Accurate and Efficient Quantum Chemistry Calculations for Noncovalent Interactions in Many-Body Systems

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Ka Un Lao, M.S.

Graduate Program In Chemistry

The Ohio State University

2016

Dissertation Committee:

John M. Herbert, Advisor

Sherwin J. Singer

Marcos M. Sotomayor

© Copyright by

Ka Un Lao

2016

Abstract

We discuss the development and application of a number of fragmentation methods focused on understanding of intermolecular interactions in different systems. The advantage of fragmentation methods is to avoid the exponential growth of required computational power for the most advanced and accurate quantum chemistry theories which preclude the application in systems with large number of atoms and molecules. In those fragmentation methods, the full chemical system is partitioned into different subsystems, circumventing the exponential scaling computational cost. How this partitioning is performed and applied appropriately is the principal emphasis of this work. One of the fragmentation methods developed by our group, called extended XSAPT, combines an efficient, iterative, monomer-based approach to computing many-body polarization interactions with a two-body version of symmetry-adapted perturbation theory (SAPT). The result is an efficient method for computing accurate intermolecular interaction energies in large non-covalent assemblies such as molecular and ionic clusters, supramolecular complexes, clathrates, or DNA–drug complexes. As in traditional SAPT, the XSAPT energy is decomposable into physically-meaningful components. Dispersion interactions are problematic in traditional low-order SAPT, and the empirical atom–atom dispersion potentials are introduced here in an attempt to improve this situation. Comparison to high-level *ab initio* benchmarks

for biologically-related dimers, water clusters, halide–water clusters, supremolecular complexes, methane clathrate hydrates, and a DNA intercalation complex illustrate both the accuracy of XSAPT-based methods as well as their limitations. The computational cost of XSAPT scales as $\mathcal{O}(N^3)$–$\mathcal{O}(N^5)$ with respect to monomer size, N, depending upon the particular version that is employed, but the accuracy is typically superior to alternative *ab initio* methods with similar scaling. Moreover, the monomer-based nature of XSAPT calculations makes them trivially parallelizable, such that wall times scale linearly with respect to the number of monomer units. XSAPT-based methods thus open the door to both qualitative and quantitative studies of non-covalent interactions in clusters, biomolecules, and condensed-phase systems.

To my grandma Chon Ngan, my mom Pek Ieng, my wife Yili,
and my loving daughter Bernadette.

Acknowledgments

First and foremost, I gratefully acknowledge the marvelous love and gracious support of my grandma and mom, and also my wife, Yili, which have allowed me to pursue my education and achieve this, the highest academic degree in chemistry. I sincerely thank Dr. John M. Herbert, my advisor, for insightful discussions, mentoring, and all-around helping me reach this point in my academic career. I especially thank my lab mate, Master Marc Coons, for being my amazing friend. I also would like to recognize and thank the entire Herbert group (both past and present members) for help and support along the way. Additionally, I express my deep gratitude to the Ohio State University for fostering my graduate school education as well as for granting me the Ohio State Presidential Fellowship in my final year of graduate school. Finally, I appreciate the Ohio Supercomputing Center for gratuitous amounts of computer time.

Vita

1985	Born, Macau
2003	Colégio Diocesano de São José (5ª), Macau
2007	B.S. Chemistry, Minor in Physics, National Tsing Hua University, Hsinchu, Taiwan
2009	M.S. Chemistry, National Tsing Hua University, Hsinchu, Taiwan
2011–2012	Graduate Administrative Associate, The Ohio State University
2012–2015	Graduate Research Associate, The Ohio State University
2015	1st place award in Albert L. Henne Research Competition, The Ohio State University
2015	ACS COMP Division Chemical Computing Group Research Excellence Award for Graduate Student, ACS COMP Division
2015	Phi Tau Phi (PTP) Mid-America Scholarship Award, PTP Mid-America chapter
2016	1st place oral award in Mathematical and Physical Sciences area at The 30th Edward F. Hayes Graduate Research Forum, The Ohio State University
2015–present	Presidential Fellow, The Ohio State University

Publications

PhD Publications

(25) K. U. Lao and J. M. Herbert. **Accurate prediction of lattice energies of molecular crystals with extended symmetry-adapted perturbation theory (XSAPT).** *in preparation*

(24) K. U. Lao, A. Pomogaeva, D. M. Chipman, and J. M. Herbert. **Experiment vs. state-of-the-art** *ab initio* **theory for binding affinities of solution-phase ions.** *in preparation*

(23) K. U. Lao and J. M. Herbert. **Achieving high-accuracy intermolecular interactions by combining extended symmetry-adapted perturbation theory (XSAPT) with coupled Kohn-Sham dispersion.** *in preparation*

(22) K. U. Lao and J. M. Herbert. **Atomic orbital implementation of extended symmetry-adapted perturbation theory (XSAPT) and benchmark calculations for large supramolecular complexes.** *to be submitted*

(21) K. U. Lao and J. M. Herbert. **A Simple Correction for Non-Additive Dispersion within Extended Symmetry-Adapted Perturbation Theory.** *to be submitted*

(20) K. U. Lao, K.-Y. Liu, R. M. Richard, and J. M. Herbert. **Understanding the many-body expansion for large systems: II. Accuracy considerations.** *submitted*

(19) K. U. Lao and J. M. Herbert. **Energy decomposition analysis with a well-defined charge-transfer term for interpreting intermolecular interactions.** *submitted*

(18) K. U. Lao, R. Schäffer, G. Jansen, and J. M. Herbert. **Accurate description of intermolecular interactions involving ions using symmetry-adapted perturbation theory.** *J. Chem. Theory Comput.* **11**, 2473 (2015).

(17) K. U. Lao and J. M. Herbert. **Accurate and efficient quantum chemistry calculations of non-covalent interactions in many-body systems: The XSAPT family of methods.** *J. Phys. Chem. A* **119**, 235 (2015). [Feature Article, ACS Editors' Choice, and Featured on the cover of JPCA (Jan. 15 2015)]

(16) Y. Shao *et al.* **Advances in molecular quantum chemistry contained in the Q-Chem 4 program package.** *Mol. Phys.* **113**, 184 (2015).

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

(14) R. M. Richard, K. U. Lao, and J. M. Herbert. **Understanding the many-body expansion for large systems. I. Precision considerations.** *J. Chem. Phys.* **141**, 014108 (2014).

(13) K. U. Lao and J. M. Herbert. **Symmetry-adapted perturbation theory with Kohn-Sham orbitals using non-empirically tuned, long-range-corrected density functionals.** *J. Chem. Phys.* **140**, 044108 (2014). (Selected by JCP as an "Editor's Choice for 2014")

(12) R. M. Richard, K. U. Lao, and J. M. Herbert. **Approaching the complete-basis limit with a truncated many-body expansion.** *J. Chem. Phys.* **139**, 224102 (2013).

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

(10) R. M. Richard, K. U. Lao, and J. M. Herbert. **Achieving the CCSD(T) basis-set limit in sizable molecular clusters: Counterpoise corrections for the many-body expansion.** *J. Phys. Chem. Lett.* **4**, 2674 (2013).

(9) K. U. Lao and J. M. Herbert. **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 (2013). [Erratum: *J. Chem. Phys.* **140**, 119901 (2014).]

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

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

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

Previous Publications

(5) C.-Y. Chiu, P.-J. Chung, K. U. Lao, C.-W. Liao, and M. H. Huang. **Facet-dependent catalytic activity of gold nanocubes, octahedra, and rhombic dodecahedra toward 4-nitroaniline reduction.** *J. Phys. Chem. C* **116**, 23757 (2012).

(4) K. U. Lao, T. Lankau, T.-I Fang, J.-W. Zou, and C.-H. Yu. **Interstitial water and the formation of low barrier hydrogen bonds: A computational model study.** *Int. J. Quantum Chem.* **112**, 1460 (2012).

(3) K. U. Lao, P.-K. Tsou, T. Lankau, and C.-H. Yu. **A computational study of organic polyradicals stabilized by chromium atoms.** *Phys. Chem. Chem. Phys.* **14**, 138 (2012).

(2) H.-L. Wu, H.-R. Tsai, Y.-T. Hung, K. U. Lao, C.-W. Liao, P.-J. Chung, J.-S. Huang, I-C. Chen, and M. H. Huang. **A comparative study of gold nanocubes, octahedra, and rhombic dodecahedra as highly sensitive SERS substrates.** *Inorg. Chem.* **50**, 8106 (2011).

(1) K. U. Lao and C.-H. Yu. **A computational study of unique properties of pillar[n]quinones: self-assembly to tubular structures and potential applications as electron acceptors and anion recognizers.** *J. Comput. Chem.* **32**, 2716 (2011).

Fields of Study

Major Field: Chemistry
Theoretical Physical Chemistry

Table of Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

The efficient and accurate description of noncovalent interactions has been an active research topic in the past decade because noncovalent interactions are important and ubiquitous in various areas of chemistry, such as the determination of DNA structures, drug binding to both protein and DNA, crystal packing, and protein folding. The description of noncovalent interactions based on computational chemistry has made much progress in recent years, due to improvements in both algorithms and computer power. An accurate, traditional quantum mechanical (QM) method, such as the "gold standard" coupled cluster singles and doubles with noniterative triples method [CCSD(T)], with chemical accuracy (errors $\leqslant 1$ kcal/mol) scales $\mathcal{O}(N_s^7)$ where N_s is the size of the whole system, in other words, the computational cost increases 128 times when the system size doubles. Thus, the high computational cost limits its applicability to the size of molecules with less than 30 atoms, and it remains out of reach for large noncovalent assemblies. For example, the binding pocket of the human immunodeficiency virus (HIV) protease + drug inhibitor system [Fig. 1.1(a)], as shown in Fig. 1.1(b), is a representative model system and consists of the drug

indinavir (opaque ball-and-stick model) along with 16 amino acids and 2 crystallographic waters (translucent tubular model). The total system size for this model system amounts to 323 atoms, which makes traditionally accurate QM methods intractable. Hence, it is hard to use such traditional QM methods in real applications, such as in drug design where the search of a compound to effectively interact with biomolecules is the central step, and there are lots of potential candidates to screen during the process. Although empirical screening technologies are efficient to screen lots of molecules simultaneously, the binding results predicted by these fast screening methods have lower accuracy because approximations have to be used to make them computationally feasible, often with only a loose connection to the real physics of intermolecular interactions. The purpose of my doctoral research has been to build a rapid and accurate QM approach for calculating interactions between molecules to be applied to drug design and molecular crystals modeling.

Fragment-based quantum chemistry methods offer a way to surmount this predicament and rely on decomposing the large supersystem (the system including all fragments) into subsystems (fragments) to greatly reduce the computational cost.[43–45] I have contributed to three aspects of fragmentation methods: Many-body expansion (MBE),[45–49] symmetry-adapted perturbation theory (SAPT),[50–53] and extended SAPT (XSAPT) methods.[44,54–63]

The simplest fragmentation method, many-body expansion (MBE), for a system of N fragments or monomers is given by

$$E = \sum_{I=1}^{N} E_I + \sum_{I<J} \Delta E_{IJ} + \sum_{I<J<K} \Delta E_{IJK} + \cdots + \Delta E_{IJK\cdots N} \ ,$$

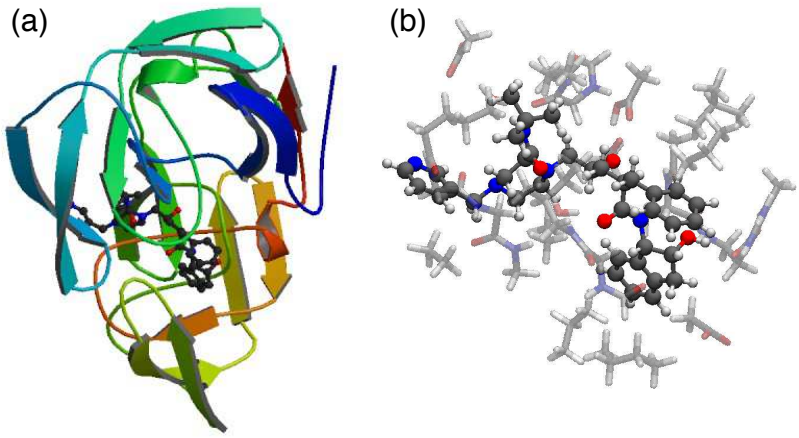

Figure 1.1: (a) Structure of the protease inhibitor indinavir bound to HIV protease, as obtained from PDB crystal structure 1HSG.[1] (b) An enlarged view of the binding pocket, consisting of indinavir (opaque ball-and-stick model) along with 16 amino acids and 2 crystallographic waters (translucent tubular model). [Panel (b) is reproduced from Ref. 2; copyright 2011 American Institute of Physics.]

where I is an index over monomers, E_I is the energy associated with monomer I, and $\Delta E_{IJ\ldots}$ represents the interaction energy between the indexed monomers. For example, ΔE_{IJ} is the dimer *interaction* energy and can be represented as $\Delta E_{IJ} = E_{IJ} - E_I - E_J$, where E_{IJ} is the energy of the IJ dimer. ΔE_{IJK} is the trimer *interaction* energy and can be represented as $\Delta E_{IJK} = E_{IJK} - E_{IJ} - E_{IK} - E_{JK} - E_I - E_J - E_K$, where E_{IJK} is the energy of the IJK trimer. A large number of previous studies using MBE methods have just focused on results truncated at three-body correction, and the high-order terms can be included by using electrostatic embedding (EE) in all subsystem calculations.[45,48,64] In reality, it is not true for systems composed

of lots of fragments, such as $(H_2O)_{40}$, and the EE-MBE does not converge even at four-body calculations.[49] Furthermore, the results are very sensitive to the charge schemes used in the electrostatic embedding.[49] The embedding charge schemes can deteriorate the MBE results involving basis sets with diffuse functions due to the overpolarization between charges and diffuse functions of basis sets.[49] Hence, one should be careful to use electrostatic-embedding scheme in the MBE calculations to capture the high-order effects. We demonstrated that a four-body expansion, MBE(4), without involving the unpredicable charge embedding schemes is required in order to obtain accurate values for both total interaction energies as well as relative energies of various isomers of water clusters. The generalized MBE (GMBE) developed by our group including the overlap fragments can resolve this problem, and it provides results of comparable accuracy at the two-body level, GMBC(2).[49] However, this is not a "free lunch" relative to MBE(4), because the size of the subsystem calculations increases in this "GMBE(2)" approach. Furthermore, the correction of basis-set superposition error (BSSE) is important to obtain accurate binding energies in chemical accuracy, and counterpoise (CP) correction is a popular method to remove the BSSE. Nevertheless, there are many extra supersystem calculations involved in the CP correction. In fact, the CP correction can be expanded using a method similar to MBE, and such a method is called many-body CP (MBCP).[46,47] The convergence of MBCP has been tested for various systems, and it suffers the same convergence issue as in MBE.[49] The generalized MBCP (GMBCP) version is also available to obtain accurate CP results although the computational cost for GMBCP is extraordinary high. This part

is described in detail in chapter 2.

Although our group has made much progress on the MBE method, there are still two drawbacks that make MBE time consuming. First, the terms beyond two-body interactions that contain many atoms and many terms have to be calculated. Second, the CP or (G)MBCP correction involving extra calcaultions has to be used to remove the BSSE. Furthermore, the supersystem calculation remains vastly cheaper than MBE(4) or GMBE(2), at least in DFT level, unless $10^3 - 10^4$ processors are available, making (G)MBE fragmentation approaches good candidates only for peta- and exa-scale computing.

The fragment-based XSAPT method was developed by our group[44,51,54–63,65] in an attempt to achieve chemical accuracy for interactions, yet remain affordable enough to be applied to systems such as the one in Fig. 1.1(b). Moreover, the XSAPT method is only a two-body method, and is inherently BSSE-free.

Our XSAPT method is based on the combination of two fragmentation methods, explicit polarization method (XPol) and symmetry-adapted perturbation theory (SAPT). This approach starts from the XPol method, and the primary function of this method is to capture many-body (MB) polarization effects (important for systems with many fragments) which are omitted in traditional force fields and computational screening functions. In a subsequent step, we apply a pairwise-additive form of SAPT to capture the rest of the interactions missing in the XPol step. However, the single-exchange approximation (SEA) used in the exchange formulas of SAPT makes SAPT unreliable for use in anionic systems.[50,51] Different levels of SAPT methods have been

used to systematically study the binding energies in different ionic systems, and the rescaled formula for SEA has been proposed to resolve the problem in exchange formulas used SEA.[50,52] Such a rescaled formula for SEA can also be used in XSAPT. The importance of rescaled formula for SEA in ionic systems is described in detail in chapter 3 and 4.

The resulting XSAPT method extends the traditional two-body SAPT method to many-body systems, and it maintains the computational cost of a two-body system. Furthermore, the XSAPT interaction energy can be decomposed into physically meaningful energy components,[56]

$$E_{\text{int}}^{\text{XSAPT}} = E_{\text{electrostatic}} + E_{\text{exchange}} + E_{\text{induction}} + E_{\text{dispersion}} \ , \tag{1.1}$$

and we can study the interplay of various energy components, some of which are attractive and some repulsive, to explore the chemical modification of target molecules. Furthermore, we have introduced many different techniques to improve the original XSAPT method.

The use of Kohn-Sham (KS) orbitals for XSAPT or "XSAPT(KS)" incorporates intramolecular correlation effects beyond mean field in a relatively low-cost way.[51,55–57] Long-range corrected (LRC) density functionals with a monomer-specific "tuning" procedure are used to obtain correct asymptotic behavior of exchange-correlation potentials in XSAPT(KS) instead of using traditional "splicing" schemes that do not correspond to any well-defined energy functional and are potentially problematic in the context of geometry optimizations.[51,60] My work for XSAPT(KS) based on LRC density functionals [*J. Chem. Phys.* **140**, 044108 (2014)] was selected by *The Journal*

of Chemical Physics as an "Editor's Choice for 2014". The detail of this method is described in chapter 5.

To correct the problematic dispersion energies in second-order perturbation theory, I replace the expensive dispersion terms in XSAPT(KS) with pairwise-additive atom−atom dispersion contributions from localized multipoles:

$$E_{\text{disp}} = -\sum_{i>j}\left(\frac{C_6^{ij}}{R_{ij}^6} + \frac{C_8^{ij}}{R_{ij}^8} + \frac{C_{10}^{ij}}{R_{ij}^{10}} + \cdots\right) \tag{1.2}$$

where R_{ij} is the interatomic distance for atoms i and j. For each pair of atoms, such as atoms i and j, C_6^{ij} is dipole−dipole dispersion coefficient, C_8^{ij} represents dipole−quadrupole dispersion coefficient, C_{10}^{ij} expresses quadrupole−quadrupole and dipole−octupole dispersion coefficients, and so on. The dispersion coefficients can be determined by appealing to *ab initio* calculations or at least partly based on fitting to a benchmark training set. The method combined with these dispersion potentials and XSAPT for non-dispersion terms is called XSAPT(KS)+D.[51,55–57] There are different generation of these dispersion potentials which are called +D1, +D2, and +D3 and are described in chapter 6, 7, and 8, respectively. My work for XSAPT(KS) with third-generation dispersion potentials ("+D3") [*J. Phys. Chem. A* **119**, 235 (2015)] was featured on the cover of *The Journal of Physical Chemistry A*.

Furthermore I have reformulated XSAPT in the atomic orbital (AO) basis.[60] This formulation avoids the four-index integral transformation that is required in the original, molecular orbital version of the method.[54,57,65] The AO-XSAPT(KS)+D method reduces the scaling from $\mathcal{O}(N_f^5)$ to $\mathcal{O}(N_f^3)$ with respect to the fragment size, N_f.[60]

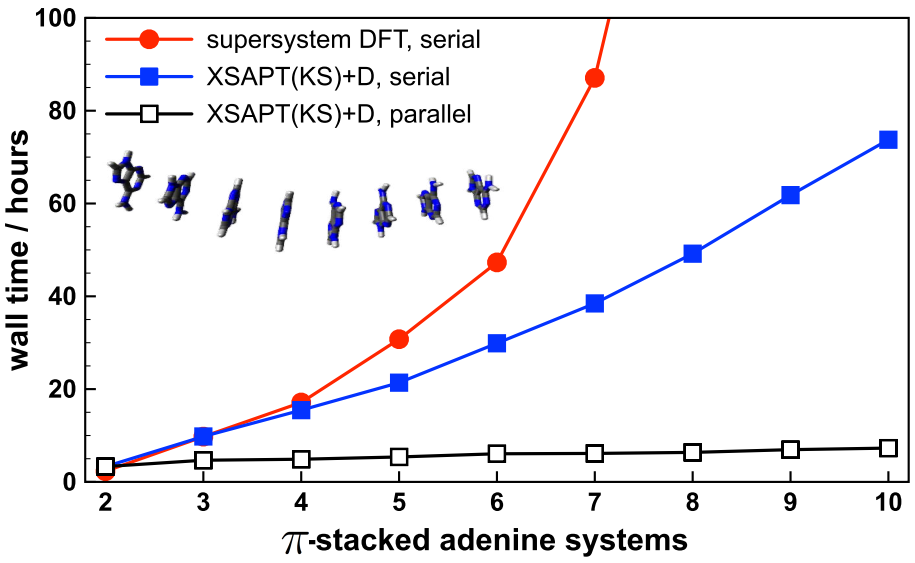

Figure 1.2: Timings for XSAPT(KS) in AO basis and supersystem DFT calculations for π-stacked $(\text{adenine})_n$ systems.

The detail of AO-XSAPT method is described in chapter 9. I lead off with data illustrating the *efficiency* of the AO-XSAPT method, as shown in Figure 1.2 which plots timings for XSAPT(KS)+D calculations in AO basis against the size of $(\text{adenine})_n$ where adenine is a DNA nucleobase. The supersystem DFT method is one of the most economical QM methods, and the timing results of this method are also shown in Figure 1.2 for comparison. Serial timings represent the total wall time required using one processor, and the scaling of XSAPT(KS)+D is $\mathcal{O}(n^2)$ where n is the number of adenine molecules. Parallel timings represent the wall time required when the calculation is simultaneously run on $n(n-1)/2$ processors for n adenine molecules. This feature of parallel mode for XSAPT is called "embarrassingly parallel" mode,

and the scaling is just $\mathcal{O}(n)$. Even in serial, XSAPT(KS)+D is just as efficient as supersystem DFT for $(\text{adenine})_2$ and is substantially more efficient for larger systems. In parallel, the wall time required for an XSAPT(KS)+D calculation on $(\text{adenine})_{10}$ is only about twice as large as that required for $(\text{adenine})_2$. Moreover, the XSAPT method is not only *efficient* as indicated at Figure 1.2 but also *accurate* as described in this thesis.

In XSAPT(KS)+D, the dispersion potentials are independent of chemical environment and the squeezing effects by neighboring atoms are missing. We present a dispersion contraction (DC) to capture the non-additive dispersion interactions in XSAPT(KS)+D. The DC effect can be calculated by dispersion energy difference based on XSAPT (with charge embedding) and SAPT (without charge embedding) but has $\mathcal{O}(N_f^5)$ scaling. The DC term can be further simplified by only including the second-order dispersion term with a double-ζ basis set where the good performance is based on fortuitous error cancellation. Then, the scaling of XSAPT(KS)+D+DC method is $\mathcal{O}(N_f^3)$ using a triple-ζ basis set for the non-dispersion terms, and $\mathcal{O}(N_f^4)$ using a double-ζ basis set for the DC corrections. The DC correction is especially important in clusters with numerous numbers of monomers, for example, the DC contribution is about 5 kcal/mol in $(H_2O)_{20}$. The DC correction for XSAPT(KS)+D is described in chapter 10.

Finally, various schemes for decomposing quantum-chemical calculations of intermolecular interaction energies into physically meaningful components can be found in the literature, but the definition of the charge-transfer (CT) contribution has proven

vexing to define in a satisfactory way and typically depends strongly on the choice of atom-centered Gaussian basis set. This is problematic in cases of dative bonding and for open-shell complexes involving cation radicals, for which one might expect significant CT. We analyze CT interactions predicted by several popular energy decomposition analyses and conclude that only the definition afforded by constrained density functional theory (cDFT) can be recommended, as it is scarcely dependent on basis set and provides results that are in accord with chemical intuition for simple model systems, and in quantitative agreement with experimental estimates of the CT energy, where available. We recommend a composite approach in which cDFT is used to define the CT component of the interaction energy and (X)SAPT defines the electrostatic, Pauli repulsion, induction, and van der Waals contributions. This approach affords stable and physically-motivated definitions for each energy component, paving the way for reliable *in silico* analysis of intermolecular interactions. The detail of this method is described in chapter 11.

CHAPTER 2

Understanding the many-body expansion for large systems

2.1 Introduction

Electronic structure methods based on molecular fragmentation are an increasingly popular way to sidestep the non-linear scaling of computational cost with respect to system size.[43,44,66] Such methods rely, at some level, on Kohn's principle of the "nearsightedness" of electronic matter,[67,68] and attempt to decompose a large calculation into a (potentially large number of) small subsystem calculations that are independent of one another and thus lend themselves to trivially-parallel distributed computing. High accuracy is reported in many applications (see Ref. 43 for a review), and thus molecular fragmentation would seem to offer the proverbial "free lunch", enabling high-level *ab initio* methods to be applied to large systems at a fraction of the cost (in wall time, memory, and disk requirements) that would otherwise be required.

Due to the high cost of obtaining benchmark results in large systems, the accuracy of fragment-based methods has primarily been evaluated in small systems (typically $\lesssim$ 25 heavy atoms) and/or at low levels of electronic structure theory (*e.g.*, self-consistent field theory with minimal or double-ζ basis sets). It is unclear whether

such benchmarks are representative of the performance that can be expected when high levels of theory are applied to larger systems. Consider that even the smallest naturally-occurring protein (the 20-residue Trp cage) has a total electronic energy approaching 10,000 hartree, which must therefore be predicted to a precision of about 0.00001% in order to achieve an accuracy of ∼1 kcal/mol in the total energy. This is the famous "weighing the captain" problem in electronic structure theory,[69,70] *i.e.*, determining the captain's weight based on measuring the ship's displacement when she is, and is not, on board. The situation is arguably somewhat worse for fragment-based methods, which require huge numbers of electronic structure calculations that must be performed at significantly higher precision than is required in conventional quantum chemistry,[48] such that a more apt analogy might be that of determining the weight of a pilot on an aircraft carrier based on measuring the displacement when the pilot sits in his plane, then determining the displacement for various combinations of aircraft and pilots on board the ship.

In part I of this series,[48] we examined the role of finite-precision arithmetic in methods based on a truncated many-body expansion (MBE), otherwise known as an "n-body expansion". The analysis in Ref. 48 focused on systematic error as system size was increased, and we found that the uncertainty in the n-body approximation to the energy is strongly influenced by the self-consistent field (SCF) convergence threshold used for the individual subsystem calculations (much more so than is the total supersystem energy), as well as by accumulation of floating-point rounding errors. Implementations of the MBE that rely on external calls to an electronic

structure program (rather than being fully integrated into such a program), suffer from an additional source of uncertainly, namely, the fact that discrepancies as small as 10^{-6} a.u. between the driver routine and the electronic structure program (due to six-digit roundoff in the electronic structure output, for example) can translate into errors of several kcal/mol, for systems not much larger than $(H_2O)_{30}$.[48] Owing to the combinatorial nature of the MBE, these problems are compounded as the system size increases, and as one moves to higher n in pursuit of greater accuracy.

Although these precision problems appear to be surmountable using a combination of tight SCF convergence thresholds, arbitrary-precision arithmetic to sum the terms in the MBE, and a consistent internal precision in all electronic structure calculations,[48] these considerations do increase the cost of the calculations relative to what has previously been discussed in the literature. The worst cases manifest only when $n \geq 4$, and for very large collections of fragments ($N \gtrsim 40$) that may be avoidable in applications to polyatomic molecules, and might be sidestepped in applications to molecular liquids via some kind of distance-based criterion for discarding or approximating well-separated n-mers. Thus, two- and three-body approaches might still prove useful, provided that good accuracy can be obtained.

Using water clusters as exemplary cases where many-body polarization effects are important, we demonstrate in the present work that a four-body expansion [MBE(4)] is required in order to obtain accurate values for both total interaction energies as well as relative energies of various isomers. On the other hand, a *generalized* many-body expansion[44,71,72] (GMBE) that utilizes overlapping subsystems can provide results of

comparable accuracy at the two-body level. This is not a "free lunch" relative to MBE(4), because the size of the subsystem calculations increases in this "GMBE(2)" approach, but the *number* of subsystem calculations is dramatically reduced, staving off numerical precision problems while maintaining the trivial parallelizability of the traditional MBE approach.

Following up on our study of precision problems in Ref. 48, here we we seek to evaluate the accuracy of various n-body approximations, using the same set of water clusters that was examined in Ref. 48. Regrettably, the overwhelming majority of our present results fail to achieve 1 kcal/mol accuracy in the total energy as approximated at the three-body level, but do reveal some interesting trends as these methods are pushed towards the large-system ($N \rightarrow \infty$) limit. Here, we focus on: (1) whether the supersystem energy at a given level of theory can be accurately approximated when the same level of theory is used for the subsystem calculations; (2) the oscillatory nature of the n-body expansion as the truncation order, n, is increased; (3) size-dependent errors in the n-body total energy and what impact these have on the prediction of relative energies; and (4) the ability of these methods to reproduce high-level benchmarks. As in our previous study,[48] our findings unearth potential pitfalls that have not been widely discussed in the rapidly-growing literature on fragment-based quantum chemistry.

2.2 Theory

2.2.1 Traditional MBE

The basic idea of the n-body expansion of the total energy E,

$$E = \sum_{I=1}^{N} E_I + \sum_{I=1}^{N} \sum_{J<I} (E_{IJ} - E_I - E_J) + \cdots \,, \tag{2.1}$$

into monomer energies (E_I), dimer energies (E_{IJ}), etc., is straightforward and has been reviewed extensively elsewhere.[43,44,66] In particular, far more compact expressions than those suggested by Eq. (2.1) can be found in our previous work.[48] To place Eq. (2.1) in the context of what is to come, note that this equation can alternatively be expressed as[46]

$$E = \sum_{I=1}^{N} \Delta E_I^{(1)} + \sum_{J=1}^{\binom{N}{2}} \Delta E_J^{(2)} + \sum_{K=1}^{\binom{N}{3}} \Delta E_K^{(3)} + \cdots \tag{2.2}$$

where

$$\Delta E_I^{(n)} = \sum_{m=0}^{n-1} (-1)^m \sum_{J \subset I} E_J^{(n-m)} \tag{2.3}$$

is a correction to the energy of the Ith n-body subsystem, of which there are a total of

$$\binom{N}{n} \equiv {}_N C_n = \frac{N!}{n!(N-n)!} \,. \tag{2.4}$$

The quantity $E_J^{(n-m)}$ in Eq. (2.3) is the energy of the Jth $(n-m)$-body sub-cluster formed from the Ith n-body cluster. As such, the second summation in Eq. (2.3) ranges from $J = 1, 2, \ldots, \binom{n}{n-m}$, and this is the meaning of the $J \subset I$ restriction in that equation.

In cases where fragmentation does not sever any covalent bonds, either of Eqs. (2.1) or (2.2) is a formally exact expression for the total energy, but their appeal comes in dropping terms beyond some given level of n-body interaction. Neglecting terms involving $(n+1)$-body and larger sub-clusters defines the so-called n-body expansion, which we will call MBE(n). A compact formula for the energy within the MBE(n) approximation is[48]

$$E^{(n)} = \sum_{m=0}^{n-1} (-1)^m \binom{N-n-1+m}{m} \sum_{K=1}^{\binom{N}{n-m}} E_K^{(n-m)} \ . \tag{2.5}$$

The summations in this equation run over all unique sub-clusters containing up to n fragments.

2.2.2 Counterpoise correction in a many-body system

It is well known that electronic structure calculations converge slowly to the basis-set limit[73] (see Ref. 46 for just one example), which arises from a combination of basis-set incompleteness and BSSE. In the present work, we address the incompleteness issue by means of complete-basis extrapolations, but the convergence of these extrapolations is sensitive to the presence of BSSE, which we therefore attempt to eliminate. The interaction energy, ΔE, subject to a generalized Boys-Bernardi counterpoise (CP) correction is[74,75]

$$\Delta E = E_{IJK\cdots N}^{IJK\cdots N} - \sum_I E_I^{IJK\cdots N} \tag{2.6}$$

where, following the convention of previous work,[46,76] the subscripts represent real molecules whereas superscripts indicate where basis functions are centered. Thus

$E_I^{IJK\cdots N}$ denotes the energy of monomer I computed in a basis set having basis functions located in the positions of monomers $I, J, K, \ldots, N$; that is, $E_I^{IJK\cdots N}$ is the energy of monomer I computed in the cluster basis, as in the original Boys-Bernardi scheme.[74] The quantity $E_{IJK\cdots N}^{IJK\cdots N}$ is the normal supersystem energy, which for brevity we will henceforth denote simply as E.

Equation (2.6) can be trivially rewritten as

$$\begin{aligned} \Delta E &= \left(E - \sum_I E_I^I \right) + \left(\sum_I E_I^I - \sum_I E_I^{IJK\cdots N} \right) \\ &= \Delta E^{(\text{uncorr})} + \delta E^{\text{CP}} \ , \end{aligned} \tag{2.7}$$

where we have separated the interaction energy ΔE into an "uncorrected" (and thus BSSE-contaminated) energy difference

$$\Delta E^{(\text{uncorr})} = E - \sum_I E_I^I \tag{2.8}$$

and subsumed all BSSE corrections into a counterpoise correction

$$\delta E^{\text{CP}} = \sum_{I=1}^{N} \delta E_I^{\text{CP}} = \sum_{I=1}^{N} \left(E_I^I - E_I^{IJK\cdots N} \right) \ . \tag{2.9}$$

The notation $\delta E_I^{\text{CP}} = E_I^I - E_I^{IJK\cdots N}$ is introduced in order to make contact with approximate counterpoise corrections introduced below, in which the quantity $E_I^{IJK\cdots N}$ is approximated by means of a MBE. Equation (2.9) looks like the traditional Boys-Bernardi correction,[74] applied in turn to each monomer.

2.2.3 Generalized MBE

The ground-state energy can also be approximated using a GMBE that employs overlapping fragments and is derived based on the set-theoretical principle of inclusion/exclusion;[71,72] application of the GMBE requires calculations on subsystems

that are formed from intersections of fragments. In an n-body GMBE, which we call GMBE(n), the approximate energy is

$$\varepsilon^{(n)} = \sum_{i=1}^{\binom{N_f}{n}} E_i^{(n)} - \sum_{i=1}^{\binom{N_f}{n}} \sum_{j>i}^{\binom{N_f}{n}} E_{i\cap j}^{(n)} + \cdots + (-1)^{\binom{N_f}{n}+1} E_{i\cap j\cap \ldots \cap \binom{N_f}{n}}^{(n)} \,. \tag{2.10}$$

We have previously called this the *intersection-corrected energy* at order n.[71] Lower case indices $i, j, \cdots$ in Eq. (2.10) refer to n-mers of fragments, whose energies are $E_i^{(n)}, E_j^{(n)}, \ldots$, and $i \cap j$ is the subsystem formed from the intersection of n-mers i and j, with energy $E_{i\cap j}^{(n)}$. For general, macromolecular applications, construction of $i \cap j$ requires severing covalent bonds and capping the severed valencies (as in a recent application of the GMBE to proteins[77]), but in this work we only consider non-covalent clusters, in an intentional effort to sidestep this complexity. Note also that in this generalized approach the number of fragments, N_f, is generally larger than the number of monomers, N. As in previous studies,[18,71,72,78,79] for the GMBE we use a distance cutoff of 3 Å between atoms to define the fragments.

The advantage of GMBE(n) relative to MBE(n) is that multiple monomers are included in a single fragment and the system is tessellated into overlapping fragments based on a simple distance criterion, but in a manner whose set-theoretical derivation prevents over- or undercounting of interactions despite the use of overlapping fragments.[71,72] Obviously, one could assign more than one monomer to a single fragment in a traditional MBE(n) calculation, but we anticipate that distance-based (rather than cardinality-based) fragmentation will make it easier to adjust the fragment definitions on-the-fly in a continuous way based on distance cutoffs, during simulations or geometry optimizations, although we leave this question to a future study. The

presence of multiple monomers per fragment means that some many-body effects are included already at the level of GMBE(1), so this is a non-trivial approximation, in contrast to MBE(1). (All MBE calculations performed here use single-monomer fragments.) The GMBE(1) method is equivalent to the *generalized energy-based fragmentation* approach of Li and co-workers.[78,79]

We find that GMBE(2) is generally sufficient to reproduce total interaction energies in non-covalent clusters, even in very challenging cases such as $F^-(H_2O)_N$.[44,71,72] At the same time, the number of subsystem calculations remains manageable as compared to the four-body approximation that we will see is necessary to obtain accurate MBE(n) results in clusters composed of polar monomers. For water clusters, the 3 Å threshold for GMBE fragmentation that is used here typically results in fragments containing 3–4 H_2O molecules. Application of GMBE(2) to the largest cluster considered here, $(H_2O)_{55}$, requires 1,469 dimer calculations (consisting of 6–9 H_2O molecules for the dimer) and 17,883 calculations on intersections (4–6 H_2O molecules each). These numbers are dwarfed by the demands of an MBE(4) calculation, which requires that subsystem calculations be performed on 341,055 tetramers (four H_2O molecules), 26,235 trimers (three H_2O molecules), and 1,486 calculations dimers. On the other hand, the GMBE(2) subsystems are larger and require ∼9 times more computational time (ωB97X-V/aug-cc-pVTZ level). Operationally, GMBE(2) sidesteps the precision problems associated with MBE(4), but at increased computational cost. Considerations of cost are revisited in Section 2.4.3.

