilde{\rho}_A(\mathbf{r})$. In fact, the multipole expansion is divergent as $R \rightarrow 0$ [22, 208–210]. More importantly, it is readily demonstrable that nonbonded close-contact distances (including typical equilibrium geometries of vdW complexes) are

not usually in the asymptotic regime where the electrostatic potential is well described (or even *qualitatively* described) by leading-order multipoles. An example is shown in Fig. 21 for different geometries of the benzene dimer.

Exact electrostatic interactions, computed using Eq. (47) without any multipole approximation, are attractive in the typical range of intermolecular π -stacking separations (3.4–3.8 Å), and they are attractive in the edge-to-face (C–H $\cdots\pi$) configuration also [211–213]. This has implications for understanding π -stacking [42, 148, 214–217], as discussed in Sect. 3.3. For now, we simply note that an atom-centered multipole approximation to E_{elst} (Fig. 21a), which should converge faster than the single-center expansion in Eq. (48), nevertheless does not agree with exact calculations for the cofacial arrangement until the face-to-face separation R is much larger than typical π -stacking distances. A leading-order approximation, using only the molecular quadrupole moments of the two benzene monomers, does not agree with exact electrostatics until even larger face-to-face separations (Fig. 21b). At typical π -stacking separations, such as $R = 3.35$ Å for graphite [218] or $R = 3.4$ Å for the parallel-displaced benzene dimer [212], exact electrostatics is attractive but quadrupolar approximations are repulsive [148, 211, 213, 215].

This breakdown in multipolar approximations is sometimes ascribed to *charge penetration* effects, meaning that multipole approximations no longer describe E_{elst} when $\rho_A(\mathbf{r})$ and $\rho_B(\mathbf{r})$ interpenetrate. Another interesting example of this effect can be found in the corannulene dimer, $(\text{C}_{20}\text{H}_{10})_2$. The corannulene monomer consists of six hexagonal rings surrounding a central pentagonal ring, the latter of which lends curvature to the structure, which has been called a “buckybowl” [219–221]. This and

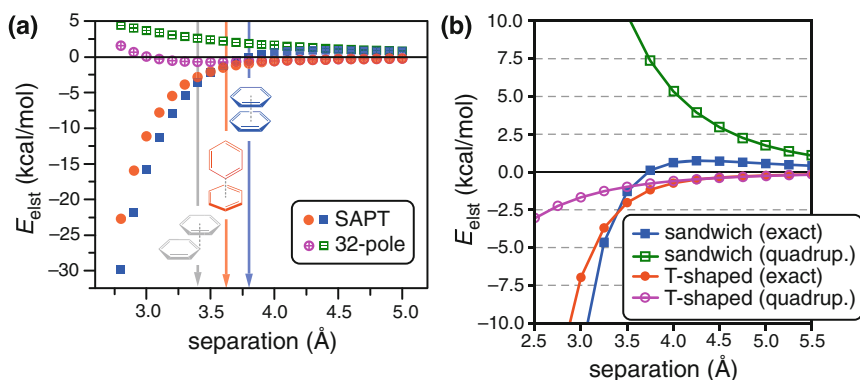


Fig. 21 (a) Electrostatic interaction energies for the cofacial (solid blue symbols) and T-shaped (solid red symbols) geometries of $(\text{C}_6\text{H}_6)_2$, computed at the HF/jun-cc-pVDZ level as a function of intermolecular separation. (For the T-shaped structure, the distance coordinate is the H-to-face separation.) Open symbols with crosses correspond to an atom-centered distributed multipole calculation at the HF/6-311G(d,p) level, including interactions up to octupole–quadrupole. Arrows indicate the minimum-energy intermolecular separations for the three structures that are shown. (Adapted with permission from Ref. [211]; copyright 2016 American Chemical Society.) (b) Exact versus quadrupolar electrostatics for $(\text{C}_6\text{H}_6)_2$, computed at the LRC- ω_{GDD} PBE/def2-TZVPD level along the same separation coordinates

other concave hydrocarbons form concentric stacks in the solid state [221–224]. (A five-member ring is not required to induce curvature [225].) The π -electron density at the bottom of these buckybowls must be compensated by electropositive terminal C–H bonds, as it is in benzene as well but the buckybowls break planar symmetry and may develop dipole moments. These range from $\mu = 2.07$ D in $\text{C}_{20}\text{H}_{10}$, based on a gas-phase measurement [226], to $\mu > 10$ D in larger buckybowls, based on calculations [227, 228]. As a result, this phenomenon has been called “flexo-electric” π -stacking [228], with the implication that the driving force for this solid-state stacking phenomenon is to align the dipole moments of the monomers. In view of exact electrostatic calculations for $(\text{C}_6\text{H}_6)_2$, however, this hypothesis should be regarded with suspicion.

To investigate this in detail, we used XSAPT+MBD calculations to examine intermolecular interactions in $(\text{C}_{20}\text{H}_{10})_2$ in a sequence of calculations in which we gradually flatten each monomer at fixed intermolecular separation, starting at the equilibrium dimer structure (with concave monomers) and ending at a structure in which both monomers are planar (so $\mu = 0$) [148]. Results of the energy decomposition along this coordinate are plotted in Fig. 22. Monomer densities $\rho_A(\mathbf{r})$ and $\rho_B(\mathbf{r})$ are also shown, using an isocontour value ($\rho = 0.001$ a.u.) that is typical for exhibiting molecular size [146]. These plots demonstrate that the monomer densities significantly interpenetrate at the equilibrium stacking distance.

Despite diminishing (and ultimately vanishing) monomer dipole moments along the planarization coordinate, electrostatic interactions remain significant. What is plotted in Fig. 22 is actually polarized electrostatics ($E_{\text{elst}} + E_{\text{ind}}$), but this sum is dominated by E_{elst} at all geometries. Even when the monomers are planar and $\mu = 0$, $E_{\text{elst}} + E_{\text{ind}}$ remains within 1 kcal/mol of its value in the equilibrium geometry. In fact, the *total* interaction energy (E_{int}) is slightly more stabilizing when the monomers are planar. Note that E_{int} does not include the penalty to deform the monomers and is simply the interaction energy for moving the rigid monomers to infinite separation. Thus, it is the deformation penalty rather than any intermolecular driving force that explains why $(\text{C}_{20}\text{H}_{10})_2$ adopts a structure with curved monomers. This *cannot* be explained based on the dipole moment induced by the curvature.

3.1.2 Steric Repulsion Is Pauli Repulsion

In the benzene dimer, E_{elst} becomes increasingly attractive at smaller values of R , as depicted in Fig. 21, whereas many textbooks would argue that short-range electrostatics should be repulsive due to interpenetration of the electron clouds on the two monomers. To debunk this myth, we consider an even simpler system: Ar_2 . Energy components for Ar_2 are plotted as a function of R in Fig. 23a [43]. Note that $E_{\text{elst}}(R) < 0$ at all values of R , indicating a strictly attractive electrostatic interaction, even for values of R that are much smaller than the equilibrium separation.

The balance of forces that enters $E_{\text{elst}}(R)$ is rather subtle. To appreciate how much so, it is necessary to decompose the total electrostatic interaction,

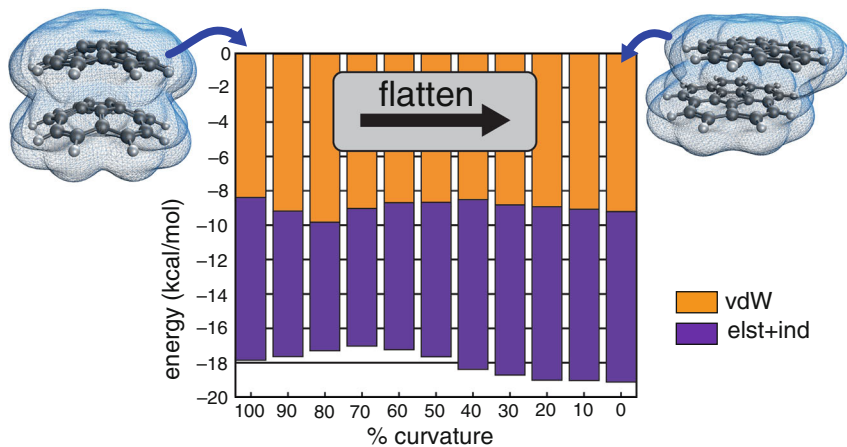


Fig. 22 XSAPT+ δ HF+MBD/def2-TZVPPD results for $(C_{20}H_{10})_2$ along a coordinate in which the monomer geometries are gradually flattened at fixed intermolecular separation, starting from the equilibrium geometry on the left and ending at a geometry in which both monomers are planar, on the right. The quantities $E_{vdW} = E_{exch} + E_{disp}$ and $E_{elst} + E_{ind}$ are provided along this coordinate. Monomer electron densities at the starting and ending geometries are plotted using an isocontour $\rho = 0.001$ a.u. (Data are from Ref. [148])

$$E_{elst} = E_{elst}^{ee} + E_{elst}^{nn} + E_{elst}^{en} \quad (49)$$

into electron–electron repulsion ($E_{elst}^{ee} > 0$), nuclei–nuclei repulsion ($E_{elst}^{nn} > 0$), and electron–nuclei attraction ($E_{elst}^{en} < 0$). Written in atomic units, the expressions for these terms are

$$E_{elst}^{nn} = \sum_{a \in A} \sum_{b \in B} \frac{Z_a Z_b}{R_{ab}} \quad (50)$$

for atomic numbers Z_a and Z_b ,

$$E_{elst}^{ee} = \int \frac{\rho_A(\mathbf{r}_1) \rho_B(\mathbf{r}_2)}{\|\mathbf{r}_1 - \mathbf{r}_2\|} d\mathbf{r}_1 d\mathbf{r}_2 \quad (51)$$

for the electron–electron repulsion, and

$$E_{elst}^{en} = - \sum_{a \in A} \int \frac{Z_a \rho_B(\mathbf{r})}{\|\mathbf{r} - \mathbf{R}_a\|} d\mathbf{r} - \sum_{b \in B} \int \frac{Z_b \rho_A(\mathbf{r})}{\|\mathbf{r} - \mathbf{R}_b\|} d\mathbf{r} \quad (52)$$

for the attractive contribution. Signs in Eq. (52) reflect the convention that electron densities are output from electronic structure programs as strictly positive quantities, representing probability densities of negatively charged electrons [43]. Although

$E_{\text{elst}}^{\text{nn}}$ and $E_{\text{elst}}^{\text{ee}}$ become increasingly repulsive at shorter intermolecular separation, $E_{\text{elst}}^{\text{en}}$ becomes more attractive as the separation is reduced and the electrons in $\rho_A(\mathbf{r})$ gain better access to the nuclei on B and *vice versa*.

For Ar_2 at $R=3.76$ Å, which is close to the equilibrium separation, Fig. 23b quantifies the three terms in Eq. (49). The attractive and repulsive contributions are each $\sim 90 E_h$ in magnitude yet sum to a mere $-0.000164 E_h = -35.9 \text{ cm}^{-1}$, which is the total value of E_{elst} [43]. In this particular case, the increased attraction of $E_{\text{elst}}^{\text{en}}$ at short range wins out over increased repulsion of $E_{\text{elst}}^{\text{ee}} + E_{\text{elst}}^{\text{nn}}$, although the effect is very small. It may be larger in other systems, as in $(\text{C}_6\text{H}_6)_2$ (see Fig. 21), although the sign is not guaranteed one way or the other and examples can be found where $E_{\text{elst}} > 0$ at short range [213].

Thus, Coulomb repulsion does not account for steric repulsion. Instead, short-range repulsion originates with the exchange energy (E_{exch}), as seen for Ar_2 in Fig. 23a. This is *Pauli repulsion*, originating from the antisymmetry requirement. Within the SAPT formalism, that requirement is not built into the direct-product wavefunction $|\psi_0^A\rangle |\psi_0^B\rangle$. This is more than a semantic or a model-dependent distinction, however. Results above demonstrate that interpenetration of two charge densities $\tilde{\rho}_A(\mathbf{r})$ and $\tilde{\rho}_B(\mathbf{r})$ need not lead to a repulsive interaction, when the electron–nuclear interactions are considered. Thus, a distinct physical phenomenon is needed to explain steric repulsion.

One may conceptualize Pauli repulsion in the following way. If the isolated monomer wavefunctions ψ_0^X (for $X = \text{A}$ or B) are computed by means of a variational principle, as in any SCF calculation, then ψ_0^X will adopt whatever shape minimizes its energy in isolation. Introduction of another molecule brings with it an antisymmetry constraint, so that the energy of the constrained system must be higher than that of the unconstrained system, for any finite intermolecular

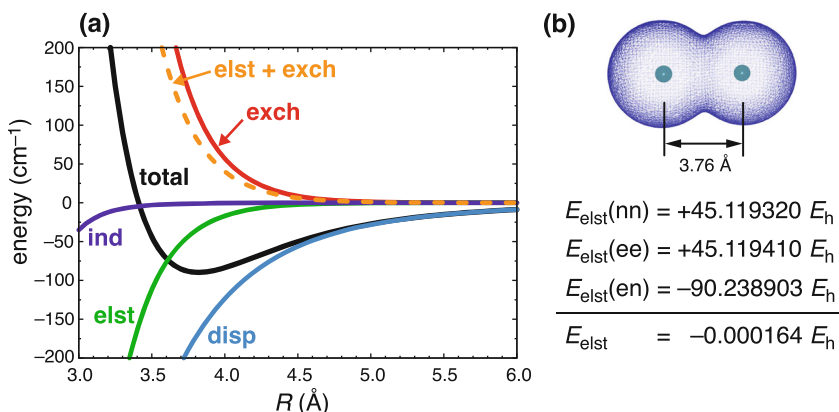


Fig. 23 (a) Total interaction potential $E_{\text{int}}(R)$ for Ar_2 (in black) and its SAPT decomposition (in color), computed at the SAPT0/aug-cc-pVTZ level. (Adapted with permission from Ref. [43]; copyright 2021 American Chemical Society.) (b) Isocontours $\rho=0.001$ a.u. of the HF electron density and components of E_{elst} , computed for $R=3.76$ Å

separation. In other words, the antisymmetry constraint forces a deformation of the wavefunction away from the minimum-energy form that it adopted in isolation and that deformation is associated with an energy penalty (steric or Pauli repulsion). An alternative but equivalent view is that steric repulsion represents an orthogonalization penalty [229]. Since the MOs for molecules A and B are computed in isolation, they will not be mutually orthogonal and must be deformed (orthogonalized) in order to construct a common Slater determinant that satisfies the Pauli principle (antisymmetry requirement) for the composite AB system.

It is worth noting that some EDA schemes choose to orthogonalize the MOs and construct an antisymmetric Slater determinant $|\psi_0^A \psi_0^B\rangle$ prior to computing electrostatic interactions [9]. This differs from the direct-product $|\psi_0^A\rangle |\psi_0^B\rangle$ that is used in SAPT. Where electrostatic interactions are computed using an antisymmetrized reference state, the antisymmetrized electrostatics should be compared to $E_{\text{elst}} + E_{\text{exch}}$ from SAPT. For Ar_2 , Fig. 23a shows that this sum is also repulsive, as the Pauli repulsion (E_{exch}) more than compensates for the attractiveness of E_{elst} .

3.1.3 Dispersion Is Size Extensive

There is a tendency to describe dispersion as a weak interaction. (Sometimes, the phrase “van der Waals interactions” is used to mean dispersion, but we avoid that terminology.) Dismissing dispersion as a weak effect is prevalent in the literature on drug discovery and noncovalent enzyme inhibition, where “vdW interactions” are sometimes described as a sub-kcal/mol phenomenon [230]. This is true for the rare-gas dimers that are often used to introduce London dispersion, as evident from the Ar_2 data in Fig. 23. However, the magnitude of E_{disp} scales with system size and need not be small in large molecular systems, even as compared to covalent bond energies of ~ 100 kcal/mol. Assuming typical values for atomic C_6 coefficients, a rough estimate suggests that two 100-atom organic molecules at 5 Å separation might exhibit a mutual dispersion interaction $E_{\text{disp}} \sim -60$ kcal/mol [5]. Entropic effects tend to reduce interaction energies through entropy/enthalpy compensation, such that $|\Delta G_{\text{bind}}^\circ| < 20$ kcal/mol even for ~ 100 -atom host–guest complexes for which $|\Delta U_{\text{bind}}^\circ|$ exceeds 100 kcal/mol [186].

Nevertheless, in order to understand the balance of intermolecular forces it is important to get the interaction energies right. Qualitative effects of dispersion on organic reactivity have been discussed [6], aided by *in silico* computational modeling in which dispersion interactions can be turned off in order to isolate their effects. This has led to the fascinating idea of “steric attraction” [6, 7, 231–235], in which a bulky substituent also contributes substantially to the polarizability, enhancing E_{disp} to the point that it may outcompete E_{exch} and the larger substituent becomes the less repulsive one. This phenomenon is only possible because dispersion is size extensive.

To illustrate how the magnitude of the dispersion interaction generally increases with molecular size, Fig. 24 shows two examples relevant to drug discovery, namely,

a DNA intercalation ligand (the antitumor agent ellipticine) and also the protease inhibitor indinavir in a model of the HIV protease binding site. These two systems have become standard benchmarks for noncovalent quantum chemistry [236, 237]. In both cases, the XSAPT+MBD interaction energy lies rather close to a benchmark CCSD(T)/CBS value [236]. The interesting observation in the present context is the magnitude of the dispersion energy. For the DNA intercalation complex, which is an obvious example of π -stacking where dispersion might be expected to play a prominent role, the XSAPT+MBD dispersion energy is -71 kcal/mol [41]. Meanwhile, the HIV–indinavir system has no obvious π – π interactions, yet the dispersion contribution is nearly twice as large, at -135 kcal/mol [42]. The model system for HIV–indinavir contains 323 atoms as compared to 157 atoms for the DNA–ellipticine model, and the dispersion energy is more attractive in the larger model system.