2.2.4 Approximate many-body counterpoise corrections

There has long been a perceived inadequacy with the Boys-Bernardi counterpoise correction, although the evidence for this inadequacy is debatable.[80] The Valiron-Mayer function counterpoise (VMFC) approach[81] was later introduced as an attempt to quantify and correct this inadequacy. It is based on the exact energy expression in Eqs (2.1) and (2.2), with a counterpoise correction applied order-by-order in the MBE. The resulting BSSE correction can be written

$$\delta E^{\mathrm{VMFC}} = \delta E^{\mathrm{CP}} + \sum_{I<J} \left(\Delta E_{IJ}^{IJ} - \Delta E_{IJ}^{IJK\cdots N} \right) + \cdots \tag{2.11}$$

where $\Delta E_{IJ}^{XY\cdots}$, for example, is the two-body interaction between fragments I and J, computed in a basis with functions on centers $XY\cdots$. Analogous expressions exist for the three-body and higher-order terms.[46,76] Note that the monomer term is equivalent to the original Boys-Bernardi correction, δE^{CP} in Eq. (2.9).

In large clusters, where one might expect more significant BSSE effects, even the original Boys-Bernardi correction is expensive to evaluate, and evaluation of VMFC becomes intractable even more rapidly. Indeed, the cost of these calculations has impeded any final reckoning on the (in)adequacy of the Boys-Bernardi approach. Several approximations have been proposed that are consistent with a truncated MBE, and in particular we have suggested an n-body approximation to Eq. (2.9) that we have called a many-body counterpoise correction truncated at order n, or MBCP(n).[46,47] This approximation consists in applying an n-body expansion to $E_I^{IJK\cdots N}$ in Eq. (2.9).

For $n = 2$–4, the MBCP(n) approximations for monomer I are

$$\delta E_I^{\mathrm{MBCP}(2)} = (N-1)E_I^I - \sum_{J\neq I}^{N} E_I^{IJ} \ , \tag{2.12}$$

$$\delta E_I^{\mathrm{MBCP}(3)} = \delta E_I^{\mathrm{MBCP}(2)} + (N-2)\sum_{J\neq I}^{N} E_I^{IJ} - \sum_{J\neq I}^{N}\sum_{\substack{K>J\\K\neq I}}^{N} E_I^{IJK} - \tfrac{1}{2}(N-2)(N-1)E_I^I \ , \tag{2.13}$$

and

$$\begin{aligned}\delta E_I^{\mathrm{MBCP}(4)} = \delta E_I^{\mathrm{MBCP}(3)} &+ (N-3)\sum_{J\neq I}^{N}\sum_{\substack{K>J\\K\neq I}}^{N} E_I^{IJK} - \sum_{J\neq I}^{N}\sum_{\substack{K>J\\K\neq I}}^{N}\sum_{\substack{L>K\\L\neq I}}^{N} E_I^{IJKL} \\ &- \tfrac{1}{2}(N-3)(N-2)\sum_{J\neq I}^{N} E_I^{IJ} + \tfrac{1}{6}(N-3)(N-2)(N-1)E_I^I \ .\end{aligned} \tag{2.14}$$

Analogous to Eq. (2.9), the overall MBCP(n) counterpoise correction is then

$$\delta E^{\mathrm{MBCP}(n)} = \sum_{I=1}^{N} \delta E_I^{\mathrm{MBCP}(n)} \ , \tag{2.15}$$

A somewhat different counterpoise technique, termed by "many-ghost many-body expansion", has been suggested recently,[82] which consists in applying a MBE to each of the terms containing ghost functions in Eq. (2.11). In practice, the higher-order terms are found to be negligible and only the leading term was retained in Ref. 82; this leading term is precisely what we call MBCP(n). We use this correction together with an n-body expansion of the energy in a method that we call MBE(n)+MBCP(n), meaning that n-body approximations are applied separately to both $\Delta E^{(\mathrm{uncorr})}$ and δE^{CP} in Eq. (2.7).

In a similar vein, Kamiya *et al.*[76] introduced an n-body approximation to the VMFC procedure, which we call VMFC(n). The MBCP(2) and VMFC(2) corrections are equivalent but the two approaches differ starting at $n = 3$, with VMFC(n)

involving a significantly larger number of calculations. Based on comparisons to complete-basis benchmarks in Ref. 46, for systems no larger than $F^-(H_2O)_{10}$, it is not clear that the increased expense of VMFC(3) is justified, but those are the largest VMFC(3) calculations reported to date and results in larger clusters may paint a different picture.

Alternatively, δE^{CP} in Eq. (2.9) can be approximated using the GMBE. When truncated at order n, we will refer to this as a generalized many-body counterpoise correction, GMBCP(n), the formula for which is introduced here for the first time:

$$\delta E^{\mathrm{GMBCP}(n)} = \sum_{I=1}^{N} \left[\sum_{i=1}^{\binom{N_f}{n}} E_I^i - \sum_{i=1}^{\binom{N_f}{n}} \sum_{j>i}^{\binom{N_f}{n}} E_I^{i\cap j} + \cdots + (-1)^{\binom{N_f}{n}+1} E_I^{i\cap j\cap\cdots\cap\binom{N_f}{n}} \right] . \quad (2.16)$$

The notation here requires some explanation. For each monomer I, the first term inside the brackets ($\sum_i E_I^i$) represents the energy of I computed in all of the possible n-mer basis sets that contain this monomer. In the second term, the energy of monomer I is computed in all possible two-fragment intersection basis sets that contain monomer I, and so on. Like MBCP(n), there is only one "real" monomer in any of these subsystem calculations, with the other fragments serving simply as centers for basis functions. VMFC(n), in contrast, requires calculations with up to $n-1$ real monomers in an n-body basis.

2.3 Computational details

In some previous studies, the MBE has been used in a "multi-level" or "stratified" way, with different n-body terms computed at different levels of theory.[83–87] In other cases, efforts have been made to identify a subsystem level of theory that will reproduce

high-accuracy supersystem benchmarks, computed at an altogether different level of theory.[88] In contrast, the aim here is to test the accuracy and convergence of the n-body expansion itself, so we will exclusively compare n-body calculations to supersystem benchmarks computed at the same level of theory.

In the first part of this study, we examine a sequence of water clusters, $(H_2O)_{N=6-55}$, whose structures are putative global minima on the TIP4P potential surface at each cluster size.[89] The same set of clusters was used in our previous investigation of precision considerations for the MBE,[48] and as in that study, we use the affordable B3LYP/cc-pVDZ level of theory since our goal is to understand the size- and n-body-dependence of errors in total interaction energies. In view of a recommendation by Bettens *et al.*[90] that the MBE should never be used with augmented basis sets (due to the appearance of serious BSSE problems when diffuse functions are added), we will also test aug-cc-pVDZ and 6-31+G(d,2p). The SG-1 quadrature grid[91] was used for all B3LYP calculations. A detailed study in Ref. 48 suggests that while MBE results are sensitive to certain numerical thresholds, as discussed below, they are only very weakly-dependent on the choice of grid.

Various forms of electrostatic embedding (EE) have been used in an attempt to hasten the convergence (and thus improve the accuracy) of the n-body expansion,[43] and in some cases EE-MBE results are surprisingly insensitive to the precise details of the embedding charges,[45,64,92] at least in systems such as small water clusters. Here, we test the effects of including supersystem-derived atom-centered point charges.

computed at the B3LYP/cc-pVDZ level. (This is not ideal for production calculations, as it requires the very supersystem calculation that we are trying to avoid, but is used here for testing purposes and in the end we will conclude that embedding charges provide very little advantage anyway.) Mulliken charges, Hirshfeld charges,[93] charges from natural population analysis (NPA),[94] charges derived from the electrostatic potential ("ChElPG"),[95] and charges from the empirical CM5 model.[96]

To avoid precision issues associated with the reading and writing of electronic structure input and output files, all calculations are carried out completely internally within a locally-modified version of Q-Chem,[97] where the subsystem calculations are parallelized via the message-passing interface. (This will be released with Q-Chem v. 4.4.) By performing all MBE calculations internally within Q-Chem, we avoid the need to compute and subtract out the self-interaction of the embedding charges. Although this may seem like a trivial point, errors due to finite precision can accumulate rapidly in the self-energy calculation, especially for the larger systems considered here, if the self-interaction calculation is not performed using all digits of double-precision arithmetic.[48] GMBE and GMBCP calculations were performed using our driver program, Fragme∩t,[71,72,77] to prepare and execute the necessary Q-Chem input files in full double precision. Note that this requires reading Q-Chem's binary scratch files.

In the second part of this study, we focus on establishing, and then attempting to reproduce, high-accuracy benchmarks for a set of $(H_2O)_{20}$ clusters representing four different structural motifs. In Ref. 3, structures for 20 different isomers of

$(H_2O)_{20}$ are reported, representing the five lowest-lying isomers in each of four families of structures, according to basin-hopping Monte Carlo calculations. For reasons of cost, we consider only 8 of these structures, representing the highest- and lowest-energy structures of the five reported for each structural motif. (Structures of the 8 that we consider are provided in the Supplementary Material.) Benchmark energetics are computed using second-order Møller-Plesset perturbation theory (MP2) within the resolution-of-identity (RI) approximation,[98–100] RIMP2, as implemented in Q-Chem.[101] The aug-cc-pVTZ (aTZ) and aug-cc-pVQZ (aQZ) basis sets are used in conjunction with their corresponding auxiliary basis sets,[100,102] and only valence orbitals are correlated. We assume that the Hartree-Fock/aQZ energy is sufficiently close to the complete basis set (CBS) limit to be used without extrapolation, whereas we extrapolate the RIMP2 correlation energies using a two-point ("aTZ/aQZ") X^{-3} scheme.[73] Standard Boys-Bernardi counterpoise corrections are applied,[74] which for a many-body cluster means that we compute each monomer energy using the full cluster basis,[75] as in Eq. (2.9).

In principle these RIMP2/CBS benchmarks could be extended to the CCSD(T)/CBS level by means of a triples correction

$$\delta_{\mathrm{CCSD(T)}} = E_{\mathrm{CCSD(T)}} - E_{\mathrm{MP2}} \,, \tag{2.17}$$

and good results based on two- and three-body approximations to Eq. (2.17) have been demonstrated.[46,56,103,104] The $\delta_{\mathrm{CCSD(T)}}$ for $(H_2O)_{20}$ has been estimated to be around -2.7 kcal/mol,[18,57] but in the interest of performing a larger number of exploratory calculations, we omit this correction from the MP2 calculations reported here. As an

alternative, we will report calculations using the ωB97X-V density functional,[105] as it has been shown to provide accurate non-covalent interaction energies, when used in conjunction with a triple-ζ basis set.[52,57,105] In particular, for $(H_2O)_{20}$, counterpoise-corrected ωB97X-V/aTZ calculations differ from CCSD(T)/CBS benchmarks only by an average of 1.7 kcal/mol.[59] Per the recommendation in Ref. 105, we used used an Euler-Maclaurin-Lebedev grid with 75 radial points and 302 angular points to integrate the the semi-local parts of ωB97X-V, and the SG-1 grid[91] for the nonlocal correlation part.

2.4 Results and discussion

In Section 2.4.1 below, we focus on how well the MBE reproduces a supersystem calculation at the same level of theory. It will be seen that this issue is complicated by the oscillatory (with respect to n) and size-extensive (with respect to N) nature of the resulting errors. In Section 2.4.2, we focus on the ability of the MBE to reproduce high-accuracy benchmarks. There, we will see that error cancellation between MBE(n) and MBCP(n) plays a major role in the accuracy of the results, whereas the good performance of GMBE(2)+GMBCP(2) does not rely on cancellation of errors.

2.4.1 Absolute performance of the many-body expansion

Errors versus system size

Previously, we observed that the total energy predicted by an n-body expansion can change dramatically as a function of both the SCF convergence criterion, τ_{SCF}, as well

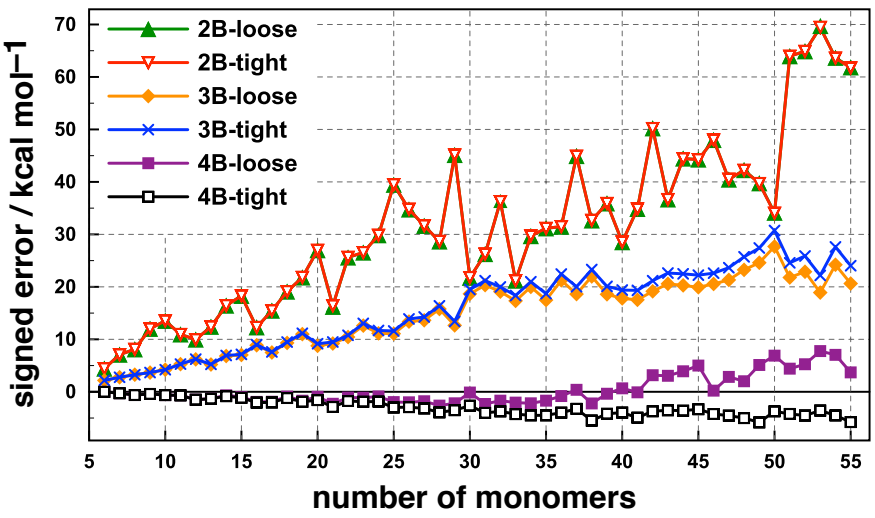

Figure 2.1: Signed errors in two-, three-, and four-body total interaction energies for a series of $(H_2O)_{6-55}$ clusters relative to supersystem results at the B3LYP/cc-pVDZ level. The "loose" thresholds are $\tau_{\text{SCF}} = 10^{-5}$ a.u. and $\tau_{\text{ints}} = 10^{-9}$ a.u., whereas "tight" thresholds are $\tau_{\text{SCF}} = 10^{-7}$ a.u. and $\tau_{\text{ints}} = 10^{-14}$ a.u.

as the integral screening threshold, τ_{ints}.[45,48] Figure 2.1 presents a more thorough, size-dependent analysis of the effects of numerical thresholds, comparing results obtained with "loose" thresholds (defined here as $\tau_{\text{SCF}} = 10^{-5}$ a.u. and $\tau_{\text{ints}} = 10^{-9}$ a.u.) to those garnered from "tight" thresholds ($\tau_{\text{SCF}} = 10^{-7}$ a.u. and $\tau_{\text{ints}} = 10^{-14}$ a.u.). A statistical summary of the mean absolute errors (MAEs) is given in Table 2.1. Throughout this work, we define

$$\text{error} = E^{(n)} - E_{\text{supersystem}} \ , \tag{2.18}$$

so that negative error indicates that the n-body approximation is overbound with respect to the benchmark.

Embedding	MAE (kcal/mol)			
	$n = 1$	$n = 2$	$n = 3$	$n = 4$
	— MBE (loose thresholds) —			
None	–	31.31 (1.04)	14.73 (0.48)	2.11 (0.07)
	— MBE (tight thresholds) —			
None	–	31.31 (1.04)	15.88 (0.51)	3.01 (0.10)
Mulliken	–	5.38 (0.20)	14.15 (0.45)	2.14 (0.07)
ChElPG	–	47.14 (1.45)	12.68 (0.40)	0.79 (0.03)
NPA	–	57.74 (1.77)	12.57 (0.39)	0.75 (0.03)
Hirshfeld	–	6.92 (0.25)	14.29 (0.46)	2.24 (0.07)
CM5	–	30.59 (0.92)	12.91 (0.41)	1.11 (0.04)
	— GMBE (tight thresholds) —			
None	26.68 (0.78)	0.68 (0.02)	–	–

Table 2.1: MAEs and (in parenthesis) MAEs per monomer, for (G)MBE(n) calculations of clusters in the range $(H_2O)_{6-55}$, with energetics computed at the B3LYP/cc-pVDZ level.

From Fig. 2.1, the difference between MBE(2) using loose and tight thresholds cannot be differentiated, and is only barely discernible for MBE(3), except for the largest clusters. When using a four-body expansion, however, the accumulation of roundoff is enough to change the sign of the error for $N \gtrsim 40$. In the largest cluster, $(H_2O)_{55}$, the differences between results with loose and tight thresholds are -0.11 kcal/mol (two-body), 3.40 kcal/mol (three-body), and -9.47 kcal/mol (four-body). This observation complements the propagation-of-errors analysis that we presented in Ref. 48, and underscores the fact that each subsystem energy in the MBE is multiplied by a binomial coefficient that is growing factorially with respect to both n and N [see Eq. (2.5)]. All remaining calculations in this work will be performed using the tight thresholds. For a B3LYP/cc-pVDZ calculation of $(H_2O)_{55}$, this requires about 2.4

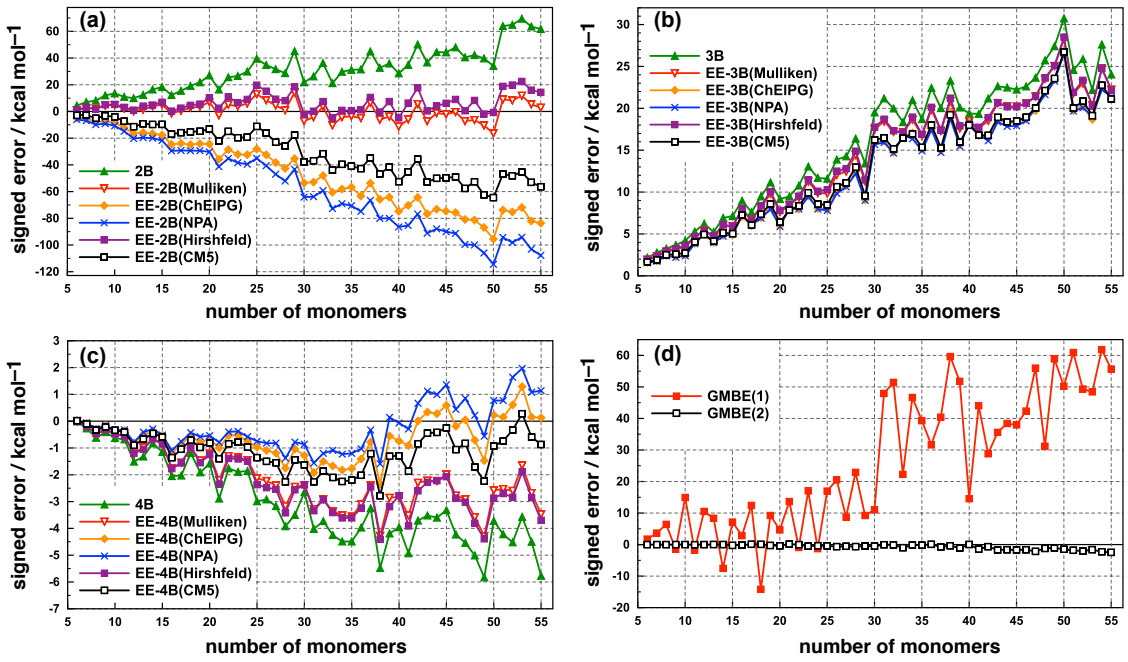

Figure 2.2: Signed errors in total interaction energies for a series of water clusters using (a) two-body, (b) three-body, and (c) four-body expansions, at the B3LYP/cc-pVDZ level, either with or without electrostatic embedding. In (d), we plot the signed errors for GMBE(1) and GMBE(2). Error is measured with respect to a supersystem calculation at the same level of theory. Note that the vertical energies scales are different in each panel.

times more computer time as compared to the looser thresholds.

The other obvious message from these data is that, even with tight thresholds, errors at the MBE(2) and MBE(3) level are unacceptably large, even for clusters containing only a few water molecules. Electrostatic embedding is designed to rectify this, so in Fig. 2.2 we plot size-dependent errors in MBE(n) using a variety of different embedding charges. [We use the notation "EE-nB" to mean an embedded version of MBE(n).] In stark contrast to the conventional wisdom that embedding charges

should improve the accuracy of truncated MBEs, no such trend is observed at the two-body level, where the errors can change dramatically (even in sign) depending upon the choice of point charges. In the three- and four-body cases, use of any one of five different choices for the embedding charges does lead to a systematic reduction in the error across all cluster sizes, and the results are consistent with previous observations that the results are rather insensitive to the particulars of how the embedding charges are chosen.[45,64,92] Moreover, they are consistent with the notation that the details of the embedding should matter less as n increases and ever-large sub-clusters are described quantum-mechanically.[64] On the other hand, in the three-body case the error reduction engendered by the embedding charges is quite small relative to the overall error with respect to the supersystem benchmark. Only in the four-body case—which we would like to avoid, by virtue of its disastrous combinatorics—do we see a meaningful reduction in the errors by virtue of the embedding charges.

Of all the EE-MBE(n) methods examined in Fig. 2.2, only the four-body expansion with NPA, ChElPG, or CM5 charges achieves so-called "chemical accuracy" of $\sim$ 1 kcal/mol. Note that the error *per monomer* (Table 2.1) is a size-intensive quantity, unlike the total errors in Fig. 2.2, and it has been argued that for dynamical studies of cluster evolution, it is sufficient to achieve a "dynamical accuracy" per monomer equal to 10% of the average molecular kinetic energy at room temperature, $(3/2)k_BT$.[82] This amounts to an accuracy threshold of 0.09 kcal/mol, and unfortunately none of the two- or three-body expansions achieves even this level of accuracy, although all of the four-body methods do, including the MBE(4) with no embedding charges at

all. This is a useful fact to note, given that evaluation of the Coulomb self-energy of the embedding charges is a serious obstacle to obtaining reproducible, high-precision results when the n-body expansion is implemented via an external driver program or script.[48]

Several previous studies have noted that the "success" of n-body expansions often relies heavily on error cancellation,[47,76,83,90] and the EE-2B data in Fig. 2.2(a) suggest that MBE(2) with either Mulliken or Hirshfeld embedding represents a "Pauling point"[106] at the two-body level. This is almost certainly not for any physically-meaningful reason, given that the average Mulliken charge on oxygen is $-0.26e$ in the case of $(H_2O)_{40}$, and $-0.22e$ when Hirshfeld charges are used, as compared to $-0.80e$ for ChElPG (which is similar to force-field charges designed to reproduce the presumed H_2O dipole moment) and $-0.99e$ for NPA (consistent with chemical intuition).

Finally, results for GMBE(n) are plotted in Fig. 2.2(d). As noted above, the GMBE includes some many-body effects even for $n = 1$, but these turn out to be insufficient to afford reasonable accuracy even in small clusters, at least with the 3 Å threshold for fragment formation that is used here. Using the same threshold, however, GMBE(2) results are outstanding, with a MAE per monomer of only 0.02 kcal/mol across the whole range of cluster sizes, $N = 6$–55. This is more accurate than any of the MBE(4) methods, and this accuracy does not require the use of embedding charges. While it is likely true that the total error in GMBE(2) calculations is size-extensive, the plot in Fig. 2.2(d) suggests that the rate of growth with N is so

Embedding	MAE per monomer (kcal/mol)		
	cc-pVDZ	aug-cc-pVDZ	6-31+G(d,2p)
None	0.51	0.06	0.11
Mulliken	0.45	0.07	0.05
ChElPG	0.40	0.37	0.09
NPA	0.39	0.56	0.11
Hirshfeld	0.46	0.04	0.05
CM5	0.41	0.23	0.06

Table 2.2: Mean absolute errors per monomer for EE-MBE(3) calculations of $(H_2O)_{6-55}$ at the B3LYP level in several basis sets. Supersystem charges computed at the B3LYP/cc-pVDZ are used for charge embedding.

small that this extensivity is unlikely to prove problematic for values of N that are likely to be used in practical calculations, and GMBE(2) can safely be employed to predict *total* cluster binding energies.

Basis-set dependence

Much of the rationale for examining water clusters in this work is these have been popular test cases for various fragment-based methods, and in the Supplementary Material we compare our EE-2B and EE-3B results to a large number of previous benchmarks from the literature. Care must be taken in comparing these literature values either to one another or to the numbers reported here, because it is uncommon to find a detailed specification of how the thresholds were set in a given calculation, and these settings can significantly alter the results, as can the choice of electrostatic embedding, in the two-body case.

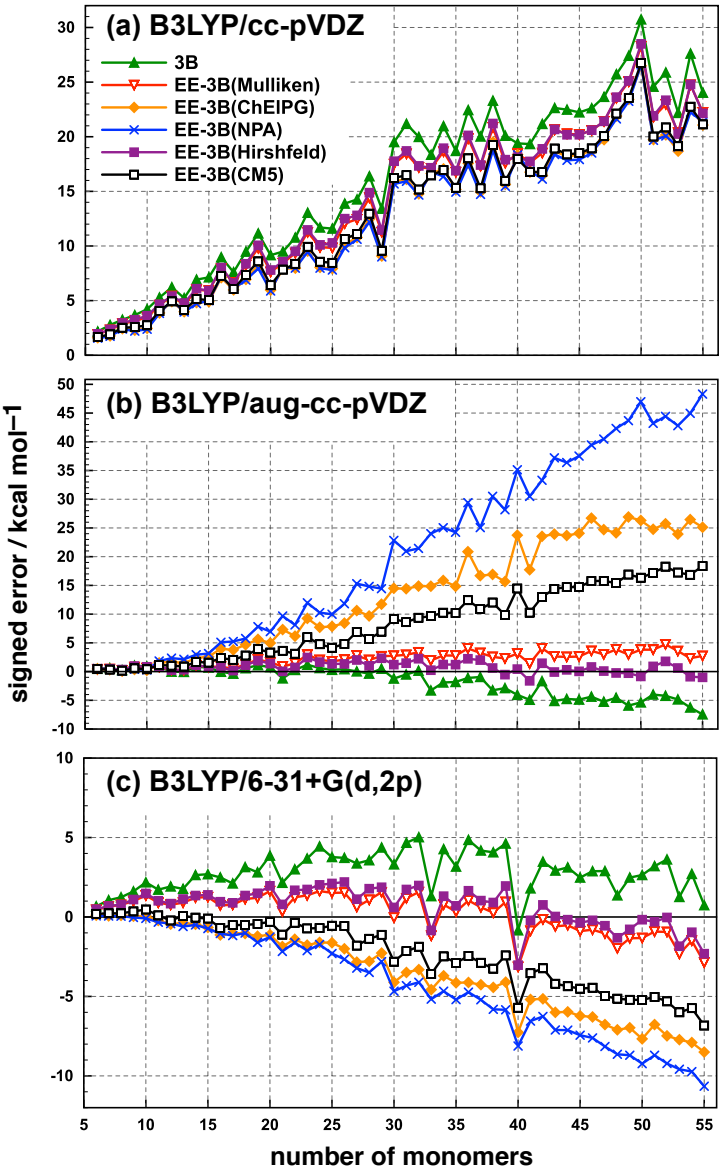

Figure 2.3: Signed errors in the MBE(3) total interaction energies for $(H_2O)_{6-55}$, using the B3LYP functional with three different basis sets. Error is measured with respect to a supersystem calculation at the same level of theory.

That said, a survey of the literature suggests that the large MBE(2) errors documented above are not unprecedented. Our MBE(3) results are noticeably worse than most literature values, however. With only a few exceptions, the literature reports errors of $\lesssim 3$ kcal/mol for clusters in the size range $N = 16$–32, with numerous reports of errors < 1 kcal/mol in basis sets such as STO-3G, 3-21G, and 6-31G* (see Table S2).

In an attempt to understand this discrepancy, we have examined size-dependent trends in the MBE(3) errors in three different basis sets, as plotted in Fig. 2.3 with error statistics are summarized in Table 2.2. In contrast to B3LYP/cc-pVDZ results, where the choice of embedding charges made only a minor difference in the error, results using aug-cc-pVDZ show a dramatic dependence on the choice of embedding charges. Starting from the MBE(3) results without charge embedding [green symbols in Fig. 2.3(b)], the various EE-MBE(3) plots diverge in an order that correlates precisely with the average embedding charge on the oxygen atoms, which is small for the Hirshfeld and Mulliken embeddings but nearly $-1e$ for NPA charges. Use of the latter leads to errors that approach 50 kcal/mol for the largest clusters, whereas errors are < 10 kcal/mol when no embedding charges are used at all. This strongly suggests that overpolarization caused by point charges interacting with diffuse basis functions is the reason for the dramatic dependence on the choice of embedding charges when the aug-cc-pVDZ basis set is used.

Very accurate results for small water clusters ($N < 10$) were reported in Ref. 84 using EE-MBE(3) at the B3LYP/6-31+G(d,2p) level, so we have applied the

MAE	Embedding	Signed Error (kcal/mol)			
		MBE(2)	MBE(3)	MBE(4)	GMBE(2)
cc-pVDZ	none	28.56	19.39	−3.96	−0.79
cc-pVDZ	ChElPG	−74.74	18.36	−0.74	—
aug-cc-pVDZ	none	105.00	−4.05	−9.26	0.23
aug-cc-pVDZ	ChElPG	−22.06	23.74	−75.35	—
cc-pVTZ	none	20.09	32.59	−9.85	0.05
cc-pVTZ	ChElPG	−68.78	28.71	−6.50	—
aug-cc-pVTZ	none	105.54	−8.92	9.26	0.33
aug-cc-pVTZ	ChElPG	−0.40	4.19	−23.40	—
6-31G(d,2p)	none	74.18	5.83	−0.82	0.07
6-31G(d,2p)	ChElPG	−32.47	5.77	1.60	—
6-31+G(d,2p)	none	84.24	−0.84	5.19	0.73
6-31+G(d,2p)	ChElPG	6.68	−7.26	5.65	—

Table 2.3: Signed errors for $(H_2O)_{40}$ using B3LYP in various basis sets. (Embedding charges are from a supersystem calculation at the level of B3LYP/cc-pVDZ.)

same methodology to the $(H_2O)_{6-55}$ sequence, with errors reported in Fig. 2.3(c). MAEs in Table 2.2 show that each choice of embedding charges except NPA leads to reduction in the errors, and even the errors using NPA charges are comparable in magnitude (though different in sign) to those without embedding. The errors are also much smaller using 6-31+G(d,2p) as compared to either of the other two basis sets, although total errors remain $\gg 1$ kcal/mol for the largest clusters.

This dramatic sensitivity to basis sets is the most surprising conclusion drawn from Fig. 2.3, all the more so in view of the fact that the three basis sets used here

are not *so* dissimilar. A recent review of fragment-based methods[43] mentions the need for a comprehensive study of how the accuracy of each method depends upon the level of electronic structure theory used for the subsystem calculations, yet few systematic studies exist in the literature.[45,47,48,90] Here, we use B3LYP in conjunction with six different double- and triple-ζ basis sets to study the performance of (G)MBE(n) for $(H_2O)_{40}$, both in the absence of embedding charges and when ChElPG embedding charges are used. Signed errors are reported in Table 2.3.

For MBE(2), results are improved by employing charge embedding in all basis sets except cc-pVDZ and cc-pVTZ. For MBE(3), results are improved by employing charge embedding in all basis sets except aug-cc-pVDZ and 6-31+G(d,2p). For MBE(4), results are improved only for cc-pVDZ and cc-pVTZ by employing charge embedding. In short, the results are rather erratic, and in none of the six basis sets tested does charge embedding consistently reduce the errors for two-, three-, and four-body expansions. Moreover, in the absence of embedding charges we observe a monotonic decrease in the errors (going from n = 2–4) only for cc-pVDZ and 6-31G(d,2p). This is consistent with the observation by Ouyang *et al.*[90] that only in the absence of diffuse functions does one obtain monotone convergence of the MBE, although our results for cc-pVTZ are an exception demonstrating that the absence of diffuse functions alone does not guarantee monotone convergence. In summary, and in view of all the data presented so far, it is difficult to predict whether MBE(n) results will be improved or degraded by MBE($n+1$) for a given basis set or charge embedding scheme.

Fortunately, GMBE(2) performs well for all six basis sets tested, and is superior to (or in one case, comparable to) MBE(4) results with or without charge embedding. Recent work has shown that the performance of GMBE(2) does depend on the nature of the embedding charges,[77] so in this study we choose *not* to use these charges in GMBE calculations.

Convergence of the expansion

Ultimately, one needs a metric for gauging the accuracy of a given approach that does not require a supersystem calculation, else the utility of the MBE is lost. Given the apparent extensivity of errors in total interaction energies, it is not immediately clear how (or even whether) small cluster benchmarks can be used to assess accuracy in applications to much larger clusters. A reasonable metric is to examine how the errors converge (or fail to converge) with respect to the truncation order, n, but unfortunately the convergence behavior of the MBE is complicated,[82,90,107–110] as seen above. For atomic clusters, the MBE is oscillatory and slowly-convergent, often precluding truncation.[107–109] For small molecular clusters, the situation may not be so dire, as the oscillations tend to settle down much more quickly,[83,109,111] possibly owing to the weaker nature of intermolecular interactions as compared to interatomic interactions.

Consider the n-body interaction energy $\Delta E_I^{(n)}$ in Eq. (2.3). As a simple, qualitative model, let us assume that the error in each subsystem energy, $E_J^{(n-m)}$, is the

same, and denote this error as δE. Then the total error in $\Delta E_I^{(n)}$ would be

$$\begin{aligned} \delta\left(\Delta E_I^{(n)}\right) &= (\delta E)\sum_{m=0}^{n-1}(-1)^m\binom{n}{m} \\ &= (-1)^{n+1}(\delta E)\ . \end{aligned} \tag{2.19}$$

This suggests there is reason to expect that errors in MBE(n) calculations may oscillate as a function of N, although the result is only rigorous if the errors are truly identical for each n, which likely requires that $|\Delta E_I^{(n)}| \approx |\Delta E_J^{(n+1)}|$. If the magnitude of the $(n+1)$-body interactions are significantly different than those of the n-body interactions, then one cannot say with certainty whether the error will oscillate or not.

In light of this analysis, let us reconsider the $n = 2$ results in Fig. 2.2(a). According to Eq. (2.19), the error should be negative, and indeed large, negative errors are obtained using CM5, ChElPG, and NPA charges, but on the other hand large *positive* errors are obtained in the absence of charge embedding. (For reasons ultimately having to do with error cancellation, the Hirshfeld and Mulliken embeddings lead to errors much closer to zero, and which change sign as a function of cluster size.) These observations suggest that the analysis in Eq. (2.19) is overly simplistic in this case, consistent with the fact that $|\Delta E_I^{(1)}| \gg |\Delta E_I^{(2)}|$.

On the other hand, errors for EE-MBE(3) [Fig. 2.2(b)] are all positive, consistent with Eq. (2.19), and for EE-MBE(4) [Fig. 2.2(c)], most of the errors are negative, and those that are positive are relatively small. Thus, we conclude there is reason to expect that the MBE converges (if at all) in an oscillatory way, behavior that may prove troublesome for high-accuracy applications if the oscillations do not decay

rapidly enough. In the present calculations, this does not seem to be the case even by $n = 4$, necessitating five-body calculations to check for convergence. As noted in Ref. 47, and in Section 2.4.2 below, these oscillations may play a role in error cancellation. BSSE is a significant component of these oscillations.[90]

2.4.2 High-accuracy calculations of relative energies

To this point we have focused on the intricacies of replicating a low level of theory in a large system. We now turn our attention to an investigation of the accuracy of the MBE as compared to high-quality RIMP2/CBS and ωB97X-V/aTZ benchmarks, specifically for relative energies of $(H_2O)_{20}$ isomers.

Accuracy of the benchmarks

Total interaction energies for eight different $(H_2O)_{20}$ clusters are plotted in Fig. 2.4. In these calculations, we attempted to reach the RIMP2/CBS using T/Q extrapolation of RIMP2/aug-cc-pVXZ data, both with and without counterpoise corrections. These corrections are quite large, on average 21 kcal/mol for the aTZ basis set and 10 kcal/mol for the aQZ basis set, and the extrapolated RIMP2/CBS values in the absence of counterpoise correction are shifted to larger interaction energies by $\approx$ 3.5 kcal/mol relative to an extrapolation of the counterpoise-corrected values. The counterpoise-corrected RIMP2/CBS extrapolations differ by an average of 2.7 kcal/mol as compared to counterpoise-corrected CCSD(T)/CBS benchmarks.[57,59]

The ωB97X-V functional[105] combined with triple-ζ basis sets gives good performance for interaction energies in $(H_2O)_{20}$ with a MAE of 1.7 kcal/mol as compared

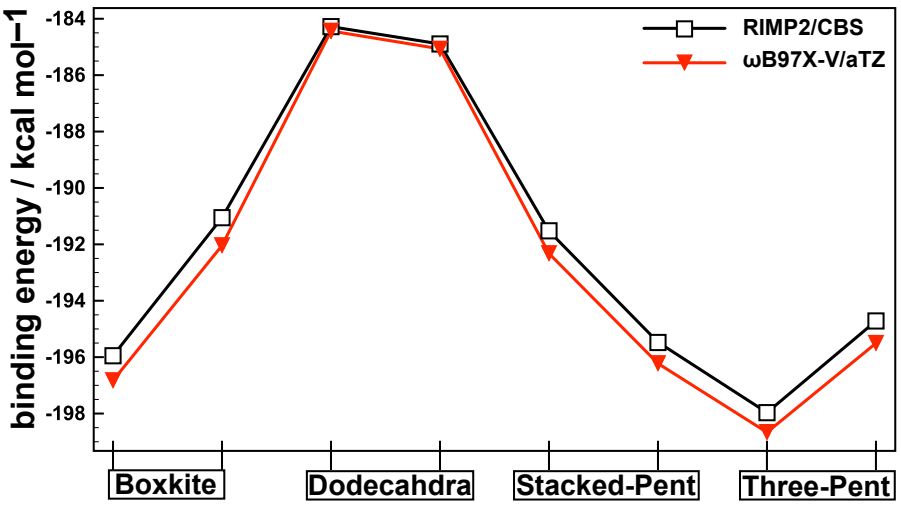

Figure 2.4: Total interaction energies for eight different isomers of $(H_2O)_{20}$, including two from each of the four families of structural motifs.[3]

to CCSD(T)/CBS benchmarks.[59] In the absence of counterpoise correction, ωB97X-V/aTZ values are shifted to higher interaction energies by $\approx$ 2.8 kcal/mol, relative to counterpoise-corrected results. In contrast to the RIMP2/aTZ case, counterpoise corrections at the ωB97X-V/aTZ level are $<$ 1 kcal/mol, consistent with more rapid basis-set convergence of DFT as compared to MP2. Counterpoise-corrected ωB97X-V/aTZ interaction energies agree with (counterpoise-corrected) RIMP2/CBS results to within 1 kcal/mol, as shown in Fig. 2.4, suggesting that these two methods are reliable levels of theory for relative energies of $(H_2O)_{20}$ isomers. In the next section, we examine how well n-body approximations can reproduce these benchmarks.

Accuracy of the (G)MBE

In Fig. 2.5, we examine the errors in three- and four-body versions of MBE(n)+MBCP(n) as compared to counterpoise-corrected RIMP2/CBS and ωB97X-V/aTZ benchmarks. We find that MBE(2)+MBCP(2) performs quite poorly, with errors $>$ 40 kcal/mol, so this approach will not be discussed here. In contrast, the GMBE(2)+GMBCP(2) performs well but the GMBCP(2) correction is very demanding in terms of the number of subsystem calculations that are required, hence we have only used this method in the aTZ basis set, and will not extrapolate to the CBS limit.

In each of the MBE(3)+MBCP(3) calculations shown in Fig. 2.5, the MAE is $>$ 1 kcal/mol, and this three-body approach consistently underestimates the interaction energies in the two dodecahedral isomers of $(H_2O)_{20}$, as compared to those in the other three structural motifs. (In a previous study, we noted that BSSE effects are quite different for the dodecahedrons as compared to the other families of isomers.[47]) This is somewhat disturbing, in that it suggests that a three-body expansion might compare favorably to benchmark calculations in one region of the potential surface, only to perform in a less-favorable for other isomers of the same cluster. On the other hand, the MBE(4)+MBCP(4) and GMBE(2)+GMBCP(2) methods perform consistently well for all four classes of $(H_2O)_{20}$ isomers, with MAEs $<$ 0.3 kcal/mol.

In an attempt to separate the role of BSSE corrections from the performance of the underlying n-body approximation to the total, uncorrected interaction energy, we plot in Fig. 2.6 the errors in MBE(3), MBE(4), and GMBE(2) interaction energies for the aTZ basis set, the aQZ basis set, and the CBS extrapolation, without applying

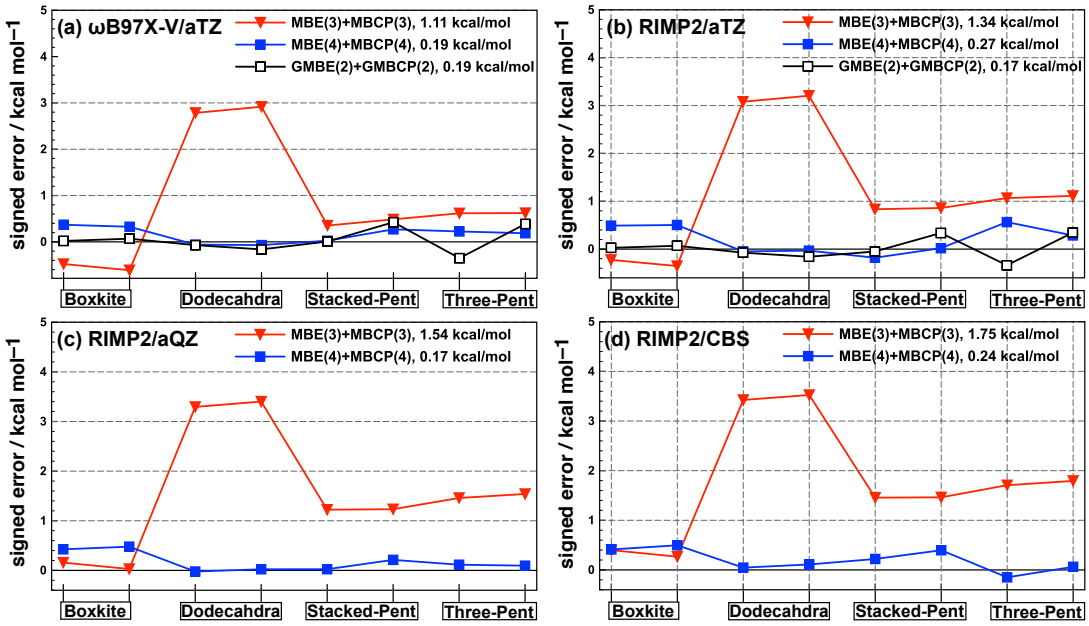

Figure 2.5: Signed errors in relative energies of $(H_2O)_{20}$ isomers at the indicated levels of theory, using the MBE(n)+MBCP(n) approach for $n = 3$ and 4. MAEs for each fragmentation method are also shown, with respect to supersystem benchmarks computed at the same level of theory. All supersystem benchmarks are counterpoise-corrected, hence the MBE approximation with MBCP correction, or GMBE with GMBCP correction, is the appropriate comparison.

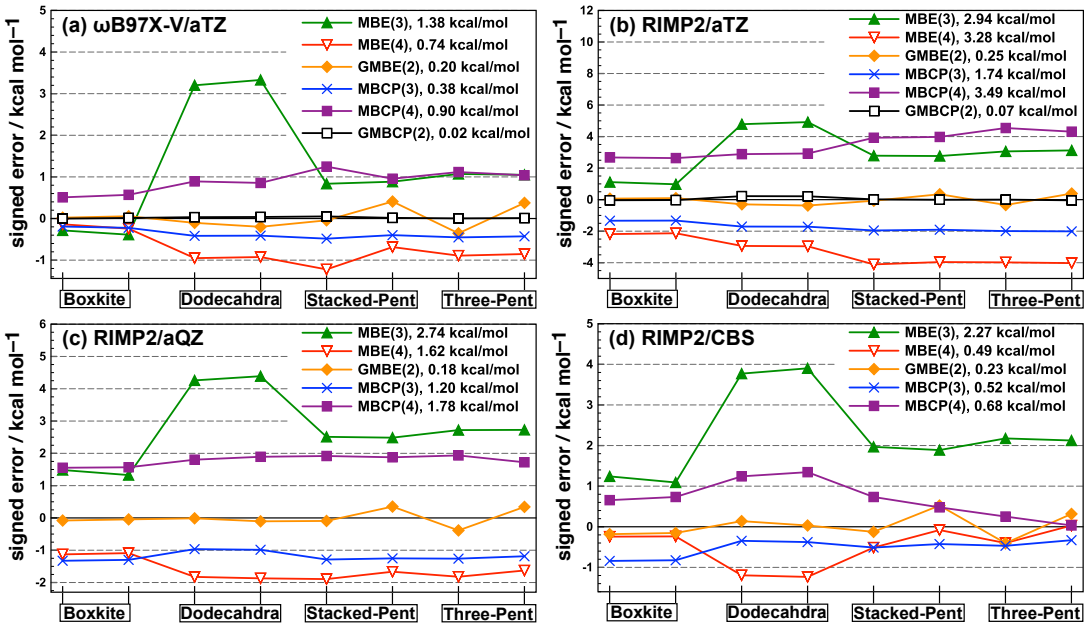

Figure 2.6: Signed errors in relative energies of $(H_2O)_{20}$ isomers at the indicated levels of theory. Errors in MBE(n) and GMBE(2) results are defined relative to a supersystem calculation that has *not* been corrected for BSSE, and errors in the MBCP(n) and GMBCP(2) counterpoise corrections are defined relative to the "full" (supersystem) counterpoise correction. Note that the vertical energy scales are different in each panel.

counterpoise corrections to either the supersystem calculation or the n-body calculation. (This is therefore a test of how well a BSSE-contaminated n-body expansion reproduces the results of a BSSE-contaminated supersystem calculation.) In the same figure, we also isolate the BSSE calculations and compare the MBCP(3), MBCP(4), and GMBCP(2) corrections to the "full" counterpoise correction, the latter being defined as the difference between traditional counterpoise-corrected and uncorrected supersystem calculations.

Previously, Ouyang *et al.*[90] have suggested that, for reasons related to BSSE, MBE calculations should be performed using either the full cluster basis set (which is intractable in large systems), or else using basis sets of at least aug-cc-pVTZ quality, and these are the only basis sets considered in Fig. 2.6. Here, RIMP2 results indicate that the performance of MBE(n) increases systematically as the basis set is improved, especially for $n = 4$, whereas the errors start small for GMBE(2) calculations—even as compared to MBE(4)—and are not substantially different between aTZ, aQZ, and CBS. GMBE(2) is also the best-performing of these three methods at the ωB97X-V level.

Likewise for the counterpoise corrections, the GMBCP(2) errors are smaller than those observed for MBCP(3) or MBCP(4) in both the RIMP2 and the DFT calculations. Because GMBE(2), without counterpoise correction, agrees well with BSSE-contaminated supersystem results, while GMBCP(2) agrees well with the supersystem counterpoise correction, we conclude that the composite GMBE(2)+GMBCP(2)

Method	MAE (kcal/mol)		
	MBE(3)+ MBCP(3)	MBE(4)+ MBCP(4)	GMBE(2)+ GMBCP(2)
	— interaction energy —		
ωB97X-V/aTZ	1.11	0.19	0.19
RIMP2/aTZ	1.34	0.27	0.17
RIMP2/aQZ	1.54	0.17	–
RIMP2/CBS	1.75	0.24	–
	— relative energy —		
ωB97X-V/aTZ	1.03	0.16	0.45
RIMP2/aTZ	1.05	0.42	0.41
RIMP2/aQZ	1.01	0.16	–
RIMP2/CBS	0.98	0.40	–

Table 2.4: MAEs in relative energies of $(H_2O)_{20}$ isomers and in total interaction energies for the same isomers.

method offers good accuracy for the right reasons, and does not rely on error cancellation. In contrast, MBE(3) tends to underestimate the BSSE-contaminated interaction energy while MBCP(3) overestimates the counterpoise correction, while at the four-body level the reverse is true, consistent with the oscillatory nature of the MBE. In these cases, the composite MBE(n)+MBCP(n) method is relying on error cancellation between the neglect of higher-order terms in the MBE and spurious BSSE effects. This example is a more incisive analysis of a similar error cancellation that we first noted in Ref. 47.

In studies of non-covalent clusters, relative energies are usually more important than total interaction energies, the latter of which are only measurable in small clusters. Having demonstrated that error cancellation plays a pivotal role in the accuracy

of MBE(n)+MBCP(n) calculations, it is conceivable that this could be turned into a feature, by exploiting error cancellation to obtain high accuracy at affordable cost. As such, we report MAEs for relative energies of the eight $(H_2O)_{20}$ clusters in Table 2.4, where we also compare them to MAEs in the total interaction energies.

Interestingly, the error statistics for total interaction energies are not significantly different from those for relative isomer energies, vindicating the use of the former as the metric by which MBEs have been assessed in our work and many previous studies. (We rationalize this observation in terms of the fact that the total interaction energy is, from a certain point of view, merely the relative energy between two very different points on the potential energy surface, namely, a deep well versus an exit channel.) The MBE(3)+MBCP(3) approach does not lead to sub-kcal/mol accuracy, for which the four-body analogue is required, or alternatively once can bypass the need for four-body calculations and use GMBE(2)+GMBCP(2) instead, whose favorable performance does not appear to rest on error cancellation.

More exhaustive testing is required in order to obtain a comprehensive understanding of which subsystem levels of theory can be expected to perform well as compared to high-level supersystem benchmarks. Ideally, one would like to establish a protocol (in terms of the choice of embedding, the level of truncation, the size of the fragments, and any BSSE corrections) such that the accuracy of a given electronic structure model for calculations in small clusters would be in some way indicative of its accuracy when applied to large clusters via the MBE. This is needed

especially since RIMP2/CBS, for example, is a reasonable level of theory for water clusters but would not be for applications to systems where a sizable fraction of the interaction energy comes from dispersion. The same can be said for the M11-based[112] density-functional models that are used for water clusters in Ref. 88, whose success for dispersion-bound systems appears to rest on a delicate cancellation of errors.[113] Counterpoise-corrected ωB97X-V/aTZ, on the other hand, appears to be accurate across a broad spectrum of non-covalent interactions,[52,57,105] and the MBE(4)+MBCP(4) and GMBE(2)+GMBCP(2) approximations to this supersystem method accurately reproduce both total interaction energies and relative isomers energies for water clusters. Given a sufficient number of cores over which to distribute such calculations, the time-to-solution can be made quite small.

2.4.3 Computational cost

It is often tacitly assumed that fragment-based methods are always computationally less expensive than the corresponding supersystem calculations, but in terms of aggregate computer time (rather than time-to-solution or "wall time"), this if often not the case.[77] In this section, we consider the cost of the (G)MBE calculations reported here, with the caveat that we have made no attempt to perform any sort of thresholding, by means of which subsystem calculations involving spatially-distant monomers might be neglected or approximated. Preliminary tests suggests that a significant fraction of the subsystem calculations are negligible in many cases, and that this fraction increases with n, but since the purpose of this work is to evaluate

the intrinsic accuracy of the (G)MBE we retain all subsystems. Given the cancellation of errors that is often inherent in applications of the MBE,[48] it seems wise to establish exact (G)MBE(n) benchmarks before proceeding to discard terms.

Table 2.5 summarizes the number of subsystem calculations required for a MBE(4) + MBCP(4) or GMBE(2)+GMBCP(2) calculation, for both $(H_2O)_{20}$ and $(H_2O)_{55}$. For $N = 20$, GMBE(2) affords a modest reduction in the number of subsystem calculations, as compared to MBE(4), but the reduction is quite dramatic for $N = 55$. This is especially true for the counterpoise correction, which is far more expensive than a simple calculation of the supersystem energy. For $N = 55$, use of GMBCP(2) to approximate $E_I^{IJK\cdots N}$ generates 205 dimers (with 5–8 ghost molecules each) and 1,430 intersections (3–5 ghost molecules). Since there are N separate monomer energies to correct, the full GMBCP(2) calculation consists of 11,275 dimers and 78,650 intersections. In contrast, for MBCP(4) there are 1,364,220 tetramers (with three ghost molecules each), with a large number of smaller subsystems as well.

The most time-consuming calculations for MBE(4)+MBCP(4) are tetramers in the tetramer basis set and monomers in the tetramer basis set, which are approximately equally expensive at the DFT level because they require all the same electron repulsion integrals. (Although the cost analysis changes somewhat for correlated wave functions, we will see below that timing profiles for DFT and RIMP2 are actually rather similar.) For MBE(4)+MBCP(4) there are over 1.7 million terms requiring a tetramer basis set, whereas the full GMBE(2)+MBCP(2) calculation requires fewer than 110,000 subsystem calculations. On the other hand, the subsystems are

size	$(H_2O)_{20}$				$(H_2O)_{55}$			
	MBE(4)[a]	MBCP(4)[b]	GMBE(2)[a]	GMBCP(2)[b]	MBE(4)[a]	MBCP(4)[b]	GMBE(2)[a]	GMBCP(2)[b]
$n = 4$	4,845	19,380	–	–	341,055	1,364,220	–	–
$n = 3$	1,140	3,420	–	–	26,235	78,705	–	–
$n = 2$	190	380	–	–	1,485	2,970	–	–
$n = 6$–9	–	–	150	1,110	–	–	1,469	11,275
$n = 4$–6	–	–	4,113	16,040	–	–	17,883	78,650
Total	6,175	23,180	4,263	17,150	368,775	1,445,895	19,352	89,925

[a]Calculations involving n monomers in an n-mer basis.

[b]Calculations involving one monomer in an n-mer basis.

Table 2.5: Number of subsystem calculations required for several different fragment-based approaches, for two different cluster sizes.

Fragment Method	ωB97X-V/ aTZ	RIMP2/ aTZ
MBE(4)	9	8
MBCP(4)	3	2
GMBE(2)	53	74
GMBCP(2)	18	34

Table 2.6: Ratio of the total aggregate computing time for a given fragment-based calculation to that required for the supersystem calculation that it is meant to approximate, rounded up to the nearest integer. The system is the same $(H_2O)_{20}$ isomer used to generate the timing data in Fig. 2.7.

larger in the latter case, and for $(H_2O)_{20}$ at the ωB97X-V/aTZ level, the average computer time per subsystem job is 601 s for MBE(4)+MBCP(4) but 5534 s for GMBE(2)+GMBCP(2). These figures reflect the total aggregate computing time across all processors rather than wall time. In practice, we carry out these calculations across as many as 500 processors, a number that is limited by our available resources rather than by a lack of scalability, since the fragment-based calculation should scale well at least to the regime where the number of processors is comparable to the number of subsystem calculations required.

Figure 2.7 shows actual timing data for one ωB97X-V/aTZ or RIMP2/aTZ calculation on $(H_2O)_{20}$, where we have separated the cost of the total energy calculation from that of the counterpoise correction. In terms of total computer time aggregated across all processors, a MBE(4) calculation is 8–9 times more expensive than simply performing a calculation on the entire cluster, depending on the level of theory, and a GMBE(2) calculation is up to 74 times more expensive! From another point of view,

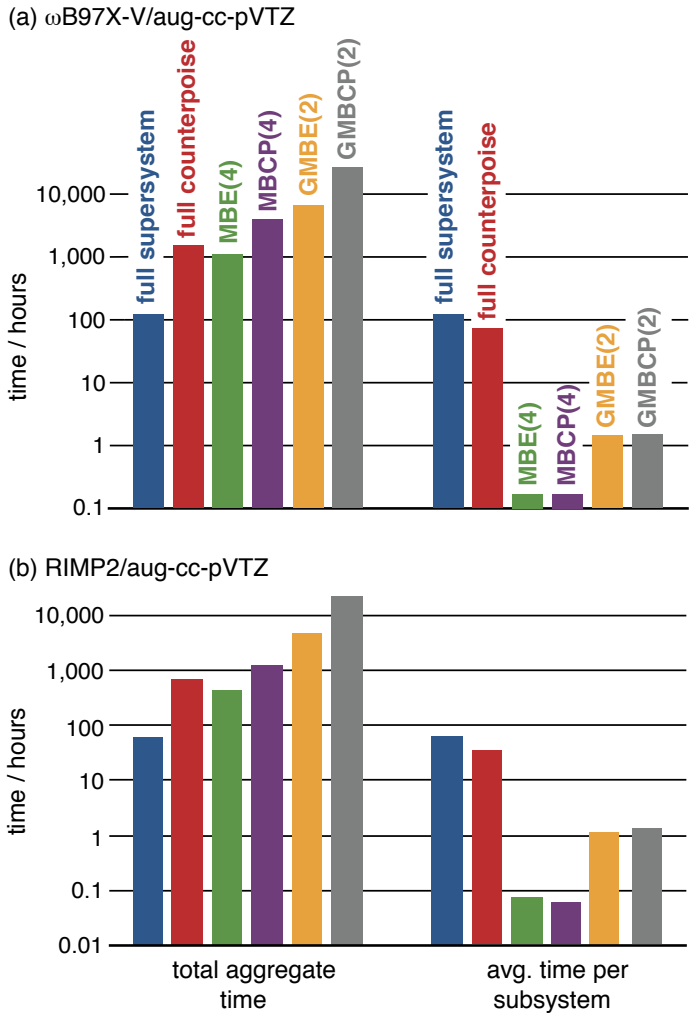

Figure 2.7: Left data sets: total computer time (summing all processors) for counterpoise corrected ωB97X-V/aTZ and RIMP2/aTZ calculations on $(H_2O)_{20}$, with timings for the counterpoise correction listed separated from those for the $(H_2O)_{20}$ total energy. Right data set: total computer time divided by the number of subsystem calculations required. The bar labeled "full supersystem" represents a DFT calculation on the entire cluster, which does not decompose into trivially-parallelizable subsystems, but the corresponding "full counterpoise" correction can be sub-divided by a factor of $N = 20$, and this reduction is reflected in the "full counterpoise" bar in the data set on the right. Supersystem calculations were multithreaded across 20 cores, whereas (G)MBE calculations were run in serial but the subsystem jobs were distributed across as many as 500 processors.

however, these ratios (which are listed in Table 2.6) provide a rough estimate of the number of processors that would be required in order to make the time-to-solution the same for both the supersystem and the fragment-based calculation, assuming that all jobs are run in serial mode. Looking at it in this way, the highly accurate GMBE(2) method will outperform the supersystem approach already on fewer than 100 processors.

As a measure of how favorable the wall times could be made, given the availability of a very large number of processors, the data on the right side of Fig. 2.7 show the average time per subsystem calculation. (We obtain this simply by dividing the total aggregate computer time by the total number of subsystem calculations from Table 2.5, reasoning that the largest and most expensive subsystem calculations are also the most numerous. The full Boys-Bernardi counterpoise correction can be trivially parallelized into $N = 20$ separate calculations, and the timings on the right side of Fig. 2.7 reflect this.) Speed-ups approaching factors of 700–800 for MBE(4), and 60–80 for GMBE(2), are possible, although to realize these will require use of a number of processors comparable to the number of subsystem calculations, *i.e.*, $\sim 10^4$ or so. Scalability should be good into this regime, meaning that (G)MBE-based calculations are good applications for peta- and exa-scale computing.

2.5 Conclusion

We report a systematic study of the accuracy of truncated MBEs as applied to water clusters, $(H_2O)_{N=6-55}$. Elaborating upon a previous study of finite-precision problems

associated with these methods,[48] we demonstrate that for systems containing 30–40 fragments or more, results from three- and four-body expansions are quite sensitive to the numerical thresholds used in the subsystem calculations, such that common default settings in an electronic structure program may afford supersystem energies that differ by several kcal/mol as compared to results obtained using very tight thresholds. The same is not true of the supersystem calculation itself, but arises from the very large number of subsystem energies that must be summed. It is therefore important to use tight thresholds in all subsystem calculations in order to minimize propagation of errors and ensure that the result is not an artifact of roundoff error.

These problems are dramatically worse for the four-body expansion than they are for the three-body expansion, which is unfortunate because only the former appears to achieve what has been called "dynamical accuracy" for cluster calculations,[82] defined as 10% of mean thermal energy at room temperature, or 0.09 kcal/mol per monomer. Although atom-centered embedding point charges are often assumed to accelerate convergence of the MBE, we find no compelling evidence that this is true at the three- or four-body level, while two-body results can be highly erratic depending on the particular choice of embedding charges.

In contrast, a *generalized* two-body expansion based on overlapping fragments containing 3–4 water molecules per fragment achieves an accuracy of 0.02 kcal/mol per monomer, without resorting to embedding charges. The fact that high accuracy can be achieved without charge embedding is significant for several reasons. First,

precision problems associated with computing the Coulomb self-energy of the embedding charges present a serious challenge to any implementation of the MBE that uses a script or driver program that is external to the quantum chemistry program itself, because the magnitude of this term can easily exceed 10^5 hartree in examples such as the ones considered here, so that six- or eight-digit roundoff in the output files of the electronic structure program can make a significant difference.[48] Second, applications to more heterogeneous clusters probably requires that the embedding charges be iterated to self-consistency, a process that destroys the variational nature of the subsystem SCF calculations and significantly complicates the formulation of analytic energy gradients for the MBE. Finally, the use of embedding charges leads to dramatic variations in the accuracy of the MBE from one basis set to the next, especially in the presence of diffuse functions where overpolarization problems are severe.

Having established the general characteristics of the (G)MBE using modest levels of theory, we examined isomers of $(H_2O)_{20}$ at several more respectable levels, namely RIMP2/CBS and counterpoise-corrected ωB97X-V/aTZ. To achieve high accuracy, one must deal with BSSE present in the n-body calculations, and we have introduced many-body counterpoise corrections for both the traditional MBE (in Refs. 46 and 47) and for the GMBE (in the present work). The three-body approach with three-body counterpoise corrections, MBE(3)+MBCP(3), fails to achieve 1 kcal/mol accuracy, which can be achieved using the corresponding four-body calculation. As above, however, the GMBE approach—GMBE(2)+GMBCP(2), in this case—performs well

at the two-body level, with a MAE < 0.3 kcal/mol as compared to counterpoise-corrected supersystem benchmarks. Separate analysis of the counterpoise corrections themselves suggest that this success does not rest on error cancellation, whereas for MBE(4)+MBCP(4) some error cancellation is indeed involved, wherein BSSE compensates for five-body and higher-order terms that are neglected.

Together, this study and a previous one[48] raise serious concerns about the efficacy, or at least the generality, of the n-body expansion. These difficulties arise not only from the factorial increase in the number of subsystem calculations, with respect to both n and N, but also from the fact that to maintain consistent precision, tighter SCF convergence thresholds and arbitrary precision floating-point arithmetic are required. Non-monotone convergence with respect to n, coupled with size-extensive errors in total interaction energies, mean that higher-order terms become more important as system size grows. Together, these problems certainly detract from the "free lunch" reputation of fragment-based methods. On the other hand, GMBE(2)+GMBCP(2) has a number of advantages, as suggested above, and in conjunction with the ωB97X-V/aTZ level of theory offers an accurate and stable method for application to large systems, which can be competitive in cost on only a few hundred processors. Performance will improve given a thresholding procedure for discarding irrelevant subsystem calculations *a priori*, and we hope to report on systematic tests of such a procedure in the future.

CHAPTER 3

Breakdown of the single-exchange approximation in third-order symmetry-adapted perturbation theory[3.1]

3.1 Introduction

Symmetry-adapted perturbation theory (SAPT) is a systematically-improvable hierarchy of methods for direct calculation of intermolecular interaction energies,[114–117] in which the interaction energy decomposes naturally into a sum of physically-meaningful components: electrostatics, induction, dispersion, and exchange-repulsion. This is appealing because one can construct physically-motivated potential energy surfaces or force fields by separately fitting the individual components of a SAPT calculation. Applications to clusters,[118–120] molecular crystals,[121–124] bulk liquids,[125] and biomolecules[126] have been reported recently.

Starting from a zeroth-order Hamiltonian equal to the sum of two monomer Fock operators, the SAPT approach is based on a symmetrized Rayleigh-Schrödinger (SRS) perturbation expansion[114,115] with respect to three perturbations: the intermolecular interaction potential, $\hat{V}$, and the two monomer fluctuation potentials, $\hat{W}_{\mathrm{A}}$ and $\hat{W}_{\mathrm{B}}$.

[3.1] This chapter appeared as a full article in the *Journal of Physical Chemistry A*, in 2012, volume 116, pages 3042–3047.

The interaction energy can be expressed as

$$E_{\text{int}} = \sum_{i=1}^{\infty} \sum_{j=0}^{\infty} \left(E_{\text{pol}}^{(ij)} + E_{\text{exch}}^{(ij)} \right) , \tag{3.1}$$

where i indicates the order in perturbation theory with respect to $\hat{V}$ and j indicates the order with respect to both $\hat{W}_{\text{A}}$ and $\hat{W}_{\text{B}}$. The correction terms $E_{\text{pol}}^{(ij)}$ are known collectively as the *polarization expansion*, and these are precisely the same terms that would appear in ordinary Rayleigh-Schrödinger perturbation theory. The polarization expansion contains electrostatic, induction and dispersion interactions, but in the SRS expansion each term $E_{\text{pol}}^{(ij)}$ has a corresponding exchange term, $E_{\text{exch}}^{(ij)}$, that arises from the antisymmetrizer $\hat{\mathcal{A}}_{\text{AB}}$ that is introduced in order to project away the Pauli-forbidden components of the interaction energy.[115]

The operator $\hat{\mathcal{A}}_{\text{AB}}$ generates all possible permutations (exchanges) of electrons on monomers A and B, which makes it difficult to derive closed-form, programmable expressions for the exchange corrections. To the best of our knowledge, this has been accomplished[127] only for the first-order exchange correction, $E_{\text{exch}}^{(10)}$, and all higher-order exchange terms are evaluated within the *single-exchange approximation* (SEA).[114] (Because the resulting formulas depend on the square of the dimer overlap matrix, typically denoted by $\mathbf{S}$, the SEA is sometimes known as the "S^2 approximation".)

The SEA is thought to be quite robust at equilibrium geometries.[128] As an example, we cite a recent SAPT study of dimers consisting of nonpolar monomers.[129] There, the ratio $E_{\text{exch}}^{(10)}/E_{\text{exch}}^{(10)}(S^2)$, where the numerator is exact and the denominator invokes the SEA, was used as a scaling factor for the higher-order exchange terms. It was found that this scaling adds no more than 0.03 kcal/mol to binding energies

at equilibrium geometries, and even at much shorter distances, where the intermolecular interaction becomes repulsive, the scaling adds no more than 1 kcal/mol. The largest contribution to this additional 1 kcal/mol comes from scaling the third-order exchange–induction term, $E^{(30)}_{\text{exch-ind}}(S^2)$.[129]

A formula for the third-order exchange interaction $E^{(30)}_{\text{exch}}(S^2)$ was reported only a few years ago,[130] and has not yet been explored as thoroughly as have the lower-order SAPT corrections. Because $E^{(30)}_{\text{exch-ind}}$ ought to be more significant for polar molecules, as opposed to the nonpolar ones considered in Ref. 129, we have undertaken a careful analysis of the role of the $E^{(30)}_{\text{exch-ind}}$ term in SAPT calculations on polar dimers. The results, reported here, demonstrate that the SEA results in significant errors leading to qualitatively-incorrect potential energy surfaces at distances shorter than the equilibrium intermolecular distance.

3.2 Computational methods

We study potential energy curves (PECs) for five different dimers. Four these are polar systems: $(H_2O)_2$, $F^-(H_2O)$, $Cl^-(H_2O)$, and $HO^-(H_2O)$. In addition, we consider the helium dimer as a "control experiment", since induction plays almost no role in the binding of He_2. Third-order SAPT calculations for He_2 and $(H_2O)_2$ near their equilibrium geometries have been reported previously by Patkowski *et al.*,[130,131] but here we extend these calculations to full PECs.

Rigid monomer geometries were adopted for all computations since the frozen-monomer approximation works reasonably well for PECs.[120,121] Monomer geometries were obtained from MP2/aug-cc-pVTZ optimizations and then PECs were constructed along the distance coordinate, R, between the two heavy atoms. As benchmarks, we estimated the complete-basis set (CBS) interaction energy at the CCSD(T) level based on a two-point ("T,Q") extrapolation of the MP2 energy and an estimate of the triples correction in a smaller basis set. (Details can be found in the Supporting Information.) All SAPT calculations were performed in the aug-cc-pVTZ basis set using the SAPT 2008.2 program,[132] with integrals generated by the ATMOL program.[133]

The different levels of SAPT applied in this work are defined as follows:[117]

$$\begin{aligned} E_{\mathrm{SAPT0}} &= E_{\mathrm{elst}}^{(10)} + E_{\mathrm{exch}}^{(10)} + E_{\mathrm{ind,resp}}^{(20)} + E_{\mathrm{exch\text{-}ind,resp}}^{(20)} \\ &\quad + E_{\mathrm{disp}}^{(20)} + E_{\mathrm{exch\text{-}disp}}^{(20)} \qquad (3.2a) \\ E_{\mathrm{SAPT2}} &= E_{\mathrm{SAPT0}} + E_{\mathrm{elst,resp}}^{(12)} + E_{\mathrm{exch}}^{(11)} + E_{\mathrm{exch}}^{(12)} \\ &\quad + {}^{t}E_{\mathrm{ind}}^{(22)} + {}^{t}E_{\mathrm{exch\text{-}ind}}^{(22)} \qquad (3.2b) \\ E_{\mathrm{SAPT}} &= E_{\mathrm{SAPT2}} + E_{\mathrm{elst,resp}}^{(13)} + E_{\mathrm{disp}}^{(21)} + E_{\mathrm{disp}}^{(22)} \qquad (3.2c) \\ E_{\mathrm{SAPT2+3}} &= E_{\mathrm{SAPT}} + E_{\mathrm{ind}}^{(30)} + E_{\mathrm{exch\text{-}ind}}^{(30)} + E_{\mathrm{disp}}^{(30)} \\ &\quad + E_{\mathrm{exch\text{-}disp}}^{(30)} + E_{\mathrm{ind\text{-}disp}}^{(30)} + E_{\mathrm{exch\text{-}ind\text{-}disp}}^{(30)} \qquad (3.2d) \end{aligned}$$

The subscripts denote electrostatic (elst), exchange (exch), induction (ind), and dispersion (disp) components, and "resp" indicates that the component in question includes the Hartree-Fock response of each monomer to the static electric field of its

interacting partner. This response, which amounts to orbital relaxation, is obtained by solving coupled-perturbed Hartree-Fock equations. The superscript "t" in ${}^{t}E^{(22)}_{\rm ind}$ indicates that this is the "true" correlation part of $E^{(22)}_{\rm ind}$, not included in $E^{(20)}_{\rm ind,resp}$, and the corresponding "true" correlation part of $E^{(22)}_{\rm exch\text{-}ind}$ is estimated as

$$ {}^{t}E^{(22)}_{\rm exch\text{-}ind} = {}^{t}E^{(22)}_{\rm ind} \left(\frac{E^{(20)}_{\rm exch\text{-}ind,resp}}{E^{(20)}_{\rm ind,resp}} \right) \; . \tag{3.3} $$

With the exception of He_2, the dimers that we investigate are dominated by induction effects, so it will be convenient to define

$$ \Delta E^{(k)}_{\rm ind} = E^{(k0)}_{\rm ind} + E^{(k0)}_{\rm exch\text{-}ind} \tag{3.4a} $$

$$ \Delta E^{(k)}_{\rm ind,resp} = E^{(k0)}_{\rm ind,resp} + E^{(k0)}_{\rm exch\text{-}ind,resp} \tag{3.4b} $$

for $k > 1$. The quantity $\Delta E^{(k)}_{\rm ind}$ represents the contribution to the total induction energy that arises at kth order in $\hat{V}$, in the absence of monomer correlation,[117] and $\Delta E^{(k)}_{\rm ind,resp}$ is the analogous orbital-relaxed quantity.

The Hartree-Fock interaction energy, $E^{\rm HF}_{\rm int}$, can be approximated within SAPT, and through second- and third-order in $\hat{V}$ these approximations are[130]

$$ E^{[2]}_{\rm SAPT\text{-}HF} = E^{(10)}_{\rm elst} + E^{(10)}_{\rm exch} + \Delta E^{(2)}_{\rm ind,resp} \tag{3.5a} $$

$$ E^{[3]}_{\rm SAPT\text{-}HF} = E^{[2]}_{\rm SAPT\text{-}HF} + \Delta E^{(3)}_{\rm ind} \; . \tag{3.5b} $$

Since induction interactions converge slowly for polar molecules, and because one may argue that a dimer Hartree-Fock calculation contains induction and exchange-induction effects through infinite order,[134–136] a "hybrid" SAPT approach is recommended for polar molecules.[130,131] In this approach, a dimer Hartree-Fock calculation

is used to evaluate the energy difference

$$\delta E_{\rm int}^{(2)} = E_{\rm int}^{\rm HF} - E_{\rm SAPT\text{-}HF}^{[2]} \ , \tag{3.6}$$

which is then taken as an estimate of induction effects beyond second order.[130] This term can then be added as a correction to any of the SAPT methods that are second-order in $\hat{V}$, including SAPT0, SAPT2, and SAPT. [What we call $\delta E_{\rm int}^{(2)}$ has alternatively been called the $\delta E_{\rm int}^{\rm HF}$ correction[117] and the δ(HF) correction.[137]] Although this correction brings in higher-order induction and exchange-induction effects, it also contains spurious unphysical terms,[127,130,136] most notably, exchange deformation of the orbitals.[136]

3.3 Results and discussion

Qualitatively similar results are obtained for each of the four $X\cdots H_2O$ systems considered here (X = F^-, Cl^-, HO^-, and H_2O), and we will focus largely on $F^-(H_2O)$ as an example. However, all of the quantities that we plot for $F^-(H_2O)$ are plotted, for each $X\cdots H_2O$ system and also for He_2, in the Supporting Information.

PECs computed for $F^-(H_2O)$, using all four versions of SAPT defined in Eq. (3.2), are plotted in Fig. 3.1. Each of these methods underbinds this dimer at near-equilibrium distances, as compared to the benchmark CCSD(T)/CBS result, but the SAPT0, SAPT, and SAPT2 curves at least exhibit reasonable shapes. However, in the case of SAPT2+3—which, in principle, is the highest-level SAPT method that is employed here—the PEC is reasonable only for $R > R_{\rm eq}$, where $R_{\rm eq}$ denotes the minimum-energy intermolecular distance computed at the CCSD(T)/CBS level.