3.2 Comparison of Dispersion Models

Having established that SAPT calculations can reveal fundamental physics regarding the origins of intermolecular forces, and that the magnitude of dispersion increases with system size, we next want to consider the dispersion interactions

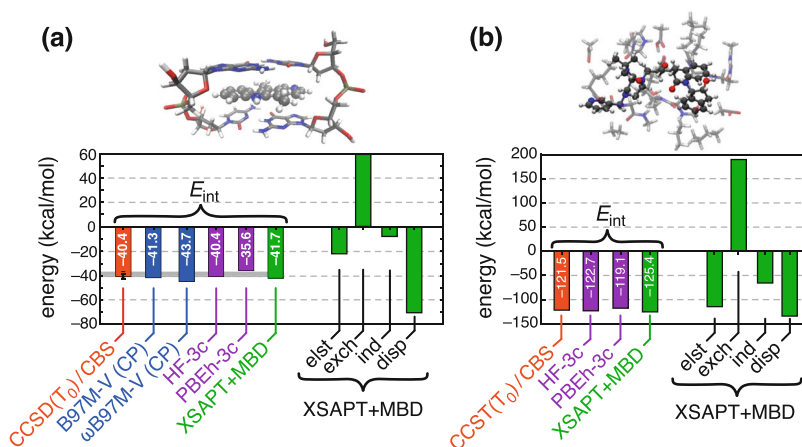


Fig. 24 Interaction energies and XSAPT+MBD decompositions for two pharmacologically relevant complexes: (a) ellipticine intercalated into DNA (157 atoms) and (b) indinavir in the binding pocket of HIV-2 protease (323 atoms), with the inhibitor (indinavir) shown using opaque atoms. In (a), CP-corrected DFT/def2-TZVPPD results are provided for two different functionals, whereas XSAPT+MBD calculations use the def2-hpTZVPP basis set for both systems; these calculations are from Ref. [41]. Semiempirical HF-3c and PBEh-3c results are from Ref. [42]. (Adapted with permission from Ref. [42] but updated to reflect new CCSD(T₀)/CBS benchmarks from Ref. [236]; copyright 2021 American Chemical Society)

more carefully. The second-order expression obtained by London is [43, 145, 238, 239]

$$E_{\text{disp,AB}}^{(2)} = - \sum_{m>0} \sum_{n>0} \frac{|\langle \psi_m^A \psi_n^B | (\hat{H}_{AB} - \hat{H}_A - \hat{H}_B) | \psi_0^A \psi_0^B \rangle|^2}{(\epsilon_m^A - \epsilon_0^A)(\epsilon_n^B - \epsilon_0^B)}. \quad (53)$$

This corresponds to the dispersion energy in SAPT0 (Eq. 8a). As discussed previously, it is typically much too attractive, e.g., by 80–100% for π -stacked benzene dimer [75, 240]. For larger systems, errors in second-order perturbation theory are even more egregious [38, 71].

For that reason, we substitute a dispersion model in XSAPT+MBD, which performs extremely well as compared to dispersion energies obtained from SAPT2+(3) calculations, as discussed previously in the context of Fig. 6. We consider that SAPT2+(3) constitutes a benchmark level of theory for the energy components because total interaction energies computed at that level agree with CCSD(T) results to sub-kcal/mol accuracy [25–27, 57, 58]. Good agreement between XSAPT+MBD and SAPT2+(3) energy components thus lends credence to the idea that the former does not benefit significantly from error cancellation, so that individual energy components within the SAPT formalism can be swapped out for better models, e.g., substituting the non-perturbative MBD model for dispersion.

It is interesting to compare these benchmark-quality dispersion energies to those obtained from the popular ALMO-EDA2 scheme [13], which is DFT-based. We select the PBE0+D3 functional for this assessment, as it performs well for total interaction energies on a variety of noncovalent datasets [1]. ALMO-EDA2 provides a means to separate induction from the total (CP-corrected) interaction energy [82], and the remaining energy components are grouped together as the “frozen density” part of ΔE_{int} , which is

$$\Delta E_{\text{FRZ}} = E_{\text{elst}} + E_{\text{exch}} + E_{\text{disp}}. \quad (54)$$

To isolate the dispersion energy, it is recommended [13] to compute ΔE_{FRZ} using both the target functional (PBE0+D3 in this case) and also with a “dispersionless” functional that is assumed to describe dispersion poorly or not at all (e.g., HF theory). The difference between ΔE_{FRZ} computed with either functional defines the ALMO-EDA2 dispersion energy [13]:

$$E_{\text{disp}}^{\text{ALMO-EDA2}} = \Delta E_{\text{FRZ}}^{\text{target}} - \Delta E_{\text{FRZ}}^{\text{disp'less}}. \quad (55)$$

If the target functional is a hybrid then HF theory is suggested for computing $\Delta E_{\text{FRZ}}^{\text{disp'less}}$ [13].

Application of this procedure to the S22 dataset, however, leads to ALMO-EDA2 dispersion energies that are systematically less attractive than SAPT2+(3) values, as documented in Fig. 25. These sizable discrepancies may arise because the recipe in

Eq. (55) assumes that every correlation effect is dispersion, when the “dispersionless” functional is selected to be HF exchange (only). Correlation corrections to electrostatics or Pauli repulsion might reduce the attractiveness of what is counted as dispersion. To investigate this, we experimented with using HF exchange plus PBE correlation (the HF-PBE functional) as the “dispersionless” functional in Eq. (55) [1]. This affords dispersion energies that are even farther from SAPT2+(3)-values, as shown also in Fig. 25.

Figure 26 provides an alternative presentation of these data, plotting $E_{\text{disp}}^{\text{MBD}}$ and $E_{\text{disp}}^{\text{ALMO-EDA2}}$ versus $E_{\text{disp}}^{\text{SAPT2+(3)}}$. A linear fit of the MBD data affords a slope of ≈ 1.0 with an intercept of -0.4 kcal/mol, which we regard as good agreement with the benchmarks. ALMO-EDA2 dispersion energies also correlate reasonably well with SAPT2+(3) benchmarks, although less so when HF-PBE is used as the “dispersionless” functional. Given that removing the semilocal correlation effects from ALMO-EDA2 dispersion seems to worsen agreement with benchmark dispersion energies, it is unclear whether ALMO-EDA2 energy components can be

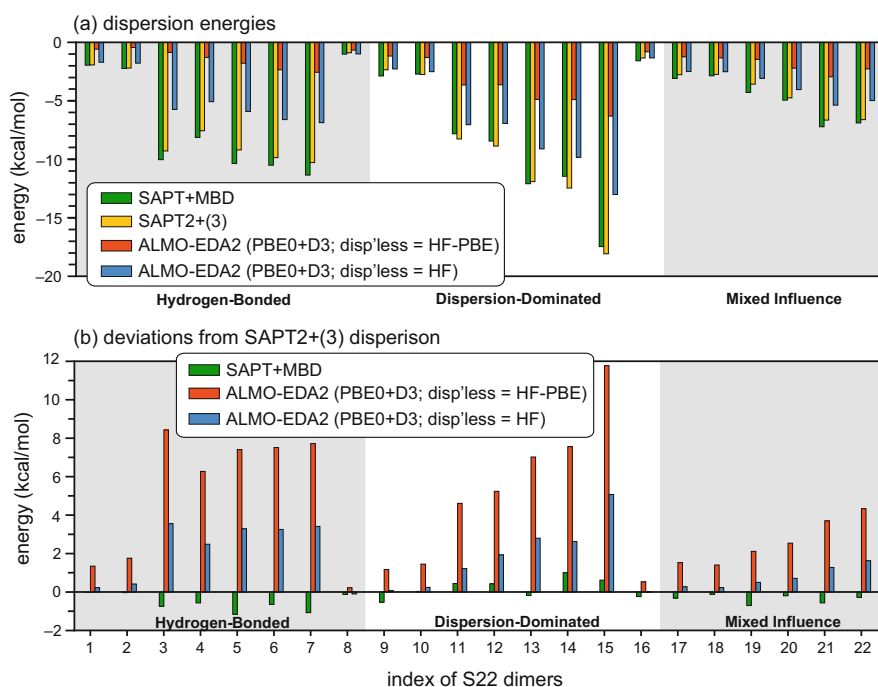


Fig. 25 (a) Dispersion energies for the S22 dataset computed using various methods. (b) Differences between the SAPT(KS)+MBD or ALMO-EDA2 dispersion energies and SAPT2+(3)/aug-cc-pVTZ benchmarks from Ref. [151]. ALMO-EDA2 dispersion energies are computed using Eq. (55) with PBE0+D3 as the target functional and either HF or HF-PBE as the dispersionless functional. (Adapted with permission from Ref. [1]; copyright 2024 Elsevier)

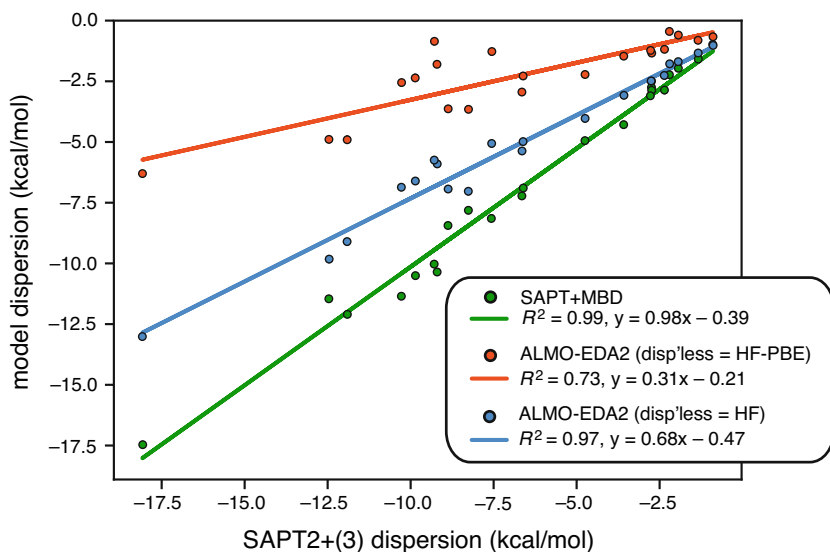


Fig. 26 Correlation between dispersion energies computed using either SAPT(LRC- ω_{GDD} PBE) + MBD or else ALMO-EDA2 applied to PBE0+D3 (on the vertical axis), and benchmarks computed at the SAPT2+(3)/aug-cc-pVTZ level (on the horizontal axis). For ALMO-EDA2, two different “dispersionless” XC functionals are examined for separation of E_{disp} from ΔE_{FRZ} via Eq. (55). These are the same data as in Fig. 25. (Adapted with permission from Ref. [1]; copyright 2024 Elsevier)

compared directly to SAPT values. Consult Ref. [1] for some additional comparisons between dispersion models.

3.3 Arene Stacking Interactions

In a series of publications, we have used XSAPT+MBD calculations to explore the fundamental nature of π - π interactions [42, 148, 215–217], which have also been investigated using conventional SAPT [212–214, 241–243]. Although the benzene dimer has long been the archetypal model system for studying such interactions, it is not a good model for larger polycyclic aromatic hydrocarbon (PAH) dimers [148], for reasons discussed in Sect. 3.3.2. Nevertheless, understanding substituent effects on stacking energies in dimers of benzene and small heterocycles remains an important guide for synthetic chemistry [244–250]. Below, we use XSAPT+MBD to explore the origins of parallel-displaced π -stacking (Sect. 3.3.1) and then address an old question of whether π - π interactions are a unique type of interaction that is distinct from “ordinary” dispersion (Sect. 3.3.3).

3.3.1 Offset π -Stacking

Three canonical configurations of $(\text{C}_6\text{H}_6)_2$ are depicted in Fig. 27a–c. These three geometries lie within ≈ 1 kcal/mol of one another [251], with the parallel-displaced structure being slightly lower in energy than the cofacial “sandwich” structure, which is itself a saddle point between symmetry-equivalent parallel-displaced geometries [212, 240]. The T-shaped geometry is essentially iso-energetic with the parallel-displaced π -stacked structure, and may be a saddle point between tilted T-shaped structures (not shown), albeit with a negligibly small energy barrier of ~ 0.2 kcal/mol [212, 215, 252]

The conventional point of view in the organic chemistry literature is that π – π interactions are controlled by quadrupole–quadrupole interactions between the arene rings [253–255], recognizing that the benzene monomer has a rather large quadrupole moment for a small molecule [256]. This results from the π -electron density on its faces that leads to electropositive C–H bonds, as depicted in Fig. 27. The offset-stacking phenomenon, whereby two proximal arenes are much more likely to be found in a parallel-displaced arrangement than in a face-to-face-configuration [257, 258], is also typically explained in terms of quadrupolar electrostatics [253, 254]. The argument is that quadrupole–quadrupole interactions are repulsive in the face-to-face configuration but attractive in the edge-to-face arrangement, which is certainly true according to multipolar electrostatic calculations (Fig. 21) but misrepresents exact electrostatics.

This quadrupolar electrostatic viewpoint on π -stacking exemplifies how electrostatic considerations dominate the thinking in organic chemistry and how leading-order multipolar interactions are often used as surrogates for electrostatics. Not only does the multipolar picture disagree with exact electrostatics at π -stacking distances (predicting even the wrong sign for E_{elst} at $R = 3.4$ Å separation), it also cannot explain the origin of parallel-displaced π -stacking in the $(\text{C}_6\text{H}_6)\cdots(\text{C}_6\text{F}_6)$ heteromolecular dimer, which is predicted by CCSD(T) calculations [259–261] and illustrated in Fig. 27. For a symmetric top molecule, the quadrupole moment

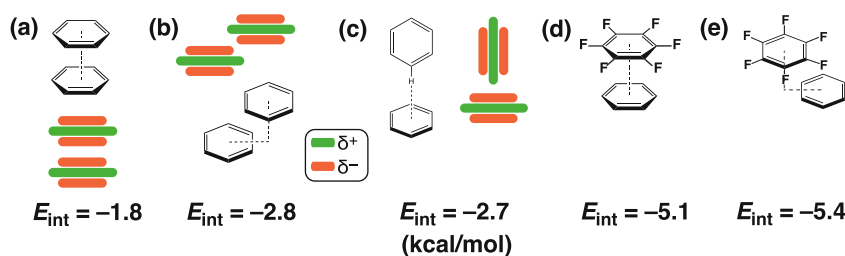


Fig. 27 Three canonical configurations of $(\text{C}_6\text{H}_6)_2$ and $(\text{C}_6\text{F}_6)\cdots(\text{C}_6\text{H}_6)$: (a) eclipsed-cofacial (“sandwich”) isomer of $(\text{C}_6\text{H}_6)_2$, (b) slip-stacked isomer of $(\text{C}_6\text{H}_6)_2$, (c) T-shaped isomer of $(\text{C}_6\text{H}_6)_2$, (d) cofacial isomer of $(\text{C}_6\text{H}_6)\cdots(\text{C}_6\text{F}_6)$, and (e) offset-stacked isomer of $(\text{C}_6\text{H}_6)\cdots(\text{C}_6\text{F}_6)$. Cartoon charge distributions are shown for the benzene dimer structures. CCSD(T)/CBS interaction energies are from Ref. [251] for $(\text{C}_6\text{H}_6)_2$ and from Ref. [260] for $(\text{C}_6\text{H}_6)\cdots(\text{C}_6\text{F}_6)$

tensor has only one independent component (call it Θ_{zz}) [207, 262], which is opposite in sign but nearly equal in magnitude in C_6F_6 and C_6H_6 : $\Theta_{zz}(C_6F_6) = +9.50 \pm 0.51$ B versus $\Theta_{zz}(C_6H_6) = -8.69 \pm 0.51$ B [256]. As a result, the quadrupole–quadrupole interaction is *attractive* in face-to-face orientations of $(C_6H_6) \cdots (C_6F_6)$ [215]. Nevertheless, accurate calculations suggest that $(C_6H_6) \cdots (C_6F_6)$ also exhibits offset π -stacking [215, 259–261].

We pause here for an aside on nomenclature and attribution. In the supramolecular organic chemistry literature, the quadrupolar electrostatic idea described above is often used synonymously with the term “Hunter–Sanders model” [217]. Although widely repeated, this misrepresents force-field models developed to explain the configurational preferences of $(C_6H_6)_2$ and other systems [257, 263–265]. Even so, Hunter himself would eventually come to describe π – π interactions in quadrupolar terms [253]. The idea that π -stacking has a quadrupolar electrostatic basis remains widespread in the synthetic chemistry literature, as discussed elsewhere [217] and evidenced by recent reviews [254, 255]. In view of this continuing misconception, we have argued that cartoon charge distributions such as those as in Fig. 27a–c are misleading because they do not adequately capture the change in quadrupolar interactions for $(C_6H_6) \cdots (C_6F_6)$ [217].

While it may be ubiquitous, the quadrupolar picture of π -stacking is also irredeemably incorrect. Charge penetration leads to a qualitative breakdown in the multipole approximation at typical vdW contact distances, including typical face-to-face π -stacking separations (3.4–3.8 Å) as seen in Fig. 21. Not only is the multipole expansion divergent as $R \rightarrow 0$ [22, 208–210], but leading-order molecular multipole approximations and even distributed (atom-centered) multipole expansions may afford the wrong sign for E_{elst} at these intermolecular distances.

To make this explicit, Fig. 28 juxtaposes two-dimensional contour plots of exact and quadrupolar electrostatics for $(C_6H_6)_2$ [217]. One of the coordinates in these plots is the intermolecular distance R , representing the face-to-face separation for parallel arrangements of the dimer and face-to-center distance for perpendicular orientations. The other coordinate is a lateral displacement that generates slip stacking in the cofacial arrangement, while in perpendicular orientations it moves the dimer between T-shaped and L-shaped geometries. The exact E_{elst} calculation in Fig. 28 uses SAPT electrostatics, computed at the LRC- ω_{GDD} PBE/def2-TZVPD level in order to be consistent with XSAPT+MBD calculations to follow, while the quadrupolar calculation uses the quadrupole tensor Θ for C_6H_6 that is obtained at the same level of theory.

Consistent with organic chemistry lore, quadrupole–quadrupole repulsion in the face-to-face orientation would push the system away from the sandwich geometry and into an offset-stacked arrangement (Fig. 28a), while quadrupolar attraction in the perpendicular orientation favors a T-shaped geometry over an L-shaped one (Fig. 28d). However, these results are inconsistent with exact electrostatic calculations. For the face-to-face arrangement, contours of (exact) E_{elst} are relatively flat along the lateral displacement coordinate (Fig. 28b), at least for typical π -stacking distances, indicating that electrostatics is largely ambivalent toward offset-stacking.

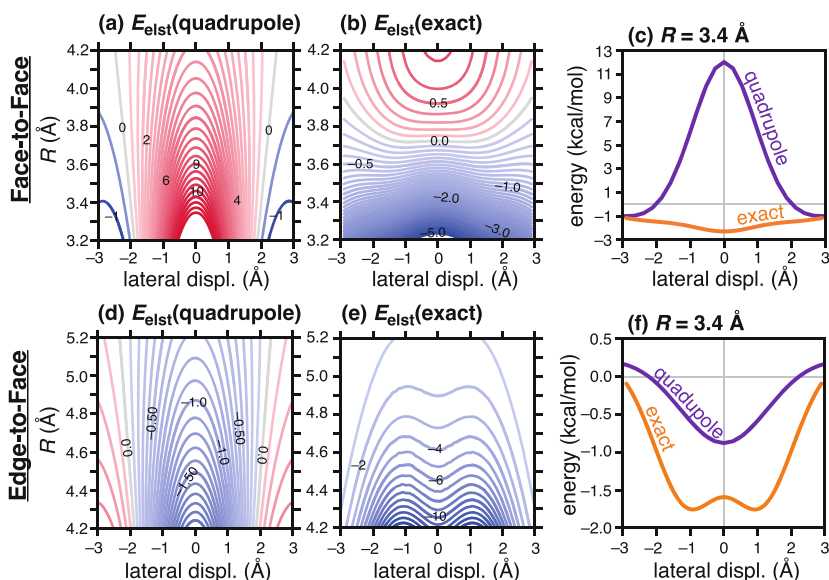


Fig. 28 Electrostatic interaction profiles for $(\text{C}_6\text{H}_6)_2$. Top row: parallel orientation, plotting (a) the quadrupole–quadrupole interaction along the face-to-face distance (R) and lateral displacement (slip-stacking) coordinates; (b) exact electrostatics along the same two coordinates, computed using full charge densities; and (c) one-dimensional slices through (a) and (b) at $R = 3.4 \text{ \AA}$, corresponding to the parallel-displaced minimum-energy geometry. Bottom row: perpendicular (edge-to-face) configuration, plotting (d) quadrupolar electrostatics along the center-to-face distance R and lateral displacement coordinates; (e) exact electrostatics; and (f) one-dimensional slices through (a) and (b) at $R = 5.0 \text{ \AA}$, corresponding to the T-shaped saddle point. Contours are drawn in intervals of (a) 0.5 kcal/mol, (b) 0.1 kcal/mol, (d) 0.1 kcal/mol, and (e) 1.0 kcal/mol. (Most data are from Ref. [217], replotted here with some additional calculations)

This is emphasized in Fig. 28c, which plots one-dimensional slices at the minimum-energy stacking distance, $R = 3.4 \text{ \AA}$. If anything, exact electrostatics favors the sandwich configuration rather than the offset-stacked geometry, albeit very weakly. Meanwhile in the edge-to-face geometry, exact electrostatics favors an L-shaped geometry with a $1 \text{ \AA}$ offset, which is seldom observed [257], rather than a T-shaped geometry with no offset.

The quadrupolar electrostatic explanation is clearly deficient, and we turn to XSAPT+MBD to explain the physical origins of slip stacking. In Fig. 29, we examine lateral displacement potentials for the parallel and perpendicular arrangements of $(\text{C}_6\text{H}_6)_2$. The total interaction potential exhibits a T-shaped minimum for the edge-to-face benzene dimer (Fig. 29a) and symmetric parallel-displaced minima in the face-to-face arrangement (Fig. 29b). The same critical points are exhibited by the *vdW* potential,

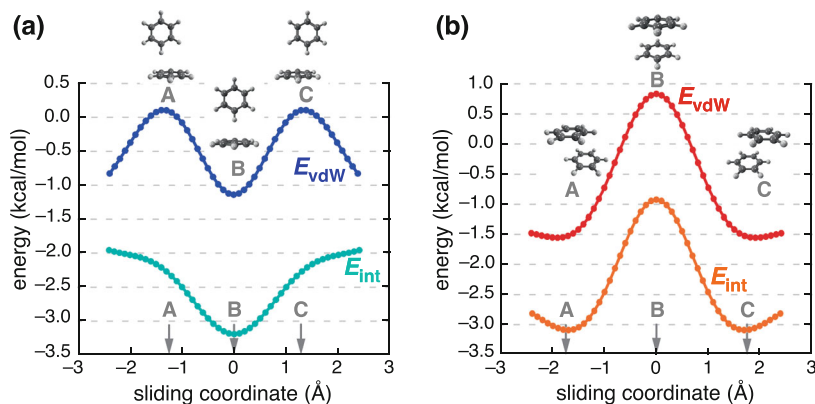


Fig. 29 Total interaction potentials versus vdW potentials ($E_{\text{vdW}} = E_{\text{exch}} + E_{\text{disp}}$) for sliding one benzene monomer past another in either (a) an edge-to-face orientation or (b) a face-to-face configuration. Arrows at the bottom of each panel indicate the position coordinate of the structures that are shown. Intermolecular distances are consistent with either (a) the T-shaped structure of $(\text{C}_6\text{H}_6)_2$, with a 5.0 Å face-to-center distance, or (b) the parallel-displaced minimum of $(\text{C}_6\text{H}_6)_2$, with a 3.4 Å face-to-face stacking distance. All calculations were performed using XSAPT+ δHF +MBD/def2-TZVPPD. (Adapted from Ref. [215] under a CC BY 3.0 license)

$$E_{\text{vdW}} = E_{\text{exch}} + E_{\text{disp}}. \quad (56)$$

This terminology recognizes that van der Waals' modifications to the ideal gas equation contained the effects of both attractive interactions (E_{disp}) and finite molecular size (E_{exch}). The data in Fig. 29 suggest that E_{vdW} is qualitatively sufficient to explain the origin of offset π -stacking in $(\text{C}_6\text{H}_6)_2$. It arises from a competition between Pauli repulsion, which favors an offset in order to reduce density overlap, and dispersion that favors spatially proximal atoms. Electrostatic interactions are not required for this geometric preference to emerge.

That is not to say that there is no role for electrostatics in π -stacking or that the vdW potential alone sets the stacking parameters. Although the attractive electrostatic contours for cofacial benzene dimer (Fig. 28b) have little influence on whether slip stacking occurs or not, they are essential for obtaining the correct face-to-face separation because they provide a stabilizing interaction at short range. Absent E_{elst} , one would obtain a stacking distance $R_{\text{min}} \approx 3.9$ Å for the parallel-displaced configuration [217], significantly larger than the correct value $R_{\text{min}} = 3.4$ Å.

To examine this in a system where electrostatics plays an even more important role, energy profile contours were computed for the nitrobenzene dimer and plotted in Fig. 30. By itself, the vdW potential affords a minimum-energy separation of $R_{\text{min}} \approx 3.8$ Å, whereas the correct value (considering all energy components) is about 3.4 Å. This difference is directly attributable to electrostatic interactions that are attractive at these distances, contrary to the quadrupolar electrostatic picture. In the case of $(\text{C}_6\text{H}_5\text{NO}_2)_2$, short-range electrostatics is attractive regardless of whether the

monomer dipole moments are aligned parallel or antiparallel; see Fig. 30c for the parallel configuration and Fig. 30f for the antiparallel arrangement.

Although the slip-stacking phenomenon arises strictly from E_{vdW} , electrostatics-driven reduction in the minimum-energy separation has general implications for π - π interactions that are explored in Sect. 3.3.3. Those implications depend critically on the fact that electrostatics is attractive at π -stacking distances, a phenomenon that is not captured—even qualitatively—by low-order multipole expansions. This is demonstrated for $(\text{C}_6\text{H}_5\text{NO}_2)_2$ in Fig. 31, which presents the results of an approximate electrostatic calculation that uses only the dipole and quadrupole moments of the two monomers. These multipolar electrostatic surfaces are qualitatively different from the exact ones. In the parallel-dipole arrangement, the multipolar approximation predicts strong electrostatic repulsion, in geometries consistent with π -stacking (Fig. 31a), whereas short-range electrostatics is always attractive when it is computed using full monomer charge densities (Fig. 30c). For the antiparallel-dipole arrangement, the multipolar approximation changes sign near zero displacement (Fig. 31b), which is also completely inconsistent with an exact calculation where there is no repulsive region at all (Fig. 30f).

A recent survey of protein crystal structures confirms this viewpoint [258]. It updates an early census of neighboring phenylalanines by Hunter et al. [257] but demonstrates that the “vdW picture” of π -stacking (driven by E_{vdW}) better explains the prevalence of parallel-displaced cofacial stacking. As we speculated earlier [215], the vdW picture provides a much more satisfactory explanation for why offset stacking is so pervasive in protein crystal structures, across disparate

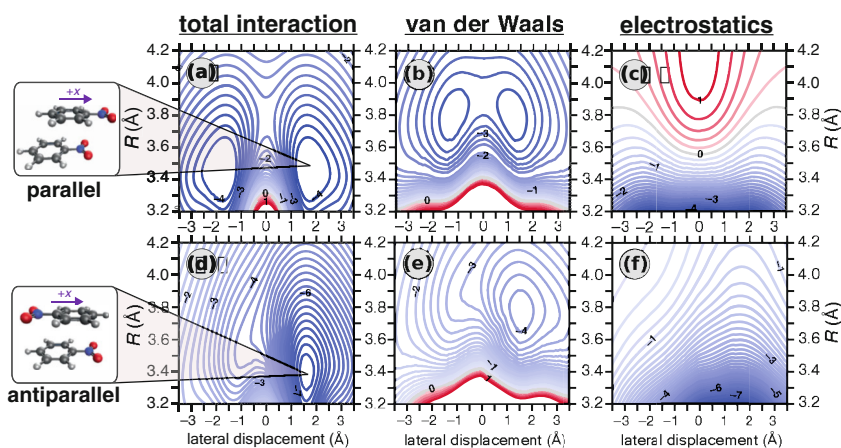


Fig. 30 Potential energy scans for $(\text{C}_6\text{H}_5\text{NO}_2)_2$ in configurations where the monomer dipole moments are either (a)–(c) parallel-aligned, or else (d)–(f) antiparallel, as shown in the structures on the left. The vertical coordinate is the cofacial separation R and the horizontal coordinate represents lateral displacement along the dipolar axis. Contours are drawn every 0.2 kcal/mol with blue and red indicating negative and positive values, respectively; selected contours are labeled in kcal/mol. (Adapted with permission from Ref. [217]; copyright 2025 American Chemical Society)

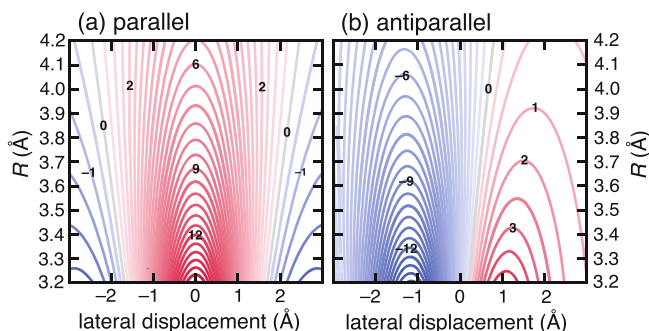


Fig. 31 Multipolar electrostatic potentials for cofacial $(C_6H_5NO_2)_2$ in orientations where the monomer dipole moments are either (a) parallel or (b) antiparallel, as in Fig. 30. In either case, the electrostatic potential is approximated using monomer dipole and quadrupole moments computed at a level of theory that is consistent with XSAPT+MBD electrostatics. The corresponding exact results, using full charge densities, can be found in Fig. 30c for the parallel configuration and in Fig. 30f for the antiparallel arrangement. (Adapted with permission from Ref. [217]; copyright 2025 American Chemical Society)

electrostatic environments. The reason is that E_{vdW} consists of only short-range intermolecular forces, which are omnipresent for arene–arene interactions regardless of the electrostatic environment. In contrast, it is more difficult to rationalize how an explanation based on longer-range electrostatic interactions could persist across the myriad electrostatic environments that are encountered within different proteins.

3.3.2 π – π Interactions in Larger Systems

Examples considered in Sect. 3.3.1 exploit XSAPT’s energy decomposition but don’t require its reduced scaling. A larger example along the same lines is benzene atop a $C_{94}H_{24}$ graphene nanoflake, for which XSAPT+MBD results are shown in Fig. 32 [42]. Here, eclipsed-cofacial stacking (analogous to the sandwich structure of benzene dimer) represents a saddle point between symmetry-equivalent slip-stacked minima, where an offset of approximately 1.4 Å lowers the energy by about 1 kcal/mol.

The vdW explanation for offset stacking persists in more exotic systems as well. For example, we have examined slip stacking in models of the covalent organic framework (COF) known as “COF1” [266], models for which are illustrated in Fig. 33a [216]. COFs are layered materials with potential applications in catalysis, but those applications require that the pore structure be open. Although the structures are likely amorphous or polycrystalline, characterized by microdomains, there has been much discussion regarding whether solvent molecules can occlude the pores [267–269] or whether even the solvent-free structure might be characterized by slippage between the layers that could block the pores [270, 271]. Using the nomenclature of graphene [272, 273], the question is whether these materials exhibit

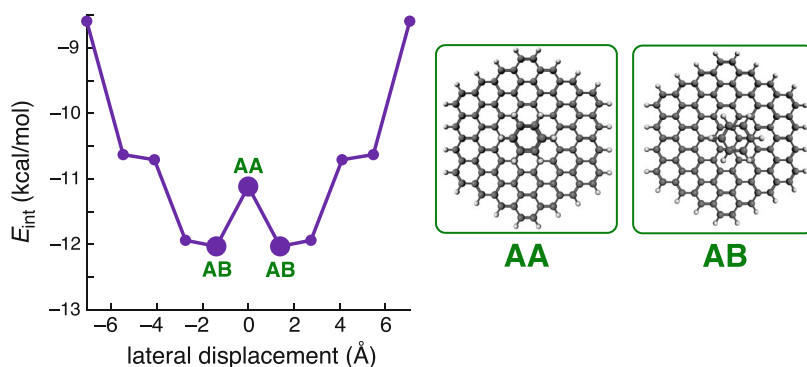


Fig. 32 Total interaction energy (E_{int}) for lateral displacement of C_6H_6 atop a $\text{C}_{94}\text{H}_{24}$ graphene nanoflake, in a cofacial orientation, computed at the XSAPT+MBD/def2-TZVPPD level. Critical points are labeled using the AA/AB nomenclature for graphene, as illustrated on the right. (Data are from Ref. [148])

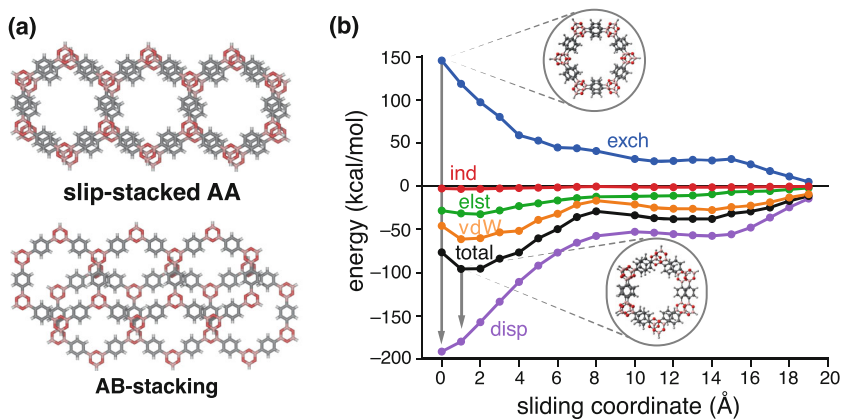


Fig. 33 (a) Three-unit dimer model of COF1, showing a slip-stacked version of AA-stacking (upper structure) as well as an AB-stacked structure in which the pores are occluded. (b) XSAPT+MBD/def2-ma-TZVP interaction potential and energy components, for lateral displacement of a one-unit dimer model at a fixed separation of 3.28 Å, corresponding to the parallel-displaced value of R_{min} .

“AA stacking” (perfectly cofacial, with pores open) or “AB stacking” (with alternating offset layers that will occlude the pores) [267–269, 274–280].

These possibilities are illustrated in Fig. 33a and can be investigated with XSAPT+MBD, whose low cost is crucial because the smallest monomer unit of COF1 is a 12-member ring composed of six-membered rings, containing 102 atoms. The XSAPT+MBD/def2-ma-TZVP calculations presented in Fig. 33b, for the dimer of that fundamental unit, are updates from previous calculations using def2-SVPD [216], taking advantage of improved memory management in the new code to

improve the basis-set quality. These energy profiles indicate a small lateral offset of $\lesssim 1.5$ Å, much smaller than the pore size. This is more consistent with benzene dimer and other small π -stacked systems than it is with true AB-stacking. As seen in Fig. 33b, the slight offset in the (COF1)₂ model originates mostly from E_{vdW} , although E_{elst} makes a nontrivial contribution to the well depth.

3.3.3 Are π - π Interactions Unique?

We next use XSAPT+MBD to consider a question that has been asked repeatedly in the literature [281–283], namely, do π - π interactions constitute a unique form of dispersion, different from that encountered in other molecules? CCSD(T) calculations on the coronene dimer (C₂₄H₁₂)₂ and perhydrocoronene dimer (C₂₄H₃₆)₂, as models of graphene and graphane, respectively, suggest that E_{int} is more than 30% larger per carbon atom in the [24]graphene model as compared to [24]graphane [284]. This effect is not captured in B97+D/6-31G(d,p) calculations, which predict similar interaction energies per carbon atom for much larger $[n]$ graphene and $[n]$ graphane models [285].

Close examination of stacking energies for acene dimers and their saturated perhydroacene analogues suggests an explanation for enhanced stacking energies in $[n]$ graphenes [148]. Figure 34 compares stacking energies for the acene and perhydroacene dimers, where the former are considered in both cofacial (parallel-displaced) and perpendicular (T-shaped) geometries. For $n \geq 2$ rings, the π -stacked acene is systematically more stable than either the C–H $\cdots\pi$ (T-shaped) acene dimer or the stacked perhydroacene dimer, by an amount that increases with the number of rings. In contrast, all three structures have very similar interaction energies for $n = 1$ rings (i.e., the benzene and cyclohexane dimers), demonstrating that results for benzene dimer are not representative of larger PAHs. These differences are confirmed in calculations of C₆H₆ atop C₉₆H₂₄, which exhibit a clear preference for cofacial π -stacking over perpendicular C–H $\cdots\pi$ structures [148].

Remarkably, stacking energies for the perhydroacene dimers with one to five cyclohexane rings are a nearly identical match to data for T-shaped acene dimers with the same number of carbon atoms [148]. This suggests that there is something special about cofacial π - π interactions specifically. To understand what it might be, we consider isosurfaces of the reduced density gradient (RDG) [286],

$$\text{RDG}(\mathbf{r}) = \left(\frac{1}{2(3\pi^2)^{1/3}} \right) \frac{\|\hat{\nabla}\rho(\mathbf{r})\|}{\rho(\mathbf{r})^{4/3}}, \quad (57)$$

which has been used extensively as an indicator of noncovalent interactions [286–291]. This works because noncovalent interactions are characterized by relatively large oscillations in the density (manifesting as large values of $\|\hat{\nabla}\rho(\mathbf{r})\|$) in regions where the density itself is small, due to disturbances in the tails of the monomer electron densities originating in the presence of the other monomer.

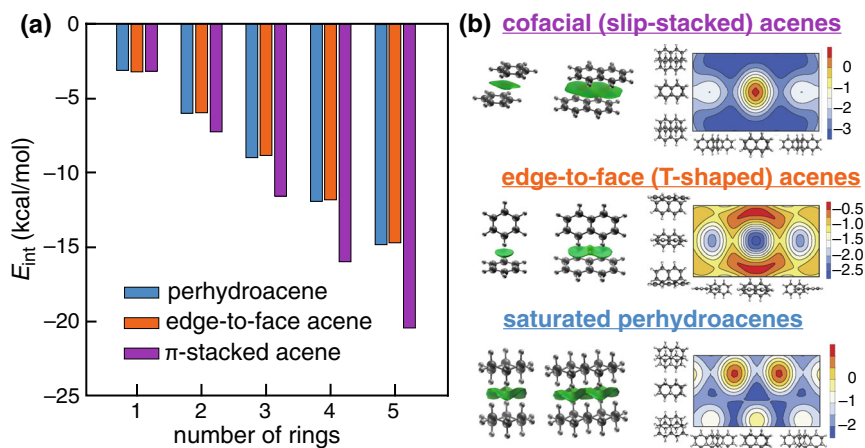


Fig. 34 (a) Interaction energies computed using XSAPT+ δ HF+MBD/def2-TZVPPD for perhydroacene dimers and for acene dimers, with the latter considered in either a parallel-displaced π -stacked or a T-shaped edge-to-face arrangement. (Adapted with permission from Ref. [148]; copyright 2020 the PCCP Owner Societies.) (b) Illustration of these three systems for $n = 1$ and $n = 2$ rings including isosurfaces of RDG($\mathbf{r}$) (in green) and lateral displacement potential energy profiles. The energy scales in (b) are in kcal/mol. (Adapted with permission from Ref. [42]; copyright 2021 American Chemical Society)

Isosurfaces of RDG($\mathbf{r}$) for the acene and perhydroacene dimers with $n = 1$ or 2 rings are plotted in Fig. 34b [148]. We observe that the RDG isosurface is corrugated for both the T-shaped acene dimer and the perhydroacene dimer, with stronger interactions when the C–H bonds of one monomer point into the middle of the ring of its partner. In contrast, the RDG isosurfaces for the cofacial acene dimers are relatively flat, meaning that strong interactions persist upon lateral displacement.