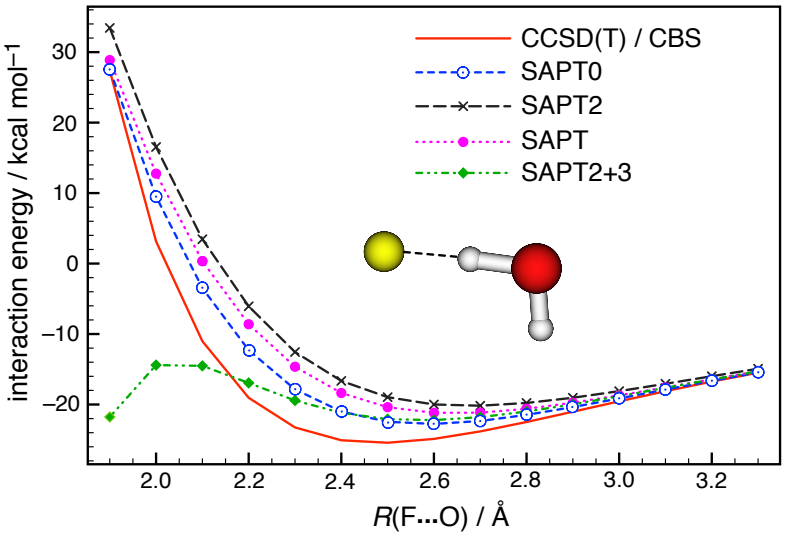

Figure 3.1: Comparison of different levels of SAPT [as defined in Eq. (3.2)] for the $F^-(H_2O)$ system. The H_2O geometry is frozen in these potential energy scans.

[For $F^-(H_2O)$, $R_{eq} = 2.5$ Å.] At shorter distances, the SAPT2+3 curve turns over, becoming attractive rather than repulsive.

The reason for this failure can be discerned by decomposing the interaction energy into different components, some of which are plotted in Fig. 3.2. Attractive interactions in $F^-(H_2O)$ are dominated by induction and we observe that for $R < 1.8$ Å, the third-order induction correction, $\Delta E^{(3)}_{\rm ind}$, is even larger than $\Delta E^{(2)}_{\rm ind}$. Although a variety of third-order terms are present in a SAPT2+3 calculation [see Eq. (3.2d)], if we add $\Delta E^{(3)}_{\rm ind}$ to an otherwise qualitatively-correct SAPT0 calculation, the result is a PEC that is completely wrong at short R. This result is shown in Fig. 3.3 for all four ion–water complexes considered here. These calculations identify $\Delta E^{(3)}_{\rm ind}$ as the

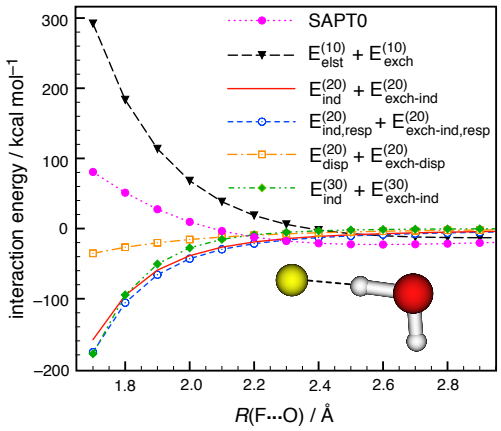

Figure 3.2: SAPT0 decomposition of the interaction energy for $F^-(H_2O)$ into different orders of electrostatic (elst), induction (ind), and dispersion (disp) components, each with a corresponding exchange (exch) contribution. For the second-order induction and exchange–induction components, results are shown both with and without orbital relaxation (resp) corrections.

origin of the problem.

In Fig. 3.4, we address the convergence of the SAPT approximations to the supermolecular Hartree-Fock interaction energy for $F^-(H_2O)$ and also for $(H_2O)_2$. For either system, both $E^{[2]}_{\text{SAPT-HF}}$ and $E^{[3]}_{\text{SAPT-HF}}$ are excellent approximations to $E^{\text{HF}}_{\text{int}}$ for $R \geq R_{\text{eq}}$. For intermolecular distances a bit shorter than R_{eq}, the addition of $\Delta E^{(3)}_{\text{ind}}$ to $E^{[2]}_{\text{SAPT-HF}}$ (which defines the third-order approximation, $E^{[3]}_{\text{SAPT-HF}}$) successfully accounts for the difference between $E^{\text{HF}}_{\text{int}}$ and $E^{[2]}_{\text{SAPT-HF}}$. Thus the third-order approximation to $E^{\text{HF}}_{\text{int}}$ is basically converged for distances ranging from a bit shorter than R_{eq}, out to $R = \infty$. For $(H_2O)_2$ near its equilibrium geometry, this convergence was noted previously by Patkowski *et al*,[130,131] but even for the anion–water systems

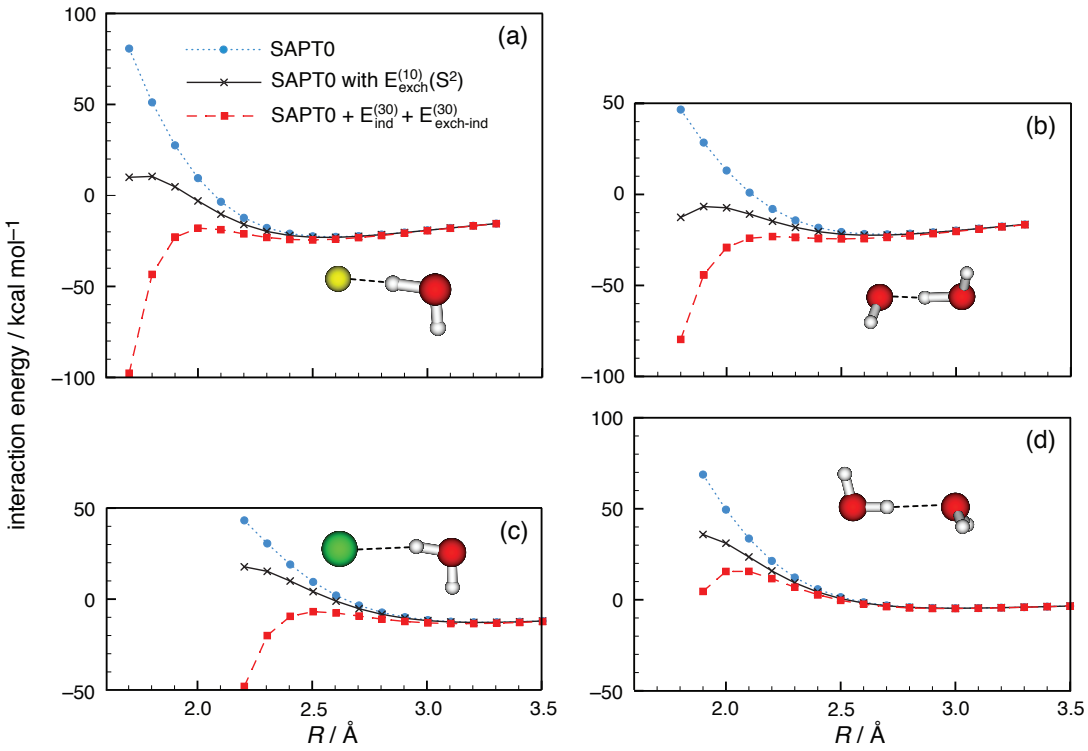

Figure 3.3: SAPT0 potential curves for (a) $F^- \cdots H_2O$, (b) $HO^- \cdots H_2O$, (c) $Cl^- \cdots H_2O$, and (d) $H_2O \cdots H_2O$, as a function of the distance R between the two heavy atoms. (Both the vertical and horizontal scales are the same in all four panels.) Also shown are the potential curves that result when $\Delta E^{(3)}_{\rm ind}$ is added to a SAPT0 calculation and when the exact $E^{(10)}_{\rm exch}$ term in SAPT0 is replaced by its single-exchange approximation, $E^{(10)}_{\rm exch}(S^2)$. Results for helium dimer are not shown because all three curves are indistinguishable on the energy scale used in this figure.

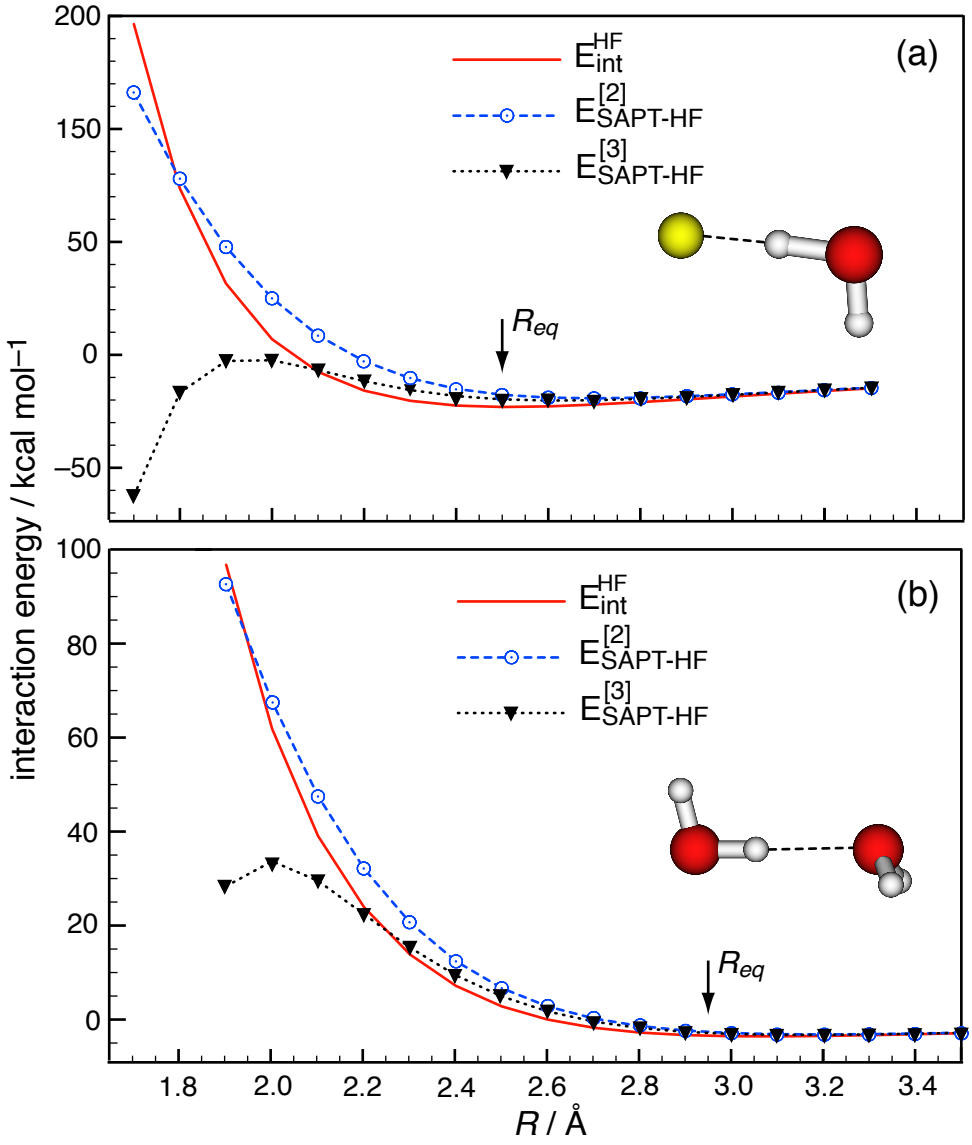

Figure 3.4: Distance dependence of the supermolecular Hartree-Fock interaction energy ($E_{\rm int}^{\rm HF}$) and its second-order ($E_{\rm SAPT\text{-}HF}^{[2]}$) and third-order ($E_{\rm SAPT\text{-}HF}^{[3]}$) SAPT approximations, for (a) $F^-(H_2O)$ and (b) $(H_2O)_2$. In each case, the coordinate R is the distance between the heavy atoms, and the arrow indicates the CCSD(T)/CBS minimum-energy geometry. Note that the two panels use different energy scales.

considered here, where the induction effects are larger, we find that $E^{[3]}_{\text{SAPT-HF}}$ is an excellent approximation to $E^{\text{HF}}_{\text{int}}$ near the equilibrium geometry and also at larger values of R.

In contrast, for $R \ll R_{\text{eq}}$ the *second*-order approximation $E^{[2]}_{\text{SAPT-HF}}$ is clearly *not* converged for $F^-(H_2O)$, although the PEC defined by the $E^{[2]}_{\text{SAPT-HF}}$ method is at least qualitatively correct. For $(H_2O)_2$, where the binding energy and the induction effects are much smaller, $E^{[2]}_{\text{SAPT-HF}}$ is a much better approximation to the Hartree-Fock interaction energy. For both systems, addition of $\Delta E^{(3)}_{\text{ind}}$ to $E^{[2]}_{\text{SAPT-HF}}$, which defines the third-order approximation $E^{[3]}_{\text{SAPT-HF}}$, improves upon $E^{[2]}_{\text{SAPT-HF}}$ at distances somewhat shorter than R_{eq} but eventually the $E^{[3]}_{\text{SAPT-HF}}$ curves turn over, even in the charge-neutral water dimer. This behavior is clearly a manifestation of the divergence of $\Delta E^{(3)}_{\text{ind}}$ at short intermolecular separation, which again points to this quantity as the culprit responsible for the qualitatively wrong PECs obtained at the SAPT2+3 level of theory.

One might object that in our calculations the third-order induction terms do not include orbital relaxation (response) effects. At near-equilibrium geometries, Patkowski *et al.*[131] have shown that orbital relaxation increases the third-order induction interaction by up to 50% for dimers such as $(H_2O)_2$ that are composed of polar monomers. Anion–water complexes were not considered by Patkowski *et al.*, but for $F^-(H_2O)$ the *second*-order induction corrections are compared, with and without orbital relaxation, in Fig. 3.2. The effect of orbital relaxation is to make the potential energy curve *more* attractive at short distances, albeit by a relatively small amount.

As such, we find it unlikely that the inclusion of orbital response at third order would correct the qualitatively wrong PECs caused by $\Delta E_{\text{ind}}^{(3)}$.

Thus the question remains: what is the problem with the $\Delta E_{\text{ind}}^{(3)}$ term? According to Fig. 3.2, the interaction energy contributed by $\Delta E_{\text{ind}}^{(3)}$ increases as intermolecular distance decreases, so it must be that $E_{\text{ind}}^{(30)}$ is not sufficiently quenched by its exchange counterpart, $E_{\text{exch-ind}}^{(30)}$, at short distance. The $\Delta E_{\text{ind}}^{(2)}$ (or $\Delta E_{\text{ind,resp}}^{(2)}$) curve has roughly the same basic shape as that for $\Delta E_{\text{ind}}^{(3)}$, although the latter diverges *slightly* more rapidly than the former as R decreases; it is difficult to ascribe any qualitative problems to this subtle difference.

Perhaps more telling are certain calculations reported in Fig. 3.3, in which we have replaced the exact $E_{\text{exch}}^{(10)}$ term in a SAPT0 calculation with its SEA, $E_{\text{exch}}^{(10)}(S^2)$. For each of the four $\text{X}\cdots\text{H}_2\text{O}$ systems that we consider (including the water dimer), this has the effect of greatly weakening the short-range repulsive interactions, and for $\text{F}^-(\text{H}_2\text{O})$ and $\text{HO}^-(\text{H}_2\text{O})$, where induction effects are largest, this modified SAPT0 potential curve even becomes attractive at sufficiently short distance. Therefore, even at first order the SEA can produce attractive PECs at short distances, although the influence of the SEA on the third-order corrections is much larger. One might expect that the SEA might also have qualitatively important effects on the *second*-order exchange at short intermolecular distances, but in fact qualitatively correct PECs are obtained using SAPT methods that are only second-order in $\hat{V}$.

For the $\text{X}\cdots\text{H}_2\text{O}$ systems, the value of the turnover point where the SAPT0 + $\Delta E_{\text{ind}}^{(3)}$ potential curve (Fig. 3.3) changes from repulsive to attractive decreases in the

System	R_{eq}/ Å	Turnover point/Å
$F^-(H_2O)$	2.5	1.9
$HO^-(H_2O)$	2.6	2.1
$Cl^-(H_2O)$	3.1	2.4
$(H_2O)_2$	2.9	2.0

Table 3.1: CCSD(T)/CBS equilibrium distances and SAPT2+3 "turnover points" where the potential becomes attractive at short distance.

order $Cl^- > HO^- > H_2O > F^-$. (The turnover points for SAPT2+3 calculations turn over in the same order; see Table 3.1.) For the anions, this is the same as the order of the ionic radii ($Cl^- > O^{2-} > F^-$), which we rationalize in terms of the fact that ions with larger radii would be expected to have larger exchange–repulsion interactions, at least for intermolecular distances not significantly smaller than the sum of the van der Waals radii. Considering the water dimer, the aforementioned turning point occurs at a smaller value of R than it does in $HO^-(H_2O)$, which we attribute to the much larger induction effects in the ionic complex. Together, these observations suggest that the large negative values of $\Delta E^{(3)}_{\rm ind}$ at small R are probably attributable to the failure of $E^{(30)}_{\rm exch\text{-}ind}(S^2)$ to quench $E^{(30)}_{\rm ind}$.

Following Ref. 129, we use the ratio $E^{(10)}_{\rm exch}/E^{(10)}_{\rm exch}(S^2)$ to estimate third-order exchange effects beyond the SEA. Thus we introduce an *ad hoc* scaling formula

$$E^{(30)}_{\rm exch\text{-}ind,scale} = E^{(30)}_{\rm exch\text{-}ind}\left(\frac{E^{(10)}_{\rm exch}}{E^{(10)}_{\rm exch}(S^2)}\right)^{\alpha} \tag{3.7}$$

where the exponent α is an empirical parameter. (In Ref. 129, only $\alpha = 1$ was considered.) Figure 3.5 shows the result when $E^{(30)}_{\rm exch\text{-}ind}$ is replaced in a SAPT2+3

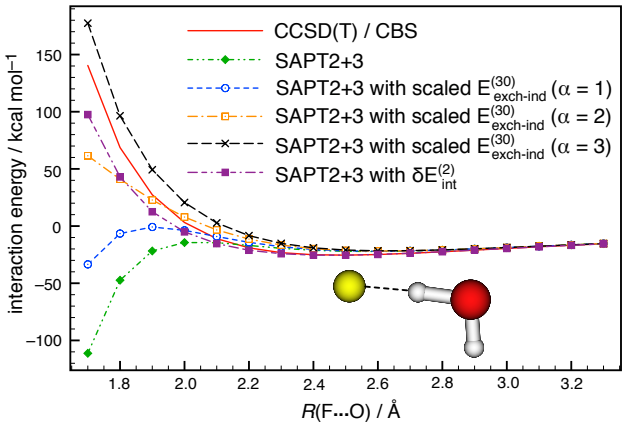

Figure 3.5: Comparison of several variants of SAPT2+3 for $F^-(H_2O)$. In three of these, the $E^{(30)}_{\text{exch-ind}}$ term has been scaled using Eq. (3.7), with different values of the parameter α, in order to approximate exchange interactions beyond the SEA. In another, we have replaced the $\Delta E^{(3)}_{\text{ind}}$ term in SAPT2+3 with the $\delta E^{(2)}_{\text{int}}$ correction defined in Eq. (3.6), in order to capture higher-order induction effects.

calculation by the scaled version in Eq. (3.7). This replacement has little effect for $R \geq R_{\text{eq}}$, which is an indication of the robustness of the SEA at equilibrium geometries and beyond. At short intermolecular distances, however, scaling using $\alpha = 3$ corrects the qualitatively incorrect SAPT2+3 potential curves. The choice $\alpha = 2$ also prevents the PEC from turning over at short distance, although the shape of the repulsive wall is not correct, and with $\alpha = 1$ the PEC still turns over at short distance. There is no sound theoretical justification for any choice of α, but the fact that $\alpha > 1$ is required to obtained a qualitatively correct PEC indicates that exchange interactions beyond the SEA are more important at third order than they are at first order, when the intermolecular distance is small.

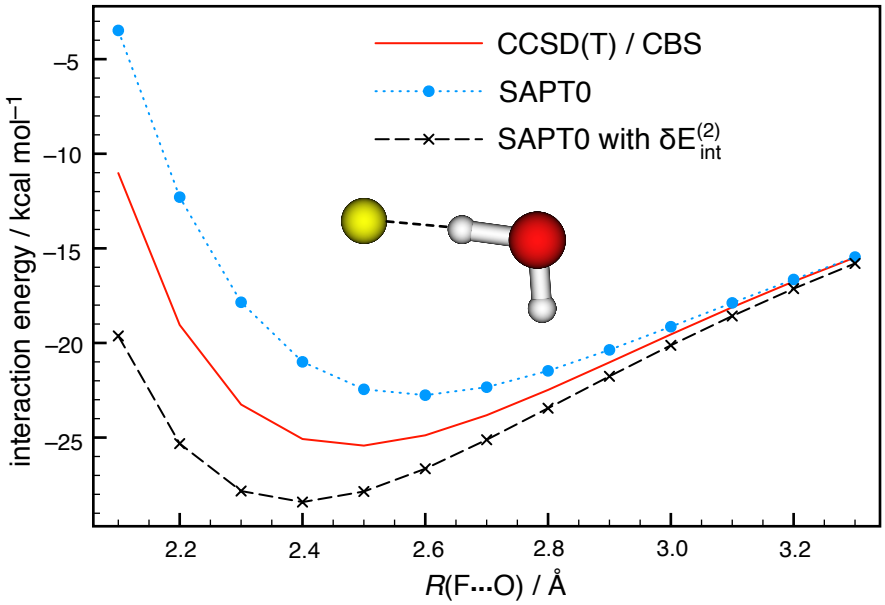

Figure 3.6: Comparison of SAPT0 for $F^-(H_2O)$ with and without the $\delta E^{(2)}_{\rm int}$ correction defined in Eq. (3.6).

On the other hand, a SAPT2+3 calculation where $\Delta E^{(3)}_{\rm ind}$ is replaced by $\delta E^{(2)}_{\rm int}$ is in *quantitative* agreement with CCSD(T)/CBS results for $R \geq R_{\rm eq}$ (see Fig. 3.5). While this SAPT2+3 + $\delta E^{(2)}_{\rm HF}$ result is not quantitative for $R < R_{\rm eq}$ (which is to be expected, owing to the spurious exchange contributions present in $\delta E^{(2)}_{\rm int}$), at the very least this curve does not turn over at small R, and remains qualitatively correct even at very short intermolecular separations. In the strongly repulsive region of the PEC, the SAPT2+3 + $\delta E^{(2)}_{\rm HF}$ result is superior to the PEC obtained simply by scaling $E^{(30)}_{\rm exch\text{-}ind}$ using $\alpha = 1$ or $\alpha = 2$.

These results serve to reinforce a previous recommendation[130,131] to use the $\delta E^{(2)}_{\rm int}$

correction term when the monomers are polar. (For nonpolar monomers, the unphysical artifacts inherent to this correction are more significant, relative to the very small induction corrections, and better results are sometimes obtained *without* using this correction.[130]) Even for polar monomers, however, one should note that the $\delta E^{(2)}_{\rm int}$ correction does not improve the results in all cases. For example, adding $\delta E^{(2)}_{\rm int}$ to SAPT0 for $F^-(H_2O)$ does not improve the result, as shown in Fig. 3.6. The previous recommendation[130,131] was based on results for less strongly-bound systems such as $(H_2O)_2$, and at geometries not much different from the equilibrium geometry.

3.4 Conclusions

We have found that the $E^{(30)}_{\rm exch\text{-}ind}$ term in SAPT, which at present can be computed only within the single-exchange approximation, fails to quench $E^{(30)}_{\rm ind}$ at short intermolecular distances. This leads to the anomalous result that potential energy curves for polar systems become *attractive* at sufficiently small intermolecular distances. For the three anion–water complexes considered here, "sufficiently small" means about 0.5 Å shorter than the equilibrium monomer separation, while for $(H_2O)_2$ the turnover occurs at $R = 2.0$ Å, as compared to $R_{\rm eq} = 2.9$ Å. Scaling $E^{(30)}_{\rm exch\text{-}ind}$ based on the ratio $E^{(10)}_{\rm exch}/E^{(10)}_{\rm exch}(S^2)$ approximates some higher-order exchange effects sufficiently well to avoid catastrophic divergence of the potential energy curve, but for highly polar systems this result serves mostly to identify the nature of the problem rather than to correct it. Further amplification of this ratio, *e.g.*, using $[E^{(10)}_{\rm exch}/E^{(10)}_{\rm exch}(S^2)]^\alpha$ for $\alpha \approx 3$, is required in order to obtain potential energy curves that are qualitatively correct at

short distances. Alternatively, calculation of the $\delta E_{\text{int}}^{(2)}$ correction cannot be avoided, even at third order.

Interestingly, even in cases where the third-order method fails catastrophically, *second*-order SAPT potential curves remain qualitatively correct even at rather small intermolecular distances. This suggests that the single-exchange approximation is more severe in the case of $E_{\text{exch-ind}}^{(30)}$ than it is for lower-order exchange interactions, at least for highly polar monomers. This points to the importance of developing post-SEA correction formulas at higher orders in SAPT. Furthermore, our results show that it inadvisable to include $E_{\text{ind}}^{(30)}$ and $E_{\text{exch-ind}}^{(30)}$ when constructing analytic potential energy surfaces based on SAPT.

CHAPTER 4

Accurate description of intermolecular interactions involving ions using symmetry-adapted perturbation theory[4.1]

4.1 Introduction

Non-covalent interactions play an important role in a broad range of chemical systems, from the aggregation of rare gases to the formation of crystal structures.[138–140] The coupled-cluster singles and doubles method with perturbative triples, CCSD(T), evaluated in the complete-basis limit, is still considered to be the gold standard for non-covalent interaction energies, as higher-order electron correlation effects typically contribute $\ll 0.1$ kcal/mol.[33,141] However, the computational scaling of CCSD(T) is $\mathcal{O}(N^7)$ with respect to total system size, N, which renders this approach unfeasible for large systems. Density functional theory (DFT), on the other hand, can routinely be applied to systems with more than 1,000 basis functions on ordinary workstations, but most conventional functionals fail to account the long-range part of the dispersion interaction.

[4.1] This chapter appeared as a full article in the *Journal of Chemical Theory and Computation*, in 2015, volume 11, pages 2473–2486.

There are several strategies to address this issue, including brute-force parameterization (as in the Minnesota density functionals[142]) as well as the addition of explicit, classical C_6/R^6 atom–atom dispersion potentials, as in the "DFT-D" approach popularized by Grimme and co-workers.[143] Nonlocal correlation functionals have also been developed to address this issue.[144–148]

Unlike these *supermolecular* methods, symmetry-adapted perturbation theory (SAPT) offers a way to compute intermolecular interactions in a physically-meaningful way, as a sum of electrostatic, induction, and dispersion interactions along with their exchange counterparts.[114–117,149,150] Importantly, zeroth-order basis set superposition error (BSSE) is absent in SAPT calculations. SAPT methods traditionally use the Hartree-Fock (HF) determinants for the monomers as reference wave functions (HF-SAPT) and a double perturbation expansion to account for both *inter*molecular Coulomb interactions (between monomers) as well as *intra*molecular electron correlation. The latter is treated as in Møller-Plesset perturbation theory[114,151] or coupled-cluster theory,[152–156] and together this double-perturbation expansion constitutes a successful approach to determining intermolecular interactions, with results that compare favorably to complete-basis CCSD(T) calculations for dimers of neutral molecules.[114–117,149,150] The computational cost is high, though considerable progress has been made recently reducing the computational cost,[157–159] and systems as large as a DNA intercalation complex have been treated in this way.[160] Alternatively, SAPT based on a DFT description of monomers has been used, in an approach known either as DFT-SAPT or SAPT(DFT).[137,150,161] In this case, asymptotic correction of

the monomer exchange-correlation potentials is necessary in order to obtain accurate energy components.[162–164]

High-order SAPT methods exhibit good quantitative accuracy for describing intermolecular interactions between *neutral* monomers,[117] and while these methods have sometimes been applied to ionic systems as well,[50,165–173] these studies have mostly focused on qualitative trends.[166,167,169] Where ionic monomers are involved, the perturbation theory is typically slowly convergent and possibly even divergent,[50] and "chemical accuracy" (error $\lesssim$ 1 kcal/mol) may not be realized in such cases.

There have been only a few attempts to benchmark SAPT for ionic systems. These include a study by Matczak,[170] who investigated 16 pairs of alkali halides (A^+X^-) whose binding energies exceed 100 kcal/mol. That study concluded that the wave function-based SAPT method is superior to CCSD(T) in the same basis set, which in our opinion is probably a manifestation of the fact that SAPT calculations are largely free of BSSE, since the CCSD(T) results were not extrapolated to the basis-set limit. In any case, neither SAPT nor CCSD(T) achieves chemical accuracy, as compared to experiment, for these alkali halides.[170] On the other hand, Ansorg *et al.*[171] studied the cation-π interaction between NH_4^+ and benzene using DFT-SAPT, and demonstrated good agreement between CCSD(T) and DFT-SAPT with a triple-ζ basis set. Two of us[174,175] have recently examined $HeCl^-$, $NeNa^+$, and Li^+F^-, with an eye toward understanding the role of the single-exchange approximation in describing the repulsive wall, while Korona[172] has performed DFT-SAPT calculations on $F^-(HF)$, $F^-(H_2O)$, and $Na^+(H_2O)$, finding good agreement with SAPT(CCSD)

energy components.

The aforementioned studies are insufficient to determine the complete-basis CCSD-(T) result or to evaluate whether SAPT can describe intermolecular interactions in ionic systems in a general way. The aim of this work is, firstly, to establish complete-basis CCSD(T) benchmarks for a broad range of cationic, anionic, and ion-pair systems. Then, these benchmark values will be used to compare various level of wave function-based SAPT, along with the recently-developed "attenuated" MP2 method that has been suggested for non-covalent interactions.[17,176–178] Finally, DFT-SAPT and several supermolecular density functionals will be tested for the same data set, including some relatively new functionals that have been suggested to be accurate for non-covalent interactions. Lao and Herbert[56,57] have recently shown that a number of density functionals that had previously been recommended for non-covalent interactions yield poor results for binding energies of halide anions in water clusters, and the more extensive tests reported here confirm that binding energies in ionic systems are challenging tests for DFT.

4.2 Methods

4.2.1 Symmetry-adapted perturbation theory

The SAPT interaction can be expressed as[115]

$$E_{\mathrm{int}}^{\mathrm{SAPT}} = \sum_{i=1}^{\infty} \sum_{j=0}^{\infty} \left(E_{\mathrm{pol}}^{(ij)} + E_{\mathrm{exch}}^{(ij)} \right) \tag{4.1}$$

where i indicates the order in perturbation theory with respect to the intermolecular potential, and j indicates the order with respect to the intramolecular electron

correlation (the "fluctuation potential", in the language of Møller-Plesset perturbation theory). The terms $E_{\rm pol}^{(ij)}$ originate from the so-called polarization expansion and contain electrostatic, induction and dispersion interactions. Each term $E_{\rm pol}^{(ij)}$ has a corresponding exchange term, $E_{\rm exch}^{(ij)}$, that arises from antisymmetry requirements.

The simplest method SAPT method, often called SAPT0, neglects intramolecular electron correlation and treats the intermolecular perturbation up to second order:

$$\begin{aligned} E_{\rm int}^{\rm SAPT0} = &\left[E_{\rm elst}^{(10)}\right]_{\rm elst} + \left[E_{\rm exch}^{(10)}\right]_{\rm exch} \\ &+ \left[E_{\rm ind,resp}^{(20)} + E_{\rm exch\text{-}ind,resp}^{(20)} + \delta E_{\rm HF}^{(2)}\right]_{\rm ind} \\ &+ \left[E_{\rm disp}^{(20)} + E_{\rm exch\text{-}disp}^{(20)}\right]_{\rm disp} . \end{aligned} \tag{4.2}$$

Following Ref. 173, the various terms in the SAPT0 energy expression are grouped, using square brackets, into electrostatic, exchange, induction, and dispersion interactions, according to the partitioning scheme used by Sherrill *et al.*[117,173] The Hartree-Fock correction term (δHF) up to second order, $\delta E_{\rm HF}^{(2)}$, incorporates polarization effects beyond second order and is defined as

$$\begin{aligned} \delta E_{\rm HF}^{(2)} = E_{\rm int}^{\rm HF} &- E_{\rm elst}^{(10)} - E_{\rm exch}^{(10)} \\ &- E_{\rm ind,resp}^{(20)} - E_{\rm exch\text{-}ind,resp}^{(20)} , \end{aligned} \tag{4.3}$$

where $E_{\rm int}^{\rm HF}$ is the counterpoise-corrected HF binding energy for the dimer. The "response" ("resp") subscripts indicate that the response correction for induction is incorporated by solving coupled-perturbed HF equations.[179,180]

Traditionally, a closed-form analytic formula for the SAPT exchange energies has been available only at first order ($E_{\rm exch}^{(10)}$), whereas higher-order exchange terms have been evaluated using the "single-exchange approximation" (SEA). Formulas obtained

from the SEA involve the square of the orbital overlap matrix ($\mathbf{S}^2$), hence this is also known as the "S^2 approximation".[114] Schäffer and Jansen[174,175] have recently extended the complete derivation to second order, *i.e.*, they have derived $E^{(20)}_{\text{exch}}$ without invoking the SEA, and results using exact second-order exchange will be examined here.

Lao and Herbert[50] have previously shown that use of the SEA in third-order exchange-induction, $E^{(30)}_{\text{exch-ind}}$, breaks down severely for ionic systems at short intermolecular distances. Following Refs. 181 and 129, we introduced an *ad hoc* scaling factor

$$p_{\text{ex}}(\alpha) = \left(\frac{E^{(10)}_{\text{exch}}}{E^{(10)}_{\text{exch}}(S^2)} \right)^{\alpha} \tag{4.4}$$

that is then used to scale $E^{(20)}_{\text{exch}-\text{ind,resp}}$ and $E^{(30)}_{\text{exch}-\text{ind,resp}}$ for correcting the deficiency of SEA. In Eq. (4.4), the numerator represents the exact first-order exchange energy while the denominator is evaluated within the SEA. The idea is to use the difference between exact and S^2 first-order exchange energies as a means to estimate the error in the SEA used in the second-order exchange-induction energies although Schäffer and Jansen have suggested the difference is smaller at first order than at second order.[174,175] In a recent study of $F^-(H_2O)$,[50] we used the exponent $\alpha = 2$ for $E^{(20)}_{\text{exch-ind,resp}}$ and $\alpha = 3$ for $E^{(30)}_{\text{exch-ind,resp}}$, the latter value having been recommended for strongly hydrogen-bonded systems where the intermolecular distances are small and thus the exchange energies are large.[173] The value value $\alpha = 1$ has been used in other studies,[129,173] and some results with $\alpha = 1$ are presented here as well.

In previous work, Lao and Herbert[50] demonstrated that the $\delta E^{(2)}_{\rm HF}$ correction improves the SAPT results for $F^-(H_2O)$ at short intermolecular distances, consistent with previous recommendations that the $\delta E^{(2)}_{\rm HF}$ correction is important for dimers composed of polar monomers.[130,131] The $\delta E^{(2)}_{\rm HF}$ term cancels the $p_{\rm ex}(\alpha)$ scaling correction for $E^{(20)}_{\text{exch-ind,resp}}$, and the errors in $E^{(20)}_{\text{exch-ind,resp}}$ arising from the SEA can thus be avoided. Based on the $p_{\rm ex}(\alpha)$ scaling factor, Parker *et al.*[173] proposed a "scaled" SAPT0 method, *s*SAPT0:

$$\begin{aligned} E^{s\text{SAPT0}}_{\rm int} = & \\ & \left[E^{(10)}_{\rm elst}\right]_{\rm elst} + \left[E^{(10)}_{\rm exch}\right]_{\rm exch} \\ & + \left[E^{(20)}_{\rm ind,resp} + p_{\rm ex}(3.0) E^{(20)}_{\text{exch-ind,resp}} + \delta E^{(2)}_{\rm HF}\right]_{\rm ind} \\ & + \left[E^{(20)}_{\rm disp} + p_{\rm ex}(3.0) E^{(20)}_{\text{exch-disp}}\right]_{\rm disp} . \end{aligned} \tag{4.5}$$

Here, the value of $\delta E^{(2)}_{\rm HF}$ computed with $p_{\rm ex}(\alpha = 1)$ is left unchanged, and the $p_{\rm ex}(\alpha)$ scaling factor is also applied to $E^{(20)}_{\text{exch-disp}}$. Parker *et al.*[173] have called *s*SAPT0/jaDZ the "bronze standard" of SAPT,[173] where the jun-cc-pVDZ (jaDZ) basis set[182] is also been recommended for use with SAPT0.[160]

Schäffer and Jansen[175] have shown that $E^{(10)}_{\rm exch}$ and $E^{(20)}_{\text{exch-ind,resp}}$ are underestimated by the SEA, while $E^{(20)}_{\text{exch-disp}}$ is generally overestimated. Thus, the $p_{\rm ex}(\alpha)$ scaling factor based on first-order exchange is suitable for use with $E^{(20)}_{\text{exch-ind,resp}}$ but not $E^{(20)}_{\text{exch-disp}}$. As such, the fact that good results are obtained using *s*SAPT0, despite the fact that $E^{(20)}_{\text{exch-disp}}$ is scaled *up*, suggests that the success of this "bronze standard" rests on error cancellation. In this study, we will compare the effects of the SEA on the $E^{(10)}_{\rm exch}$,

$E^{(20)}_{\text{exch-ind,resp}}$, and $E^{(20)}_{\text{exch-disp}}$ terms for ionic systems, and explore what is the best α value to use in scaling $E^{(20)}_{\text{exch-ind,resp}}$.