From this analysis, we conclude that π -stacking is unique among other dispersion-driven interactions, specifically due to the planarity of the arene moiety that facilitates exceptionally strong interactions when two arenes are arranged in a cofacial geometry but considerably weaker interactions (at least for PAHs) in the T-shaped geometry. The difference is that the cofacial arrangement allows for closer approach of the two arenes, which can then access short-range electrostatic interactions that are attractive in the cofacial geometry. Thus, the fact that exact electrostatics differs in sign from quadrupolar electrostatics for cofacial arenes is a crucial aspect of what makes π -stacking special, different from an ordinary dispersion interaction. Two arenes in a T-shaped geometry still interact via dispersion forces but are kept farther apart due to Pauli repulsion of the C–H $\cdots\pi$ arrangement and do not access the attractive short-range electrostatic interaction, leading to systematically weaker interaction energies as compared to those in cofacial geometries. Interactions in $[n]$ graphane models are weaker for the same reason.

Note that this explanation for what is special about π -stacking invokes planarity but not aromaticity. For arenes, the former obviously derives from the latter, yet our

analysis suggests that exceptionally strong stacking interactions should be available to systems that are planar but not aromatic. In fact, this has been observed [255, 292–295]. Our mental image is that the dispersion interactions associated with the π -electron density (or the electron density in some otherwise planar moiety) are served up on a two-dimensional platter, resembling a pizza peel, which can be rotated into an optimal coplanar configuration but whose shape cannot be deformed. If it can be, as in the perhydroacenes, then the stronger stacking interactions are lost. We refer to this as the “pizza- π ” model of stacking interactions [148]. Strong π - π interactions turn out to be more about the pizza (molecular planarity) than they are about the π (aromaticity).

3.4 Charge Transfer in Anion–Water Hydrogen Bonds

Partial covalency of hydrogen bonds has long been a topic of debate [154]. For anion–water hydrogen bonds, that debate is settled (in our view) by the observation of strong redshifts in gas-phase vibrational spectra of hydrated anion clusters [296–298]. These are the signature of $n \rightarrow \sigma_{\text{OH}}^*$ charge donation from the anion into antibonding MOs of the hydrogen-bonded O–H moiety [299, 300], which is then considerably redshifted, whereas the “free” or non-hydrogen-bonded O–H vibration is not [296–298]. Quantitative values for the fraction of an electron that is transferred vary greatly depending on the theoretical method that is used but are on the order of a few milli-electrons in the neutral water dimer [154], up to perhaps ≈ 0.2 electrons for $\text{F}^-(\text{H}_2\text{O})$ [300].

Here, we examine the role of CT in the $\text{F}^-(\text{H}_2\text{O})$ hydrogen bond using SAPT0+ δHF in conjunction with the cDFT-based procedure for separating polarization and CT that was introduced in Sect. 1.3.2 [178]. Figure 35b plots various energy components as the ion is swept radially around the H_2O molecule. Note that the C_{2v} geometry of $\text{F}^-(\text{H}_2\text{O})$, in which the ion sits at the positive end of water’s dipole moment vector, is actually a saddle point between two symmetry-equivalent minima corresponding to quasi-linear hydrogen bonds. Halide–water coordination is monodentate rather than bidentate, consistent with the interpretation of cluster ion vibrational spectroscopy [296–298], solution-phase neutron diffraction [301], and x-ray absorption experiments [302].

In contrast to the dipolar picture of anion–water interactions that remains common in textbooks [303], electrostatics alone actually favors the hydrogen-bonded structure over the C_{2v} geometry. That said, E_{exch} is maximally repulsive in the hydrogen-bonded configuration because the ion–water distance is smaller as compared to that in the C_{2v} geometry. (Note that the plots in Fig. 35 represent relaxed scans, meaning that the geometry of the complex is optimized at each fixed angle that is used to generate the one-dimensional energy profile. The small cusps that can be seen at the C_{2v} geometry are artifacts of these relaxed scans that disappear if the H_2O deformation energy is included, as explained in Ref. [178].) A plot of

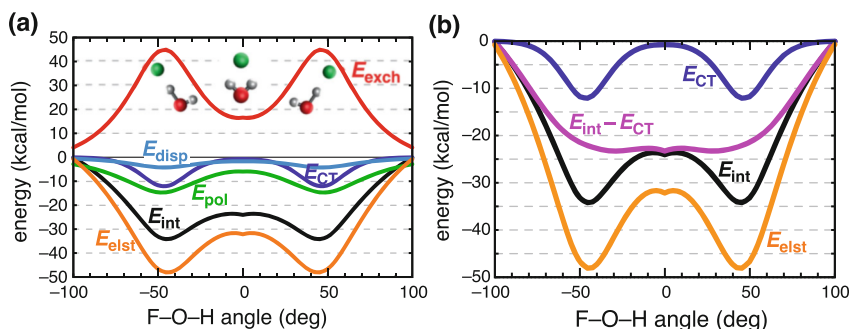


Fig. 35 (a) Energy components for $\text{F}^-(\text{H}_2\text{O})$ computed at the SAPT0+ δHF /jun-cc-pVDZ level, along a relaxed scan in which the F–O–H angle is varied while the other degrees of freedom are relaxed at the MP2/aug-cc-pVTZ level. Three representative geometries are depicted, corresponding to right and left hydrogen bonds with a C_{2v} -symmetric saddle point between them. (b) A closer look at some energy component profiles. (Reproduced with permission from Ref. [178]; copyright 2021 American Chemical Society)

$E_{\text{int}} - E_{\text{CT}}$ is nearly flat for angular displacements to either side of the C_{2v} geometry (Fig. 35b), indicating that energy components other than CT are ambivalent toward the position of the ion, so long as it resides within the angle subtended by the two O–H bonds. In contrast, E_{CT} turns on sharply for quasi-linear $\text{F}^- \cdots \text{H}-\text{O}$ geometries, implying that CT is primarily responsible for stabilizing the hydrogen-bonded configuration relative to the ion–dipole arrangement. Quasi-linearity is a defining feature of the hydrogen bond [304, 305] and arises from the strong angle dependence of E_{CT} , which can certainly be rationalized in terms of overlap between the MOs $n(\text{F})$ and $\sigma^*(\text{O}-\text{H})$, but emerges here without explicit MO considerations. The same conclusions are valid for complexes $\text{X}^-(\text{H}_2\text{O})$ with larger halide ions [178].

Encouraged by this successful description, we turned to SAPT(HF)+MBD calculations to analyze anion–water interactions for inorganic anions in aqueous solution [179]. This originates in an attempt to determine whether the solvation environment of simple anions (including halides, ClO_x^- , NO_x^- , and cyanates) is different in bulk water versus the liquid/vapor interface, and a provocative suggestion that it might not be [201]. Figure 36 shows results in the form of box-and-whisker plots representing ensemble averages of SAPT(HF)+MBD interaction energies and energy components for anion–water clusters extracted from a simulation [179]. (Each cluster contains approximately two solvation shells of water molecules that are treated as a single monomer.) For a variety of surface-active monovalent anions [306–311], we find little difference in E_{int} between the bulk and interfacial environments (Fig. 36a) [179]. Interestingly, the CT energies are also quite similar (Fig. 36b), with the possible exception of I^- , ClO^- , and SCN^- , although differences are modest even in those cases. Since E_{int} and E_{CT} are similar, it follows that polarization energies are not very different between bulk and interfacial environments [179].

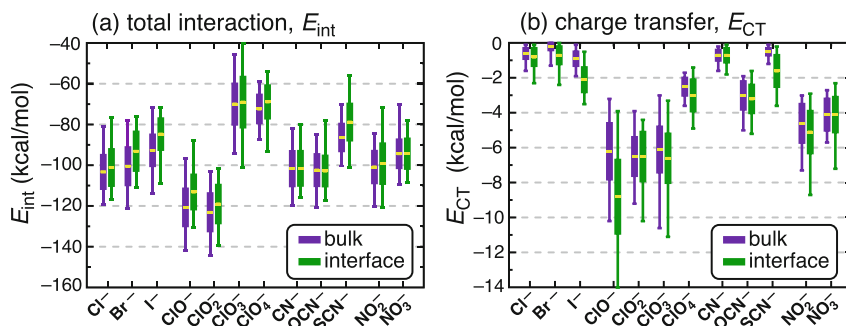


Fig. 36 Box-and-whisker plots showing mean values, standard deviations, and extremal values of (a) total interaction energies and (b) CT energies for ion–water clusters extracted from a molecular dynamics simulation. Electronic structure calculations were performed at the SAPT(HF)+MBD/6-311+G(d,p) level, and each snapshot from the simulation was binned according to whether the ion was in a bulk solvation environment or else was present at the air/water interface. (Adapted from Ref. [179] under a CC BY 4.0 license)

4 Summary

The SAPT formalism, and XSAPT with it, has quite a few moving parts that can make it intimidating to beginners, and setting up the calculations requires a variety of decisions on the user's part. In this lengthy overview of XSAPT, we have taken the time to explain carefully how those various parts work and what decisions need to be made. Philosophically, we suggest that the need for choices is a feature rather than a bug. Compared to DFT-based EDAs, which are subject to the vagaries of XC functional choice and an ambiguous definition of dispersion, (X)SAPT-based energy decomposition is well defined and improvable, in principle. Clean separability between energy components means that it is possible to test the effect that specific decisions have on individual energy components and thereby gauge the reliability of one's conclusions based on their sensitivity to this menu of choices.

That said, our preferred choice for complexes not involving ions is XSAPT +MBD/def2-ma-TZVP, with monomer wavefunctions obtained using the LRC- ω_{GDD} PBE functional, tuned in a monomer-specific way, and with CM5 embedding charges. The latter perform well for ions, and we recommend XSAPT +*ai*D3 as a starting point for ionic complexes. These methods are available in the Q-Chem code [46], with improved multithreading capabilities in recent releases.

Exemplary applications discussed herein demonstrate how the XSAPT formalism can be brought to bear on important problems in physical (and physical organic) chemistry, exposing certain fallacies regarding intermolecular interactions that persist in textbooks and in research literature. We hope that XSAPT-based calculations can start to change conventional thinking by furnishing qualitative insight based on quantitative calculations.

Acknowledgements J.M.H. thanks his former Ph.D. students, Dr. Leif Jacobson and Prof. Ka Un Lao, who developed XSAPT along with the present coauthors. The original development of XSAPT was supported by the US Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences under Award No. DE-SC0008550. Work by K.-Y.L. to improve XSAPT's efficiency was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award No. 1R43GM148095-01A1. Recent work by M.G. is supported by a Presidential Fellowship from the Ohio State University.

Competing Interests J.M.H. is part owner of Q-Chem Inc. and serves on its board of directors.

References

1. Gray M, Herbert JM (2024) Density functional theory for van der Waals complexes: Size matters. *Annu Rep Comput Chem* 20:1–61. <https://doi.org/10.1016/bs.arcc.2024.03.001>
2. Grimme S (2011) Density functional theory with London dispersion corrections. *Wiley Interdiscip Rev Comput Mol Sci* 1:211–228. <https://doi.org/10.1002/wcms.30>
3. Grimme S, Hansen A, Brandenburg JG, Bannwarth C (2016) Dispersion-corrected mean-field electronic structure methods. *Chem Rev* 116:5105–5154. <https://doi.org/10.1021/acs.chemrev.5b00533>
4. Hermann J, DiStasio Jr. RA, Tkatchenko A (2017) First-principles models for van der Waals interactions in molecules and materials: Concepts, theory, and applications. *Chem Rev* 117: 4714–4758. <https://doi.org/10.1021/acs.chemrev.6b00446>
5. Grimme S (2014) Dispersion interaction and chemical bonding. In: Frenking G, Shaik S (eds) *The chemical bond: chemical bonding across the periodic table*, chap 16. Wiley-VCH, Weinheim, pp 477–499. <https://doi.org/10.1002/9783527664658.ch16>
6. Wagner JP, Schreiner PR (2015) London dispersion in molecular chemistry—Reconsidering steric effects. *Angew Chem Int Ed Engl* 54:12274–1229. <https://doi.org/10.1002/anie.201503476>
7. Rummel L, Schreiner PR (2024) Advances and prospects in understanding London dispersion interactions in molecular chemistry. *Angew Chem Int Ed Engl* 63:e202316364. <https://doi.org/10.1002/anie.202316364>
8. Kitaura K, Morokuma K (1976) A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation. *Int J Quantum Chem* 10:325–340. <https://doi.org/10.1002/qua.560100211>
9. Bickelhaupt FM, Baerends EJ (2000) Kohn-Sham density functional theory: Predicting and understanding chemistry. In: Lipkowitz KB, Boyd DB (eds) *Reviews in Computational Chemistry*, vol 15, chap 1. Wiley-VCH, New York, pp 1–86. <https://doi.org/10.1002/9780470125922.ch1>
10. Francisco E, Pendás AM (2017) Energy partition analyses: Symmetry-adapted perturbation theory and other techniques. In: de la Roza AO, DiLabio GA (eds) *Non-Covalent Interactions in Quantum Chemistry and Physics*, chap 2. Elsevier, Amsterdam, pp 27–64. <https://doi.org/10.1016/B978-0-12-809835-6.00003-7>
11. Zhao L, von Hopffgarden M, Andrada DM, Frenking G (2018) Energy decomposition analysis. *Wiley Interdiscip Rev Comput Mol Sci* 8:e1345. <https://doi.org/10.1002/wcms.1345>
12. Su P, Tang Z, Wu W (2020) Generalized Kohn-Sham energy decomposition analysis and its applications. *Wiley Interdiscip Rev Comput Mol Sci* 10:e1460. <https://doi.org/10.1002/wcms.1460>
13. Mao Y, Loipersberger M, Horn PR, Das A, Demerdash O, Levine DS, Veccham SP, Head-Gordon T, Head-Gordon M (2021) From intermolecular interaction energies to observable shifts to component contributions and back again: A tale of variational energy decomposition

- analysis. *Annu Rev Phys Chem* 72:641–666. <https://doi.org/10.1146/annurev-physchem-090419-115149>
14. Thirman J, Head-Gordon M (2015) An energy decomposition analysis for second-order Møller–Plesset perturbation theory based on absolutely localized molecular orbitals. *J Chem Phys* 143:084124. <https://doi.org/10.1063/1.4929479>
 15. Thirman J, Head-Gordon M (2017) Efficient implementation of energy decomposition analysis for second-order Møller–Plesset perturbation theory and application to anion- π interactions. *J Phys Chem A* 121:717–728. <https://doi.org/10.1021/acs.jpca.6b11516>
 16. Bistoni G (2020) Finding chemical concepts in the Hilbert space: Coupled cluster analyses of noncovalent interactions. *Wiley Interdiscip Rev Comput Mol Sci* 10:e1442. <https://doi.org/10.1002/wcms.1442>
 17. Andrés J, Ayers PW, Boto RA, Carbó-Dorca R, Chermette H, Cioslowski J, Contreras-García J, Cooper DL, Frenking G, Gatti C, Heidar-Zadeh F, Joubert L, Pendás AM, Matito E, Mayer I, Misquitta AJ, Mo Y, Pilmé J, Popelier PLA, Rahm M, Ramos-Cordoba E, Salvador P, Schwarz WHE, Shahbazian S, Silvi B, Solà M, Szalewicz K, Tognetti V, Weinhold F, Zins EL (2019) Nine questions on energy decomposition analysis. *J Comput Chem* 40:2248–2283. <https://doi.org/10.1002/jcc.26003>
 18. Andrada DM, Foroutan-Nejad C (2020) Energy components in energy decomposition analysis (EDA) are path functions; why does it matter? *Phys Chem Chem Phys* 22:22459–22464. <https://doi.org/10.1039/D0CP04016A>
 19. Poater J, Andrada DM, Solà M, Foroutan-Nejad C (2022) Path-dependency of energy decomposition analysis & the elusive nature of bonding. *Phys Chem Chem Phys* 24:2344–2348. <https://doi.org/10.1039/D1CP04135E>
 20. Clark T (2023) How deeply should we analyze non-covalent interactions? *J Mol Model* 29:66. <https://doi.org/10.1007/s00894-023-05460-4>
 21. Rocher-Casterline BE, Ch’ng LC, Mollner AK, Reisler H (2011) Communication: Determination of the bond dissociation energy (D_0) of the water dimer, $(\text{H}_2\text{O})_2$, by velocity map imaging. *J Chem Phys* 134:211101. <https://doi.org/10.1063/1.3598339>
 22. Jeziorski B, Moszynski R, Szalewicz K (1994) Perturbation theory approach to intermolecular potential energy surfaces of van der Waals complexes. *Chem Rev* 94:1887–1930. <https://doi.org/10.1021/cr00031a008>
 23. Szalewicz K (2012) Symmetry-adapted perturbation theory of intermolecular forces. *Wiley Interdiscip Rev Comput Mol Sci* 2:254–272. <https://doi.org/10.1002/wcms.86>
 24. Jansen G (2014) Symmetry-adapted perturbation theory based on density functional theory for noncovalent interactions. *Wiley Interdiscip Rev Comput Mol Sci* 4:127–144. <https://doi.org/10.1002/wcms.1164>
 25. Hohenstein EG, Sherrill CD (2012) Wavefunction methods for noncovalent interactions. *Wiley Interdiscip Rev Comput Mol Sci* 2:304–326. <https://doi.org/10.1002/wcms.84>
 26. Parker TM, Burns LA, Parrish RM, Ryno AG, Sherrill CD (2014) Levels of symmetry adapted perturbation theory (SAPT). I. Efficiency and performance for interaction energies. *J Chem Phys* 140:094106. <https://doi.org/10.1063/1.4867135>
 27. Schriber J, Cheney D, Sherrill CD (2023) Levels of symmetry adapted perturbation theory (SAPT). II. Convergence of interaction energy components. *ChemRxiv*. <https://doi.org/10.26434/chemrxiv-2023-ftt1v>
 28. Patkowski K (2020) Recent developments in symmetry-adapted perturbation theory. *Wiley Interdiscip Rev Comput Mol Sci* 10:e1452. <https://doi.org/10.1002/wcms.1452>
 29. Gao J (1998) A molecular-orbital derived polarization potential for liquid water. *J Chem Phys* 109:2346–2354. <https://doi.org/10.1063/1.476802>
 30. Xie W, Song L, Truhlar DG, Gao J (2008) The variational explicit polarization potential and analytical first derivative of energy: Towards a next generation force field. *J Chem Phys* 128:234108. <https://doi.org/10.1063/1.2936122>