The SAPT2 method extends SAPT0 by including *intra*molecular electron correlation up to second order for electrostatic, exchange, and induction interactions (only). It is therefore only suitable for systems whose binding is dominated by electrostatics. The performance of SAPT2 is similar to MP2 and its energy expression is

$$E^{\text{SAPT2}}_{\text{int}} = E^{\text{SAPT0}}_{\text{int}} + \left[E^{(12)}_{\text{elst,resp}}\right]_{\text{elst}} + \left[E^{(11)}_{\text{exch}} + E^{(12)}_{\text{exch}}\right]_{\text{exch}} + \left[{}^{t}E^{(22)}_{\text{ind}} + {}^{t}E^{(22)}_{\text{exch-ind}}\right]_{\text{ind}} . \tag{4.6}$$

The $E^{(11)}_{\text{exch}}$ and $E^{(12)}_{\text{exch}}$ terms are still based on the SEA. The superscripts "t" in ${}^{t}E^{(22)}_{\text{ind}}$ indicates that this is the "true" correlation part of $E^{(22)}_{\text{ind}}$ not included in $E^{(20)}_{\text{ind,resp}}$. The corresponding "true" correlation part of $E^{(22)}_{\text{exch-ind}}$ is estimated as

$${}^{t}E^{(22)}_{\text{exch-ind}} = {}^{t}E^{(22)}_{\text{ind}} \left(\frac{E^{(20)}_{\text{exch-ind,resp}}}{E^{(20)}_{\text{ind,resp}}} \right) , \tag{4.7}$$

where $E^{(20)}_{\text{exch-ind,resp}}$ is based on the SEA, therefore the ${}^{t}E^{(22)}_{\text{exch-ind}}$ is also based on the SEA. Although the δMP2 term (as shown below) often cancels the SEA employed in ${}^{t}E^{(22)}_{\text{exch-ind}}$, we will discuss the influence of the SEA for ${}^{t}E^{(22)}_{\text{exch-ind}}$ in normal SAPT calculations that do not employ the δMP2 correction term. Furthermore, the addition of the δMP2 term offers a means to correct the SEA in $E^{(20)}_{\text{exch-disp}}$, as shown below.

Three higher-level SAPT methods will also be studied: SAPT2+, SAPT2+(3), and SAPT2+3. These approaches include intramolecular electron correlation for dispersion up to second order; this is equivalent to MP4-level dispersion and scales

as $\mathcal{O}(N^7)$. The corresponding energy expressions are[117]

$$E_{\rm int}^{\rm SAPT2+} = E_{\rm int}^{\rm SAPT2} + \left[E_{\rm disp}^{(21)} + E_{\rm disp}^{(22)}\right]_{\rm disp} , \tag{4.8}$$

$$E_{\rm int}^{\rm SAPT2+(3)} = E_{\rm int}^{\rm SAPT2+} + \left[E_{\rm elst,resp}^{(13)}\right]_{\rm elst} + \left[E_{\rm disp}^{(30)}\right]_{\rm disp} , \tag{4.9}$$

and

$$\begin{aligned} E_{\rm int}^{\rm SAPT2+3} = E_{\rm int}^{\rm SAPT2+(3)} \\ + \left[E_{\rm ind,resp}^{(30)} + E_{\rm exch\text{-}ind,resp}^{(30)}\right]_{\rm ind} \\ + \left[E_{\rm exch\text{-}disp}^{(30)} + E_{\rm ind\text{-}disp}^{(30)} + E_{\rm exch\text{-}ind\text{-}disp}^{(30)}\right]_{\rm disp} . \end{aligned} \tag{4.10}$$

The SAPT2+3 method includes the coupling between induction and dispersion. The second-order correction $\delta E_{\rm HF}^{(2)}$ is replaced by the third-order $\delta E_{\rm HF}^{(3)}$ correction in SAPT2+3,

$$\begin{aligned} \delta E_{\rm HF}^{(3)} = E_{\rm int}^{\rm HF} - \big(E_{\rm elst}^{(10)} + E_{\rm exch}^{(10)} + E_{\rm ind,resp}^{(20)} + E_{\rm exch\text{-}ind,resp}^{(20)} \\ + E_{\rm ind,resp}^{(30)} + E_{\rm exch\text{-}ind,resp}^{(30)}\big) . \end{aligned} \tag{4.11}$$

The $E_{\rm exch\text{-}ind,resp}^{(30)}$, $E_{\rm exch\text{-}disp}^{(30)}$, and $E_{\rm exch\text{-}ind\text{-}disp}^{(30)}$ terms are still based on the SEA, which is probably acceptable since $E_{\rm exch\text{-}disp}^{(30)}$ is much smaller than $E_{\rm exch\text{-}disp}^{(20)}$ and $E_{\rm exch\text{-}ind\text{-}disp}^{(30)}$ is mostly cancelled out by the corresponding $E_{\rm ind\text{-}disp}^{(30)}$ term. Furthermore, the $\delta E_{\rm HF}^{(3)}$ terms often cancel the SEA used in $E_{\rm exch\text{-}ind,resp}^{(30)}$. Parker *et al.*[173] have called SAPT2+/aug-cc-pVDZ the "silver standard" in SAPT.

The "δMP2" correction that was mentioned above is intended to account for missing terms such as high-order coupling between induction and dispersion.[173] This correction as defined as

$$\delta E_{\rm MP2} = E_{\rm int}^{\rm MP2} - E_{\rm int}^{\rm SAPT2} , \tag{4.12}$$

where $E_{\rm int}^{\rm MP2}$ is the CP-corrected MP2 binding energy for the dimer, and $E_{\rm int}^{\rm SAPT2}$ contains the δHF correction, according to Eqs. (4.2) and (4.6). (Somewhat arbitrarily, the δMP2 is grouped with the induction terms in the energy decomposition proposed in Ref. 173.) The δMP2 term can be incorporated into SAPT2+, SAPT2+(3), or SAPT2+3, affording methods that we will call SAPT2+δMP2, SAPT2+(3)-δMP2, and SAPT2+3-δMP2. This is equivalent to replacing $E_{\rm int}^{\rm SAPT2}$ in Eqs. (4.8)–(4.10) with $E_{\rm int}^{\rm MP2}$. In other words, the three methods mentioned above, with the addition of $\delta E_{\rm MP2}$, are effectively supermolecular MP2 plus a few additional SAPT terms. Parker *et al.*[173] have called SAPT2+(3)-δMP2/aug-cc-pVTZ the "gold standard" in SAPT. One can also compute dispersion with doubles amplitudes from a coupled-cluster doubles (CCD) calculation,[183,184] and for the three higher-order wave function-based SAPT methods this addendum leads to methods called SAPT2+(CCD), SAPT2+(3)(CCD), and SAPT2+3(CCD).

Finally, we will consider DFT-SAPT calculations in which the interaction energy is

$$\begin{aligned} E_{\rm int}^{\text{DFT-SAPT}} = &\left[E_{\rm elst}^{(1)}\right]_{\rm elst} + \left[E_{\rm exch}^{(1)}\right]_{\rm exch} \\ &+ \left[E_{\rm ind}^{(2)} + E_{\text{exch-ind}}^{(2)} + \delta E_{\rm HF}^{(2)}\right]_{\rm ind} \\ &+ \left[E_{\rm disp}^{(2)} + E_{\text{exch-disp}}^{(2)}\right]_{\rm disp} . \end{aligned} \tag{4.13}$$

Here, the single superscript index indicates that intramolecular electron correlation is not included perturbatively, but rather via DFT. Induction and dispersion energies in DFT-SAPT are determined at the coupled-perturbed static and frequency-dependent Kohn-Sham levels, respectively.

4.2.2 Data sets and benchmarks

We first computed binding energies for the pre-existing "IHB15" data set of 15 ionic hydrogen-bonded dimers,[39] in which the ion is either acetate, guanidinium, methyl ammonium, or imidazolium. These calculations were performed at the SAPT2+(3)/ aug-cc-pVTZ level to establish benchmarks for the individual energy components, as in previous benchmark studies.[37,51,56] The "gold standard" of SAPT methods,[173] namely, SAPT2+(3)-δMP2, was also used to establish the accuracy of these benchmarks. CCSD(T) energies extrapolated to the complete basis-set (CBS) limit are also available for the IHB15 data set.[39] The binding energies in this data set average -17.42 kcal/mol for the three anionic systems and -20.06 kcal/mol for the twelve cationic systems, computed at the CCSD(T)/CBS level. The mean absolute error (MAE) for SAPT2+(3)/aug-cc-pVTZ, relative to these CCSD(T)/CBS benchmarks, is 0.71 kcal/mol for the anions, 0.18 kcal/mol for the cations, and 0.28 kcal/mol overall.

Adding the δMP2 correction to SAPT2+(3)/aug-cc-pVTZ [to obtain the method that we call SAPT2+(3)-δMP2] reduces the MAE of the IHB15 data set to 0.16 kcal/mol (0.19 kcal/mol for the anions and 0.15 kcal/mol for the cations). The mean absolute contribution of the δMP2 correction is 0.90 kcal/mol for the anions and 0.20 kcal/mol for the cations. Notably, this correction is more significant for the anions and suggests that it would be useful to possess a data set with more than three different anions. For this reason, we have assembled a new database "AHB21" of CCSD(T)/CBS benchmark for 21 hydrogen-bonded dimers consisting of an anion and

Table 4.1: CCSD(T)/CBS binding energy benchmarks for the AHB21 and CHB6 ion–neutral data sets.

No.	System	Binding Energy/ kcal mol^{-1}
	——AHB21——	
1	$F^-(NH_3)$	−17.79
2	$F^-(H_2O)$	−32.50
3	$F^-(HF)$	−65.68
4	$Cl^-(NH_3)$	−8.98
5	$Cl^-(H_2O)$	−15.61
6	$Cl^-(HF)$	−25.52
7	$Cl^-(H_2S)$	−14.35
8	$Cl^-(HCl)$	−41.79
9	$OH^-(NH_3)$	−17.03
10	$OH^-(H_2O)$	−37.31
11	$N_3^-(NH_3)$	−7.97
12	$N_3^-(H_2O)$	−14.13
13	$N_3^-(HF)$	−26.01
14	$N_3^-(H_2S)$	−11.07
15	$SH^-(NH_3)$	−8.62
16	$SH^-(H_2O)$	−15.73
17	$SH^-(HF)$	−26.24
18	$HCO_2^-(CH_3NH_2)$	−12.80
19	$HCO_2^-(CH_3OH)$	−20.65
20	$HCO_2^-(H_2O)$	−21.03
21	$HCO_2^-(HF)$	−31.40
	——CHB6——	
22	$Li^+(H_2O)$	−34.43
23	$Na^+(H_2O)$	−23.83
24	$K^+(H_2O)$	−17.83
25	$Li^+(C_6H_6)$	−39.09
26	$Na^+(C_6H_6)$	−25.63
27	$K^+(C_6H_6)$	−19.90

a neutral molecule. The anions in AHB21 are F^-, Cl^-, N_3^-, SH^-, and $HCOO^-$, and the dimers assembled from these anions are shown in Table 4.1 along with benchmark CCSD(T)/CBS binding energies, which range from −65.68 kcal/mol for $F^-(HF)$ to −7.97 kcal/mol for $N_3^-(NH_3)$. (Coordinates for all three of the data sets introduced here can be found in the Supporting Information, SI.)

The three most strongly-bound complexes in the AHB21 set are $F^-(HF)$, $Cl^-(HCl)$, and $OH^-(H_2O)$, with binding energies of −65.68, −41.79, and −37.31 kcal/mol, respectively. Consistent with solution-phase experiments for $(FHF)^-$,[185] each of these species is characterized by a proton that is shared equally between two heavy atoms (single minimum on the potential surface for proton transfer between the two X^- moities). The strength of the $X^- \cdots H^+$ interaction and the symmetry of the complex drives the the non-covalent interaction closer to a *covalent* interaction, resulting in a much larger binding energy as compared to the rest of the AHB21 data set. This characteristic also makes the definition of fragments ambiguous in these three systems, which may therefore be difficult test cases for SAPT and other fragment-based methods.

To complement the AHB21 set of anions, we have also assembled a cation-binding data set that we designate as "CHB6", which includes three alkali–water and three alkali–benzene complexes (see Table 4.1). Binding energies for CHB6 range from −39.09 kcal/mol for $Li^+(C_6H_6)$ to −17.83 kcal/mol for $K^+(H_2O)$.

Finally, we wish to consider cation/anion pairs. Zahn *et al.*[4] recently reported CCSD(T)/CBS binding energies for an "IL-2013" database consisting of 236 different

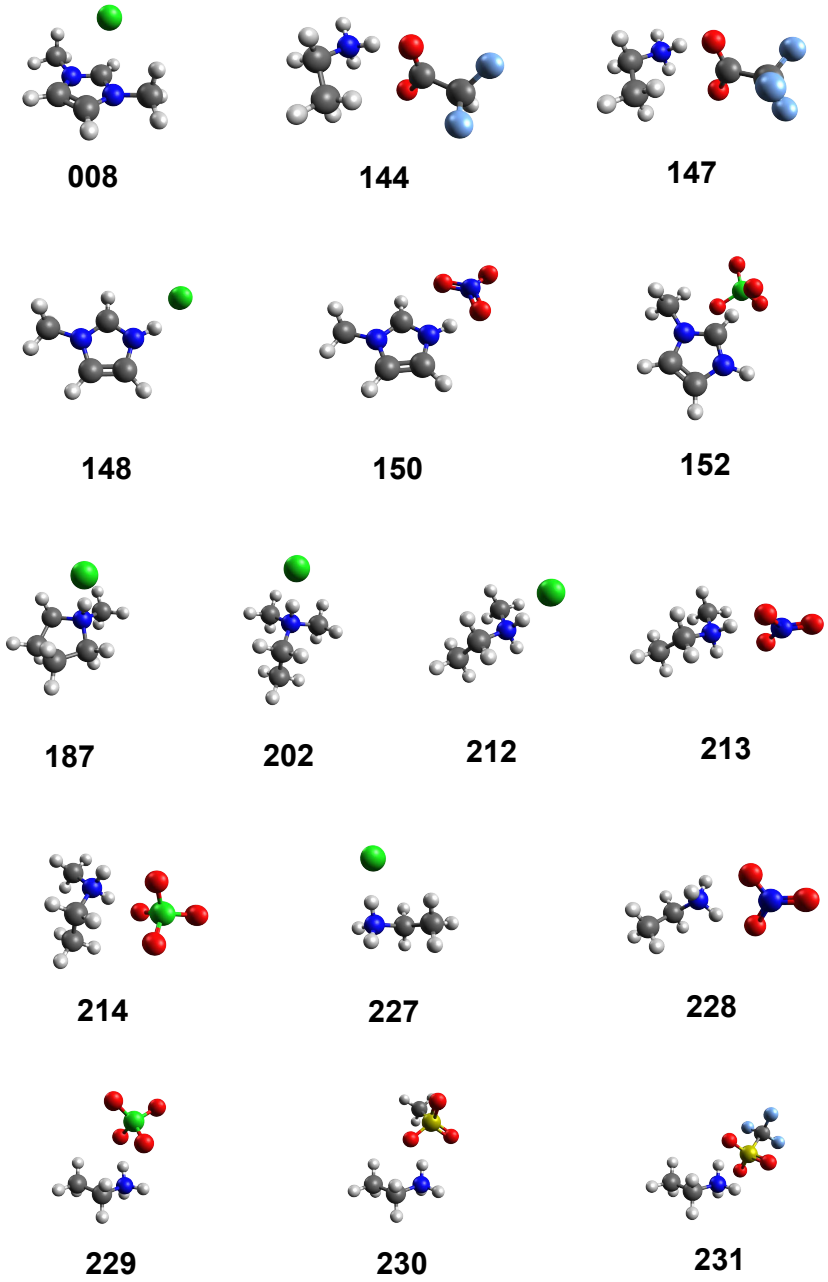

Figure 4.1: Structures of the cation/anion complexes in the IL16 data set, which were taken (along with the numbering scheme) from Ref. 4. The coloring system is as follows: white spheres (hydrogen), gray (carbon), dark blue (nitrogen), red (oxygen), pale blue (fluorine), green (chlorine), yellow (sulfur).

Table 4.2: CCSD(T)/CBS binding energy benchmarks for the IL16 data set.

System[a]	Binding Energy/ kcal mol^{-1}
IL-008	−100.41
IL-144	−120.80
IL-147	−116.91
IL-148	−105.01
IL-150	−104.44
IL-152	−87.42
IL-187	−114.00
IL-202	−113.51
IL-212	−114.91
IL-213	−112.75
IL-214	−104.47
IL-227	−118.19
IL-228	−112.02
IL-229	−106.53
IL-230	−110.98
IL-231	−102.37

[a]Nomenclature is taken from Ref. 4 and structures are shown in Fig. 4.1.

pairs of cations and anions commonly encountered in ionic liquids (IL). The smallest 16 of these ion-pair structures are shown in Fig. 4.1 and constitute what we will call the "IL16" data set. CCSD(T)/CBS binding energies for IL16 are listed in Table 4.2 and range from −87.42 to −120.80 kcal/mol. Note that the binding energies in AHB21, CHB6, and IL16 data sets are considerably larger, on average, as compared to more established databases of neutral dimers, such as S22 or S66.[10,186]

4.2.3 Computational methods

All geometries in the AHB21 and CHB6 data sets were optimized with the resolution-of-identity MP2 method using the aug-cc-pVTZ (aTZ) basis set, except that the def2-TZVPP basis set was used for the alkali atoms. [Similarly, in what follows def2-QZVPP is used for the alkali atoms in cases where aug-cc-pVQZ (aQZ) is used for the remaining atoms.] These geometry optimizations employed Q-Chem v. 4.2.[97] The geometries of the IL16 data set were taken from Ref. 4. Subsequent single-point MP2 and CCSD(T) calculations were carried out using the beta5 version of the Psi4 program,[187] and CCSD(T)-F12 calculations for $HCOO^-$ complexes were performed using the Orca 3.0.2 program.[188] The CCSD(T)-F12 and MP2-F12 calculations for complexes in IL16 were computed using the Molpro 2012.1 program.[189]

Regarding the accuracy of CCSD(T)/CBS benchmarks in general for non-covalent interactions, note two recent studies by Hobza and co-workers that have gone beyond this level for small dimers.[33,141] In one of these studies, CCSD(T) binding energies were found to differ by no more than 0.1 kcal/mol (and typically much less) from coupled-cluster results containing connected pentuple excitations (CCSDTQP).[141]

Another study considered contributions arising from core correlation, relativistic effects, and higher-order excitations at the level of CCSDT(Q), yet the sum total of these effects altered CCSD(T)/CBS binding energies by no more than 0.04 kcal/mol.[33]

For dimers in the AHB21 data sets that involve F^-, Cl^-, OH^-, N_3^-, and SH^-, as well as for the alkali–water complexes in CHB6, the CCSD(T) correlation energy in the CBS limit was estimated using a two-point (aTZ/aQZ) extrapolation scheme.[190] This correlation energy was then added to the Hartree-Fock/aQZ energy to obtain the CCSD(T)/CBS energy.

For the $HCOO^-$ complexes in AHB21 data set, CCSD(T)-F12 correlation energies in the CBS limit were estimated using a two-point extrapolation (cc-pVDZ-F12/cc-pVTZ-F12 with the corresponding near-complete auxiliary basis sets cc-pVDZ-F12-CABS and cc-pVTZ-F12-CABS).[191,192] This correlation energy was added to the Hartree-Fock/cc-pVTZ-F12 energy including the HF-F12 basis set correction[193,194] to estimate the CCSD(T)/CBS energy.

For the three alkali–benzene complexes, the MP2 correlation energies in the CBS limit were estimated using a two-point (aTZ/aQZ) extrapolation, and this correlation energy was added to the Hartree-Fock/aQZ energy to estimate the MP2/CBS energy. In this case, we then add a triples correction

$$\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}} = E_{\mathrm{CCSD(T)}} - E_{\mathrm{MP2}} \; . \tag{4.14}$$

The basis set used to evaluate this correction consists of def2-TZVPP for the alkali atoms and the "heavy augmented" (haTZ) basis set for the remaining atoms, in which the diffuse functions on hydrogen in aTZ are removed.

For the 16 complexes in IL16 data set, the same MP2 extrapolation scheme as described above has been used to calculate the MP2/CBS energy. A triples correction, $\delta_{\mathrm{MP2-F12}}^{\mathrm{CCSD(T)\text{-}F12}}$, was estimated as energy difference between CCSD(T)-F12 and MP2-F12 methods using the aDZ basis set. We expect that the quality of these CCSD(T)/CBS benchmarks is better than that of the original IL-2013 benchmarks,[4] in which the MP2 energy was extrapolated to CBS limit using basis sets lacking in diffuse functions (cc-pVTZ/cc-pVQZ), and moreover the $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ correction was estimated using a double-ζ basis set with no diffuse functions. The MAE between our new CCSD(T)/CBS benchmarks (Table 4.2) and the old ones (Ref. 4) is 0.19 kcal/mol, with a maximum deviation of 0.41 kcal/mol.

To determine the usefulness of wave function-based SAPT for ionic systems, we will compare the benchmark CCSD(T)/CBS binding energies to SAPT2+, SAPT2+(3), and SAPT2+3 results, both with and without CCD amplitudes for dispersion. These SAPT calculations were performed using Dunning's aug-cc-pVXZ basis sets.[195,196] (As mentioned above, Ahlrichs' def2-SVP, def2-TZVPP, and def2-QZVPP basis sets are used for the alkali atoms,[197] but we will continue to abbreviate these basis sets as aDZ, aTZ, and aQZ.) For alkali–benzene dimers, the aQZ calculations strain our computational resources, so in these cases the heavy-aug-cc-pVQZ (haQZ) basis set is used in place of the full aQZ basis, where diffuse functions on hydrogen are omitted. Note that the significantly diminished BSSE in SAPT calculations makes extrapolation to the CBS limit less important than it is for supersystem methods, and in fact previous work has suggested that DFT-SAPT/aQZ calculations are essentially

converged to the basis set limit.[42] Results presented here for the AHB21 data sets show only minor differences between SAPT/aTZ and SAPT/aQZ results, although the SAPT/aTZ outliers are slightly larger for the CHB6 data set. In any case, because the IL16 dimers are somewhat larger than those in the AHB21 and CHB6 data sets, we omit the high-level SAPT/aQZ calculations for this particular data set.

The SAPT0/jaDZ and *s*SAPT0/jaDZ methods will also be evaluated for our three data sets. (The jaDZ basis set was recommended for SAPT0 calculations by Sherrill and co-workers.[160,173] The same basis has also been called aug-cc-pVDZ′ and removes a subset of the diffuse functions from aDZ.) Truncations of the virtual space based on MP2 natural orbitals were used to reduce the cost of the SAPT2+, SAPT2+(3), and SAPT2+3, as described in Refs. 159 and 184. Density fitting, as implemented in Psi4, was used in all SAPT calculations. The SAPT2 and CP-corrected MP2 binding energies required for the δMP2 correction were also performed with the Psi4 program.[187] The exact second-order (non-SEA) terms $E^{(20)}_{\text{exch-ind,resp}}$ and $E^{(20)}_{\text{exch-disp}}$ were evaluated using a locally-modified version of the Molpro 2012.1 program.[189]

DFT-SAPT calculations employ Kohn-Sham (KS) orbitals determined using the PBE0AC exchange-correlation (XC) potential.[163] A hybrid XC kernel consisting of 25% exact exchange and 75% of the adiabatic local density approximation[198] was used to solve coupled-perturbed static and frequency-dependent KS equations for the second-order contributions. The shift parameter in the asymptotic correction[199] was computed in each case using PBE0/aQZ calculations for the neutral molecules and PBE0/def2-QZVPP for the cations. The anionic systems were left without asymptotic

correction, since the XC potentials in these cases are short-ranged and do not decay as $1/r$.[200] The $\delta E^{(2)}_{\rm HF}$ is included in DFT-SAPT to estimate polarization effects beyond second order. All DFT-SAPT calculations were performed using a locally-modified version of the MOLPRO 2012.1 program.[189] Again due to the larger size of the IL16 dimers, the HF-SAPT and DFT-SAPT calculations with exact second-order exchange were limited to the aTZ basis set in these cases.

We will also take the opportunity to test some density-functional methods for AHB21, CHB6, and IL16. Amongst density functionals, M06-2X,[142] ωB97X-D,[201] ωB97X-D3,[13] LC-VV10,[147] ωB97X-V,[105] and B97M-V[202] will be tested, because they have shown good performance for non-covalent interactions in neutral systems.[13,57,105,202] Their performance for ionic systems is suspect, however, in view of previous calculations on $SO_4^{2-}(H_2O)_n$, $F^-(H_2O)_n$, and $Cl^-(H_2O)_n$ clusters,[56,105,203] where the root mean square deviations for binding energies were in some cases as large as 4.8 kcal/mol (M06-2X), 1.3 kcal/mol (ωB97X-D), and 2.7 kcal/mol (LC-VV10), although both ωB97X-V and B97M-V exhibit errors no larger than 0.5 kcal/mol.[202] To the best of our knowledge, the ωB97X-D3 method has not been used to calculate the binding energies in ionic systems, but will be tested here. The aTZ basis set (with def2-QZVPP for alkali atoms, as usual) was used in all DFT calculations, along with a ($N_r = 99$,$N_\Omega = 590$) Euler-Maclaurin-Lebedev quadrature grid for the semi-local functionals and a (75,302) grid for the non-local contributions to LC-VV10, ωB97X-V, and B97M-V.

Finally, we test a Coulomb-attenuated MP2 method (att-MP2).[17,176–178] This approach splits the two-electron Coulomb operator into short- and long-range components,[204] then neglects the long-range part in an effort to eliminate BSSE. We use the aTZ basis set for all att-MP2 calculations (def2-TZVPP for the alkali atoms), since the attenuation parameter was fitted using aTZ and since att-MP2/aTZ calculations yield accurate intermolecular interaction energies in neutral dimers, and att-MP2/aQZ results offer only a very tiny improvement that does not justify the increased cost.[17]

MP2/aTZ, MP2/aQZ, and MP2/CBS binding energies are also reported, using a two-point extrapolation scheme. All supersystem calculations are counterpoise corrected except for the DFT and att-MP2 calculations. (DFT/aTZ results are likely close to the basis-set limit already, and att-MP2 is designed to eliminate BSSE so it does not make sense to apply a counterpoise correction in this case.)

4.3 Results and discussion

4.3.1 Accuracy of the S^2 approximation

Given the importance of exchange effects beyond the SEA for anions,[50] we first examine the accuracy of the exchange, exchange-induction, and exchange-dispersion interactions within the S^2 approximation. Table 4.3 presents the errors in these energy components for our three data sets, where "error" is defined with respect to the corresponding exact (non-SEA) energy components. (These results are obtained using the aTZ basis, but similar behavior is observed using aDZ and aQZ, and these

data can be found in the SI.) In keeping with the observations of Schäffer and Jansen for neutral systems and a few ions,[174,175] we find that the SEA underestimates the first-order exchange energy and the second-order exchange-induction energy, whereas $E^{(20)}_{\text{exch-disp}}(S^2)$ is too large.

For the cationic dimers, the S^2 approximation engenders very little error in either $E^{(1)}_{\text{exch}}(S^2)$, $E^{(2)}_{\text{exch-ind}}(S^2)$, or $E^{(2)}_{\text{exch-disp}}(S^2)$. For the anions, however, the $E^{(10)}_{\text{exch}}(S^2)$ and $E^{(20)}_{\text{exch-ind}}(S^2)$ terms exhibit large errors, with MAEs of 1 kcal/mol and maximum errors as large as 6–9 kcal/mol in the case of Cl^-(HCl). On the other hand, $E^{(20)}_{\text{exch-disp}}(S^2)$ is quite accurate for anions, with maximum errors $\lesssim 0.5$ kcal/mol. For the ion pairs, the MAE for the $E^{(10)}_{\text{exch}}(S^2)$ term is similar to that obtained for the AHB21 data set. On the other hand, the MAE for $E^{(20)}_{\text{exch-ind}}(S^2)$ in IL16 is is about 0.4 kcal/mol larger than in AHB21. Maximum errors in $E^{(10)}_{\text{exch}}(S^2)$ and $E^{(20)}_{\text{exch-ind}}(S^2)$ are generally smaller for IL16 as compared to AHB21, as the ion-pair data set does not contain any of the symmetric shared-proton complexes, $X^- \cdots H^+ \cdots X^-$, that are so difficult for SAPT-based methods. The $E^{(20)}_{\text{exch-disp}}(S^2)$ term is very accurate for IL16, with maximum error of only 0.11 kcal/mol.

For the wave function-based HF-SAPT method, we will scale $E^{(20)}_{\text{exch-ind,resp}}(S^2)$ by $p_{\text{ex}}(\alpha)$, where an optimal value $\alpha = 2.04$ was determined by minimizing the MAE (with respect to exact values of $E^{(20)}_{\text{exch-ind,resp}}$) for the AHB21 data set. This optimal value is consistent with the value $\alpha = 2$ that we used for $F^-(H_2O)$ in a previous study,[50] and for the full AHB21 data set, scaling with $\alpha = 2.04$ affords a MAE of only 0.07 kcal/mol for the $E^{(20)}_{\text{exch-ind,resp}}(S^2)$ term, with a maximum error of 0.3 kcal/mol for

Table 4.3: Mean absolute errors[a] (MAEs) and maximum errors in the single-exchange (S^2) approximation.[b]

Exchange Term	error / kcal mol^{-1}								
	AHB21			**CHB6**			**IL16**		
	MAE	maximum		MAE	maximum		MAE	maximum	
		value	system		value	system		value	system
			——**HF-SAPT**——						
$E^{(10)}_{\text{exch}}(S^2)$	1.05	6.35	Cl^-(HCl)	0.03	0.05	Li^+(H_2O)	0.98	1.71	IL-202
$E^{(20)}_{\text{exch-ind,resp}}(S^2)$	1.05	8.44	Cl^-(HCl)	0.06	0.10	Li^+(C_6H_6)	1.28	2.59	IL-008
$p_{\text{ex}}(\alpha = 2.04)E^{(20)}_{\text{exch-ind,resp}}(S^2)$	0.07	0.31	F^-(HF)	0.00	0.01	K^+(C_6H_6)	0.08	0.15	IL-213
$E^{(20)}_{\text{exch-disp}}(S^2)$	0.07	0.35	F^-(HF)	0.00	0.01	K^+(C_6H_6)	0.03	0.06	IL-202
${}^tE^{(22)}_{\text{exch-ind}}(S^2)$	0.11	0.49	F^-(HF)	0.00	0.01	K^+(H_2O)	0.15	0.25	IL-202
$p_{\text{ex}}(\alpha = 2){}^tE^{(22)}_{\text{exch-ind}}(S^2)$	0.01	0.07	F^-(HF)	0.00	0.00	Li^+(H_2O)	0.02	0.04	IL-213
			——**DFT-SAPT**——						
$E^{(1)}_{\text{exch}}(S^2)$	1.15	6.31	Cl^-(HCl)	0.03	0.06	Li^+(H_2O)	1.11	1.77	IL-202
$E^{(2)}_{\text{exch-ind}}(S^2)$	1.33	9.28	Cl^-(HCl)	0.07	0.10	Li^+(H_2O)	1.69	3.36	IL-008
$p_{\text{ex}}(\alpha = 2.07)E^{(2)}_{\text{exch-ind}}(S^2)$	0.09	0.83	Cl^-(HCl)	0.00	0.01	K^+(C_6H_6)	0.10	0.16	IL-202
$E^{(2)}_{\text{exch-disp}}(S^2)$	0.11	0.52	F^-(HF)	0.00	0.01	K^+(C_6H_6)	0.07	0.11	IL-202

[a]Error is defined with respect to the exact first- or second-order exchange energy. [b]aTZ basis set is used here but aDZ and aQZ results can be found in the SI.

F^-(HF). The optimal scaling value for DFT-SAPT calculations is similar ($\alpha = 2.07$), and errors in the scaled DFT-SAPT value of $E^{(2)}_{\text{exch-ind}}(S^2)$ are also small. For the IL16 data set, an optimal value $\alpha = 2.20$ was obtained for both HF-SAPT and DFT-SAPT calculations; the MAE for the scaled value of $E^{(20)}_{\text{exch-ind,resp}}(S^2)$ with $\alpha = 2.20$ is only 0.1 kcal/mol.

Typically, the ${}^tE^{(22)}_{\text{exch-ind}}$ term in Eq. (4.7) employs the SEA, since $E^{(20)}_{\text{exch-ind,resp}}$ in Eq. (4.7) is based on the SEA as well. For the AHB21 and IL16 data sets, the MAEs in ${}^tE^{(22)}_{\text{exch-ind}}(S^2)$ based on the SEA are 0.11 and 0.15 kcal/mol, respectively, as compared to exact results. Scaling with $\alpha = 2$ reduces both the mean and maximum errors nearly to zero; see Tables 4.3. Errors for the cation data set are even smaller.

To summarize, the S^2 approximation appears to be suitable for use in cation–neutral systems, and also for the $E^{(20)}_{\text{exch-disp}}$ term in anion–neutral and ion-pair systems, but not for $E^{(10)}_{\text{exch}}$ or $E^{(20)}_{\text{exch-ind,resp}}$ in anion–neutral and ion pair binding. Since exact second-order exchange formulas[174,175] have not yet been widely implemented in quantum chemistry codes, we recommend scaling the SEA versions of these terms by $p_{\text{ex}}(\alpha)$ in Eq. (4.4), with $\alpha = 2$. This approach will be taken in the subsequent SEA-based SAPT calculations in this work.

For calculations based on HF-SAPT, we also use a scaling factor of $p_{\text{ex}}(\alpha = 2)$ for $E^{(20)}_{\text{exch-ind,resp}}(S^2)$, and then this scaled value is used to obtain ${}^tE^{(22)}_{\text{exch-ind}}$ according to Eq. (4.7). For the DFT-SAPT methods, the closed-form analytic formula is used for $E^{(2)}_{\text{exch-disp}}$ and $E^{(2)}_{\text{exch-ind}}$, since the SEA and scaled SEA values exhibit somewhat larger maximum errors. Furthermore, the exact (non-SEA) $\delta E^{(2)}_{\text{HF}}$ correction will be used to

obtain total binding energies with DFT-SAPT. For the sSAPT0 method, $p_{\mathrm{ex}}(3.0)$ is used for both $E^{(20)}_{\text{exch-ind,resp}}$ and $E^{(20)}_{\text{exch-disp}}$ in Eq. (4.5), as proposed in Ref. 173. The δHF term can correct the SEA used in exchange-induction, and the δMP2 term can further correct the SEA used in $E^{(20)}_{\text{exch-disp}}(S^2)$ and ${}^tE^{(22)}_{\text{exch-ind}}(S^2)$.

4.3.2 AHB21 anion–neutral data set

Existing benchmarks for neutral systems suggest that SAPT0, the simplest SAPT method, performs well in conjunction with the jaDZ basis set in non-hydrogen-bonded systems, owing to favorable error cancellation.[160,173] For hydrogen-bonded systems, SAPT0/jaDZ exhibits a MAE of 1.26 kcal/mol and a maximum error of 6.68 kcal/mol, across four different data sets examined by Parker *et al.*[173] Almost all of these systems are overbound at the SAPT0/jaDZ level, hence the errors are reduced by empirical scaling, *i.e.*, the sSAPT0/jaDZ method of Eq. (4.5). This is the "bronze standard" of SAPT, and it affords a MAE of 0.71 kcal/mol and a maximum error of 1.55 kcal/mol for the same, neutral data set.[173]

The AHB21 systems exhibit even stronger hydrogen bonds, and for this data set the MAE for SAPT0/jaDZ is 2.01 kcal/mol and the maximum error is 9.54 kcal/mol. The most difficult systems are F^-(HF), Cl^-(HCl), and OH^-(H_2O), for which the binding energies are overestimated by 9.54, 6.24, and 6.51 kcal/mol, respectively. Scaling of the exchange interactions only helps a little: the sSAPT0/jaDZ method affords a MAE of 1.58 kcal/mol and a maximum error of 7.58 kcal/mol for AHB21. These results suggest that low-order SAPT methods are not suitable for use in anionic systems with large binding energies.