31. Wang Y, Mazack MJM, Truhlar DG, Gao J (2016) Explicit polarization theory. In: Cui Q, Meuwly M, Ren P (eds) *Many-Body Effects and Electrostatics in Biomolecules*, chap 2. Pan Stanford, Boca Raton, FL, pp 33–64. <https://doi.org/10.1201/b21343-5>
32. Jacobson LD, Herbert JM (2011) An efficient, fragment-based electronic structure method for molecular systems: Self-consistent polarization with perturbative two-body exchange and dispersion. *J Chem Phys* 134:094118. <https://doi.org/10.1063/1.3560026>
33. Herbert JM, Jacobson LD, Lao KU, Rohrdanz MA (2012) Rapid computation of intermolecular interactions in molecular and ionic clusters: Self-consistent polarization plus symmetry-adapted perturbation theory. *Phys Chem Chem Phys* 14:7679–7699. <https://doi.org/10.1039/C2CP24060B>
34. Lao KU, Herbert JM (2012) Accurate intermolecular interactions at dramatically reduced cost: XPol+SAPT with empirical dispersion. *J Phys Chem Lett* 3:3241–3248. <https://doi.org/10.1021/jz301015p>
35. Jacobson LD, Richard RM, Lao KU, Herbert JM (2013) Efficient monomer-based quantum chemistry methods for molecular and ionic clusters. *Annu Rep Comput Chem* 9:25–58. <https://doi.org/10.1016/B978-0-444-62672-1.00002-9>
36. Lao KU, Herbert JM (2013) An improved treatment of empirical dispersion and a many-body energy decomposition scheme for the explicit polarization plus symmetry-adapted perturbation theory (XSAPT) method. *J Chem Phys* 139:034107. <https://doi.org/10.1063/1.4813523>
37. Lao KU, Herbert JM (2015) Accurate and efficient quantum chemistry calculations of noncovalent interactions in many-body systems: The XSAPT family of methods. *J Phys Chem A* 119:235–253. <https://doi.org/10.1021/jp5098603>
38. Lao KU, Herbert JM (2018) Atomic orbital implementation of extended symmetry-adapted perturbation theory (XSAPT) and benchmark calculations for large supramolecular complexes. *J Chem Theory Comput* 14:2955–2978. <https://doi.org/10.1021/acs.jctc.8b00058>
39. Lao KU, Herbert JM (2018) A simple correction for nonadditive dispersion within extended symmetry-adapted perturbation theory (XSAPT). *J Chem Theory Comput* 14:5128–5142. <https://doi.org/10.1021/acs.jctc.8b00527>
40. Carter-Fenk K, Lao KU, Liu KY, Herbert JM (2019) Accurate and efficient *ab initio* calculations for supramolecular complexes: Symmetry-adapted perturbation theory with many-body dispersion. *J Phys Chem Lett* 10:2706–2714. <https://doi.org/10.1021/acs.jpclett.9b01156>
41. Liu KY, Carter-Fenk K, Herbert JM (2019) Self-consistent charge embedding at very low cost, with application to symmetry-adapted perturbation theory. *J Chem Phys* 151:031102. <https://doi.org/10.1063/1.5111869>
42. Carter-Fenk K, Lao KU, Herbert JM (2021) Predicting and understanding non-covalent interactions using novel forms of symmetry-adapted perturbation theory. *Acc Chem Res* 54:3679–3690. <https://doi.org/10.1021/acs.accounts.1c00387>
43. Herbert JM (2021) Neat, simple, and wrong: Debunking electrostatic fallacies regarding noncovalent interactions. *J Phys Chem A* 125:7125–7137. <https://doi.org/10.1021/acs.jpca.1c05962>
44. Patkowski K (2017) Benchmark databases of intermolecular interaction energies: Design, construction, and significance. *Annu Rep Comput Chem* 13:3–91. <https://doi.org/10.1016/bs.arcc.2017.06.004>
45. Shao Y, Gan Z, Epifanovsky E, Gilbert ATB, Wormit M, Kussmann J, Lange AW, Behn A, Deng J, Feng X, Ghosh D, Goldey M, Horn PR, Jacobson LD, Kaliman I, Khaliullin RZ, Kúš T, Landau A, Liu J, Proynov EI, Rhee YM, Richard RM, Rohrdanz MA, Steele RP, Sundstrom EJ, Woodcock III HL, Zimmerman PM, Zuev D, Albrecht B, Alguire E, Austin B, Beran GJO, Bernard YA, Berquist E, Brandhorst K, Bravaya KB, Brown ST, Casanova D, Chang CM, Chen Y, Chien SH, Closser KD, Crittenden DL, Diedenhofen M, DiStasio Jr. RA, Do H, Dutoi AD, Edgar RG, Fatehi S, Fusti-Molnar L, Ghysels A, Golubeva-Zadorozhnaya A, Gomes J, Hanson-Heine MWD, Harbach PHP, Hauser AW, Hohenstein EG, Holden ZC, Jagau TC, Ji H, Kaduk B, Khistyayev K, Kim J, Kim J, King RA, Klunzinger P, Kosenkov D, Kowalczyk T, Krauter CM, Lao KU, Laurent AD, Lawler KV, Levchenko SV, Lin CY, Liu F,

- Livshits E, Lochan RC, Luenser A, Manohar P, Manzer SF, Mao SP, Mardirossian N, Marenich AV, Maurer SA, Mayhall NJ, Neuscamman E, Oana CM, Olivares-Amaya R, O'Neill DP, Parkhill JA, Perrine TM, Peverati R, Prociuk A, Rehn DR, Rosta E, Russ NJ, Sharada SM, Sharma S, Small DW, Sodt A, Stein T, Stück D, Su YC, Thom AJW, Tsuchimochi T, Vanovschi V, Vogt L, Vydrov O, Wang T, Watson MA, Wenzel J, White A, Williams CF, Yang J, Yeganeh S, Yost SR, You ZQ, Zhang IY, Zhang X, Zhao Y, Brooks BR, Chan GKL, Chipman DM, Cramer CJ, Goddard III WA, Gordon MS, Hehre WJ, Klamt A, Schaefer III HF, Schmidt MW, Sherrill CD, Truhlar DG, Warshel A, Xu X, Aspuru-Guzik A, Baer R, Bell AT, Besley NA, Chai JD, Dreuw A, Dunietz BD, Furlani TR, Gwaltney SR, Hsu CP, Jung Y, Kong J, Lambrecht DS, Liang W, Ochsenfeld C, Rassolov VA, Slipchenko LV, Subotnik JE, Van Voorhis T, Herbert JM, Krylov AI, Gill PMW, Head-Gordon M (2015) Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol Phys* 113:184–215. <https://doi.org/10.1080/00268976.2014.952696>
46. Epifanovsky E, Gilbert ATB, Feng X, Lee J, Mao Y, Mardirossian N, Pokhilko P, White AF, Coons MP, Dempwolff AL, Gan Z, Hait D, Horn PR, Jacobson LD, Kaliman I, Kussmann J, Lange AW, Lao KU, Levine DS, Liu J, McKenzie SC, Morrison AF, Nanda KD, Plasser F, Rehn DR, Vidal ML, You ZQ, Zhu Y, Alam B, Albrecht BJ, Aldossary A, Alguire E, Andersen JH, Athavale V, Barton D, Begam K, Behn A, Bellonzi N, Bernard YA, Berquist EJ, Burton HGA, Carreras A, Carter-Fenk K, Chakraborty R, Chien AD, Closser KD, Cofer-Shabica V, Dasgupta S, de Wergifosse M, Deng J, Didenhofen M, Do H, Ehlert S, Fang PT, Fatehi S, Feng Q, Friedhoff T, Gayvert J, Ge Q, Gidofalvi G, Goldey M, Gomes J, González-Espinoza CE, Gulania S, Gunina AO, Hanson-Heine MWD, Harbach PHP, Hauser A, Herbst MF, Hernández Vera M, Hodecker M, Holden ZC, Houck S, Huang X, Hui K, Huynh BC, Ivanov M, Jász A, Ji H, Jiang H, Kaduk B, Kähler S, Khistyayev K, Kim J, Kis G, Klunzinger P, Koczor-Benda Z, Koh JH, Kosenkov D, Koulis L, Kowalczyk T, Krauter CM, Kue K, Kunitsa A, Kus T, Ladjánszki I, Landau A, Lawler KV, Lefrançois D, Lehtola S, Li RR, Li YP, Liang J, Liebenthal M, Lin HH, Lin YS, Liu F, Liu KY, Loipersberger M, Luenser A, Manjanath A, Manohar P, Mansoor E, Manzer SF, Mao SP, Marenich AV, Markovich T, Mason S, Maurer SA, McLaughlin PF, Menger MFSJ, Mewes JM, Mewes SA, Morgante P, Mullinax JW, Oosterbaan KJ, Paran G, Paul AC, Paul SK, Pavošević F, Pei Z, Prager S, Proynov EI, Rák A, Ramos-Cordoba E, Rana B, Rask AE, Rettig A, Richard RM, Rob F, Rossomme E, Scheele T, Scheurer M, Schneider M, Sergueev N, Sharada SM, Skomorowski W, Small DW, Stein CJ, Su YC, Sundstrom EJ, Tao Z, Thirman J, Tornai GJ, Tsuchimochi T, Tubman NM, Veccham SP, Vydrov O, Wenzel J, Witte J, Yamada A, Yao K, Yeganeh S, Yost SR, Zech A, Zhang IY, Zhang X, Zhang Y, Zuev D, Aspuru-Guzik A, Bell AT, Besley NA, Bravaya KB, Brooks BR, Casanova D, Chai JD, Coriani S, Cramer CJ, Cserey G, DePrince III AE, DiStasio Jr. RA, Dreuw A, Dunietz BD, Furlani TR, Goddard III WA, Hammes-Schiffer S, Head-Gordon T, Hehre WJ, Hsu CP, Jagau TC, Jung Y, Klamt A, Kong J, Lambrecht DS, Liang W, Mayhall NJ, McCurdy CW, Neaton JB, Ochsenfeld C, Parkhill JA, Peverati R, Rassolov VA, Shao Y, Slipchenko LV, Stauch T, Steele RP, Subotnik JE, Thom AJW, Tkatchenko A, Truhlar DG, Van Voorhis T, Wesolowski TA, Whaley KB, Woodcock III HL, Zimmerman PM, Faraji S, Gill PMW, Head-Gordon M, Herbert JM, Krylov AI (2021) Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *J Chem Phys* 155:084801. <https://doi.org/10.1063/5.0055522>
47. Moszynski R (2007) Theory of intermolecular forces: An introductory account. In: Sokalski WA (ed) *Molecular Materials and Specific Interactions—Modeling and Design*. Springer, New York, pp 1–152. https://doi.org/10.1007/1-4020-5372-X_1
48. Jeziorski B, Bulski M, Piela L (1976) First-order perturbation treatment of the short-range repulsion in a system of many closed-shell atoms or molecules. *Int J Quantum Chem* 10:281–297. <https://doi.org/10.1002/qua.560100208>
49. Schäffer R, Jansen G (2012) Intermolecular exchange-induction energies without overlap expansion. *Theor Chem Acc* 131:1235. <https://doi.org/10.1007/s00214-012-1235-6>

50. Schäffer R, Jansen G (2013) Single-determinant-based symmetry-adapted perturbation theory without single-exchange approximation. *Mol Phys* 111:2570–2584. <https://doi.org/10.1080/00268976.2013.827253>
51. Hirschfelder JO (1967) Perturbation theory for exchange forces, I. *Chem Phys Lett* 1:325–329. [https://doi.org/10.1016/0009-2614\(67\)80007-1](https://doi.org/10.1016/0009-2614(67)80007-1)
52. Jeziorski B, Moszynski R, Ratkiewicz A, Rybak S, Szalewicz K, Williams HL (1993) SAPT: A program for many-body symmetry-adapted perturbation theory calculations of intermolecular interaction energies. In: Clementi E (ed) *Methods and Techniques in Computational Chemistry: METECC-94*, vol B, chap 3. STEF, Cagliari, pp 79–129
53. Szalewicz K, Patkowski K, Jeziorski B (2005) Intermolecular interactions via perturbation theory: From diatoms to biomolecules. In: Wales DJ (ed) *Intermolecular Forces and Clusters II, Structure and Bonding*, vol 116. Springer, Berlin, pp 43–117. https://doi.org/10.1007/430_004
54. Szalewicz K, Jeziorski B (2022) Physical mechanisms of intermolecular interactions from symmetry-adapted perturbation theory. *J Mol Model* 28:273. <https://doi.org/10.1007/s00894-022-05190-z>
55. Gray M, Herbert JM (2021) Simplified tuning of long-range corrected density functionals for symmetry-adapted perturbation theory. *J Chem Phys* 155:034103. <https://doi.org/10.1063/5.0059364>
56. Moszynski R, Jeziorski B, Ratkiewicz A, Rybak S (1993) Many-body perturbation theory of electrostatic interactions between molecules: Comparison with full configuration interaction for four-electron dimers. *J Chem Phys* 99:8856–8869. <https://doi.org/10.1063/1.465554>
57. Patkowski K, Szalewicz K, Jeziorski B (2006) Third-order interactions in symmetry-adapted perturbation theory. *J Chem Phys* 125:154107. <https://doi.org/10.1063/1.2358353>
58. Patkowski K, Szalewicz K, Jeziorski B (2010) Orbital relaxation and the third-order induction energy in symmetry-adapted perturbation theory. *Theor Chem Acc* 127:211–221. <https://doi.org/10.1007/s00214-010-0748-0>
59. Gray M, Bowling PE, Herbert JM (2022) Systematic examination of counterpoise correction in density functional theory. *J Chem Theory Comput* 18:6742–6756. <https://doi.org/10.1021/acs.jctc.2c00883>
60. Altun A, Neese F, Bistoni G (2019) HFLD: A nonempirical London dispersion-corrected Hartree–Fock method for the quantification and analysis of noncovalent interaction energies of large molecular systems. *J Chem Theory Comput* 15:5894–5907. <https://doi.org/10.1021/acs.jctc.9b00425>
61. Garcia J, Szalewicz K (2020) Ab initio extended Hartree–Fock plus dispersion method applied to dimers with hundreds of atoms. *J Phys Chem A* 124:1196–1203. <https://doi.org/10.1021/acs.jpca.9b11900>
62. Bistoni G, Altun A, Wang Z, Neese F (2024) Local energy decomposition analysis of London dispersion effects: From simple model dimers to complex biomolecular assemblies. *Acc Chem Res* 57:1411–1420. <https://doi.org/10.1021/acs.accounts.4c00085>
63. Conway A, Murrell JN (1974) The exchange energy between two neon atoms. *Mol Phys* 27: 873–878. <https://doi.org/10.1080/00268977400100791>
64. Cohen JS, Pack RT (1974) Modified statistical method for intermolecular potentials. Combining rules for higher van der Waals coefficients. *J Chem Phys* 61:2372–2382. <https://doi.org/10.1063/1.1682318>
65. Hepburn J, Scoles G, Penco R (1975) A simple but reliable method for the prediction of intermolecular potentials. *Chem Phys Lett* 36:451–456. [https://doi.org/10.1016/0009-2614\(75\)80278-8](https://doi.org/10.1016/0009-2614(75)80278-8)
66. Ahlrichs R, Penco R, Scoles G (1977) Intermolecular forces in simple systems. *Chem Phys* 19: 119–130. [https://doi.org/10.1016/0301-0104\(77\)85124-0](https://doi.org/10.1016/0301-0104(77)85124-0)
67. Douketis C, Scoles G, Marchetti S, Zen M, Thakkar AJ (1982) Intermolecular forces via hybrid Hartree–Fock–SCF plus damped dispersion (HFD) energy calculations. An improved spherical model. *J Chem Phys* 76:3057–3063. <https://doi.org/10.1063/1.443345>

68. Hobza P, Zahradník R (1983) Van der Waals molecules: Quantum chemistry, physical properties, and reactivity. *Int J Quantum Chem* 23:325–338. <https://doi.org/10.1002/qua.560230128>
69. Douketis C, Hutson JM, Orr BJ, Scoles G (1984) Anisotropic intermolecular forces from Hartree-Fock plus damped dispersion (HFD) calculations. *Mol Phys* 52:763–781. <https://doi.org/10.1080/00268978400101541>
70. Cybulski SM, Lytle ML (2007) The origin of the deficiency of the supermolecule second-order Møller-Plesset approach for evaluating interaction energies. *J Chem Phys* 127:141102. <https://doi.org/10.1063/1.2795693>
71. Nguyen B, Chen GP, Agee MM, Burow AM, Tang M, Furche F (2020) Divergence of many-body perturbation theory for noncovalent interactions of large molecules. *J Chem Theory Comput* 16:2258–2273. <https://doi.org/10.1021/acs.jctc.9b01176>
72. Szabo A, Ostlund NS (1977) The correlation energy in the random phase approximation: Intermolecular forces between closed-shell systems. *J Chem Phys* 67:4351–4360. <https://doi.org/10.1063/1.434580>
73. Řezáč J, Riley KE, Hobza P (2011) S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures. *J Chem Theory Comput* 7:2427–2438. <https://doi.org/10.1021/ct2002946>
74. Papajak E, Zheng J, Xu X, Leverentz HR, Truhlar DG (2011) Perspectives on basis sets beautiful: Seasonal plantings of diffuse basis functions. *J Chem Theory Comput* 7:3027–3034. <https://doi.org/10.1021/ct200106a>
75. Sherrill CD (2009) Computations of noncovalent π interactions. In: Lipkowitz KB, Cundari TR (eds) *Reviews in Computational Chemistry*, vol 26, chap 1. Wiley-VCH, Weinheim, pp 1–38. <https://doi.org/10.1002/9780470399545.ch1>
76. Hohenstein EG, Sherrill CD (2010) Density fitting of intramonomer correlation effects in symmetry adapted perturbation theory. *J Chem Phys* 133:014101. <https://doi.org/10.1063/1.3451077>
77. Gray M, Herbert JM (2022) Comprehensive basis-set testing of extended symmetry-adapted perturbation theory and assessment of mixed-basis combinations to reduce cost. *J Chem Theory Comput* 18:2308–2330. <https://doi.org/10.1021/acs.jctc.1c01302>
78. Heßelmann A, Jansen G, Schütz M (2005) Density-functional theory symmetry-adapted intermolecular perturbation theory with density fitting: A new efficient method to study intermolecular interaction energies. *J Chem Phys* 122:014103. <https://doi.org/10.1063/1.1824898>
79. Lao KU, Herbert JM (2014) Symmetry-adapted perturbation theory with Kohn-Sham orbitals using non-empirically tuned, long-range-corrected density functionals. *J Chem Phys* 140:044108. <https://doi.org/10.1063/1.4862644>
80. Khaliullin RZ, Cobar EA, Lochan RC, Bell AT, Head-Gordon M (2007) Unravelling the origin of intermolecular interactions using absolutely localized molecular orbitals. *J Phys Chem A* 111:8753–8765. <https://doi.org/10.1021/jp073685z>
81. Williams HL, Mas EM, Szalewicz K, Jeziorski B (1995) On the effectiveness of monomer-, dimer-, and bond-centered basis functions in calculations of intermolecular interaction energies. *J Chem Phys* 103:7374–7391. <https://doi.org/10.1063/1.470309>
82. Khaliullin RZ, Head-Gordon M, Bell AT (2006) An efficient self-consistent field method for large systems of weakly-interacting fragments. *J Chem Phys* 124:204105. <https://doi.org/10.1063/1.2191500>
83. Rybak S, Jeziorski B, Szalewicz K (1991) Many-body symmetry-adapted perturbation theory of intermolecular interactions. H_2O and HF dimers. *J Chem Phys* 95:6576–6601. <https://doi.org/10.1063/1.461528>
84. Misquitta AJ, Bukowski R, Szalewicz K (2000) Spectra of Ar-CO_2 from *ab initio* potential energy surfaces. *J Chem Phys* 112:5308–5319. <https://doi.org/10.1063/1.481120>

85. Williams HL, Chabalowski CF (2001) Using Kohn–Sham orbitals in symmetry-adapted perturbation theory to investigate intermolecular interactions. *J Phys Chem A* 105:646–659. <https://doi.org/10.1021/jp003883p>
86. Jansen G, Heßelmann A (2001) Comment on “Using Kohn–Sham orbitals in symmetry-adapted perturbation theory to investigate intermolecular interactions”. *J Phys Chem A* 105: 11156–11157. <https://doi.org/10.1021/jp0112774>
87. Misquitta AJ, Szalewicz K (2002) Intermolecular forces from asymptotically corrected density functional description of monomers. *Chem Phys Lett* 357:301–306. [https://doi.org/10.1016/S0009-2614\(02\)00533-X](https://doi.org/10.1016/S0009-2614(02)00533-X)
88. Heßelmann A, Jansen G (2002) First-order intermolecular interaction energies from Kohn–Sham orbitals. *Chem Phys Lett* 357:464–470. [https://doi.org/10.1016/S0009-2614\(02\)00538-9](https://doi.org/10.1016/S0009-2614(02)00538-9)
89. Heßelmann A, Jansen G (2002) Intermolecular induction and exchange-induction energies from coupled-perturbed Kohn–Sham density functional theory. *Chem Phys Lett* 362:319–325. [https://doi.org/10.1016/S0009-2614\(02\)01097-7](https://doi.org/10.1016/S0009-2614(02)01097-7)
90. Perdew JP, Burke K (1996) Comparison shopping for a gradient-corrected density functional. *Int J Quantum Chem* 57:309–319. [https://doi.org/10.1002/\(SICI\)1097-461X\(1996\)57:3%3C309::AID-QUA4%3E3.0.CO;2-1](https://doi.org/10.1002/(SICI)1097-461X(1996)57:3%3C309::AID-QUA4%3E3.0.CO;2-1)
91. Tozer DJ, Handy NC (1998) Improving virtual Kohn–Sham orbitals and eigenvalues: Application to excitation energies and static polarizabilities. *J Chem Phys* 109:10180–10189. <https://doi.org/10.1063/1.477711>
92. Perdew JP, Parr RG, Levy M, Balduz, Jr. JL (1982) Density-functional theory for fractional particle number: Derivative discontinuities of the energy. *Phys Rev Lett* 49:1691–1694. <https://doi.org/10.1103/PhysRevLett.49.1691>
93. Levy M, Perdew JP, Sahni V (1984) Exact differential equation for the density and ionization energy of a many-particle system. *Phys Rev A* 30:2745–2748. <https://doi.org/10.1103/PhysRevA.30.2745>
94. Perdew JP (1985) What do the Kohn–Sham orbitals energies mean? How do atoms dissociate? In: Dreizler RM, da Providência J (eds) *Density Functional Methods in Physics*. NATO Advanced Science Institute Series, vol 123. Springer, pp 265–308. https://doi.org/10.1007/978-1-4757-0818-9_10
95. Perdew JP, Levy M (1997) Comment on “Significance of the highest occupied Kohn–Sham eigenvalue”. *Phys Rev B* 56:16021–16028. <https://doi.org/10.1103/PhysRevB.56.16021>
96. van Leeuwen R, Baerends EJ (1994) Exchange-correlation potential with correct asymptotic behavior. *Phys Rev A* 49:2421–2431. <https://doi.org/10.1103/PhysRevA.49.2421>
97. Grüning M, Gritsenko OV, van Gisbergen SJA, Baerends EJ (2001) Shape corrections to exchange-correlation potentials by gradient-regulated seamless connection of model potentials for inner and outer region. *J Chem Phys* 114:652–660. <https://doi.org/10.1063/1.1327260>
98. Gaiduk AP, Staroverov VN (2009) How to tell when a model Kohn–Sham potential is not a functional derivative. *J Chem Phys* 131:044107. <https://doi.org/10.1063/1.3176515>
99. Baer R, Livshits E, Salzner U (2010) Tuned range-separated hybrids in density functional theory. *Annu Rev Phys Chem* 61:85–109. <https://doi.org/10.1146/annurev.physchem.012809.103321>
100. Tsuneda T, Hirao K (2014) Long-range correction for density functional theory. *Wiley Interdiscip Rev Comput Mol Sci* 4:375–390. <https://doi.org/10.1002/wcms.1178>
101. Herbert JM (2023) Density-functional theory for electronic excited states. In: García-Iriepa C, Marazzi M (eds) *Theoretical and Computational Photochemistry: Fundamentals, Methods, Applications and Synergy with Experimental Approaches*, chap 3. Elsevier, Amsterdam, pp 69–118. <https://doi.org/10.1016/B978-0-323-91738-4.00005-1>
102. Rohrdanz MA, Herbert JM (2008) Simultaneous benchmarking of ground- and excited-state properties with long-range-corrected density functional theory. *J Chem Phys* 129:034107. <https://doi.org/10.1063/1.2954017>

103. Rohrdanz MA, Martins KM, Herbert JM (2009) 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* 130:054112. <https://doi.org/10.1063/1.3073302>
104. Alam B, Morrison AF, Herbert JM (2020) Charge separation and charge transfer in the low-lying excited states of pentacene. *J Phys Chem C* 124:24653–24666. <https://doi.org/10.1021/acs.jpcc.0c07932>
105. Modrzejewski M, Rajchel Ł, Chalasinski G, Szczesniak MM (2013) Density-dependent onset of the long-range exchange: A key to donor–acceptor properties. *J Phys Chem A* 117:11580–11586. <https://doi.org/10.1021/jp4088404>
106. Mandal A, Herbert JM (2025) Simplified tuning of long-range corrected time-dependent density functional theory. *J Phys Chem Lett* 16:2672–2680. <https://doi.org/10.1021/acs.jpclett.5c00086>
107. Karolewski A, Kronik L, Kümmel S (2013) Using optimally tuned range separated hybrid functionals in ground-state calculations: Consequences and caveats. *J Chem Phys* 138:204115. <https://doi.org/10.1063/1.4807325>
108. Phipps MJS, Fox T, Tautermann CS, Skylaris CK (2015) Energy decomposition analysis approaches and their evaluation on prototypical protein–drug interaction patterns. *Chem Soc Rev* 44:3177–3211. <https://doi.org/10.1039/c4cs00375f>
109. Herbert JM (2019) Fantasy versus reality in fragment-based quantum chemistry. *J Chem Phys* 151:170901. <https://doi.org/10.1063/1.5126216>
110. Řezáč, J, Huang Y, Hobza P, Beran GJO (2015) Benchmark calculations of three-body intermolecular interactions and the performance of low-cost electronic structure methods. *J Chem Theory Comput* 11:3065–3079. <https://doi.org/10.1021/acs.jctc.5b00281>
111. Alkan M, Xu P, Gordon MS (2019) Many-body dispersion in molecular clusters. *J Phys Chem A* 123:8406–8416. <https://doi.org/10.1021/acs.jpca.9b05977>
112. Yang J, Hu W, Usvyat D, Matthews D, Schütz M, Chan GKL (2014) Ab initio determination of the crystalline benzene lattice energy to sub-kilojoule/mole accuracy. *Science* 345:640–643. <https://doi.org/10.1126/science.1254419>
113. Kennedy MR, McDonald AR, DePrince III AE, Marshall MS, Podeszwa R, Sherrill CD (2014) Communication: Resolving the three-body contribution to the lattice energy of crystalline benzene: Benchmark results from coupled-cluster theory. *J Chem Phys* 140:121104. <https://doi.org/10.1063/1.4869686>
114. Borca CH, Glick ZL, Metcalf DP, Burns LA, Sherrill CD (2023) Benchmark coupled-cluster lattice energy of crystalline benzene and assessment of multi-level approximations in the many-body expansion. *J Chem Phys* 158:234102. <https://doi.org/10.1063/5.0159410>
115. Chen Y, Li H (2010) Intermolecular interaction in water hexamer. *J Phys Chem A* 114:11719–11724. <https://doi.org/10.1021/jp104822e>
116. Cisneros GA, Wikfeldt KT, Ojamäe L, Lu J, Xu Y, Torabifard H, Bartók AP, Csányi G, Molinero V, Paesani F (2016) Modeling molecular interactions in water: From pairwise to many-body potential energy functions. *Chem Rev* 116:7501–7528. <https://doi.org/10.1021/acs.chemrev.5b00644>
117. Jacobson LD, Williams CF, Herbert JM (2009) The static-exchange electron-water pseudopotential, in conjunction with a polarizable water model: A new Hamiltonian for hydrated-electron simulations. *J Chem Phys* 130:124115. <https://doi.org/10.1063/1.3089425>
118. Breneman CM, Wiberg KB (1990) Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J Comput Chem* 11:361–373. <https://doi.org/10.1002/jcc.540110311>
119. Holden ZC, Richard RM, Herbert JM (2013) Periodic boundary conditions for QM/MM calculations: Ewald summation for extended Gaussian basis sets. *J Chem Phys* 139:244108. <https://doi.org/10.1063/1.4850655>

120. Marenich AV, Jerome SV, Cramer CJ, Truhlar DG (2012) 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* 8:527–541. <https://doi.org/10.1021/ct200866d>
121. Hirshfeld FL (1977) XVII. Spatial partitioning of charge density. *Isr J Chem* 16:198–201. <https://doi.org/10.1002/ijch.197700033>
122. Tkatchenko A, DiStasio Jr. RA, Car R, Scheffler M (2012) Accurate and efficient method for many-body van der Waals interactions. *Phys Rev Lett* 108:236402. <https://doi.org/10.1103/PhysRevLett.108.236402>
123. Tkatchenko A, Ambrosetti A, DiStasio Jr. RA (2013) Interatomic methods for the dispersion energy derived from the adiabatic connection fluctuation-dissipation theorem. *J Chem Phys* 138:074106. <https://doi.org/10.1063/1.4789814>
124. Ambrosetti A, Reilly AM, DiStasio Jr. RA, Tkatchenko A (2014) Long-range correlation energy calculated from coupled atomic response functions. *J Chem Phys* 140:18A508. <https://doi.org/10.1063/1.4865104>
125. Carter-Fenk K, Herbert JM (2023) Appraisal of dispersion damping functions for the effective fragment potential method. *Mol Phys* 121:e2055504. <https://doi.org/10.1080/00268976.2022.2055504>
126. Heßelmann A (2011) Comparison of intermolecular interaction energies from SAPT and DFT including empirical dispersion contributions. *J Phys Chem A* 115:11321–11330. <https://doi.org/10.1021/jp205031e>
127. Schriber JB, Sirianni DA, Smith DGA, Burns LA, Sitkoff D, Cheney DL, Sherrill CD (2021) Optimized damping parameters for empirical dispersion corrections to symmetry-adapted perturbation theory. *J Chem Phys* 154:234107. <https://doi.org/10.1063/5.0049745>
128. Wallace AM, Sherrill CD (2024) Optimization of damping function parameters for -D3 and -D4 dispersion models for Hartree–Fock based symmetry-adapted perturbation theory. *J Chem Phys* 161:114115. <https://doi.org/10.1063/5.0219185>
129. Grimme S, Antony J, Ehrlich S, Krieg H (2010) A consistent and accurate *ab initio* parameterization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J Chem Phys* 132:154104. <https://doi.org/10.1063/1.3382344>
130. Caldeweyher E, Bannwarth C, Grimme S (2017) Extension of the D3 dispersion coefficient model. *J Chem Phys* 147:034112. <https://doi.org/10.1063/1.4993215>
131. Grimme S (2006) Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J Comput Chem* 27:1787–1799. <https://doi.org/10.1002/jcc.20495>
132. Sedlak R, Janowski T, Pitoňák M, Řezáč J, Pulay P, Hobza P (2013) Accuracy of quantum chemical methods for large noncovalent complexes. *J Chem Theory Comput* 9:3364–3374. <https://doi.org/10.1021/ct400036b>
133. Goldey M, Head-Gordon M (2012) Attenuating away the errors in inter- and intramolecular interactions from second-order Møller–Plesset calculations in the small aug-cc-pVDZ basis set. *J Phys Chem Lett* 3:3592–3598. <https://doi.org/10.1021/jz301694b>
134. Dobson JF (2014) Beyond pairwise additivity in London dispersion interactions. *Int J Quantum Chem* 114:1157–1161. <https://doi.org/10.1002/qua.24635>
135. Tkatchenko A, Scheffler M (2009) Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. *Phys Rev Lett* 102:073005. <https://doi.org/10.1103/PhysRevLett.102.073005>
136. von Lilienfeld OA, Tkatchenko A (2010) Two- and three-body interatomic dispersion energy contributions to binding in molecules and solids. *J Chem Phys* 132:234109. <https://doi.org/10.1063/1.3432765>
137. Proynov E, Liu F, Gan Z, Wang M, Kong J (2015) Density-functional approach to the three-body dispersion interaction based on the exchange dipole moment. *J Chem Phys* 143:084125. <https://doi.org/10.1063/1.4929581>
138. Al-Saidi WA, Voora VK, Jordan KD (2012) An assessment of the vdW-TS method for extended systems. *J Chem Theory Comput* 8:1503–1513. <https://doi.org/10.1021/ct200618b>

139. Chu X, Dalgarno A (2004) Linear response time-dependent density functional theory for van der Waals coefficients. *J Chem Phys* 121:4083–4088. <https://doi.org/10.1063/1.1779576>
140. De Proft F, Vivas-Reyes R, Peeters A, Van Alsenoy C, Geerlings P (2003) Hirshfeld partitioning of the electron density: Atomic dipoles and their relation with functional group properties. *J Comput Chem* 24:463–469. <https://doi.org/10.1002/jcc.10241>
141. Krishtal A, Senet P, Yan M, Van Alsenoy C (2006) A Hirshfeld partitioning of polarizabilities of water clusters. *J Chem Phys* 125:034312. <https://doi.org/10.1063/1.2210937>
142. Olasz A, Vanommeslaeghe K, Krishtal A, Vesszprémi T, Van Alsenoy C, Geerlings P (2007) The use of atomic intrinsic polarizabilities in the evaluation of the dispersion energy. *J Chem Phys* 127:224105. <https://doi.org/10.1063/1.2805391>
143. Brinck T, Murray JS, Politzer P (1993) Polarizability and volume. *J Chem Phys* 98:4305–4306. <https://doi.org/10.1063/1.465038>
144. Szabó P, Góger S, Charry J, Karimpour MR, Fedorov DV, Tkatchenko A (2022) Four-dimensional scaling of dipole polarizability in quantum systems. *Phys Rev Lett* 128:070602. <https://doi.org/10.1103/PhysRevLett.128.070602>
145. Longuet-Higgins HC (1965) Intermolecular forces. *Discuss Faraday Soc* 40:7–18. <https://doi.org/10.1039/DF9654000007>
146. Herbert JM (2021) Dielectric continuum methods for quantum chemistry. *Wiley Interdiscip Rev Comput Mol Sci* 11:e1519. <https://doi.org/10.1002/wcms.1519>
147. Ambrosetti A, Ferri N, DiStasio Jr. RA, Tkatchenko A (2016) Wavelike charge density fluctuations and van der Waals interactions at the nanoscale. *Science* 351:1171–1176. <https://doi.org/10.1126/science.aae0509>
148. Carter-Fenk K, Herbert JM (2020) Reinterpreting π -stacking. *Phys Chem Chem Phys* 22:24870–24886. <https://doi.org/10.1039/D0CP05039C>
149. Bryenton KR, Johnson ER (2023) Many-body dispersion in model systems and the sensitivity of self-consistent screening. *J Chem Phys* 158:204110. <https://doi.org/10.1063/5.0142465>
150. Jurečka P, Šponer J, Černý J, Hobza P (2006) 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* 8:1985–1993. <https://doi.org/10.1039/B600027D>
151. Flick JC, Kosenkov D, Hohenstein EG, Sherrill CD, Slipchenko LV (2012) 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* 8:2835–2843. <https://doi.org/10.1021/ct200673a>
152. Grafova L, Pitonak M, Řezáč J, Hobza P (2010) Comparative study of selected wave function and density functional methods for noncovalent interaction energy calculations using the extended S22 data set. *J Chem Theory Comput* 6:2365–2376. <https://doi.org/10.1021/ct1002253>
153. Zheng J, Xu X, Truhlar DG (2011) Minimally augmented Karlsruhe basis sets. *Theor Chem Acc* 128:295–305. <https://doi.org/10.1007/s00214-010-0846-z>
154. Ronca E, Belpassi L, Tarantelli F (2014) A quantitative view of charge transfer in the hydrogen bond: The water dimer case. *ChemPhysChem* 15:2682–2687. <https://doi.org/10.1002/cphc.201402321>
155. Stone AJ (2017) Natural bond orbitals and the nature of the hydrogen bond. *J Phys Chem A* 121:1531–1534. <https://doi.org/10.1021/acs.jpca.6b12930>
156. Mao Y, Ge Q, Horn PR, Head-Gordon M (2018) On the computational characterization of charge-transfer effects in noncovalently bound molecular complexes. *J Chem Theory Comput* 14:2401–2417. <https://doi.org/10.1021/acs.jctc.7b01256>
157. Taylor R (2024) Aerogen bond, halogen bond, chalcogen bond, pnictogen bond, tetrel bond, triel bond... why so many names? *Crys Growth Des* 24:4003–4012. <https://doi.org/10.1021/acs.cgd.4c00303>
158. Cavallo G, Metrangolo P, Milani R, Pilati T, Priimagi A, Resnati G, Terraneo G (2016) The halogen bond. *Chem Rev* 116:2478–2601. <https://doi.org/10.1021/acs.chemrev.5b00484>