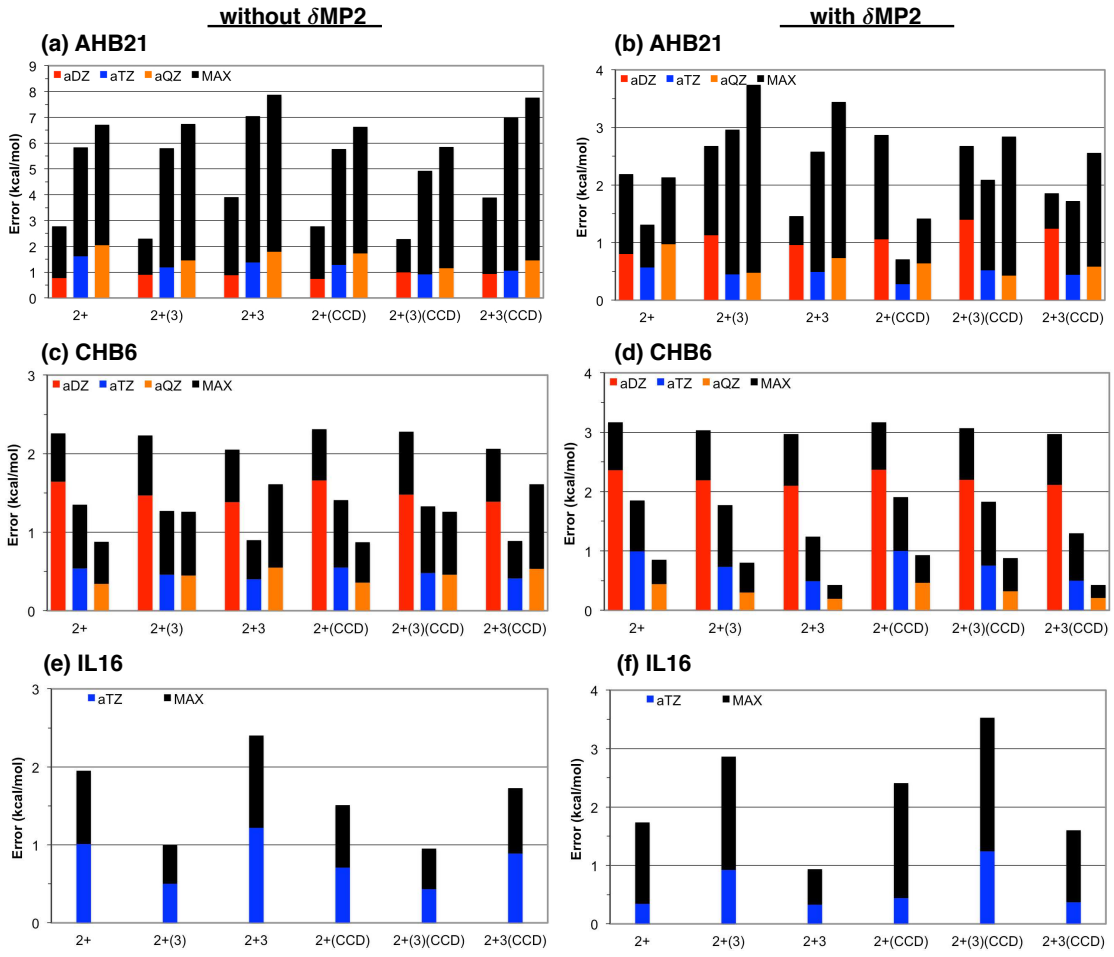

Figure 4.2: Mean absolute errors (MAEs, in red, blue, and orange) and maximum errors (in black) for different levels of SAPT as applied to the AHB21 and CHB6 databases of hydrogen-bonded ion–neutral dimers. For all terms based on the SEA, ${}^{t}E^{(22)}_{\text{exch-ind}}$ is scaled by $p_{\text{ex}}(\alpha = 2)$, $E^{(20)}_{\text{exch-ind}}$ is corrected by δHF, and $E^{(20)}_{\text{exch-disp}}$ is still based on the SEA. Panels on the right show the results when the δMP2 correction [Eq. (4.12)] is included, whereas this correction is omitted in the panels on the left. Note that different panels use different vertical scales.

Figure 4.2(a) shows AHB21 error statistics for six high-level SAPT methods: SAPT2+, SAPT2+(3) and SAPT2+3, each with and without CCD amplitudes for dispersion. (Binding energies for each individual system at each level of theory are available in the SI.) Consistent with the trends observed by Parker *et al.*[173] for neutral hydrogen-bonded systems, we find that both the mean and maximum errors *increase* as the basis set is enlarged. [The only exception to this trend is a very slight decrease in the MAE going from aDZ to aTZ at the SAPT2+(3)(CCD) level.] The best-performing methods are SAPT2+/aDZ (MAE = 0.77 kcal/mol) and SAPT2+(CCD)/aDZ (MAE = 0.74 kcal/mol), and the maximum error for these two methods is 2.8 kcal/mol, for the Cl^-(HCl) system. If the aTZ or aQZ basis set is employed, then SAPT2+(3)(CCD) affords the best performance. Use of CCD dispersion reduces the MAE when the aTZ or aQZ basis is used, although it actually causes a slight increase in the MAE when the basis set is aDZ. In the AHB21 data set, the silver standard SAPT method, SAPT2+/aDZ, performs better than the bronze standard, *s*SAPT0/jaDZ.

Replacing $E^{(20)}_{\text{exch-ind,resp}}$ with a scaled version thereof amounts to adding a term $[p_{\text{ex}}(\alpha) - 1]E^{(20)}_{\text{exch-ind,resp}}$ to SAPT. The value of this scaled-exchange correction, along with that of $[p_{\text{ex}}(\alpha) - 1]^t E^{(22)}_{\text{exch-ind}}$, is plotted in Fig. 4.3(a) for each dimer in AHB21 and for $\alpha = 2$. These corrections are all similar in the aDZ, aTZ, and aQZ basis sets, so only aTZ results are shown in Fig. 4.3. (The analogous plots for the aDZ and aQ basis sets can be found in the SI. The values $\alpha = 1$ and $\alpha = 3$ also afford similar results, although the best-performing value of α is somewhat sensitive to the choice

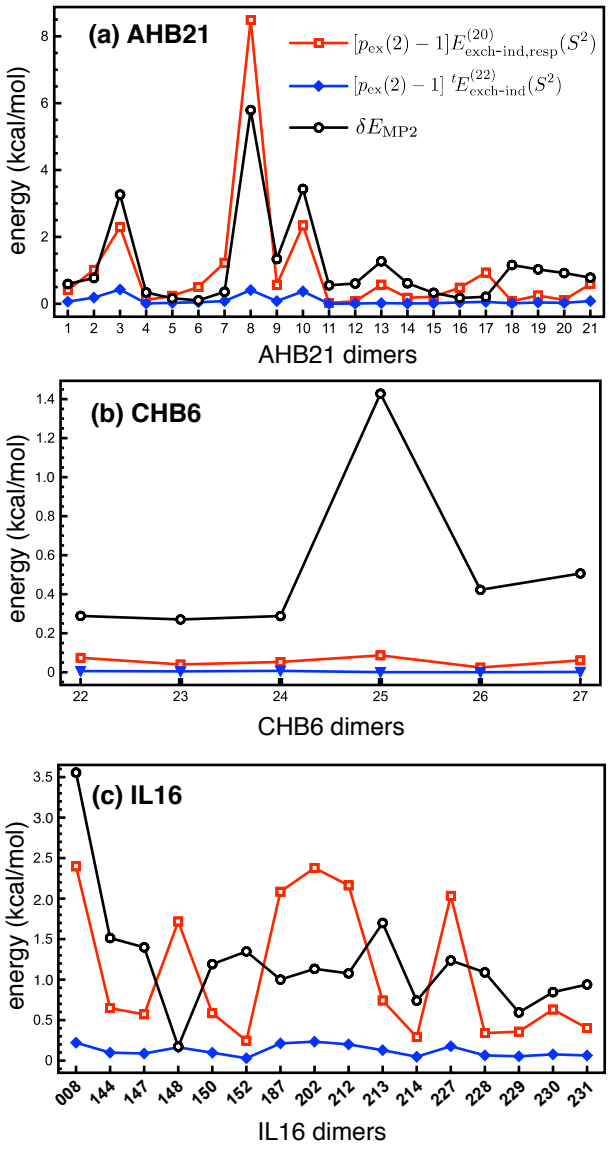

Figure 4.3: Magnitude of the δMP2 correction [Eq. (4.12)], as well as the two second-order scaled-exchange corrections, using the aTZ data set. (Note that the various panels use different vertical scales.) The quantity $p_{ex}(\alpha)$ is the scaling factor defined in Eq. (4.4). The numbering system for the various dimers corresponds to that given in Table 4.1 (for AHB21 and CHB6) and Fig. 4.1 (for IL16).

of α. MAEs for AHB21 computed at the SAPT2+3-δMP2 in various basis sets using either $\alpha = 1$ or $\alpha = 3$ can be found in the SI.

The exchange correction for $E^{(20)}_{\text{exch-ind,resp}}$ is significant especially for F^-(HF), Cl^-(HCl), and OH^-(H_2O), where it contributes 2.3, 8.5, and 2.3 kcal/mol, respectively. These three systems are better described as $X^- \cdots H^+ \cdots X^-$, in which a proton is shared equally between two anions, leading to $H^+ \cdots X^-$ distances that are in the range of typical covalent bond lengths: 1.144 Å for X = F, 1.558 Å for X = Cl, and 1.359 Å for X = OH. The SAPT formalism, however, requires us to describe the system as $X^- \cdots HX$, and the short inter-monomer distance leads to a severe breakdown of the SEA (see Table 4.3). As such, $E^{(10)}_{\text{exch}}$ is very different from $E^{(10)}_{\text{exch}}(S^2)$ for these three systems, and the scaled-exchange correction is large. However, the breakdown of the S^2 approximation or the large exchange correction for $E^{(20)}_{\text{exch-ind,resp}}$ can be mitigated by incorporating $\delta E^{(2)}_{\text{HF}}$, as has been pointed out before.[50] The exchange correction for ${}^tE^{(22)}_{\text{exch-ind}}$ is not large, with a maximum value amongst the AHB21 dimers of only 0.41 kcal/mol, for Cl^-(HCl).

The large errors documented for AHB21 in Fig. 4.2(a) can be remedied by adding the δMP2 correction [Eq. (4.12)]. The magnitude of this correction for each AHB21 dimer is shown in Fig. 4.3(a), from which we see that the δMP2 correction is also largest for the three $X^- \cdots H^+ \cdots X^-$ systems, namely, 3.27, 5.79, and 3.43 kcal/mol for X = F, Cl, and OH, respectively. The large magnitude of this correction is another indication of breakdown of the perturbation series such that SAPT may not yield accurate energy components in such cases, as discussed also by Parker *et*

al.[173] Total binding energies can still be reproduced accurately if the δMP2 correction is included, as shown in Fig. 4.2(b) and discussed below. Note that SAPT with the δMP2 correction is really just supermolecular MP2 with a few additional SAPT terms, and this correction can cancel out the SEA used in $E^{(20)}_{\text{exch-disp}}$ and ${}^{t}E^{(22)}_{\text{exch-ind}}$.

Comparing Fig. 4.2(b) to Fig. 4.2(a), we see that the δMP2 correction significantly reduces the errors at all levels of SAPT. The SAPT-δMP2 methods with the aTZ and aQZ basis sets outperform the corresponding methods with the aDZ basis, with SAPT-δMP2/aTZ being slightly more accurate than the corresponding aQZ method, except in the case of SAPT2+(3)-δMP2 and SAPT2+(3)(CCD)-δMP2. Amongst all of the SAPT methods evaluated here, the SAPT2+(CCD)-δMP2/aTZ approach provides the most accurate results for the AHB21 data set, with a MAE of 0.28 kcal/mol and a maximum error of 0.71 kcal/mol, which occurs in the case of $F^-(H_2O)$. The superior performance of SAPT2+(CCD)-δMP2/aTZ is in line with previous conclusions,[173] although the errors for these anion–neutral dimers are slightly larger than errors reported for neutral systems in Ref. 173. Note that SAPT2+(CCD)-δMP2 has also been called MP2(CCD), since it amounts to a supermolecular MP2 calculation supplemented with dispersion corrections from CCD amplitudes.[173]

In the four hydrogen-bonded systems examined by Parker *et al.*,[173] the MAE in SAPT2+(3)-δMP2/aTZ was 0.24 kcal/mol and was slightly worse than results at the SAPT2+(CCD)-δMP2/aTZ level (0.22 kcal/mol) and SAPT2+3-δMP2/aTZ level (0.21 kcal/mol). Nevertheless, those authors designate SAPT2+(3)-δMP2/aTZ

as the "gold standard" of SAPT, since the performance is similar but the computational cost is ∼50% less than SAPT2+(CCD)-δMP2/aTZ. However, for the AHB21 data set, the SAPT2+(3)-δMP2/aTZ exhibits a MAE of 0.45 kcal/mol and a maximum error of 2.96 kcal/mol. This MAE is about twice as large as that obtained at the SAPT2+(CCD)-δMP2/aTZ level, and as such we recommend SAPT2+(CCD)-δMP2/aTZ as the method of choice when a balanced description of both neutral and anionic hydrogen-bonded systems is required. The SAPT2+3-δMP2/aTZ method performs similar to SAPT2+(3)-δMP2/aTZ, with a MAE of 0.49 kcal/mol and maximum error of 2.6 kcal/mol.

Statistical errors in AHB21 binding energies obtained from DFT-SAPT calculations *without* the SEA are summarized in Table 4.4. The MAEs of DFT-SAPT decrease as the basis set is enlarged, becoming as small as 0.91 kcal/mol in the aQZ basis. The wave function-based SAPT2+(CCD)-δMP2/aTZ, SAPT2+(3)-δMP2/aTZ, and SAPT2+3-δMP2/aTZ methods thus exhibit smaller MAEs and smaller maximum errors for AHB21, as compared to DFT-SAPT, although the latter approach is less expensive and can be made to scale as $\mathcal{O}(N^5)$ with density fitting.[205]

For comparison, error statistics for AHB21 using MP2 and DFT methods are also listed in Table 4.4. The M06-2X and LC-VV10 methods, which have previously been recommended for DFT calculations of non-covalent interactions, afford MAEs in excess of 1 kcal/mol, as in previous studies of $F^-(H_2O)_n$ and $Cl^-(H_2O)_n$ clusters.[56,57] The ωB97X-D, ωB97X-D3, ωB97X-V, and B97M-V methods afford much

Table 4.4: Mean absolute errors[a] (MAEs) and maximum errors in binding energies for the AHB21 data set.

Method	error / kcal mol^{-1}		
	MAE	maximum	
		value	system
SAPT2+(CCD)-δMP2/aTZ[b]	0.28	0.71	$F^-(H_2O)$
DFT-SAPT[c]/aDZ	1.93	5.73	F^-(HF)
DFT-SAPT[c]/aTZ	1.03	6.15	Cl^-(HCl)
DFT-SAPT[c]/aQZ	0.91	6.97	Cl^-(HCl)
DFT-SAPT[c]/CBS	0.89	7.56	Cl^-(HCl)
M06-2X/aTZ	1.08	3.97	F^-(HF)
LC-VV10/aTZ	1.18	3.55	Cl^-(HCl)
ωB97X-D/aTZ	0.27	0.64	$OH^-(H_2O)$
ωB97X-D3/aTZ	0.32	0.98	$OH^-(H_2O)$
ωB97X-V/aTZ	0.32	0.99	$OH^-(H_2O)$
B97M-V/aTZ	0.29	1.29	Cl^-(HCl)
att-MP2/aTZ	0.47	2.47	Cl^-(HCl)
MP2/aTZ	0.67	1.96	F^-(HF)
MP2/aQZ	0.45	1.88	Cl^-(HCl)
MP2/CBS	0.40	2.64	Cl^-(HCl)

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

[b] Best-performing SAPT method for this data set.

[c] Includes exact second-order exchange.

smaller MAEs (0.3 kcal/mol), performance that is comparable to SAPT2+(CCD)-δMP2/aTZ. This result is consistent with the good performance of ωB97X-D/aTZ for other hydrogen-bonded systems.[206] As with most of the SAPT methods, the $X^- \cdots H^+ \cdots X^-$ systems remain challenging, and constitute the largest source of error for each density functional, as well as for each MP2 method.

Regarding "attenuated" MP2 (att-MP2) methods,[176,204] we find that the att-MP2/aTZ results are superior to MP2/aTZ, consistent with results of a previous study.[17] However, MP2/aQZ and MP2/CBS results are slightly better still. The reason may be that the Coulomb attenuation parameter in att-MP2 that was reported in Ref. 17 (and used here) was optimized against the S66 data set,[10] which contains only charge-neutral monomers and therefore this attenuation parameter may not strike an ideal balance between neutral and ionic systems.

4.3.3 CHB6 cation–neutral data set

Errors in high-level SAPT methods applied to CHB6 are shown in Fig. 4.2(c), and are much smaller than those obtained for AHB21 [Fig. 4.2(a)]. his is consistent with results for Hobza's IHB15 data set,[39] for which the SAPT2+(3)/aTZ method exhibits large errors for the anionic dimers but performs much better for the cations. Unlike AHB21 results, for CHB6 the aTZ and aQZ basis sets afford much smaller errors as compared to results obtained using aDZ.

The scaled-exchange and δMP2 corrections for CHB6 are plotted in Fig. 4.3(b). (As for AHB21, the aDZ and aQZ results are given in the SI.) The two exchange corrections are close to zero, and the δMP2 correction is also not large, with a maximum

value of 1.43 kcal/mol for $Li^+(C_6H_6)$. In this outlier, the distance from Li^+ to the center of the benzene ring is only 1.91 Å.

In contrast to the AHB21 dimers, for these cation–neutral systems the addition of δMP2 tends to increase the errors, except for SAPT2+(3)/aQZ and SAPT2+3/aQZ (with or without CCD dispersion); see Fig. 4.2(d). However, the δMP2 correction does cause SAPT to exhibit monotonic convergence towards the CBS limit, so that errors with respect to CCSD(T)/CBS results are smallest in the largest basis sets. Notably, this monotonic convergence is *not* observed in the AHB21 data set; see Fig. 4.2(b). The cations–neutral complexes in CHB6 are also better behaved with respect to increasing the level of SAPT, *i.e.*

$$\text{SAPT2+} \;\rightarrow\; \text{SAPT2+(3)} \;\rightarrow\; \text{SAPT2+3}\,.$$

However, the more robust treatment of dispersion based on CCD amplitudes has little effect.

The best SAPT methods for the CHB6 data set are SAPT2+3/aQZ and SAPT2+3 (CCD)/aQZ, which exhibit MAEs of 0.19 and 0.21 kcal/mol, respectively, and errors no larger than 0.43 kcal/mol [for $Li^+(H_2O)$]. The so-called gold and silver SAPT standards, SAPT2+(3)-δMP2/aTZ and SAPT2+/aDZ (respectively), afford MAEs of 0.73 and 1.64 kcal/mol, respectively. The SAPT2+(CCD)-δMP2/aTZ method, which we identified as the best approach for the AHB21 data set, exhibits a 1 kcal/mol MAE for CHB6, with a maximum error of 1.91 kcal/mol. SAPT2+3-δMP2/aTZ, identified by Parker *et al.*[173] as the best SAPT method for neutral hydrogen-bonded systems, still affords moderate errors for CHB6 data set with a MAE of 0.49 kcal/mol and

Table 4.5: Mean absolute errors[a] (MAEs) and maximum errors in binding energies for the CHB6 data set.[b]

Method	error / kcal mol^{-1}		
	MAE	maximum	
		value	system
SAPT2+3-δMP2/aQZ[c]	0.19	0.43	$Li^+(H_2O)$
DFT-SAPT[d]/aDZ	1.70	2.40	$Na^+(C_6H_6)$
DFT-SAPT[d]/aTZ	0.80	1.73	$K^+(C_6H_6)$
DFT-SAPT[d]/aQZ	0.56	1.10	$K^+(C_6H_6)$
DFT-SAPT[d]/CBS	0.47	1.05	$Na^+(C_6H_6)$
M06-2X/aTZ	1.30	3.27	$Li^+(C_6H_6)$
LC-VV10/aTZ	1.62	4.60	$Li^+(C_6H_6)$
ωB97X-D/aTZ	0.76	0.97	$Li^+(H_2O)$
ωB97X-D3/aTZ	0.32	0.77	$K^+(H_2O)$
ωB97X-V/aTZ	0.71	2.25	$Li^+(C_6H_6)$
B97M-V/aTZ	0.85	1.55	$K^+(C_6H_6)$
att-MP2/aTZ	0.88	2.05	$Na^+(C_6H_6)$
MP2/aTZ	0.90	1.20	$K^+(C_6H_6)$
MP2/aQZ	0.31	0.55	$Na^+(C_6H_6)$
MP2/CBS	0.22	0.64	$K^+(C_6H_6)$

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

[b] For the alkali atoms, def2-SVP (aDZ), def2-TZVPP (aTZ), and def2-QZVPP (aQZ) basis sets are used.

[c] Best-performing SAPT method for this data set.

[d] Includes exact second-order exchange

maximum error of 1.24 kcal/mol. Hence, the extra coupling terms between induction and dispersion in SAPT2+3 as compared to SAPT2+(3) improve the binding description in cation–neutral complexes. For such systems, we therefore recommend the highest possible level of SAPT combined with the largest affordable basis set, preferably SAPT2+3/aQZ.

DFT-SAPT (with exact second-order exchange) shows monotonic convergence with respect to expansion of the basis set (see Table 4.5), as observed also for the AHB21 data set. However, the MAE at the DFT-SAPT/aQZ level is 0.56 kcal/mol, three times worse than the best wave function-based SAPT method, SAPT2+3-δMP2/aQZ, and slightly worse than SAPT2+3-δMP2/aTZ also. On the other hand, DFT-SAPT/aQZ does outperform the so-called silver- and gold standards of wave function-based SAPT.

Table 4.5 also reports error statistics for DFT and MP2-based methods as applied to the CHB6 data set. The M06-2X and LC-VV10 methods afford large MAEs (1.30 and 1.62 kcal/mol, respectively), much worse than the performance of these functionals for benchmark non-covalent interactions involving neutral molecules.[57] The ωB97X-D, ωB97X-V, and B97M-V methods give similar errors, with MAEs of 0.7–0.9 kcal/mol. Amongst the functionals tested here, ωB97X-D3 affords the smallest MAE (0.32 kcal/mol) for the CHB6 data set. However, the MP2/aQZ and MP2/CBS methods afford even smaller MAEs, 0.31 and 0.22 kcal/mol, respectively. As with AHB21, convergence of MP2 results with respect to expansion of the basis set is also monotonic for CHB6.

Table 4.6: Mean absolute errors[a] (MAEs) and maximum errors in SAPT binding energies for the IL16 data set.

Method	error / kcal mol^{-1}			
	without δMP2[b]		with δMP2[c]	
	MAE	max	MAE	max
SAPT2+	1.01	1.95	0.34	1.74
SAPT2+(3)	0.50	1.00	0.92	2.86
SAPT2+3	1.22	2.40	0.33	0.94
SAPT2+(CCD)	0.71	1.51	0.44	2.41
SAPT2+(3)(CCD)	0.43	0.95	1.24	3.53
SAPT2+3(CCD)	0.89	1.73	0.37	1.60

[a] With respect to CCSD(T)/CBS benchmarks. [b] $^{t}E^{(22)}_{\text{exch-ind}}$ is scaled by $p_{\text{ex}}(\alpha = 2)$, $E^{(20)}_{\text{exch-ind}}$ is corrected by δHF, and $E^{(20)}_{\text{exch-disp}}$ is evaluated using the SEA. [c] $E^{(20)}_{\text{exch-ind}}$ is corrected by δHF, and both $^{t}E^{(22)}_{\text{exch-ind}}$ and $E^{(20)}_{\text{exch-disp}}$ are corrected by δMP2.

4.3.4 IL16 ion-pair data set

For the IL16 ion-pair systems, SAPT0/jaDZ exhibits a MAE of 3.38 kcal/mol with a maximum error of 5.68 kcal/mol. The "bronze standard" sSAPT0/jaDZ increases the MAE to 5.20 kcal/mol with a maximum error of 8.27 kcal/mol. Therefore, in contrast to the case of AHB21, where scaling of $E^{(20)}_{\text{exch-disp}}(S^2)$ reduces the MAE from 2.01 and 1.58'kcal/mol, for IL16 sSAPT0 does not afford good error cancellation and errors remain large. The silver standard of SAPT, SAPT2+/aDZ, performs much better, with a MAE of 0.62 kcal/mol and a maximum error of 1.87 kcal/mol.

Errors in high-level SAPT methods, with and without including the δMP2 correction, for the IL16 data set are shown in Table 4.6. By using CCD amplitudes

for the dispersion corrections, errors in SAPT methods without δMP2 are reduced while errors in SAPT methods with δMP2 are increased. The δMP2 term worsens the performance of SAPT2+(3) and SAPT2+(3)(CCD); however, it improves the performance of the rest of the high-level SAPT methods. The best method for these ion-pair systems is SAPT2+3-δMP2, with a MAE of 0.33 kcal/mol. The gold standard SAPT2+(3)-δMP2 gives a MAE of 0.92 kcal/mol, almost three times worse than SAPT2+3-δMP2. The extra coupling terms between induction and dispersion in SAPT2+3 as compared to SAPT2+(3) are important in ion-pair complexes. The best SAPT method for AHB21 data set, SAPT2+(CCD)-δMP2/aTZ, performs reasonably well, with a MAE of 0.44 kcal/mol.

The scaled-exchange and δMP2 corrections are plotted in Fig. 4.3(c) for the IL16 data set. The average contributions of $[p_{\mathrm{ex}}(\alpha = 2) - 1]E^{(20)}_{\mathrm{exch\text{-}ind,resp}}$, $[p_{\mathrm{ex}}(\alpha = 2) - 1]^t E^{(22)}_{\mathrm{exch\text{-}ind}}$, and δMP2 are 1.10, 0.12, and 1.22 kcal/mol, respectively. For comparison, the same three values for AHB21 are 0.99, 0.10, and 1.13 kcal/mol, respectively, that is, basically comparable if slightly smaller. (Note that the average binding energy in AHB21 is -22.49 kcal/mol whereas in IL16 it is -87.42 kcal/mol.) We take this as an indication that the anion exerts greater influence on the magnitude of these values than does the cation.

The performance of DFT-SAPT, DFT, and MP2-based methods for IL16 is summarized in Table 4.7. Amongst these supersystem methods, DFT-SAPT/aTZ exhibits a rather large MAE, 2.18 kcal/mol. The M06-2X and LC-VV10 functionals perform better for IL16 than for AHB21 or CHB6 data sets (MAEs < 0.5 kcal/mol), which

Table 4.7: Mean absolute errors[a] (MAEs) and maximum errors in binding energies for the IL16 data set.

Method	error / kcal mol^{-1}		
	MAE	maximum	
		value	system
SAPT2+3-δMP2/aTZ[b]	0.33	0.94	IL-008
DFT-SAPT[c]/aDZ	3.57	4.80	IL-150
DFT-SAPT[c]/aTZ	2.18	3.67	IL-150
M06-2X/aTZ	0.47	0.95	IL-152
LC-VV10/aTZ	0.29	0.85	IL-144
ωB97X-D/aTZ	1.31	3.05	IL-008
ωB97X-D3/aTZ	0.89	2.32	IL-008
ωB97X-V/aTZ	0.79	1.31	IL-187
B97M-V/aTZ	0.43	1.06	IL-148
att-MP2/aTZ	0.33	1.43	IL-008
MP2/aTZ	0.92	1.67	IL-008
MP2/aQZ	0.47	1.15	IL-008
MP2/CBS	0.31	0.81	IL-008

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

[b]Best-performing SAPT method for this data set.

[c]Includes exact second-order exchange

seems odd given the larger binding energies for IL16, whereas ωB97X-D, ωB97X-D3, and ωB97X-V (MAEs of 0.8–1.3 kcal/mol) exhibit larger errors for IL16 than for AHB21 or CHB6. B97M-V and att-MP2 exhibit MAEs < 0.5 kcal/mol for IL16 that are similar to their MAEs for AHB21 although slightly larger than their MAEs for CHB6. The MP2 method shows monotonic convergence towards the CBS limit, with a MAE for MP2/CBS of 0.31 kcal/mol.

4.3.5 Summary

Overall errors for all 43 dimers contained in AHB21, CHB6, and IL16, are listed in Table 4.8, as obtained using a variety of methods. The SAPT0/jaDZ and bronze standard[173] sSAPT0/jaDZ methods afford large errors in these dimers that involve ions, with overall MAEs of almost 3 kcal/mol. The DFT-SAPT/aTZ method affords an overall MAE of 1.43 kcal/mol, roughly consistent with its MAE of 0.92 kcal/mol that has previously been reported for neutral hydrogen-bonded systems.[173] As such, none of these three methods can be recommended for sub-kcal/mol accuracy in strongly-bound complexes.

The silver standard[173] SAPT2+/aDZ and gold standard[173] SAPT2+(3)-δMP2/aTZ methods afford more moderate errors, with overall MAEs of 0.67 and 0.44 kcal/mol, respectively. The SAPT2+(CCD)-δMP2/aTZ method, which gave the smallest MAE for the AHB21 data set on its own, affords an MAE of 0.44 kcal/mol for this composite data set, while the best method for IL16 on its own (SAPT2+3-δMP2/aTZ) provides a similar MAE of 0.43 kcal/mol for the composite set. These methods therefore appear to be accurate and consistent for complexes involving ions.

Table 4.8: Overall mean absolute errors[a] (MAEs) and maximum errors in binding energies for the composite AHB21 + CHB6 + IL16 data set.[b]

Method	error / kcal mol^{-1}		
	MAE	maximum	
		value	system
SAPT0/jaDZ	2.60	9.54	F^-(HF)
sSAPT0/jaDZ	3.15	8.27	IL-227
SAPT2+/aDZ	0.84	2.78	Cl^-(HCl)
SAPT2+(3)-δMP2/aTZ	0.67	2.96	F^-(HF)
SAPT2+(CCD)-δMP2/aTZ	0.44	2.41	IL-008
SAPT2+3-δMP2/aTZ	0.43	2.58	F^-(HF)
DFT-SAPT[c]/aTZ	1.43	6.15	Cl^-(HCl)
M06-2X/aTZ	0.88	3.97	F^-(HF)
LC-VV10/aTZ	0.91	4.60	$Li^+(C_6H_6)$
ωB97X-D/aTZ	0.73	3.05	IL-008
ωB97X-D3/aTZ	0.69	2.32	IL-008
ωB97X-V/aTZ	0.55	2.25	$Li^+(C_6H_6)$
B97M-V/aTZ	0.42	1.55	$K^+(C_6H_6)$
att-MP2/aTZ	0.47	2.47	Cl^-(HCl)
MP2/aTZ	0.79	1.96	F^-(HF)
MP2/aQZ	0.44	1.88	Cl^-(HCl)
MP2/CBS	0.34	2.64	Cl^-(HCl)

[a]With respect to CCSD(T)/CBS benchmarks.
[b]For the alkali atoms in CHB6, def2-SVP (aDZ), def2-TZVPP (aTZ), and def2-QZVPP (aQZ) basis sets are used. [c]Includes exact second-order exchange

The SAPT2+(CCD)-δMP2/aTZ affords MAEs of 0.28, 1.00, and 0.44 kcal/mol for the AHB21, CHB6, and IL16 data sets, respectively, whereas the MAEs engendered by SAPT2+3-δMP2/aTZ are 0.49, 0.49, and 0.33 kcal/mol, respectively. The latter method therefore offers the more consistent accuracy across different charge states. The performance of SAPT2+3-δMP2/aTZ is also quite good for neutral complexes,[173] hence this level of theory is recommended for high-accuracy SAPT calculations in ionic complexes.

Amongst DFT methods, the newly developed ωB97X-V and B97M-V functionals afford small overall MAEs of 0.55 and 0.42 kcal/mol, respectively. Furthermore, B97M-V has the smallest maximum error among all methods used in this work, included SAPT- and MP2-based approaches. Thus, B97M-V offers accurate binding energies for both neutral[202] and ionic systems. The att-MP2/aTZ is superior to MP2/aTZ and MP2/aQZ, and it is also recommended as an efficient method to study binding energies for both neutral[17] and ionic systems. The best method for ionic systems in this work is MP2/CBS with a MAE of 0.34 kcal/mol, although large basis sets are required to eliminate BSSE. Moreover, MP2/CBS succeeds here due to the importance of electrostatics; in systems where dispersion is more important, this method is less accurate.[186,207]

4.4 Conclusions

In this work, we assembled three data sets of non-covalent dimers in which one or both partners is an ion, and reported benchmark CCSD(T)/CBS results for these

dimers. We call these data sets AHB21 (consisting of 21 strongly hydrogen-bonded anion–neutral complexes), CHB6 (containing six cation–neutral dimers), and IL16 (composed of 16 cation–anion dimers composed of common ions found in ionic liquids).

We have systematically evaluated the accuracy of the so-called "S^2" or single-exchange approximation that is used in the first-order exchange, second-order exchange-induction, and second-order exchange-dispersion energy components. The accuracy of various levels of SAPT has been evaluated for these three data sets, and in addition the accuracy of some DFT and MP2 methods has been assessed as well. Our findings are summarized below.

1. For these data sets, the 1^{st}-order exchange and 2^{nd}-order exchange-induction components are underestimated within the SEA, whereas the 2^{nd}-order exchange-dispersion component is overestimated. This is consistent with previous conclusions for neutral molecules and selected ions.[174,175]

2. For the complexes in AHB21 and IL16, the $E^{(10)}_{\rm exch}(S^2)$ and $E^{(20)}_{\rm exch\text{-}ind,resp}(S^2)$ terms exhibit large errors, whereas the $E^{(20)}_{\rm exch\text{-}disp}(S^2)$ and ${}^tE^{(22)}_{\rm exch\text{-}ind}(S^2)$ terms are quite accurate as compared to exact (non-SEA) results. *Ad hoc* scaling, based upon the ratio between $E^{(10)}_{\rm exch}$ with and without the SEA [$p_{\rm ex}(\alpha = 2)$ in Eq. (4.5)] can be used to correct the deficiency in $E^{(20)}_{\rm exch\text{-}ind,resp}(S^2)$ and thereby reduce the MAE for AHB21 from 1.0 to 0.1 kcal/mol, and the MAE for IL16 from 1.5 to 0.1 kcal/mol, as compared to the exact value of $E^{(20)}_{\rm exch\text{-}ind,resp}$.

3. For cations (CHB6 data set), the S^2 approximation is quite accurate.

4. The error engendered by the SEA in the $E^{(20)}_{\text{exch-ind,resp}}$ and $E^{(30)}_{\text{exch-ind,resp}}$ terms can be compensated by addition of the δHF correction. The δMP2 correction, defined as the difference between $E^{\text{MP2}}_{\text{int}}$ and $E^{\text{SAPT2}}_{\text{int}}$ [Eq. (4.12)], can compensate for the SEA in $E^{(20)}_{\text{exch-disp}}$ and ${}^{t}E^{(22)}_{\text{exch-ind}}$.

5. The δMP2 correction is large for anions and ion-pairs, but smaller for cations. This is due primarily to the fact that the higher-order induction-dispersion coupling terms are large where anions are involved, and these terms are captured by the δMP2 correction.

6. The magnitudes of the exchange-scaling correction, $\left[p_{\text{ex}}(\alpha)-1\right]E^{(20)}_{\text{exch-ind,resp}}$, and the δMP2 correction, point to the breakdown of the perturbation expansion and thus indicate cases where the individual SAPT energy components may no longer be reliable. However, total binding energies may still be accurately reproduced by SAPT methods, if the δHF and δMP2 corrections are included.

7. For the anions in AHB21, the "gold", "silver", and "bronze standards" of wave function-based SAPT[173] [*i.e.*, SAPT2+(3)-δMP2/aTZ, SAPT2+/aDZ, and sSAPT0/jaDZ, respectively] afford MAEs of 0.45, 0.77, and 1.58 kcal/mol, respectively.

8. The best SAPT method for the AHB21 anions is SAPT2+(CCD)-δMP2/aTZ, which exhibits a MAE of 0.28 kcal/mol and a maximum error of 0.71 kcal/mol. This method is equivalent to a supermolecular MP2 calculation with dispersion corrections that employ CCD amplitudes, and is therefore also known as

MP2(CCD). The MP2(CCD) method performs better than the regular MP2/aTZ method with a MAE of 0.67 kcal/mol, since dispersion corrections based on CCD amplitudes have been included in MP2(CCD). This method also performs well for describing hydrogen bonds between neutral monomers,[173] and is recommended in cases where a balanced description of both neutral and anionic systems is required.

9. For the CHB6 data set, the "gold" and "silver" SAPT standards[173] afford MAEs of 0.73 and 1.64 kcal/mol, respectively.

10. The best SAPT method for the CHB6 data set is SAPT2+3-δMP2/aQZ, with a MAE of 0.2 kcal/mol and a maximum error of 0.4 kcal/mol. We recommend use of the highest-level of SAPT available, combined with the largest feasible basis set, for cation-binding calculations.

11. For the ion-pairs in IL16, the "gold", "silver", and "bronze standards" of wave function-based SAPT[173] [*i.e.*, SAPT2+(3)-δMP2/aTZ, SAPT2+/aDZ, and sSAPT0/jaDZ, respectively] afford MAEs of 0.92, 0.62, and 5.20 kcal/mol, respectively. Thus the "silver standard" performs slightly better than the "gold standard" for ion pairs.

12. The best SAPT method for the IL16 ion pairs is SAPT2+3-δMP2/aTZ, which affords a MAE of 0.33 kcal/mol and a maximum error of 0.94 kcal/mol.

13. The extra coupling terms between induction and dispersion in SAPT2+3 as compared to SAPT2+(3) are essential to generate good binding energies in cation-binding and ion-pair complexes.

14. The SAPT2+(CCD)-δMP2/aTZ method affords MAEs of 0.28, 1.00, and 0.44 kcal/mol for the AHB21, CHB6, and IL16 data sets, respectively. Combining these three data sets, the overall MAE is 0.44 kcal/mol. The corresponding MAEs for SAPT2+3-δMP2/aTZ is 0.49, 0.49, and 0.33 kcal/mol, respectively, with an overall MAE of 0.43 kcal/mol. Although SAPT2+(CCD)-δMP2/aTZ is quite accurate for anionic systems, SAPT2+3-δMP2/aTZ provides a more balanced description for a variety of ionic systems. Furthermore, the overall performance of SAPT2+3-δMP2/aTZ is better than the aforementioned "gold standard" SAPT2+(3)-δMP2/aTZ method, with a MAE of 0.67 kcal/mol. Hence, we put forward the SAPT2+3-δMP2/aTZ approach as an alternative "gold standard" for ionic complexes.

15. SAPT methods can succeed in strongly-bound systems, so long as the supersystem δHF correction [Eq. (4.3) or (4.11)] and the supersystem δMP2 correction [Eq. (4.12)] are applied. These are no longer "pure" SAPT approaches, however, as they require supermolecular HF and MP2 calculations, and the decomposition into energy components may therefore become problematic.