159. Řezáč J, de la Lande A (2017) On the role of charge transfer in halogen bonding. *Phys Chem Chem Phys* 19:791–803. <https://doi.org/10.1039/C6CP07475H>
160. Thirman J, Engelage E, Huber SM, Head-Gordon M (2018) Characterizing the interplay of Pauli repulsion, electrostatics, dispersion and charge transfer in halogen bonding with energy decomposition analysis. *Phys Chem Chem Phys* 20:905–915. <https://doi.org/10.1039/C7CP06959F>
161. Clark T, Heßelmann A (2018) The coulombic σ -hole model describes bonding in $CX_3I \cdots Y^-$ complexes completely. *Phys Chem Chem Phys* 20:22849–22855. <https://doi.org/10.1039/C8CP03079K>
162. Kellett CW, Kennepohl P, Berlinguette CP (2020) π covalency in the halogen bond. *Nat. Commun.* 11:3310. <https://doi.org/10.1038/s41467-020-17122-7>
163. Inscoe B, Rathnayake H, Mo Y (2021) Role of charge transfer in halogen bonding. *J Phys Chem A* 125:2944–2953. <https://doi.org/10.1021/acs.jpca.1c01412>
164. Murray JS, Politzer P (2021) Can counter-intuitive halogen bonding be Coulombic? *ChemPhysChem* 22:1201–1207. <https://doi.org/10.1002/cphc.202100202>
165. Mo Y, Danovich D, Shaik S (2022) The roles of charge transfer and polarization in non-covalent interactions: A perspective from ab initio valence bond methods. *J Mol Model* 28:274. <https://doi.org/10.1007/s00894-022-05187-8>
166. Feynman RP (1939) Forces in molecules. *Phys Rev* 56:340–343. <https://doi.org/10.1103/PhysRev.56.340>
167. Politzer P, Murray JS (2015) A unified view of halogen bonding, hydrogen bonding and other σ -hole interactions. In: Scheiner S (ed) *Noncovalent forces. Challenges and Advances in Computational Chemistry and Physics*, vol 19, chap 10. Springer International Publishing, Switzerland, pp 291–321. https://doi.org/10.1007/978-3-319-14163-3_10
168. Clark T, Politzer P, Murray JS (2015) Correct electrostatic treatment of noncovalent interactions: The importance of polarization. *Wiley Interdiscip Rev Comput Mol Sci* 5:169–177. <https://doi.org/10.1002/wcms.1210>
169. Clark T (2017) Polarization, donor–acceptor interactions, and covalent contributions in weak interactions: A clarification. *J Mol Model* 23:297. <https://doi.org/10.1007/s00894-017-3473-y>
170. Clark T, Murray JS, Politzer P (2018) A perspective on quantum mechanics and chemical concepts in describing noncovalent interactions. *Phys Chem Chem Phys* 20:30076–30082. <https://doi.org/10.1039/c8cp06786d>
171. Anderson PW (1972) More is different: Broken symmetry and the nature of the hierarchical structure of science. *Science* 177:393–396. <https://doi.org/10.1126/science.177.4047.393>
172. Thouless DJ (1960) Stability conditions and nuclear rotations in the Hartree-Fock theory. *Nucl Phys* 21:225–232. [https://doi.org/10.1016/0029-5582\(60\)90048-1](https://doi.org/10.1016/0029-5582(60)90048-1)
173. Lao KU, Herbert JM (2016) Energy decomposition analysis with a stable charge-transfer term for interpreting intermolecular interactions. *J Chem Theory Comput* 12:2569–2582. <https://doi.org/10.1021/acs.jctc.6b00155>
174. Misquitta AJ (2013) Charge transfer from regularized symmetry-adapted perturbation theory. *J Chem Theory Comput* 9:5313–5326. <https://doi.org/10.1021/ct400704a>
175. Řezáč J, de la Lande A (2015) Robust, basis-set independent method for the evaluation of charge-transfer energy in noncovalent complexes. *J Chem Theory Comput* 11:528–537. <https://doi.org/10.1021/ct501115m>
176. Stone AJ, Misquitta AJ (2009) Charge-transfer in symmetry-adapted perturbation theory. *Chem Phys Lett* 473:201–205. <https://doi.org/10.1016/j.cplett.2009.03.073>
177. Kaduk B, Kowalczyk T, Van Voorhis T (2012) Constrained density functional theory. *Chem Rev* 112:321–370. <https://doi.org/10.1021/cr200148b>
178. Herbert JM, Carter-Fenk K (2021) Electrostatics, charge transfer, and the nature of the halide–water hydrogen bond. *J Phys Chem A* 125:1243–1256. <https://doi.org/10.1021/acs.jpca.0c11356>

179. Herbert JM, Paul SK (2021) Interaction energy analysis of monovalent inorganic anions in bulk water versus air/water interface. *Molecules* 26:6719. <https://doi.org/10.3390/molecules26216719>
180. Horn PR, Head-Gordon M (2015) Polarization contributions to intermolecular interactions revisited with fragment electric-field response functions. *J Chem Phys* 143:114111. <https://doi.org/10.1063/1.4930534>
181. Ochieng SA, Patkowski K (2023) Accurate three-body noncovalent interactions: The insights from energy decomposition. *Phys Chem Chem Phys* 25:28621–28637. <https://doi.org/10.1039/D3CP03938B>
182. van Meer R, Gritsenko OV, Baerends EJ (2014) Physical meaning of virtual Kohn–Sham orbitals and orbital energies: An ideal basis for the description of molecular excitations. *J Chem Theory Comput* 10:4432–4441. <https://doi.org/10.1021/ct500727c>
183. Rappoport D, Furche F (2010) Property-optimized Gaussian basis sets for molecular response calculations. *J Chem Phys* 133:134105. <https://doi.org/10.1063/1.3484283>
184. Sure R, Grimme S (2013) Corrected small basis set Hartree-Fock method for large systems. *J Comput Chem* 34:1672–1685. <https://doi.org/10.1002/jcc.23317>
185. Grimme S, Brandenburg JG, Bannwarth C, Hansen A (2015) Consistent structures and interactions by density functional theory with small atomic orbital basis sets. *J Chem Phys* 143:054107. <https://doi.org/10.1063/1.4927476>
186. Grimme S (2012) Supramolecular binding thermodynamics by dispersion-corrected density functional theory. *Chem Eur J* 18:9955–9964. <https://doi.org/10.1002/chem.201200497>
187. Sure R, Grimme S (2015) Comprehensive benchmark of association (free) energies of realistic host–guest complexes. *J Chem Theory Comput* 11:3785–3801. <https://doi.org/10.1021/acs.jctc.5b00296>
188. Mardirossian N, Head-Gordon M (2017) Thirty years of density functional theory in computational chemistry: An overview and extensive assessment of 200 density functionals. *Mol Phys* 115:2315–2372. <https://doi.org/10.1080/00268976.2017.1333644>
189. Lao KU, Schäffer R, Jansen G, Herbert JM (2015) Accurate description of intermolecular interactions involving ions using symmetry-adapted perturbation theory. *J Chem Theory Comput* 11:2473–2486. <https://doi.org/10.1021/ct5010593>
190. Gould T, Lebègue S, Ángyán JG, Bučko T (2016) A fractionally ionic approach to polarizability and van der Waals many-body dispersion calculations. *J Chem Theory Comput* 12:5920–5930. <https://doi.org/10.1021/acs.jctc.6b00925>
191. Hermann J, Tkatchenko A (2020) Density functional model for van der Waals interactions: Unifying many-body atomic approaches with nonlocal functionals. *Phys Rev Lett* 124:146401. <https://doi.org/10.1103/PhysRevLett.124.146401>
192. Hermann J, Stöhr M, Góger S, Chaudhuri S, Aradi B, Maurer RJ, Tkatchenko A (2023) libMBD: A general-purpose package for scalable quantum many-body dispersion calculations. *J Chem Phys* 159:174802. <https://doi.org/10.1063/5.0170972>
193. Bultinck P, Van Alsenoy C, Ayers PW, Carbó-Dorca R (2007) Critical analysis and extension of the Hirshfeld atoms in molecules. *J Chem Phys* 126:144111. <https://doi.org/10.1063/1.2715563>
194. Bučko T, Lebègue S, Hafner J, Ángyán JG (2013) Improved density dependent correction for the description of London dispersion forces. *J Chem Theory Comput* 9:4293–4299. <https://doi.org/10.1021/ct400694h>
195. Finzel K, Martín Pendás A, Francisco E (2015) Efficient algorithms for Hirshfeld-I charges. *J Chem Phys* 143:084115. <https://doi.org/10.1063/1.4929469>
196. Bučko T, Lebègue S, Ángyán JG, Hafner J (2014) Extending the applicability of the Tkatchenko-Scheffler dispersion correction via iterative Hirshfeld partitioning. *J Chem Phys* 141:034114. <https://doi.org/10.1063/1.4890003>
197. Hohenstein EG, Sherrill CD (2010) Density fitting and Cholesky decomposition approximations in symmetry-adapted perturbation theory: Implementation and application to probe the

- nature of π - π interactions in linear acenes. *J Chem Phys* 132:184111. <https://doi.org/10.1063/1.3426316>
198. Hohenstein EG, Parrish RM, Sherrill CD, Turney JM, Schaefer III HF (2011) Large-scale symmetry-adapted perturbation theory computations via density fitting and Laplace transformation techniques: Investigating the fundamental forces of DNA-intercalator interactions. *J Chem Phys* 135:174107. <https://doi.org/10.1063/1.3656681>
199. Herbert JM, Head-Gordon M (2004) Curvy-steps approach to constraint-free extended-Lagrangian *ab initio* molecular dynamics, using atom-centered basis functions: Convergence toward Born-Oppenheimer trajectories. *J Chem Phys* 121:11542–11556. <https://doi.org/10.1063/1.1814934>
200. Gray M, Bowling PE, Herbert JM (2024) Comment on “Benchmarking basis sets for density functional theory thermochemistry calculations: Why unpolarized basis sets and the polarized 6-311G family should be avoided”. *J Phys Chem A* 128:7739–7745. <https://doi.org/10.1021/acs.jpca.4c00283>
201. Paul SK, Herbert JM (2021) Probing interfacial effects on ionization energies: The surprising banality of anion–water hydrogen bonding at the air/water interface. *J Am Chem Soc* 143:10189–10202. <https://doi.org/10.1021/jacs.1c03131>
202. Moran D, Simmonett AC, Leach FE, Allen WD, Schleyer PvR, Schaefer HF (2006) Popular theoretical methods predict benzene and arenes to be nonplanar. *J Am Chem Soc* 128:9342–9343. <https://doi.org/10.1021/ja0630285>
203. Grimme S, Steinmetz M, Korth M (2007) How to compute isomerization energies of organic molecules with quantum chemical methods. *J Org Chem* 72:2118–2126. <https://doi.org/10.1021/jo062446p>
204. Kruse H, Goerigk L, Grimme S (2012) Why the standard B3LYP/6-31G* model chemistry should not be used in DFT calculations of molecular thermochemistry: Understanding and correcting the problem. *J Org Chem* 77:10824–10834. <https://doi.org/10.1021/jo302156p>
205. Morgante P, Peverati R (2020) The devil in the details: A tutorial review on some undervalued aspects of density functional theory calculations. *Int J Quantum Chem* 120:e26332. <https://doi.org/10.1002/qua.26332>
206. Gavini V, Baroni S, Blum V, Bowler DR, Buccheri A, Chelikowsky JR, Das S, Dawson W, Delugas P, Dogan M, Draxl C, Galli G, Genovese L, Giannozzi P, Giantomassi M, Gonze X, Govoni M, Gygi F, Gulans A, Herbert JM, Kokott S, Kühne TD, Liou KH, Miyazaki T, Motamarri P, Nakata A, Pask JE, Plessl C, Ratcliff LE, Richard RM, Rossi M, Schade R, Scheffler M, Schütt O, Suryanarayana P, Torrent M, Truflandier L, Windus TL, Xu Q, Yu VWZ, Perez D (2023) Roadmap on electronic structure codes in the exascale era. *Model Simul Mater Sci Eng* 31:063301. <https://doi.org/10.1088/1361-651X/acdf06>
207. Stone AJ (2013) *The Theory of Intermolecular Forces*, 2nd edn. Oxford University Press, Oxford, UK. https://doi.org/10.1007/978-3-642-58183-0_4
208. Vigné-Maeder F, Claverie P (1988) The exact multicenter multipolar part of a molecular charge distribution and its simplified representations. *J Chem Phys* 88:4934–4948. <https://doi.org/10.1063/1.454705>
209. Stone AJ (1991) Classical electrostatics in molecular interactions. In: Maksić ZB (ed) *Theoretical Treatment of Large Molecules and Their Interactions*. Springer, Berlin, pp 103–131. https://doi.org/10.1007/978-3-642-58183-0_4
210. Kaplan IG (2006) *Intermolecular Interactions: Physical Picture, Computational Methods and Model Potentials*. Wiley, Chichester, UK. <https://doi.org/10.1002/047086334X>
211. Ryno SM, Risko C, Brédas JL (2016) Noncovalent interactions and impact of charge penetration effects in linear oligoacene dimers and single crystals. *Chem Mater* 28:3990–4000. <https://doi.org/10.1021/acs.chemmater.6b01340>
212. Podeszwa R, Bukowski R, Szalewicz K (2006) Potential energy surface for the benzene dimer and perturbational analysis of π - π interactions. *J Phys Chem A* 110:10345–10354. <https://doi.org/10.1021/jp064095o>

213. Sherrill CD (2013) Energy component analysis of π interactions. *Acc Chem Res* 46:1020–1028. <https://doi.org/10.1021/ar3001124>
214. Hohenstein EG, Duan J, Sherrill DC (2011) Origin of the surprising enhancement of electrostatic energies by electron-donating substituents in substituted sandwich benzene dimers. *J Am Chem Soc* 133:13244–13247. <https://doi.org/10.1021/ja204294q>
215. Carter-Fenk K, Herbert JM (2020) Electrostatics does not dictate the slip-stacked arrangement of aromatic π - π interactions. *Chem Sci* 11:6758–6765. <https://doi.org/10.1039/D0SC02667K>
216. Gray M, Herbert JM (2023) Origins of offset-stacking in porous frameworks. *J Phys Chem C* 127:2675–2686. <https://doi.org/10.1021/acs.jpcc.2c08413>
217. Schramm B, Gray M, Herbert JM (2025) Substituent and heteroatom effects on π - π interactions: Evidence that parallel-displaced π -stacking is not driven by quadrupolar electrostatics. *J Am Chem Soc* 147:3243–3260. <https://doi.org/10.1021/jacs.4c13291>
218. Franklin RE (1951) The structure of graphitic carbon. *Acta Cryst* 4:253–261. <https://doi.org/10.1107/S0365110X51000842>
219. Baldrige KK, Siegel JS (1997) Corannulene-based fullerene fragments $C_{20}H_{10}$ - $C_{50}H_{10}$: When does a buckybowl become a buckytube? *Theor Chem Acc* 97:67–71. <https://doi.org/10.1007/s002140050238>
220. Scott LT, Bronstein HE, Preda DV, Ansems RBM, Bratcher MS, Hagen S (1999) Geodesic polyarenes with exposed concave surfaces. *Pure Appl Chem* 71:209–219. <https://doi.org/10.1351/pac199971020209>
221. Stuparu MC (2021) Corannulene: A curved polyarene building block for the construction of functional materials. *Acc Chem Res* 54:2858–2870. <https://doi.org/10.1021/acs.accounts.1c00207>
222. Hanson JC, Nordman CE (1976) The crystal and molecular structure of corannulene, $C_{20}H_{10}$. *Acta Cryst B* 32:1147–1153. <https://doi.org/10.1107/S0567740876012430>
223. Filatov AS, Scott LT, Petrukhina MA (2010) π - π interactions and solid state packing trends of polycyclic aromatic bowls in the indenocorannulene family: Predicting potentially useful bulk properties. *Crys Growth Des* 10:4607–4621. <https://doi.org/10.1021/cg100898g>
224. Chen R, Lu RQ, Shi PC, Cao XY (2016) Corannulene derivatives for organic electronics: From molecular engineering to applications. *Chin Chem Lett* 27:1175–1183. <https://doi.org/10.1016/j.ccllet.2016.06.033>
225. Kennedy MR, Burns LA, Sherrill CD (2012) Buckyplates and buckybowls: Examining the effects of curvature on π - π interactions. *J Phys Chem A* 116:11920–11926. <https://doi.org/10.1021/jp305700k>
226. Lovas FJ, McMahon RJ, Grabow JU, Schnell M, Mack J, Scott LT, Kuczkowski RL (2005) Interstellar chemistry: A strategy for detecting polycyclic aromatic hydrocarbons in space. *J Am Chem Soc* 127:4345–4349. <https://doi.org/10.1021/ja0426239>
227. Kvashnin AG, Sorokin PB, Yakobson BI (2015) Flexoelectricity in carbon nanostructures: Nanotubes, fullerenes, and nanocones. *J Phys Chem Lett* 6:2740–2744. <https://doi.org/10.1021/acs.jpcclett.5b01041>
228. Martin JW, Slavchov RI, Yapp EKY, Akroyd J, Mosbach S, Kraft M (2017) The polarization of polycyclic aromatic hydrocarbons curved by pentagon incorporation: The role of the flexoelectric dipole. *J Phys Chem C* 121:27154–27163. <https://doi.org/10.1021/acs.jpcc.7b09044>
229. Smith DW (1998) A simple molecular orbital treatment of the barrier to internal rotation in the ethane molecule. *J Chem Educ* 75:907–909. <https://doi.org/10.1021/ed075p907>
230. Zhou P, Huang J, Tian F (2012) Specific noncovalent interactions at protein–ligand interface: Implications for rational drug design. *Curr Med Chem* 19:226–238. <https://doi.org/10.2174/092986712803414150>
231. Sauers RR (1996) Steric attraction: The far side. *J Chem Educ* 73:114–116. <https://doi.org/10.1021/ed073p114>