16. The overall MAEs for SAPT0/jaDZ, bronze standard *s*SAPT0/jaDZ, and DFT-SAPT/aTZ are 2.60, 3.15, and 1.43 kcal/mol, respectively. It is not recommended to apply these approaches to strongly-bound complexes. The performance of silver standard SAPT2+/aDZ approach is slightly better, with an overall MAE of 0.84 kcal/mol for the three data sets considered here.

17. Amongst all supermolecular methods, B97M-V/aTZ is the best DFT method with an overall MAE of 0.42 kcal/mol. MP2/CBS is the best wave function method for these ionic complexes, with an overall MAE of 0.34 kcal/mol.

CHAPTER 5

Symmetry-adapted perturbation theory with Kohn-Sham orbitals using non-empirically tuned, long-range-corrected density functionals[5.1]

5.1 Introduction

Symmetry-adapted perturbation theory[115–117,149,150] (SAPT) is a popular method to calculate dimer interaction energies and to decompose them into physically meaningful components: electrostatics, induction, dispersion, and their exchange counterparts. In this approach, the Hamiltonian is partitioned into monomer Fock operators, Møller-Plesset fluctuation operators (representing intramolecular electron correlation), and the intermolecular interaction operators. Unfortunately, high-order terms in the fluctuation potentials are required in order to achieve highly accurate interaction energies, which limits the application of SAPT to dimers composed of small monomer units, or to semi-quantitative results if intramolecular electron correlation is neglected. As such, there has been considerable interest in combining the wave-function-based SAPT formalism with a low-cost density functional theory

[5.1] This chapter appeared as a full article in the *Journal of Chemical Physics*, in 2014, volume 140, pages 044108:1–8.

(DFT) description of the monomers, which would therefore include intramolecular electron correlation.[149,150]

Such a hybrid approach was first tested in 2001,[208] simply by substituting Kohn-Sham (KS) orbitals and energies levels into the SAPT formalism, without further justification. (Some formal properties of this approach were considered a short time later.[209]) This approach, wherein the KS determinant is used as the reference state for an otherwise traditional SAPT calculation, is known as SAPT(KS), and its computational cost is essentially the same as the Hartree–Fock-based approach. (The traditional second-order approach with Hartree-Fock determinants is usually called SAPT0.[117]) Results obtained using SAPT(KS) were disappointing,[208,209] however, which was ultimately attributed to the incorrect asymptotic behavior of the exchange-correlation (XC) potential, $v_{\rm xc}$, in existing density-functional approximations.[162,208]

The proper asymptotic behavior of $v_{\rm xc}$ should be[210]

$$v_{\rm xc}(r) \sim -\frac{1}{r} + \Delta_\infty \tag{5.1}$$

for large r. The limiting value of $v_{\rm xc}(r)$ as $r \to \infty$ should be[210,211]

$$\Delta_\infty = {\rm IP} + \varepsilon_{\rm HOMO} \ , \tag{5.2}$$

where IP is the lowest ionization potential and $\varepsilon_{\rm HOMO}$ is the KS eigenvalue for the highest occupied molecular orbital (HOMO). It was later demonstrated that SAPT(KS) based on an asymptotically-corrected XC functional is able to predict the electrostatics, first-order exchange, second-order induction and exchange-induction energies with good accuracy,but second-order dispersion and exchange-dispersion energies remain quite poor.[162–164]

Accurate dispersion and exchange-dispersion energies were ultimately obtained by replacing the MP2-like sum-over-states dispersion formula ("uncoupled Hartree-Fock" approximation[212]) with a formula involving frequency-dependent density susceptibilities for the monomers obtained from time-dependent coupled Kohn-Sham (TD-CKS) calculations.[213,214] This method has variously been called DFT-SAPT or SAPT(DFT).[118,137,149,150,161,205] To address the issue of the long-range behavior of $v_{\rm xc}$, two asymptotic correction (AC) schemes have been employed: the Tozer-Handy splicing scheme in conjunction with the Fermi-Amaldi asymptotic potential,[210] and the gradient-regulated asymptotic correction with the van Leeuwen-Baerends asymptotic potential.[199,215] The drawback of these AC schemes is that the corrected XC potential, $v_{\rm xc}^{\rm AC}$, is not the functional derivative of the XC energy for any exchange-correlation functional:

$$v_{\rm xc}^{\rm AC} \neq \frac{\delta E_{\rm xc}}{\delta \rho} \; . \tag{5.3}$$

Exchange-correlation potentials that are not proper functional derivatives have been called "stray" potentials. Such potentials may generate spurious forces and torques, causing potentially serious problems with geometry optimizations.[216] Only recently has an AC model potential with $v_{\rm xc}^{\rm AC} = \delta E_{\rm xc}/\delta\rho$ been proposed,[217] but this functional has not yet been tested in the context of SAPT(KS). So far, the inconsistency in Eq. (5.3) does not seem to cause problems in the context of SAPT(KS) calculations,[218] but it is necessary to resolve this issue in order to derive analytic energy gradients for the extended SAPT (XSAPT) methods developed by our group,[44,54–56,65] which are designed for fast calculations of non-covalent clusters.

As an alternative to traditional AC model potentials, long-range corrected (LRC) density functionals, also known as range-separated hybrid functionals, can improve the asymptotic behavior of the XC potential.[219–224] These functionals partition the electron–electron Coulomb operator, r_{12}^{-1}, into short-range (SR) and long-range (LR) components using the error function (erf):

$$\frac{1}{r_{12}} = \underbrace{\frac{1-\mathrm{erf}(\omega r_{12})}{r_{12}}}_{\mathrm{SR}} + \underbrace{\frac{\mathrm{erf}(\omega r_{12})}{r_{12}}}_{\mathrm{LR}} \ . \tag{5.4}$$

Here, ω is an adjustable range-separation parameter that determines the length scale ($\sim \omega^{-1}$) of SR Coulomb potential, and is often determined by fitting to some data set.[224–227] For a generalized gradient approximation (GGA) of the form

$$E_{\mathrm{xc}} = E_{\mathrm{c}} + (1 - C_{\mathrm{HF}}) E_{\mathrm{x}}^{\mathrm{GGA}} + C_{\mathrm{HF}} E_{\mathrm{x}}^{\mathrm{HF}} \tag{5.5}$$

(which is technically a hybrid functional if the coefficient of Hartree-Fock exchange, C_{HF}, is different from zero), the corresponding LRC function is[228]

$$E_{\mathrm{xc}}^{\mathrm{LRC}} = E_{\mathrm{c}} + (1 - C_{\mathrm{HF}}) E_{\mathrm{x}}^{\mathrm{SR,GGA}} + C_{\mathrm{HF}} E_{\mathrm{x}}^{\mathrm{SR,HF}} + E_{\mathrm{x}}^{\mathrm{LR,HF}} \ . \tag{5.6}$$

When the electron–molecule distance is large, the functional in Eq. (5.6) is dominated by Hartree-Fock exchange, hence the asymptotic XC potential decays as $\sim r^{-1}$. As such, LRC functionals exhibit the correct asymptotic distance dependence, however they do not reproduce the proper limiting value [Eq. (5.1)], since $\Delta_{\infty} \neq 0$.

To correct this deficiency, Baer *et al.*[229,230] proposed a physically-motivated (non-empirical) "tuning" of ω, in order to satisfy the condition

$$\varepsilon_{\mathrm{HOMO}} = -\mathrm{IP} \ , \tag{5.7}$$

and thus ensure that $\Delta_\infty = 0$. When "tuned" in this way, LRC functionals predict both fundamental gaps and and excitation energies quite accurately.[231–233]

Recently, some conventional (statistically-optimized) LRC functionals have been used in the context of SAPT(DFT) calculations, with very poor results.[218] This failure is attributed to the incorrect asymptotic limit of v_{xc}, and the authors of Ref. 218 even go so far as to suggest that the name "long-range corrected functional" is inappropriate for functionals based on Eq. (5.6). Actually, we had already shown that LRC functionals with standard, statistically-optimized range separation parameters do not correct the dispersion problems in SAPT(KS) calculations, although they do afforde slightly better results (as compared to SAPT0) for strongly hydrogen-bonded systems.[54] In the present study, we will show that the aforementioned non-empirical tuning procedure, when applied in a monomer-specific way, affords SAPT(KS) energy components in good agreement with high-level benchmarks, except for this dispersion energy, which is still poor. The dispersion energy, however, can be accurately incorporated via empirical potentials, as we will demonstrate, yielding high-accuracy SAPT(KS)+D calculations where the exchange-correlation functional is well-defined [unlike the situation in Eq. (5.3)], and is thus appropriate for the formulation of analytic energy gradients.

5.2 Theory

The traditional second-order SAPT interaction energy (SAPT0 method[117]) can be written as

$$E_{\mathrm{int,resp}}^{\mathrm{SAPT0}} = E_{\mathrm{elst}}^{(10)} + E_{\mathrm{exch}}^{(10)} + E_{\mathrm{ind,resp}}^{(20)} + E_{\text{exch-ind,resp}}^{(20)} + E_{\mathrm{disp}}^{(20)} + E_{\text{exch-disp}}^{(20)} \; . \tag{5.8}$$

The notation $E^{(nl)}$ indicates a term that is nth order in the intermolecular interaction and lth order in the monomer fluctuation potentials, hence the $E^{(n0)}$ energy components in Eq. (5.8) imply that no intramolecular electron correlation is included in SAPT0. The "response" (resp) subscripts indicate that the infinite-order response correction for induction is incorporated by solving coupled perturbed Hartree-Fock equations.[179,180] In SAPT(KS), the intramolecular correlation is included implicitly, so the superscript l is dropped and one has

$$E_{\mathrm{int}}^{\mathrm{SAPT(KS)}} = E_{\mathrm{elst}}^{(1)}(\mathrm{KS}) + \mathrm{E}_{\mathrm{exch}}^{(1)}(\mathrm{KS}) + \mathrm{E}_{\mathrm{ind,resp}}^{(2)}(\mathrm{KS}) + E_{\text{exch-ind,resp}}^{(2)}(\mathrm{KS}) + \mathrm{E}_{\mathrm{disp}}^{(2)}(\mathrm{KS}) + E_{\text{exch-disp}}^{(2)}(\mathrm{KS}) \; . \tag{5.9}$$

The computational cost of either method is about the same.

In the case of traditional (Hartree–Fock-based) SAPT methods, the induction terms using "uncoupled" and "coupled" monomer densities are identical if the infinite-order intramolecular correlation is included:[234]

$$\sum_{l=0}^{\infty} E_{\mathrm{ind}}^{(2l)} = \sum_{l=0}^{\infty} E_{\mathrm{ind,resp}}^{(2l)} \; . \tag{5.10}$$

Since intramolecular correlation is implicitly included in SAPT(KS), the difference between SAPT(KS) induction energies based on coupled versus uncoupled monomer densities is expected to be smaller than in traditional wave function-based SAPT, and the uncoupled sum-over-states formula works fairy well for the induction energy.[162,164,208,235] Since orbital relaxation is present in the CKS static response theory but not in the uncoupled sum-over-states formula, the CKS method should in principle give better results for the induction energy as compared to an uncoupled calculation, at least at large intermolecular separation.[235] In this study, the orbital relaxation for the induction and exchange-induction energies in SAPT(KS) are approximated as the energy difference between coupled and uncoupled terms calculated by SAPT0:

$$E^{(2)}_{\text{ind,resp}}(\text{KS}) = E^{(2)}_{\text{ind}}(\text{KS}) + \text{E}^{(20)}_{\text{ind,resp}} - \text{E}^{(20)}_{\text{ind}} \tag{5.11a}$$

$$E^{(2)}_{\text{exch-ind,resp}}(\text{KS}) = E^{(2)}_{\text{exch-ind}}(\text{KS}) + \text{E}^{(20)}_{\text{exch-ind,resp}} - E^{(20)}_{\text{exch-ind}} \, . \tag{5.11b}$$

Thorough comparisons to benchmark calculations, as presented below, support the accuracy and robustness of this approximation.

When truncating the perturbation series at second order, as in Eqs. (5.8) and (5.9), it is common to incorporate higher-order polarization effects by means of a correction

$$\delta E^{\text{HF}}_{\text{int,resp}} = E^{\text{HF}}_{\text{int}} - (E^{(10)}_{\text{elst}} + E^{(10)}_{\text{exch}} + E^{(20)}_{\text{ind,resp}} + E^{(20)}_{\text{exch-ind,resp}}) \, , \tag{5.12}$$

where $E^{\text{HF}}_{\text{int}}$ denotes the supermolecular (dimer) Hartree-Fock interaction energy, with counterpoise correction. The $\delta E^{\text{HF}}_{\text{int,resp}}$ term is recommended when the monomers are

polar.[50,117,130,131,170]

5.3 Computational details

To determine the usefulness of SAPT(KS) calculations using tuned LRC functionals, we will use benchmark energy components determined from high-level calculations for a variety of small dimers. High-accuracy energy components for He_2 and Ne_2 are available,[218,236–238] obtained using highly correlated wave functions evaluated near the complete basis set (CBS) limit. For Ne_2, the best available calculations[218] were obtained using the SAPT(CCSD) method.[154–156,239] For He_2, *exact* energy components, obtained with a complete account of electron correlation (equivalent to full configuration interaction), and with a basis set that is saturated using a Gaussian-type geminal (GTG), are available;[236–238] this method has been called SAPT(GTG). For the S22 data set,[186] benchmark energy components computed at the SAPT2+(3)/aug-cc-pVTZ level are taken from Ref. 37. [The SAPT2+(3) method includes terms beyond second order; see Ref. 117.] Total binding energies (though not the individual energy components), computed at the CCSD(T)/CBS level, are available for both the S22 and S66 data sets.[34,35] In addition, we will use the CCSD(T)/CBS potential energy curves for the "sandwich" (π-stacked) isomer of the benzene dimer, from Ref. 5.

In addition to examining the popular S22 and S66 data sets,[10,186] we have assembled a new set of benchmarks here. This data set, which we designate as SS41, consists of 41 small systems (dimers) taken from some existing data sets. Namely, we take the

24 dimers from the A24 data set;[33] the formamide and formic acid dimers from S22;[186] nine dimers from the S66 data set[10] (specifically, $H_2O\cdots CH_3OH$, $CH_3OH\cdots CH_3OH$, $CH_3OH\cdots CH_3NH_2$, $CH_3OH\cdots H_2O$, $CH_3NH_2\cdots CH_3OH$, $CH_3NH_2\cdots CH_3NH_2$, $C_2H_2\cdots H_2O$, and two isomers of $H_2O\cdots CH_3NH_2$); and finally, six dimers from the X40 data set[36] (specifically, $CH_4\cdots F_2$, $CH_3F\cdots CH_4$, $CH_3F\cdots CH_3F$, $HF\cdots CH_3OH$, $HF\cdots CH_3NH_2$, and $CH_3OH\cdots CH_3F$). Benchmark energy components for the SS41 dimers are reported here for the first time; these were computed at the SAPT2+(3)/aug-cc-pVQZ level using MP2 natural orbital approximations to accelerate the calculations, as described in Ref. 159. These calculations were performed using the PSI4 program,[187] and the benchmark energy components for SS41 are available in the Supplementary Material.

As in our previous work on XSAPT,[55,56] we will use the LRC-ωPBE[240] and LRC-ωPBEh[227] functionals for SAPT(KS) calculations. These functionals are based on the short-range ωPBE exchange functional,[222] augmented with 100% long-range Hartree-Fock exchange as in Eq. (5.6). The LRC-ωPBEh functional also contains 20% short-range Hartree-Fock exchange ($C_{\rm HF} = 0.2$), whereas LRC-ωPBE does not ($C_{\rm HF} = 0$). Statistically-optimized values of the range separation parameter have been suggested as either $\omega = 0.3$ bohr^{-1} or $\omega = 0.4$ bohr^{-1} for for LRC-ωPBE,[240,241] and $\omega = 0.2$ bohr^{-1} for LRC-ωPBEh.[227] A previous study of LRC functionals for SAPT(DFT) calculations used LRC-ωPBE with $\omega = 0.4$ bohr^{-1}.[218] In this work, however, the value of ω is determined by monomer-specific tuning to satisfy Eq. (5.7). [Monomer-specific AC model potentials have previously been used in the context of SAPT(DFT)

and DFT-SAPT,[137,161] but monomer-specific LRC functionals have not been used in this context.] Tuned values of ω, for each of the monomer appearing in the dimers examined herein, are available in the Supplementary Material.

We use the aug-cc-pV6Z and aug-cc-pV5Z basis sets for the SAPT(KS) calculations on He_2 and Ne_2, respectively, in order to obtain results near the CBS limit. For SAPT calculations on the SS41 data set, we use the aug-cc-pVQZ basis set. The aug-cc-pVTZ basis set was used for S22, S66, and the benzene dimer potential curve. Except for He_2 and Ne_2, all of the total binding energies computed with SAPT methods include the $\delta E^{\mathrm{HF}}_{\mathrm{int,resp}}$ correction. All SAPT calculations employ the a dimer-centered basis set,[153] meaning that monomer wave functions were converged using the dimer's basis set. With the exception of the two rare-gas dimers, we used the resolution-of-identity approximation (combined with standard auxiliary basis sets) to accelerate the wave function-based SAPT calculations, performed using PSI4.[187] SAPT(KS) calculations with the resolution-of-identity approximation were performed using a locally-modified version of Q-Chem.[242,243]

We will also report "SAPT(KS)+D" calculations that use an empirical dispersion potential in place of the $E^{(2)}_{\mathrm{disp}} + E^{(2)}_{\mathrm{exch\text{-}disp}}$ terms in second-order SAPT. This "+D" potential is taken from Ref. 244 and has been shown to afford good results for a wide variety of different systems.[56,244]

Table 5.1: Interaction energy components for He_2 at its equilibrium distance (5.6 bohr), calculated using the dimer-centered aug-cc-pV6Z basis. Values in parentheses are the percentage errors with respect to the GTG benchmarks, whereas other values are actual energies in units of cm^{-1}.

Method	$E^{(1)}_{\text{elst}}$	$E^{(1)}_{\text{exch}}$	$E^{(2)}_{\text{ind}}$	$E^{(2)}_{\text{exch-ind}}$	$E^{(2)}_{\text{disp}}$	$E^{(2)}_{\text{exch-disp}}$	Total
Exact[a]							−7.651
CCSD(T)/CBS[b]							−7.420
SAPT(GTG) benchmark[c]	−1.187	8.540	−0.196	0.177	−15.565	0.515	−7.716
LRC-ωPBE[d] ($\omega = 0.3$ bohr^{-1})	(114.18)	(131.54)	(154.71)	(149.10)	(11.38)	(41.46)	0.565
LRC-ωPBE[d] (tuning ω)	(1.67)	(1.34)	(3.44)	(11.95)	(21.33)	(30.70)	−4.664
LRC-ωPBE+D[d,e] (tuning ω)							−7.332
LRC-ωPBEh[d] ($\omega = 0.2$ bohr^{-1})	(70.17)	(94.22)	(112.15)	(107.25)	(5.89)	(23.49)	−1.435
LRC-ωPBEh[d] (tuning ω)	(0.13)	(0.30)	(1.64)	(10.30)	(20.83)	(29.74)	−4.615
LRC-ωPBEh+D[d,e] (tuning ω)							−7.211

[a] From Ref. 245.

[b] From Ref. 218.

[c] From Refs. 236–238.

[d] SAPT(KS) using the indicated density functional.

[e] SAPT(KS) using the "+D" dispersion correction from Ref. 244.

Table 5.2: Interaction energy components for Ne_2 at its equilibrium distance (3.1 Å), calculated using the dimer-centered aug-cc-pV5Z basis. Values in parentheses are the percentage errors with respect to the SAPT(CCSD) benchmarks, whereas other values are actual energies in units of cm^{-1}.

Method	$E^{(1)}_{\text{elst}}$	$E^{(1)}_{\text{exch}}$	$E^{(2)}_{\text{ind}}$	$E^{(2)}_{\text{exch-ind}}$	$E^{(2)}_{\text{disp}}$	$E^{(2)}_{\text{exch-disp}}$	Total
CCSD(T)/CBS[a]							−28.653
SAPT(CCSD)/CBS benchmark[a]	−8.997	36.317	−7.533	7.752	−60.318	3.187	−29.592
LRC-ωPBE[b] ($\omega = 0.3$ bohr^{-1})	(89.82)	(89.73)	(120.48)	(120.26)	(56.36)	(45.84)	−37.370
LRC-ωPBE[b] (tuning ω)	(4.46)	(1.63)	(9.08)	(9.84)	(30.83)	(26.61)	−48.125
LRC-ωPBE+D[b,c] (tuning ω)							−23.274
LRC-ωPBEh[b] ($\omega = 0.2$ bohr^{-1})	(55.14)	(55.23)	(77.93)	(77.68)	(54.96)	(22.88)	−46.763
LRC-ωPBEh[b] (tuning ω)	(1.99)	(3.62)	(5.45)	(6.20)	(32.00)	(24.66)	−48.256
LRC-ωPBEh+D[b,c] (tuning ω)							−22.762

[a] From Ref. 218.

[b] SAPT(KS) using the indicated density functional.

[c] SAPT(KS) using the "+D" dispersion correction from Ref. 244.

5.4 Results and discussion

He_2 and Ne_2 are popular test systems for SAPT-based methods, as they are small enough to use high-level wave function-based SAPT methods near the CBS limit. SAPT(KS) results for these two dimers, using LRC functionals, are listed in Table 5.1 for He_2 and Table 5.2 for Ne_2. Conventional (statistically-optimized) LRC functionals, by which we mean LRC-ωPBE with $\omega = 0.3$ bohr^{-1} and LRC-ωPBEh with $\omega = 0.2$ bohr^{-1}, afford large errors for all of the energy components. For example, the error in $E^{(2)}_{\rm ind}$ using LRC-ωPBE (with $\omega = 0.3$ bohr^{-1}) is about 150%. The tuning strategy, on the other hand, reduces errors in $E^{(1)}_{\rm elst}$, $E^{(1)}_{\rm exch}$, $E^{(2)}_{\rm ind}$, and $E^{(2)}_{\rm exch\text{-}ind}$ to just a few percent. In the case of the dispersion interaction, the tuning strategy improves the results for Ne_2 but has deleterious effects for He_2. Previous studies have found that SAPT(KS), using the standard second-order, uncoupled dispersion formula, afford poor dispersion energies even if the AC functionals are employed.[213,214] This observation is consistent with the results presented here.

SAPT(KS) dispersion energies can be greatly improved by computing frequency-dependent density susceptibilities for the monomers and then evaluating the dispersion energy using a generalized Casimir-Polder formalism.[161,213,214] The cost of such a calculation, however, scales no better than $\mathcal{O}(N^5)$ with respect to monomer size.[246] Alternatively, one can use an empirical dispersion potential designed for SAPT to obtain the dispersion energy,[6,16,162,244] Here, we report dispersion-corrected SAPT(KS)+D results using the empirical dispersion potential from Ref. 244 in conjunction with tuned LRC functionals.

The $E^{(1)}_{\text{elst}}$, $E^{(1)}_{\text{exch}}$, $E^{(2)}_{\text{ind}}$, and $E^{(2)}_{\text{exch-ind}}$ terms calculated by SAPT(KS) with AC XC potentials, plus an empirical dispersion potential, constitutes an approach that has been called SAPT(KS)+D.[16,162,244] Here, we apply this method not with AC model potentials but rather with LRC functionals. For He_2, SAPT(KS)+D gives very good results for the total binding energy, although errors are slightly larger for Ne_2. Because the SAPT(KS) method affords good results for the energy components $E^{(1)}_{\text{elst}}$, $E^{(1)}_{\text{exch}}$, $E^{(2)}_{\text{ind}}$, and $E^{(2)}_{\text{exch-ind}}$ in both rare-gas dimers, it seems that the problem lies with the empirical dispersion potential. The dispersion energy in SAPT calculations converges very slowly as a function of the one-particle basis set,[24] thus the parameters in the "+D" potential of Ref. 244 are fit to dispersion energies for a training set of dimers computed at the SAPT(DFT) level using the aug-cc-pVTZ basis set with additional mid-bond functions. It is possible that this basis is insufficient to afford converged dispersion energies for the Ne_2 system. We note that SAPT(KS)+D results of similar accuracy to those reported here can be obtained with AC model potentials,[16,162,235] but at the cost of sacrificing the relationship between v_{xc} and E_{xc}.

Turning now to larger systems, in Table 5.3 we present the mean unsigned errors (MUEs) and mean unsigned percentage errors (MUPEs) for SAPT(KS) and SAPT0 calculations of the SS41 data set, as compared to benchmark energy components computed at the SAPT2+(3)/aug-cc-pVQZ level. Once again, SAPT(KS) with tuned LRC functionals gives very good results for the non-dispersion energy components and is much better than the SAPT0 results where the intramolecular electron

Table 5.3: Mean unsigned errors (MUEs), in kcal/mol, and percent errors (in parentheses), for interaction energy components of SS41 data set, with respect to benchmarks computed at the SAPT2+(3)/aug-cc-pVQZ level. The errors in total binding energies are with respect to CCSD(T)/CBS results.[33–36] The $\delta E^{\mathrm{HF}}_{\mathrm{int,resp}}$ correction is added to the total binding energies of the SAPT calculations.

Method	$E^{(1)}_{\mathrm{elst}}$	$E^{(1)}_{\mathrm{exch}}$	$E^{(2)}_{\mathrm{ind}}$	$E^{(2)}_{\mathrm{exch\text{-}ind}}$	$E^{(2)}_{\mathrm{disp}}$	$E^{(2)}_{\mathrm{exch\text{-}disp}}$	Total
SAPT2+(3)[a]							0.14
							(4.97)
LRC-ωPBE[a,b]	0.12	0.20	0.09	0.14	0.23	0.09	0.22
	(4.37)	(4.33)	(3.59)	(7.25)	(6.65)	(15.93)	(10.93)
LRC-ωPBEh[a,b]	0.08	0.16	0.06	0.10	0.22	0.08	0.34
	(4.44)	(3.27)	(2.52)	(5.50)	(6.73)	(14.41)	(13.29)
SAPT0[a]	0.26	0.70	0.33	0.19	0.24	0.00[c]	0.69
	(9.04)	(9.09)	(9.56)	(9.56)	(7.37)	(0.00)[c]	(13.53)

[a] aug-cc-pVQZ basis set.

[b] SAPT(KS) results using the indicated density functional and tuning ω.

[c] SAPT0 and SAPT2+(3) share the same formula for $E^{(2)}_{\mathrm{exch\text{-}disp}}$.

correlation is completely neglected. At the SAPT0 level, errors in total binding energies range up to 5.4 kcal/mol, with the largest (absolute) error obtained for the doubly hydrogen-bonded formic acid dimer, a system where intramolecular electron correlation is known to have a large effect on the binding energy.[158]

Surprisingly, SAPT(KS) also gives quite good results for dispersion components and total binding energies for the SS41 data set. As proposed by Hobza and coworkers,[7] a complex should be considered to be dominated by electrostatics if the electrostatic component ($E^{(1)}_{\mathrm{elst}}$) is at least twice as large as the dispersion component ($E^{(2)}_{\mathrm{disp}} + E^{(2)}_{\mathrm{exch\text{-}disp}}$), and *vice versa* for a dispersion-dominated complex. Otherwise,

Table 5.4: Mean unsigned errors (MUEs), in kcal/mol, and percent errors (in parentheses), for individual energy components of the S22 data set, with respect to benchmarks computed at the SAPT2+(3)/aug-cc-pVTZ level.[37] All calculations were performed at S22 geometries.

Method	Energy Components							
	electrostatic		exchange		induction		dispersion	
LRC-ωPBE[a,b]	0.19	(3.05)	0.59	(5.65)	0.13	(3.65)	1.04	(13.90)
LRC-ωPBEh[a,b]	0.17	(3.46)	0.57	(5.64)	0.13	(3.28)	1.05	(13.97)
SAPT0[b]	0.42	(6.47)	1.25	(8.05)	0.27	(4.49)	0.72	(9.61)

[a] SAPT(KS) results using the indicated density functional and tuning ω.
[b] aug-cc-pVTZ basis set.

the complex is classified as having interactions of mixed type. According to this classification scheme, the SS41 data set contains 16 electrostatically-dominated complexes, 12 dispersion-dominated complexes, and 13 mixed-type complexes. The three largest errors (each about 20%) in SAPT(KS) dispersion components occur for the dispersion-dominated complexes, but these 20% errors only translate into ≈ 0.5 kcal/mol errors since the dispersion component is no larger than -2.4 kcal/mol [at the SAPT2+(3)/aug-cc-pVQZ level], even for the dispersion-dominated complexes. Larger molecules with more electrons are needed in order to obtain larger dispersion energies.

The S22 and S66 data sets[10,186] are popular for benchmarking non-covalent interactions. [For the latest, basis-set-consistent revisions to the CCSD(T)/CBS binding energies, see Ref. 34 for S22 and Ref. 35 for S66.] These data sets contain larger molecules, as compared to SS41; the π-stacked adenine–thymine dimer, for example, has a dispersion interaction of about 18 kcal/mol. For S22, we take benchmark energy

components from Ref. 37, where they were computed at the SAPT2+(3)/aug-cc-pVTZ level.[37]. These comparisons are listed in Table 5.4, where the electrostatic, exchange, induction, and dispersion energies are defined according to[37,117]

$$E_{\text{electrostatic}} = E^{(1)}_{\text{elst}} \tag{5.13a}$$

$$E_{\text{exchange}} = E^{(1)}_{\text{exch}} \tag{5.13b}$$

$$E_{\text{induction}} = E^{(2)}_{\text{ind,resp}} + E^{(2)}_{\text{exch-ind,resp}} + \delta E^{\text{HF}}_{\text{int,resp}} \tag{5.13c}$$

$$E_{\text{dispersion}} = E^{(2)}_{\text{disp}} + E^{(2)}_{\text{exch-disp}} \ . \tag{5.13d}$$

The SAPT(KS) and SAPT0 methods afford similar MUEs for electrostatic and induction energies, but SAPT(KS) yields smaller errors in exchange energies than SAPT0. This suggests that intramolecular electron correlation, at least for exchange energy, is important for the molecules in S22. SAPT(KS) results based on LRC-ωPBE are slightly better than those based on LRC-ωPBEh, in accordance with previous observations.[56] For the dispersion energy, on the other hand, SAPT(KS) calculations afford larger errors as compared to SAPT0, again in accordance with previous observations that the use of KS orbitals and eigenvalues is detrimental to the quality of the MP2-like sum-over-states dispersion energy.[54] The error in the dispersion energy for π-stacked adenine–thymine is about 20% or 3 kcal/mol.

Table 5.5 shows the MUEs for total binding energies with respect to CCSD(T)/CBS results for the S22 data set[34] and the S66 data set.[35] Both sets are divided to three subsets consisting of the hydrogen-bonded complexes (which are dominated by electrostatics), the dispersion-dominated complexes, and complexes with mixed influence, according to Hobza's classification scheme,[7] as discussed above. The main

Table 5.5: Mean unsigned errors (MUEs), in kcal/mol, and percent errors (in parentheses), with respect to CCSD(T)/CBS benchmarks for the S22 data set[34] and the S66 data set[35] along with subsets consisting of the hydrogen-bonded dimers, dispersion-dominated dimers, and dimers of mixed influence. All calculations were performed at S22 and S66 geometries and the $\delta E^{\mathrm{HF}}_{\mathrm{int,resp}}$ corrections are added to the total binding energies of the SAPT calculations.

Method	H-bonded		Disp.-bound		Mixed		All	
			—S22—					
LRC-ωPBE[a,b]	0.61	(3.84)	2.60	(54.74)	0.92	(25.09)	1.43	(29.11)
LRC-ωPBE+D[a,b,c]	0.48	(3.24)	1.33	(19.39)	0.56	(13.00)	0.82	(12.22)
LRC-ωPBEh[a,b]	1.22	(7.00)	2.77	(58.69)	1.13	(30.41)	1.75	(33.25)
LRC-ωPBEh+D[a,b,c]	0.43	(2.99)	1.53	(25.60)	0.28	(7.96)	0.78	(12.79)
SAPT0[b]	2.72	(16.91)	2.00	(41.17)	1.01	(26.96)	1.91	(28.93)
SAPT0+D[b,c]	3.20	(21.08)	1.39	(21.50)	0.95	(23.41)	1.83	(21.98)
SAPT2+(3)[b]	0.51	(3.62)	0.38	(4.86)	0.12	(3.05)	0.34	(3.89)
			—S66—					
LRC-ωPBE[a,b]	0.22	(2.18)	1.52	(43.99)	0.82	(22.99)	0.85	(23.05)
LRC-ωPBE+D[a,b,c]	0.32	(3.61)	0.63	(14.40)	0.35	(9.15)	0.44	(9.05)
LRC-ωPBEh[a,b]	0.57	(5.56)	1.66	(48.10)	1.02	(28.52)	1.09	(27.34)
LRC-ωPBEh+D[a,b,c]	0.21	(2.19)	0.69	(15.54)	0.45	(12.22)	0.45	(9.88)
SAPT0[b]	1.52	(14.55)	1.09	(30.34)	0.89	(24.90)	1.18	(23.19)
SAPT0+D[b,c]	1.89	(19.25)	0.90	(21.36)	0.83	(22.69)	1.22	(21.03)

[a] SAPT(KS) results using the indicated density functional and tuning ω.

[b] aug-cc-pVTZ basis set.

[c] Using the empirical dispersion potential from Ref. 244.

sources of error for the SAPT(KS) calculations are in the dispersion-dominated complexes, where the MUEs are 2.6 kcal/mol (S22) and 1.5 kcal/mol (S66). Substituting the empirical dispersion potential developed by Podeszwa *et al.*[244] in place of the SAPT(KS) dispersion energy, to obtain a SAPT(KS)+D method based on LRC functionals, the MUEs in total binding energies are reduced to 0.8 kcal/mol (S22) and 0.4 kcal/mol (S66).

It has previously been pointed out that the empirical dispersion potential developed in Ref. 244 and used here affords relatively large errors for the π-stacked uracil dimer and the π-stacked adenine–thymine complexes.[56] In fact, π-stacked complexes are underrepresented in the training set used to parameterize this potential; only the sandwich isomer of $(C_6H_6)_2$ and the pyrazine dimer are included as examples of π-stacking amongst the 79 dimers in the training set.[244] For the uracil dimer, this dispersion potential overestimates the SAPT2+(3)/aug-cc-pVTZ dispersion energy by 1.8 kcal/mol, and for adenine–thymine by 2.0 kcal/mol. If we eliminate these two problematic systems from the S22 data set, then the MUE for SAPT(KS)+D using both density functionals examined here is reduced to about 0.5 kcal/mol for both the dispersion-dominated subset and the entire set of 20 complexes (S22 minus two). Similarly for S66, the MUE for SAPT(KS)+D using both functionals is reduced to 0.4 kcal/mol when the π-stacked uracil dimer is removed from the data set.

For the SAPT0, however, the use of the dispersion potential from Ref. 244 does not obviously improve the results for either S22 or S66. Furthermore, the results for hydrogen-bonded complexes described at the SAPT0+D level are worse than

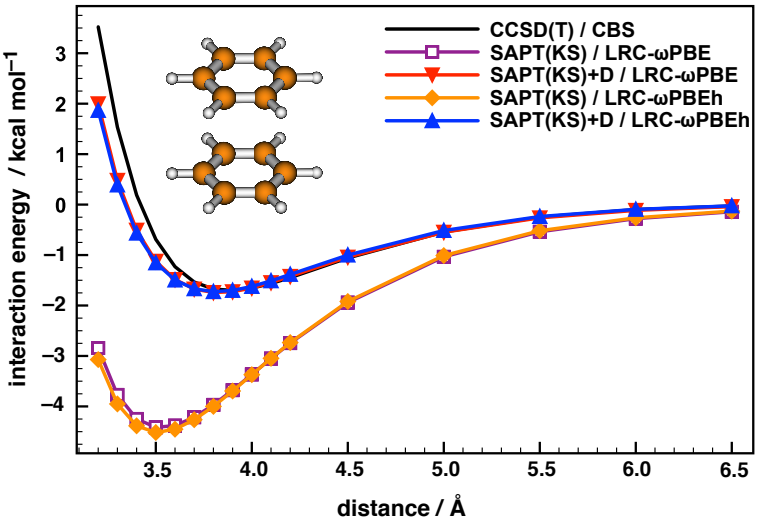

Figure 5.1: Potential energy curves for the "sandwich" isomer of $(C_6H_6)_2$ as a function of the center-to-center distance between the two benzene rings. Benchmark CCSD(T)/CBS results are taken from Ref. 5. The dimer-centered aug-cc-pVTZ basis set was used for the SAPT(KS) calculations

the SAPT0 results that use the MP2-type dispersion formula. This implies that dispersion is not the only term in SAPT0 that needs improvement: intramolecular electron correlation is important as well.

Finally, we use these SAPT(KS) methods to compute the potential energy curve for the dispersion-dominated sandwich isomer of the benzene dimer, which is regarded as a stringent test of computational methods. As shown in Fig. 5.1, SAPT(KS) methods that use second-order MP2-type dispersion greatly overestimate the interaction energy across the whole potential energy curve. The SAPT(KS)+D methods slightly underestimates the binding energy at short intermolecular distance, but are very accurate beyond the minimum-energy distance.