232. Wolters LP, Koekkoek R, Bickelhaupt FM (2015) Role of steric attraction and bite-angle flexibility in metal-mediated C–H bond activation. *ACS Catal* 5:5766–5775. <https://doi.org/10.1021/acscatal.5b01354>
233. Gryn'ova G, Corminboeuf C (2018) Steric “attraction”: Not by dispersion alone. *Beilstein J Org Chem* 14:1482–1490. <https://doi.org/10.3762/bjoc.14.125>
234. Holtrop F, Visscher KW, Jupp AR, Slootweg J (2020) Steric attraction: A force to be reckoned with. *Adv Phys Org Chem* 54:119–141. <https://doi.org/10.1016/bs.apoc.2020.08.001>
235. Hansen T, Sun X, Dalla Tiezza M, van Zeist WJ, van Stralen JNP, Geerke DP, Wolters LP, Poater J, Hamlin TA, Bickelhaupt FM (2022) C–X bond activation by palladium: Steric shielding versus steric attraction. *Chem Eur J* 28:e202201093. <https://doi.org/10.1002/chem.202201093>
236. Villot C, Ballesteros F, Wang D, Lao KU (2022) Coupled cluster benchmarking of large noncovalent complexes in L7 and S12L as well as the C₆₀ dimer, DNA–ellipticine, and HIV–indinavir. *J Phys Chem A* 126:4326–4341. <https://doi.org/10.1021/acs.jpca.2c01421>
237. Ballesteros F, Dunivan S, Lao KU (2021) Coupled cluster benchmarks of large noncovalent complexes: The L7 dataset as well as DNA–ellipticine and buckycatcher–fullerene. *J Chem Phys* 154:154104. <https://doi.org/10.1063/5.0042906>
238. London F (1937) The general theory of intermolecular forces. *Trans Faraday Soc* 33:8–26. <https://doi.org/10.1039/TF937330008B>
239. Truhlar DG (2019) Dispersion forces: Neither fluctuating nor dispersing. *J Chem Educ* 96:1671–1675. <https://doi.org/10.1021/acs.jchemed.8b01044>
240. Sinnokrot MO, Sherrill CD (2004) Highly accurate coupled cluster potential energy curves for the benzene dimer: Sandwich, T-shaped, and parallel-displaced configurations. *J Phys Chem A* 108:10200–10207. <https://doi.org/10.1021/jp0469517>
241. Hohenstein EG, Sherrill DC (2009) Effects of heteroatoms on aromatic π – π interactions: Benzene–pyridine and pyridine dimer. *J Phys Chem A* 113:878–886. <https://doi.org/10.1021/jp809062x>
242. Geng Y, Takatani T, Hohenstein EG, Sherrill CD (2010) Accurately characterizing π – π interaction energies of indole–benzene complexes. *J Phys Chem A* 114:3576–3582. <https://doi.org/10.1021/jp9099495>
243. Parrish RM, Sherrill CD (2014) 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* 136:17386–17389. <https://doi.org/10.1021/ja5101245>
244. Wheeler SE, Houk KN (2008) Substituent effects in the benzene dimer are due to direct interactions of the substituents with the unsubstituted benzene. *J Am Chem Soc* 130:10854–10855. <https://doi.org/10.1021/ja802849j>
245. Wheeler SE, Houk KN (2009) Through-space effects of substituents dominate molecular electrostatic potentials of substituted arenes. *J Chem Theory Comput* 5:2301–2312. <https://doi.org/10.1021/ct900344g>
246. Wheeler SE, Houk KN (2009) Origin of substituent effects in edge-to-face aryl–aryl interactions. *Mol Phys* 107:749–760. <https://doi.org/10.1080/00268970802537614>
247. Wheeler SE (2011) Local nature of substituent effects in stacking interactions. *J Am Chem Soc* 133:10262–10274. <https://doi.org/10.1021/ja202932e>
248. Raju RK, Bloom JW, Wheeler SE (2013) Broad transferability of substituent effects in π -stacking interactions provides new insights into their origin. *J Chem Theory Comput* 9:3479–3490. <https://doi.org/10.1021/ct400481r>
249. Wheeler SE, Bloom JW (2014) Toward a more complete understanding of noncovalent interactions involving aromatic rings. *J Phys Chem A* 118:6133–6147. <https://doi.org/10.1021/jp504415p>
250. Bootsma AN, Doney AC, Wheeler SE (2019) Predicting the strength of stacking interactions between heterocycles and aromatic amino acid side chains. *J Am Chem Soc* 141:11027–11035. <https://doi.org/10.1021/jacs.9b00936>

251. Sinnokrot MO, Valeev EF, Sherrill CD (2002) Estimates of the ab initio limit for π - π interactions: The benzene dimer. *J Am Chem Soc* 124:10887–10893. <https://doi.org/10.1021/ja025896h>
252. Bludský O, Rubeš M, Soldán P, Nachtigall P (2008) Investigation of the benzene-dimer potential energy surface: DFT/CCSD(T) correction scheme. *J Chem Phys* 128:114102. <https://doi.org/10.1063/1.2890968>
253. Hunter CA, Lawson KR, Perkins J, Urch CJ (2001) Aromatic interactions. *J Chem Soc Perkin Trans 2*:651–669. <https://doi.org/10.1039/b008495f>
254. Fagnani DE, Sotuyo A, Castellano RK (2017) π - π interactions. In: *Comprehensive Supramolecular Chemistry II*, vol 1, chap 6. Elsevier, Oxford, pp 121–148. <https://doi.org/10.1016/B978-0-12-409547-2.12485-0>
255. Molčanov K, Kojić-Prodić B (2019) Towards understanding π -stacking interactions between non-aromatic rings. *IUCrJ* 6:156–166. <https://doi.org/10.1107/S2052252519000186>
256. Battaglia MR, Buckingham AD, Williams JH (1981) The electric quadrupole moments of benzene and hexafluorobenzene. *Chem Phys Lett* 78:421–423. [https://doi.org/10.1016/0009-2614\(81\)85228-1](https://doi.org/10.1016/0009-2614(81)85228-1)
257. Hunter CA, Singh J, Thornton JM (1991) π - π interactions: The geometry and energetics of phenylalanine-phenylalanine interactions in proteins. *J Mol Biol* 218:837–846. [https://doi.org/10.1016/0022-2836\(91\)90271-7](https://doi.org/10.1016/0022-2836(91)90271-7)
258. Carter-Fenk K, Liu M, Pujal L, Loipersberger M, Tsanai M, Vernon RM, Forman-Kay JD, Head-Gordon M, Heidari-Zadeh F, Head-Gordon T (2023) The energetic origins of π - π contacts in proteins. *J Am Chem Soc* 145:24836–24851. <https://doi.org/10.1021/jacs.3c09198>
259. Gung BW, Amicangelo JC (2006) Substituent effects in C_6F_6 - C_6H_5X stacking interactions. *J Org Chem* 71:9261–9270. <https://doi.org/10.1021/jo061235h>
260. Tsuzuki S, Uchimaru T, Mikami M (2006) Intermolecular interaction between hexafluorobenzene and benzene: Ab initio calculations including CCSD(T) level electron correlation correction. *J Phys Chem A* 110:2027–2033. <https://doi.org/10.1021/jp054461o>
261. Wang W, Zhang Y, Wang YB (2017) Highly accurate benchmark calculations of the interaction energies in the complexes $C_6H_6 \cdots C_6X_6$ ($X = F, Cl, Br, \text{ and } I$). *Int J Quantum Chem* 117:e25345. <https://doi.org/10.1002/qua.25345>
262. Gray CG, Gubbins KE (1984) *Theory of Molecular Fluids*, vol I. Oxford University Press, New York. <https://doi.org/10.1093/oso/9780198556022.001.0001>
263. Hunter CA, Sanders JKM (1990) The nature of π - π interactions. *J Am Chem Soc* 112:5525–5534. <https://doi.org/10.1021/ja00170a016>
264. Vinter JG (1994) Extended electron distribution applied to the molecular mechanics of some intermolecular interactions. *J Comput Aided Mol Des* 8:653–668. <https://doi.org/10.1007/BF00124013>
265. Carter-Fenk K, Lao KU, Herbert JM (2025) Correction to “Predicting and understanding noncovalent interactions using novel forms of symmetry-adapted perturbation theory”. *Acc Chem Res* 58:1051–1053. <https://doi.org/10.1021/acs.accounts.5c00107>
266. Côté AP, Benin AI, Ockwig NW, O’Keeffe M, Matzger AJ, Yaghi OM (2005) Porous, crystalline, covalent organic frameworks. *Science* 310:1166–1170. <https://doi.org/10.1126/science.1120411>
267. Du Y, Calabro D, Wooler B, Li Q, Cundy S, Kamakoti P, Colmyer D, Mao K, Ravikovitch P (2014) Kinetic and mechanistic study of COF-1 phase change from a staggered to eclipsed model upon partial removal of mesitylene. *J Phys Chem C* 118:399–407. <https://doi.org/10.1021/jp4097293>
268. Gao J, Jiang D (2016) Covalent organic frameworks with spatially confined guest molecules in nanochannels and their impacts on crystalline structures. *Chem Commun* 52:1498–1500. <https://doi.org/10.1039/C5CC09225F>
269. Wonanke ADD, Addicoat MA (2022) Effect of unwanted guest molecules on the stacking configuration of covalent organic frameworks: A periodic energy decomposition analysis. *Phys Chem Chem Phys* 24:15494–15501. <https://doi.org/10.1039/D2CP00017B>

270. Lohse MS, Bein T (2018) Covalent organic frameworks: Structures, synthesis, and applications. *Adv Funct Mater* 28:1705553. <https://doi.org/10.1002/adfm.201705553>
271. Song Y, Sun Q, Aguila B, Ma S (2019) Opportunities of covalent organic frameworks for advanced applications. *Adv Sci* 6:1801410. <https://doi.org/10.1002/advs.201801410>
272. Birowska M, Milowska K, Majewski JA (2011) Van der Waals density functionals for graphene layers and graphite. *Acta Phys Pol A* 120:845–848. <https://doi.org/10.12693/APhysPolA.120.845>
273. Rozhkov AV, Sboychakov AO, Rakhmanov AL, Nori F (2016) Electronic properties of graphene-based bilayer systems. *Phys Rep* 648:1–104. <https://doi.org/10.1016/j.physrep.2016.07.003>
274. Lukose B, Kuc A, Frenzel J, Heine T (2010) On the reticular construction concept of covalent organic frameworks. *Beilstein J Nanotechnol* 1:60–70. <https://doi.org/10.3762/bjnano.1.8>
275. Spitler EL, Koo BT, Novotney JL, Colson JW, Uribe-Romo FJ, Gutierrez GD, Clancy P, Dichtel WR (2011) A 2D covalent organic framework with 4.7-nm pores and insight into its interlayer stacking. *J Am Chem Soc* 133:19416–19421. <https://doi.org/10.1021/ja206242v>
276. Lukose B, Kuc A, Heine T (2011) The structure of layered covalent-organic frameworks. *Chem Eur J* 17:2388–2392. <https://doi.org/10.1002/chem.201001290>
277. Koo BT, Dichtel WR, Clancy P (2012) A classification scheme for the stacking of two-dimensional boronate ester-linked covalent organic frameworks. *J Mater Chem* 22:17460–17469. <https://doi.org/10.1039/C2JM32009F>
278. Haase F, Gottschling K, Stegbauer L, Germann LS, Gutzler R, Duppel V, Vyas VS, Kern K, Dinnebier RE, Lotsch BV (2017) Tuning the stacking behaviour of a 2D covalent organic framework through non-covalent interactions. *Mater Chem Front* 1:1354–1361. <https://doi.org/10.1039/c6qm00378h>
279. Wu X, Han X, Liu Y, Cui Y (2018) Control interlayer stacking and chemical stability of two-dimensional covalent organic frameworks via steric tuning. *J Am Chem Soc* 140:16124–16133. <https://doi.org/10.1021/jacs.8b08452>
280. Geng K, He T, Liu R, Dalapati S, Tan KT, Li Z, Tao S, Gong Y, Jiang Q, Jiang D (2020) Covalent organic frameworks: Design, synthesis, and functions. *Chem Rev* 120:8814–8933. <https://doi.org/10.1021/acs.chemrev.9b00550>
281. Grimme S (2008) Do special noncovalent π – π stacking interactions really exist? *Angew Chem Int Ed Engl* 47:3430–3434. <https://doi.org/10.1002/anie.200705157>
282. Martinez CR, Iverson BL (2012) Rethinking the term “ π -stacking”. *Chem Sci* 3:2191–2201. <https://doi.org/10.1039/C2SC20045G>
283. Ninković DB, Blagojević Filipović JP, Hall MB, Brothers EN, Zarić SD (2020) What is special about aromatic–aromatic interactions? Significant attraction at large horizontal displacement. *ACS Cent Sci* 6:420–425. <https://doi.org/10.1021/acscentsci.0c00005>
284. Janowski T, Pulay P (2012) A benchmark comparison of σ/σ and π/π dispersion: The dimers of naphthalene and decalin, and coronene and perhydrocoronene. *J Am Chem Soc* 134:17520–17525. <https://doi.org/10.1021/ja303676q>
285. Fokin AA, Gerbig D, Schreiner PR (2011) σ/σ - and π/π -interactions are equally important: Multilayered graphanes. *J Am Chem Soc* 133:20036–20039. <https://doi.org/10.1021/ja206992j>
286. Boto RA, Contreras-García J, Tierny J, Piquemal JP (2016) Interpretation of the reduced density gradient. *Mol Phys* 114:1406–1414. <https://doi.org/10.1080/00268976.2015.1123777>
287. Johnson ER, Keinan S, Mori-Sánchez P, Contreras-García J, Cohen AJ, Yang W (2010) Revealing noncovalent interactions. *J Am Chem Soc* 132:6498–6506. <https://doi.org/10.1021/ja100936w>
288. Contreras-García J, Johnson ER, Keinan S, Chaudret R, Piquemal JP, Beratan DN, Yang W (2011) NCIPLOT: A program for plotting noncovalent interaction regions. *J Chem Theory Comput* 7:625–632. <https://doi.org/10.1021/ct100641a>

289. Otero-de-la-Roza A, Johnson ER, Contreras-García J (2012) Revealing non-covalent interactions in solids: NCI plots revisited. *Phys Chem Chem Phys* 14:12165–12172. <https://doi.org/10.1039/C2CP41395G>
290. Contreras-García J, Calatayud M, Piquemal JP, Recio JM (2012) Ionic interactions: Comparative topological approach. *Comput Theor Chem* 998:193–201. <https://doi.org/10.1016/j.comptc.2012.07.043>
291. Boto RA, Peccati F, Laplaza R, Quan C, Carbone A, Piquemal JP, Maday Y, Contreras-García J (2020) NCIPLOT4: Fast, robust, and quantitative analysis of noncovalent interactions. *J Chem Theory Comput* 16:4150–4158. <https://doi.org/10.1021/acs.jctc.0c00063>
292. Corminboeuf C, von Ragué Schleyer P, Warner P (2007) Are antiaromatic rings stacked face-to-face aromatic? *Org Lett* 9:3263–3266. <https://doi.org/10.1021/ol071183y>
293. Björk J, Hanke F, Palma CA, Samori P, Cecchini M, Persson M (2010) Adsorption of aromatic and anti-aromatic systems on graphene through π - π stacking. *J Phys Chem Lett* 1:3407–3412. <https://doi.org/10.1021/jz101360k>
294. Bloom JWG, Wheeler SE (2011) Taking the aromaticity out of aromatic interactions. *Angew Chem Int Ed Engl* 50:7847–7849. <https://doi.org/10.1002/anie.201102982>
295. Vernon RM, Chong PA, Tsang B, Kim TH, Bah A, Farber P, Lin H, Forman-Kay JD (2018) Pi-pi contacts are an overlooked protein feature relevant to phase separation. *eLife* 7:e31486. <https://doi.org/10.7554/eLife.31486>
296. Ayotte P, Weddle GH, Kim J, Johnson MA (1998) Vibrational spectroscopy of the ionic hydrogen bond: Fermi resonances and ion–molecule stretching frequencies in the binary $X^- \cdot H_2O$ ($X = Cl, Br, I$) complexes via argon predissociation spectroscopy. *J Am Chem Soc* 120:12361–12362. <https://doi.org/10.1021/ja981979f>
297. Robertson WH, Johnson MA (2003) Molecular aspects of halide ion hydration: The cluster approach. *Annu Rev Phys Chem* 54:173–213. <https://doi.org/10.1146/annurev.physchem.54.011002.103801>
298. Roscioli JR, Diken EG, Johnson MA, Horvath S, McCoy AB (2006) Prying apart a water molecule with anionic H-bonding: A comparative spectroscopic study of the $X^- \cdot H_2O$ ($X = OH, O, F, Cl, \text{ and } Br$) binary complexes in the 600–3800 cm^{-1} region. *J Phys Chem A* 110:4943–4952. <https://doi.org/10.1021/jp056022v>
299. Thompson WH, Hynes JT (2000) Frequency shifts in the hydrogen-bonded OH stretch in halide–water clusters. The importance of charge transfer. *J Am Chem Soc* 122:6278–6286. <https://doi.org/10.1021/ja993058q>
300. Herbert JM, Head-Gordon M (2006) Charge penetration and the origin of large O–H vibrational red-shifts in hydrated-electron clusters, $(H_2O)_n^-$. *J Am Chem Soc* 128:13932–13939. <https://doi.org/10.1021/ja064949i>
301. Sharpe AG (1990) The solvation of halide ions and its chemical significance. *J Chem Educ* 67:309–315. <https://doi.org/10.1021/ed067p309>
302. Antalek M, Pace E, Hedman B, Hodgson KO, Chillemi G, Benfatto M, Sarangi R, Frank P (2016) Solvation structure of the halides from x-ray absorption spectroscopy. *J Chem Phys* 145:044318. <https://doi.org/10.1063/1.4959589>
303. Brown TL, LeMay, Jr. HE, Bursten BE, Murphy CJ, Woodward PM, Stoltzfus MW: Chemistry: The Central Science, 13th edn. Pearson, Boston (2015)
304. Arunan E, Desiraju GR, Klein RA, Sadlej J, Scheiner S, Alkorta I, Clary DC, Crabtree RH, Dannenberg JJ, Hobza P, Kjaergaard HG, Legon AC, Mennucci B, Nesbitt DJ (2011) Defining the hydrogen bond: An account (IUPAC technical report). *Pure Appl Chem* 83:1619–1637. <https://doi.org/10.1351/PAC-REP-10-01-01>
305. Arunan E, Desiraju GR, Klein RA, Sadlej J, Scheiner S, Alkorta I, Clary DC, Crabtree RH, Dannenberg JJ, Hobza P, Kjaergaard HG, Legon AC, Mennucci B, Nesbitt DJ (2011) Definition of the hydrogen bond (IUPAC recommendations 2011). *Pure Appl Chem* 83:1637–1641. <https://doi.org/10.1351/PAC-REC-10-01-02>

306. Petersen PB, Saykally RJ (2004) Confirmation of enhanced anion concentration at the liquid water surface. *Chem Phys Lett* 397:51–55. <https://doi.org/10.1016/j.cplett.2004.08.049>
307. Petersen PB, Saykally RJ, Mucha M, Jungwirth P (2005) Enhanced concentration of polarizable anions at the liquid water surface: SHG spectroscopy and MD simulations of sodium thiocyanide. *J Phys Chem B* 109:10915–10921. <https://doi.org/10.1021/jp050864c>
308. Mucha M, Frigato T, Levering LM, Allen HC, Tobias DJ, Dang LX, Jungwirth P (2005) Unified molecular picture of the surfaces of aqueous acid, base, and salt solutions. *J Phys Chem B* 109:7617–7623. <https://doi.org/10.1021/jp0445730>
309. Ghosal S, Hemminger JC, Bluhm H, Mun BS, Hebenstreit ELD, Ketteler G, Olgetree DF, Requejo FG, Salmeron M (2005) Electron spectroscopy of aqueous solution interfaces reveals surface enhancement of halides. *Science* 307:563–566. <https://doi.org/10.1126/science.1106525>
310. Petersen PB, Saykally RJ (2006) On the nature of ions at the liquid water surface. *Annu Rev Phys Chem* 57:333–364. <https://doi.org/10.1146/annurev.physchem.57.032905.104609>
311. Otten DE, Shaffer PR, Geissler PL, Saykally RJ (2012) Elucidating the mechanism of selective ion adsorption to the liquid water surface. *Proc Natl Acad Sci USA* 109:701–705. <https://doi.org/10.1073/pnas.1116169109>