5.5 Conclusions

The increasingly-popular non-empirical tuning procedure of Baer *et al.*[230] for LRC (range-separated hybrid) functionals affords an XC potential with not only the correct asymptotic distance dependence, but also the correct limiting value as $r \to \infty$. Results presented herein demonstrate that satisfaction of this condition is very important to obtain quantitative results for the non-dispersion energy components in SAPT(KS) calculations. Although unacceptably large errors remain in the dispersion energies predicted by SAPT(KS), we expect that dispersion energies obtained from frequency-dependent density susceptibilities, computed by solving TD-CKS equations as in the SAPT(DFT) approach,[118,137,161,205] should yield quantitative results for the dispersion energies. Thus, LRC functionals should be useful as substitutes for AC model potentials in SAPT(DFT) and DFT-SAPT, while preserving the relationship $v_{\rm xc} = \delta E_{\rm xc}/\delta\rho$ that is sacrificed when AC model potentials are employed to "graft on" correct asymptotic behavior to some existing density functional. This relationship between functional and potential is crucial where analytic gradients are needed, which is a direction that our group is headed with "XSAPT" methods.[44,54–56,65]

As an alternative to SAPT(DFT), in the present work we used an empirical dispersion potential[244] to replace the second-order dispersion energy in SAPT(KS). The resulting SAPT(KS)+D method affords quantitative binding energies for the S22 and S66 data sets, save for a couple of π-stacked complexes, for which we have previously argued[56] that re-parameterization of the dispersion potential is needed. Although "conventional" LRC functionals that employ a statistically-optimized value for the

range separation parameter, which need not provide the correct limiting value of v_{xc}, have been shown to afford poor results in SAPT(DFT) calculations,[218] the tuning procedure used here affords good results for all energy components except dispersion, as compared to high-level SAPT2+(3) calculations.

CHAPTER 6

Accurate intermolecular interactions at dramatically reduced cost: XPol+SAPT with empirical dispersion[6.1]

Dispersion interactions are very important in biological systems, *e.g.*, in protein folding and in the structure of DNA.[117] In the latter case, the double helix is maintained both by dispersion-dominated π-stacking interactions within a strand and by hydrogen bonding between complementary strands. Although the H-bonding interactions are *dominated* by electrostatic effects, dispersion is still responsible for 20–30% of the base-pairing interaction,[247] and contrary to popular belief, experiments demonstrate that the stability of the double helix is mainly determined by base stacking, rather than base pairing.[248] Thus, in the theoretical description of non-covalent interactions it is crucial to employ methods that furnish an accurate description of dispersion interactions.

The "gold standard" of electronic structure theory, CCSD(T), is accurate enough for this purpose but exhibits a cost that grows as $\mathscr{O}(N^7)$ with respect to system size, N. This fact, in conjunction with the large basis sets that are required for

[6.1]This chapter appeared as a full article in the *Journal of Physical Chemistry Letters*, in 2012, volume 3, pages 3241–3248.

accurate calculations of intermolecular interaction energies [*e.g.*, to eliminate basis-set superposition error (BSSE)] limits high-level *ab initio* calculations to small systems. Furthermore, even given accurate CCSD(T) results for the intermolecular interaction energy, it is not easy to ascertain how much of this interaction is due to dispersion versus other effects such as exchange, electrostatics, or induction.

Symmetry-adapted perturbation theory (SAPT) is an alternative method to compute the intermolecular interaction energies,[116,117,149] using a monomer-based formalism rather than a supersystem calculation. As such, the SAPT interaction energy is free of BSSE, by construction, and is also decomposable into a sum of physically-meaningful contributions:[117]

$$\begin{aligned} E_{\text{int}}^{\text{SAPT}} = E_{\text{elst}}^{(1)} + E_{\text{exch}}^{(1)} + E_{\text{ind}}^{(2)} + E_{\text{exch-ind}}^{(2)} \\ + E_{\text{disp}}^{(2)} + E_{\text{exch-disp}}^{(2)} + \cdots \ . \end{aligned} \tag{6.1}$$

Here, we have explicitly listed all terms up to second order in the intermolecular interaction, with subscripts that denote electrostatic (elst), exchange (exch), induction (ind), and dispersion (disp) contributions. These low-order terms neglect intramolecular electron correlation, and for high-accuracy SAPT calculations it is essential to use a double-perturbation expansion that accounts for both inter- and intramolecular correlation.[117] Inclusion of intramolecular electron correlation, however, results in a SAPT method whose cost scales as $\mathscr{O}(N^7)$, so there is no cost savings over CCSD(T), although the SAPT result can still be used to decompose the interaction energy.

In principle, Kohn-Sham density functional theory (KS-DFT) offers a low-cost means to describe intramolecular electron correlation, and has been introduced in

the context of SAPT calculations as the so-called SAPT(KS) method.[162,208] This approach is disastrously unsuccessful, however, unless functionals with asymptotically-corrected exchange-correlation potentials are employed, but even so, the dispersion energies computed with SAPT(KS) remain quite poor.[54,162–164,208] The solution to this dilemma is to replace the MP2-like sum-over-states dispersion formula used in traditional low-order SAPT with a formula involving frequency-dependent density susceptibilities, computed using KS-DFT. This modified method has variously been called DFT-SAPT[137] or SAPT(DFT).[161] With density-fitting techniques, this method exhibits $\mathscr{O}(N^5)$ scaling yet provides an accurate description of intermolecular interactions.[137,246]

SAPT methods have mainly been used to study dimers, because calculation of non-additive three-body interactions within the SAPT formalism requires computationally expensive triple excitations.[249,250] Many-body (non-pairwise-additive) effects are large in clusters of polar molecules but are dominated by induction (*i.e.*, polarization).[8,64,84,251] In recognition of this, our group has recently introduced a low-cost, many-body generalization of the SAPT methodology that we call XPol+SAPT (XPS),[54,65] in which the variational XPol method[252] is used to generate one-body wave functions for subsequent pairwise SAPT calculations. The XPol procedure captures many-body polarization effects by means of a charge-embedded, monomer-based self-consistent field calculation whose cost scales as $\mathscr{O}(n)$ with respect to the number of monomers, n. The subsequent second-order SAPT calculations scale as either $\mathscr{O}(n^2)$ or $\mathscr{O}(n^3)$, depending on the level of approximation,[54] but in any case these

are embarrassingly parallelizable. With a suitable choice of basis set, dimer binding energies computed by XPS based on a Hartree-Fock description of the monomers [XPS(HF)] lie within 1 kcal/mol of high-level benchmarks.[54,65]

As with SAPT(KS), XPS results obtained using KS orbitals are notably inferior to those obtained using HF orbitals. In particular, dispersion energies are vastly overestimated by XPS(KS),[54] which is an artifact of the sum-over-states dispersion formula in conjunction with HOMO/LUMO gaps that are significantly smaller than HF gaps.[208] Ironically, the dispersion and exchange-dispersion terms are not only the least accurate but also the most time-consuming to compute, scaling as the fourth and fifth powers, respectively, of monomer size, whereas other second-order terms are no worse than $\mathscr{O}(N^3)$.[16] Recently, Hesselmann[16] introduced a method termed SAPT+D, in which these terms are replaced by empirical atom-atom potentials. In this work, we implement and test an analogous XPS(KS)+D method, which offers important advantages over alternative electronic structure methods for non-covalent interactions:

1. Unlike XPS(HF), it incorporates intramolecular correlation, and in a relatively low-cost way.

2. The expensive and inaccurate sum-over states dispersion formulas are replaced by simple scalar potentials.

3. Unlike SAPT+D, the method is applicable to any number of monomers.

4. The cost of XPS(KS)+D calculations is $\mathscr{O}(N^3)$ with respect to monomer size and no worse than $\mathscr{O}(n^3)$ with respect to the number of monomers, and is thus amenable to large systems.

Several asymptotic correction (AC) schemes for the KS exchange-correlation potential, v_{xc}, have been used in SAPT(KS) and SAPT(DFT),[199,210] although a drawback of these techniques is that the corrected potential, v_{xc}^{AC}, is *not* the functional derivative of the energy. As an alternative to traditional AC schemes, we utilize long-range corrected (LRC) density functionals[219,228,230] to obtain the correct asymptotic behavior. Specifically, we employ an *ansatz* that we have called LRC-ωPBEh,[227] which is based on the short-range ωPBE exchange functional.[222] Rather than using the empirically-optimized LRC-ωPBEh parameters suggested in Ref. 227, however, we re-optimize the fraction of short-range HF exchange (C_{HF}) and then apply a system-specific tuning of the range separation parameter (ω), as suggested by Baer *et al.*,[230] in order to satisfy the condition

$$\varepsilon_{\mathrm{HOMO}} = -\mathrm{IP} \tag{6.2}$$

where "IP" denotes the lowest ionization potential. For supersystems composed of well-defined monomers, as considered here, we assume that non-covalent interactions do not greatly alter the monomer IPs, hence we just need to determine ω separately for each monomer, using Eq. (6.2). The value of ω appropriate for the supersystem is the one corresponding to the lowest monomer IP. Our tests reveal that this assumption is quite robust.

The empirical dispersion potential used in this work is the one suggested for SAPT+D:[16]

$$E_{\rm disp}^{\rm SAPT+D} = -s_\beta \sum_{i\in A} \sum_{\substack{j\in B \\ (B\neq A)}} f_{\rm damp}(r_{ij}) \frac{C_{ij}}{r_{ij}^{\beta}} \; . \tag{6.3}$$

Here, i and j represent nuclei located on different monomers, and

$$f_{\rm damp}(r_{ij}) = {\rm erf}\left(\frac{\alpha \, r_{ij}}{R_i + R_j} \right) \tag{6.4}$$

is a damping function. The latter differs from the damping function typically used in dispersion-corrected DFT.[143] Note that the role of the dispersion correction is different in SAPT+D than it is in DFT+D.[16] In the latter method, the dispersion correction should only turn on at large intermolecular separation, since DFT models the short-range interactions; as such, $E_{\rm disp}^{\rm DFT+D}$ "is a model-dependent quantity with no real physical meaning",[253] and short-range damping is needed to avoid overcounting of interactions. In SAPT, however, the dispersion contribution to the energy is well-defined and should contribute at all intermolecular distances. However, short-range damping is still required to ensure that the empirical potential in Eq. (6.3) is finite when r_{ij} is small. The damping in Eq. (6.4) is much slower than that used in DFT+D.[16]

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

$$C_{ij} = (C_{6,i} \, C_{6,j})^{1/2} \; . \tag{6.5}$$

The atomic parameters $C_{6,i}$ and R_i used here are taken from those developed by Grimme.[254] The parameters $\alpha = 1.087$ and $\beta = 5.67$ are taken from SAPT+D,[16]

where they were fitted to reproduce benchmark intermolecular interaction energies. (The fact that dispersion is well-defined in SAPT suggests that these parameters may be transferable; results presented below will ultimately validate this choice.) Ultimately, β is an empirical parameter, but the fact that $\beta \neq 6$ can be understood based on the observation that higher-order multipole terms may contribute as much as 20–30% to the dispersion interaction in the middle-range region.[255] Moreover, $E^{(2)}_{\text{exch-disp}}$ also contributes to the SAPT dispersion energy, and this component varies exponentially with distance.

In this work, the parameter s_β in Eq. (6.3) was optimized for XPS(KS)+D using the S22A database,[38] which revises the energetics of the original S22 set of dimers.[186] For fixed α and β, the fit for s_β is a simple linear one that minimizes the absolute percent deviations, as this provides a more balanced fit for both weakly- and strongly-bound systems as compared to a least-squares fit.[16] It is possible that a global, nonlinear fit of all three parameters might improve the results, but this has not been attempted.

Optimal values of s_β were obtained for two different basis sets: aug-cc-pVDZ′ and def2-TZVPP (hereafter abbreviated aDZ′ and TZVPP).[117,197] The aDZ′ basis has been suggested for use in low-order SAPT0 and XPS(HF) calculations of dispersion-bound systems.[65,117,157,158] For XPS(HF) calculations on ion–water complexes, however, much better results were obtained using def2-TZVP, which exhibits only slightly worse error statistics for S22A.[54] Since then, we have found that TZVPP—which adds polarization functions to the hydrogen atoms—outperforms TZVP for S22A. As in

previous work,[54,65] we use smooth ChElPG embedding charges[54] for the XPol calculations and "projected" (pseudocanonical dimer) basis sets for the SAPT parts of the XPS calculations, along with a resolution-of-identity approximation combined with standard auxiliary basis sets. All calculations were performed using a locally modified version of the Q-Chem program.[242]

The optimized values of s_β are 0.7267 and 0.8439 for aDZ′ and TZVPP, respectively. For these values of s_β, the optimal value of $C_{\rm HF} = 0.6$. Selected error statistics for XPS-based methods, as applied to the S22A database, are listed in 6.1. The overall mean unsigned error (MUE) for XPS(KS) is larger than for XPS(HF), as observed previously,[54] due primarily to overestimation (by as much as 5 kcal/mol) of binding energies for dispersion-dominated complexes. These errors are significantly reduced by replacing the MP2-like sum-over-states dispersion formula with the empirical potential in Eq. (6.3). The best results are obtained using XPS(KS)+D/aDZ′, for which the MUE is 0.5 kcal/mol or 9%.

Note that XPS(KS) outperforms XPS(HF) for the strongly H-bonded subset of S22A, where the intermolecular interactions are dominated by electrostatics and induction. We attribute this to the effects of intramolecular electron correlation, since SAPT(KS) has previously been shown to provide accurate values for components of the intermolecular interaction other than dispersion, provided that v_{xc} is asymptotically correct.[162,235] Our results demonstrate that LRC functionals afford an alternative means to enforce correct asymptotic behavior of v_{xc}, without abandoning the relationship $v_{xc}(\mathbf{r}) = \delta E_{xc}/\delta\rho(\mathbf{r})$.

Table 6.1: Error statistics, with respect to CCSD(T)/CBS[a] benchmarks, for the S22A data set.[38]

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

[a]CBS = complete basis set. [b]Statistics listed separately for the total S22A data set, along with subsets consisting of strongly H-bonded dimers, dispersion-dominated dimers, and dimers whose interactions are of mixed influence, as classified in Ref. 38. [c]Mean unsigned error. [d]Maximum unsigned error.

The performance of CBS-extrapolated MP2 and CCSD methods, along with spin-component-scaled (SCS) versions thereof,[117] was evaluated previously for the S22A data set.[38] The relevant MUEs are:

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

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

The XPS(KS) method will not be considered further because of known problems with the SAPT(KS) treatment of dispersion, evident from results in 6.1. The XPS(KS)+D method contains parameters that were fit to the S22A database, so it is possible that S22A error statistics are overly optimistic. For a blind test of this approach, we turn to the S66 database,[10] for which error statistics are summarized in 6.2.

Table 6.2: Error statistics, with respect to CCSD(T)/CBS benchmarks, for the S66 data set.[10]

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

[a]Results taken from Ref. 10. [b]Results for the effective fragment potential (EFP), from Ref. 37

Results for S66 obtained using XPS(HF) and XPS(KS)+D, with the two basis sets considered here, are comparable to or better than CBS-extrapolated results obtained with the MP2, SCS-MP2, and CCSD methods. Particularly impressive is the XPS(KS)+D/aDZ′ method, whose MUE is just 0.27 kcal/mol and whose largest unsigned error is < 1 kcal/mol. This method does exhibit a 54% error for the benzene-ethene dimer (which is underbound by 0.78 kcal/mol), the maximum percentage error that we observe for S66 with this approach. If we eliminate benzene-ethene and also pyridine-ethene, which is underbound by 0.66 kcal/mol (35%), then the MUE is reduced to 6% and the maximum unsigned error is reduced to 20% for XPS(KS)+D/aDZ′.

Such small error in XPS(KS)+D against S66 is totally benefit from replacing the inaccurate MP2-like sum-over-states dispersion formula by the empirical dispersion potential. We can prove it by comparing the MUE of S66 calculated by XPS(KS)+D and SAPT(KS)+D. Although the s_β value in the empirical dispersion formula for the SAPT(KS)+D can be refitted against S22A database, we find that the change of s_β value is less than 0.01. Therefore, the SAPT(KS)+D method uses the same dispersion formula and parameter as XPS(KS)+D. The MUE of S66 calculated by SAPT(KS)+D is 0.32 kcal/mol and in the same range of error as XPS(KS)+D. However, XPS(KS)+D used the XPol method to capture the many-body polarization is superior than the SAPT(KS)+D in the many-body systems as pointed out at 6.1 using $(H_2O)_n$ clusters. The structures and MP2/CBS benchmark binding energies for

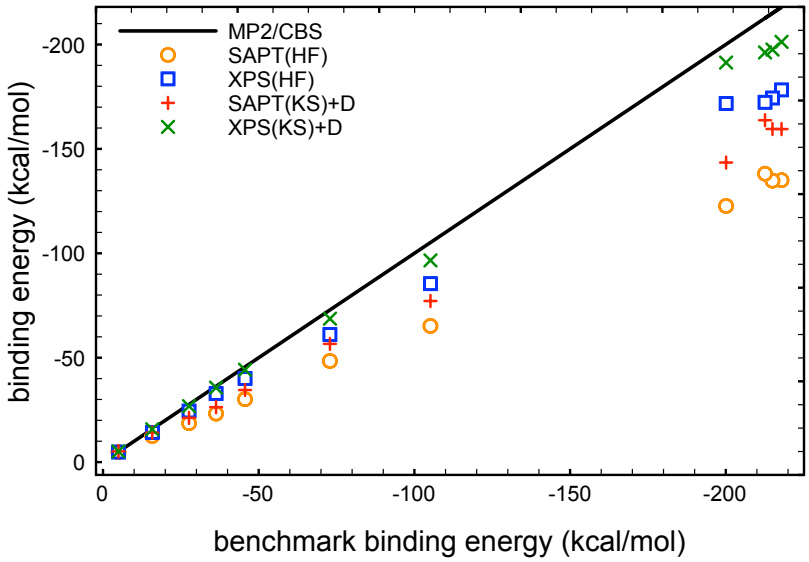

Figure 6.1: The binding energies for $(H_2O)_n$ clusters where $n = 2$, 3, 4, 5, 6, 8, 11, and 20 calculated by SAPT(HF), XPS(HF), SAPT(KS)+D, and XPS(KS)+D combined with aDZ′ basis set , as compared to MP2/CBS benchmarks.

$(H_2O)_n$ clusters for $n = 2$, 3, 4, 5, 6, 8, 11, and 20 are taken from the work of Xantheas and co-workers.[256–260] From 6.1, we note that the XPS method is significantly more accurate than the corresponding pairwise-additive SAPT method which is entirely neglected many-body effects. Furthermore, XPS(KS)+D is more accurate than the XPS(HF) method which neglects the intramolecular electron correlation and uses the MP2-like sum-over-states dispersion formula. In short, the XPS(KS)+D method definitely improves the binding-energy calculations for different sizes of clusters.

The error statistics presented above demonstrate the excellent performance of XPS(KS)+D at equilibrium geometries, but it is also important to understand how this method performs across a range of intermolecular distances. To test this, we

have chosen several challenging systems: Ar$\cdots$Ne, formic acid dimer, benzene dimer, and $X^-(H_2O)$ with X = F and Cl. In the case of Ar$\cdots$Ne, several DFT approaches thought to be accurate for non-covalent interactions predict qualitatively incorrect potential energy curves (PECs) for this system, even in cases where the binding energy evaluated at the minimum-energy geometry is accurate.[6] The binding energy of the formic acid dimer has a large contribution from intramolecular electron correlation,[158] and exhibits the maximum error in SAPT0/aDZ$'$ calculations for S22A. Benzene dimer was selected because it is a stringent test of the accuracy of dispersion interactions, and we consider both the "sandwich" isomer, which is a dispersion-dominated complex, as well as the T-shaped isomer, where quadrupolar electrostatic interactions are also significant.[38] Finally, $X^-(H_2O)$ systems exhibit much larger binding energies than those in the S22A set.

One-dimensional PECs for these systems are depicted in 6.2–6.5. The XPS(KS)+D/aDZ$'$ method yields PECs that are quite comparable to benchmark results for all of these difficult cases except $F^-(H_2O)$. For the charge-neutral systems, binding energy errors evaluated at equilibrium geometries range from 0.02 kcal/mol for Ar$\cdots$Ne to 0.77 kcal/mol for $(HCO_2H)_2$. For $Cl^-(H_2O)$ and $F^-(H_2O)$, the XPS(KS)+D/aDZ$'$ method affords larger binding energy errors: 2.07 and 2.94 kcal/mol, respectively, at the equilibrium geometries. Similar binding energies are obtained at the XPS(HF)/aDZ$'$ level, whereas XPS(HF)/TZVPP binding energies are much more accurate, consistent with previous results[54] suggesting that large basis sets are required for systems whose binding is dominated by electrostatics and induction.

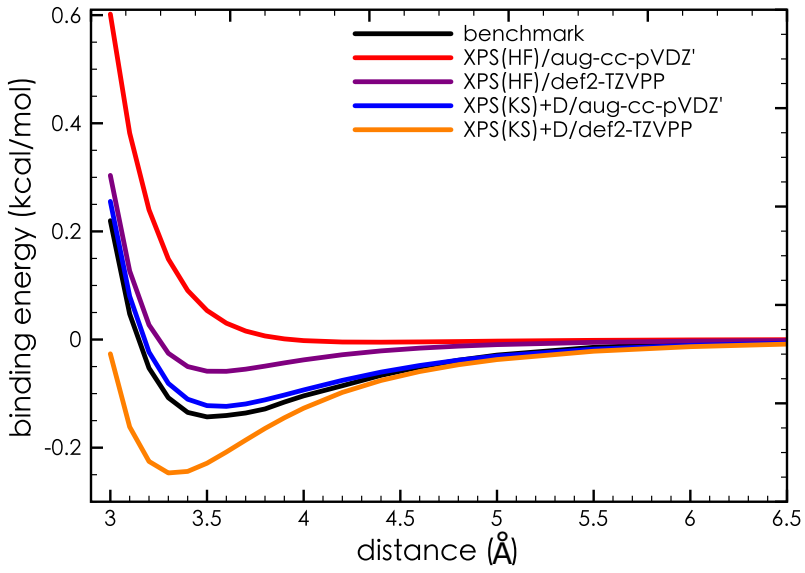

Figure 6.2: Binding energy curves for Ar$\cdots$Ne. Benchmark results were computed using a dispersion corrected version of the "dispersionless" density functional (dlDF) of Ref. 6 (dlDF+D_{as}/aTZ, in the language of Ref. 6).

Table 6.3: Error statistics, with respect to CCSD(T)/CBS benchmarks, for a set of ionic H-bonded dimers.[39]

Method	Error/kcal mol^{-1}			% Error	
	MUE	RMSE	Max	MUE	Max
XPS(HF)/aDZ′	1.26	1.61	3.31	6.3	12.4
XPS(HF)/TZVPP	0.97	1.15	1.92	6.0	10.6
XPS(KS)/aDZ′	0.85	1.03	2.10	4.4	10.7
XPS(KS)/TZVPP	0.66	0.72	1.29	3.5	7.4
XPS(KS)+D/aDZ′	1.02	1.14	1.96	5.4	14.1
XPS(KS)+D/TZVPP	0.48	0.56	1.08	2.5	6.2
MP2/cc-pVTZ[a]		1.81			
BLYP+D/def2-QZVP[a]		0.59			

[a] Counterpoise-corrected results from Ref. 39.

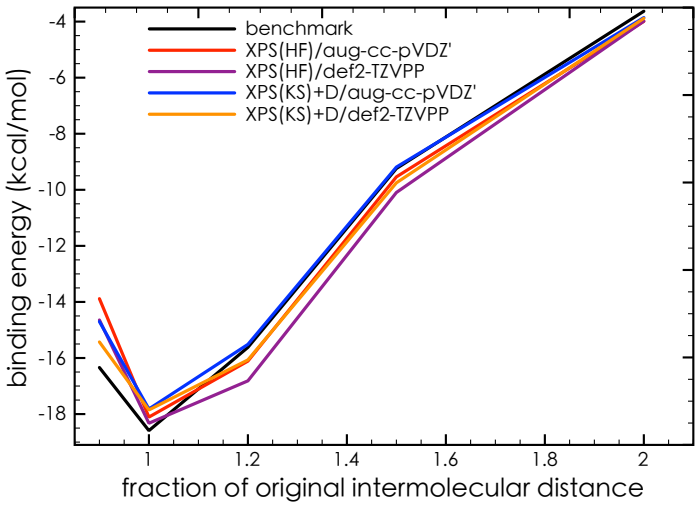

Figure 6.3: Binding energy curves for formic acid dimer. Benchmark CCSD(T)/CBS results are taken from Ref. 7. The horizontal axis is the scaled equilibrium distance between the centers of mass of the two monomers.

To examine whether these larger errors are systemic to ions in general, we have computed binding energies for a database of 15 ionic H-bonded dimers,[39] in which the ionic partner is either acetate, guanidinium, methylammonium, or imidazolium. Error statistics are listed in 6.3, and are somewhat larger than those seen for the neutral S22 and S66 data sets. Here, the importance of monomer electron correlation and triple-ζ basis sets is clear: absent either of those two factors, MUEs are $\gtrsim 1$ kcal/mol and maximum errors are 2–3 kcal/mol. This is consistent with previous XPS results indicating that the description of H-bonded and induction-bound systems generally improves in a systematic way with respect to basis-set quality and the treatment of electron correlation.[54] The XPS(KS)+D/TZVPP method delivers a root-mean-square error (RMSE) of 0.56 kcal/mol for the ionic H-bonded data set, which is significantly

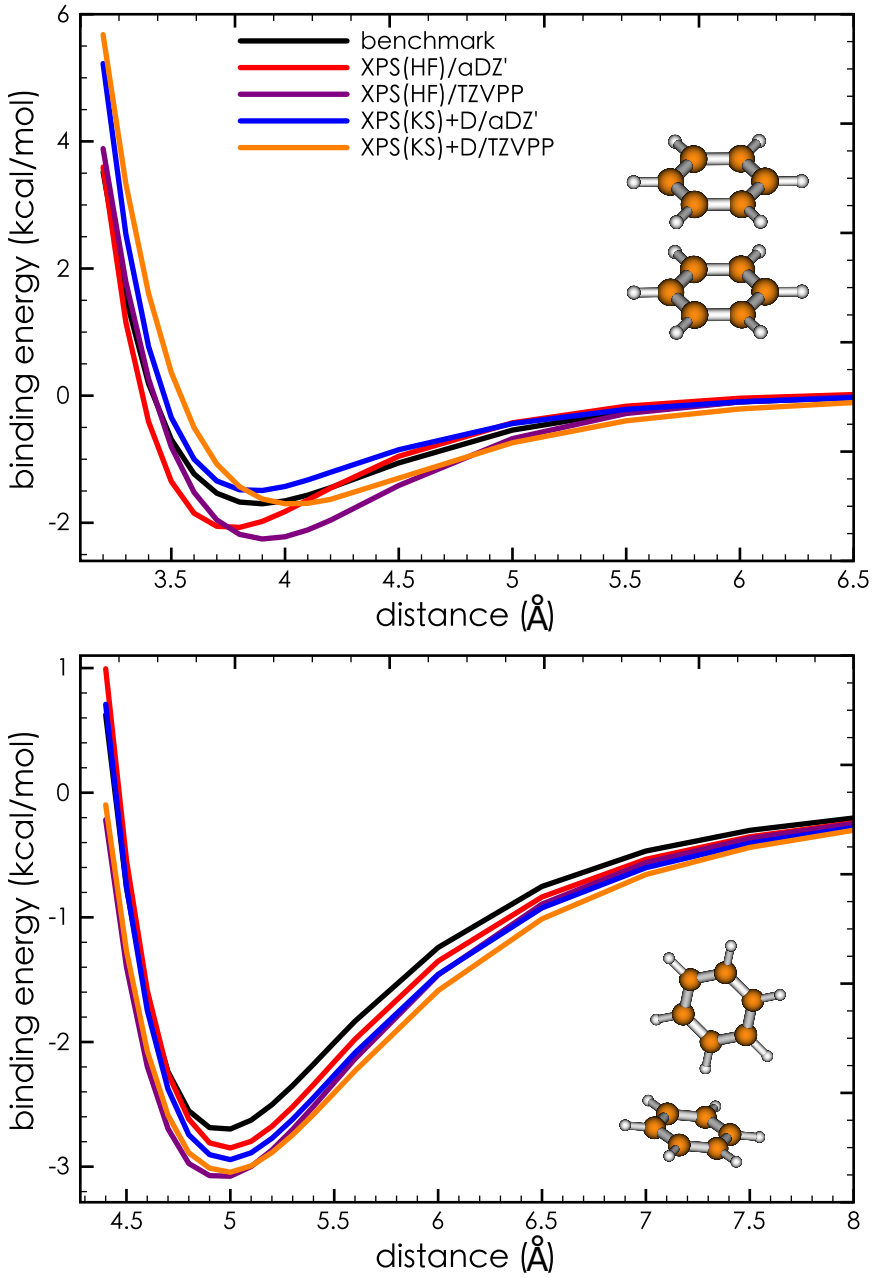

Figure 6.4: Binding energy curves for (a) the "sandwich" and (b) the "T-shaped" isomer of benzene dimer. The distance coordinate in either case is the center-to-center distance between the benzene rings. Benchmark CCSD(T)/CBS results are taken from Ref. 5.

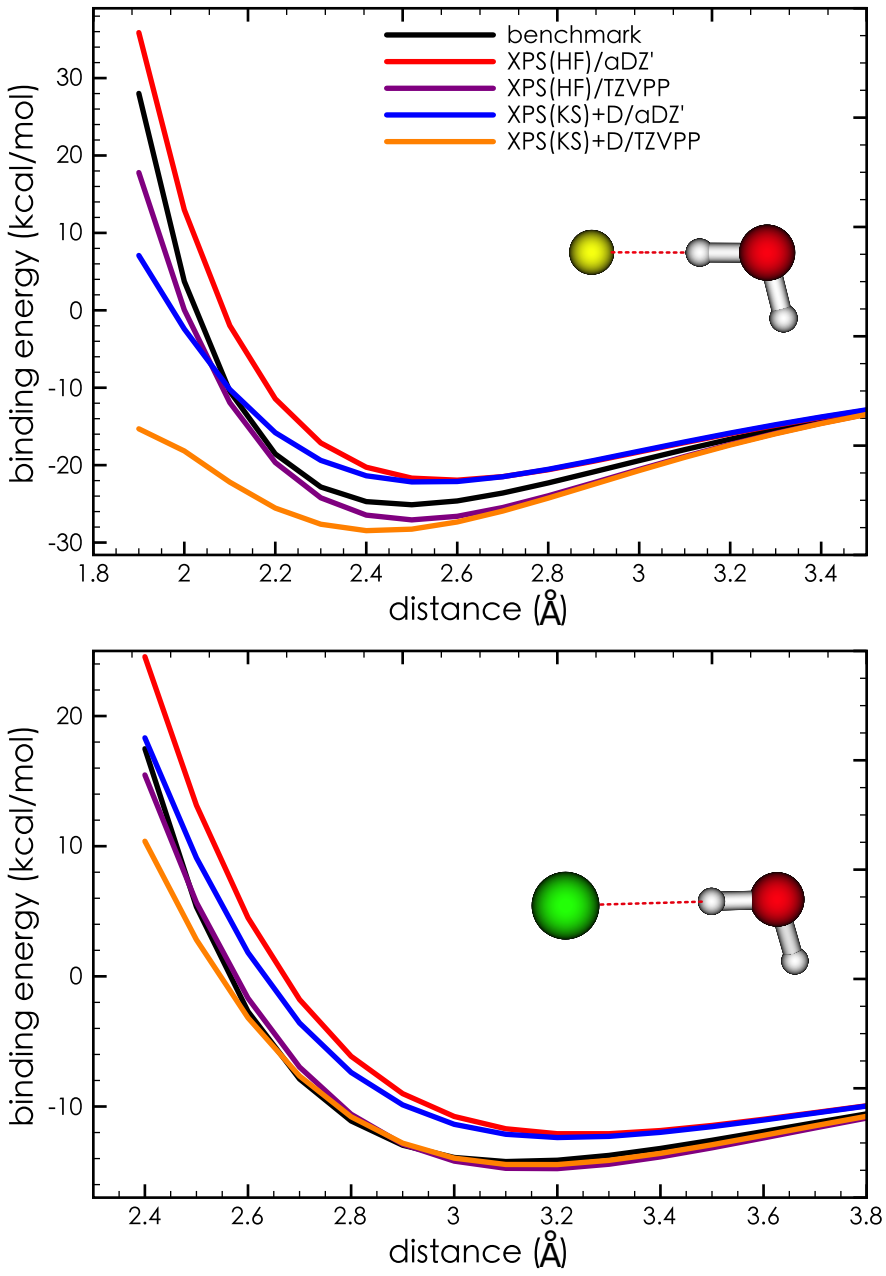

Figure 6.5: Binding energy curves for (a) $F^{-}(H_2O)$ and (b) $Cl^{-}(H_2O)$. The distance coordinate is the halide–oxygen distance and the benchmarks are CCSD(T)/CBS.

better than MP2/cc-pVTZ results and comparable to BLYP+D/def2-QZVP, but at significantly reduced cost.

Returning to $F^-(H_2O)$, we note that the PECs computed by XPS(KS)+D methods [6.5(a)] are qualitatively incorrect for intermolecular distances smaller than the equilibrium distance. Several factors might account for this. First, no ionic systems were used in fitting the dispersion potential. Second, this potential actually diverges for $r_{ij} \ll R_i + R_j$,[16] and we note that the van der Waals radius R_i for the anion is larger than that of the corresponding neutral atom, which may explain the fact that PECs for the neutral systems are accurate even at relatively short distances. Lastly, the intermolecular interaction in $F^-(H_2O)$ is known to have substantial covalent character resulting from a low-energy $FH \cdots OH^-$ diabatic state,[261] and as such this system may be especially problematic for monomer-based quantum chemistry.

As a final test of XPS(KS)+D, we have computed binding energies for nucleobase tetramers arranged in average B-DNA geometries,[262] for which MP2, SCS-MP2, and DFT-SAPT results are available.[15] MP2 is known to overestimate binding energies of π-stacked complexes,[5] and SCS-MP2 is often used to reduce this systematic error.[263,264] However, SCS-MP2 *underestimates* the binding energy of the π-stacked uracil dimer,[265] by 3 kcal/mol as compared to the CCSD(T)/CBS benchmark of 9.7 kcal/mol.[38] The SCS(MI)-MP2 method,[266] in which the two SCS parameters are optimized (in a basis-set-specific way) against the S22 database, reduces this error to 0.4 kcal/mol,[265] comparable to the MUE for the S66 database obtained at the

Table 6.4: Binding energies for nucleobase tetramers, treating each Watson-Crick base pair as a monomer unit.

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

[a]The notation WX-YZ means that WX and YZ are Watson-Crick pairs. [b]Counterpoise corrected. [c]Values from Ref. 15. [d]Includes the δ(HF) correction. [e]Includes empirical scaling of the dispersion energy, as recommended in Ref. 24.

SCS(MI)-MP2/CBS level (see 6.2).[10] The XPS(KS)+D/aDZ′ method exhibits comparable error statistics for S66.

Results for nucleobase tetramers are listed in 6.4. CCSD(T)/CBS benchmarks are only available for base pairs, but in that case comparison to DFT-SAPT suggests that the DFT-SAPT/aTZ results in 6.4 are systematically too small.[247,267] For example, the DFT-SAPT/aTZ binding energy[247] for the π-stacked AT dimer is underestimated by 1.4 kcal/mol, as compared to the CCSD(T)/CBS benchmark (11.66 kcal/mol), if the so-called δ(HF) correction[117] is used in the SAPT calculation. If it is not, then the binding energy is underestimated by 3 kcal/mol. Hence, in

perusing the DFT-SAPT/aTZ results in 6.4, one should anticipate that binding energies are underestimated by $\gtrsim 1$ kcal/mol. At the same time, the XPS(KS)+D/aDZ′ method *overestimates* the π-stacked AT binding energy by 0.8 kcal/mol. Therefore, the correct binding energies for the nucleobase tetramers probably lie between the DFT-SAPT/aTZ and XPS(KS)+D/aDZ′ results in 6.4.

This conclusion is bolstered by DFT-SAPT results in which the dispersion energy is scaled by an empirical factor of 1.051, suggested in Ref. 24 as a means of accounting for the slow basis-set convergence of this term. Scaled DFT-SAPT/aTZ binding energies lie between the raw DFT-SAPT/aTZ results and XPS(KS)+D/aDZ′ results, except for CG-CG where the scaled DFT-SAPT and XPS(KS)+D results are essentially identical. Together, the nucleobase calculations presented here place plausible bounds on the tetramer binding energies, and lend support to the accuracy of the scaled DFT-SAPT/aTZ approach.[24]

As compared to DFT-SAPT, however, XPS(KS)+D has important advantages. First, it avoids the supermolecular HF calculation that is required in order to compute the δ(HF) correction. Second, the cost of XPS(KS)+D scales as $\mathscr{O}(N^3)$ with respect to dimer size, whereas whereas DFT-SAPT (with density fitting) scales as $\mathscr{O}(N^5)$. Finally, XPS(KS)+D can be extended to any number of monomer units, at $\mathscr{O}(n^3)$ cost, whereas DFT-SAPT is restricted to dimers only. As such, the nucleobase tetramer calculations in 6.4 must be computed using Watson-Crick base pairs as the monomer units, which adds significantly to the cost of DFT-SAPT.

Finally, we want to comment on the computational scaling of the XPS method.

6.6 shows actual timings for stacking adenine clusters using the XPol, XPS(KS), and XPS(KS)+D methods, as compared to timings for supersystem DFT calculations. Since there are $n(n-1)/2$ (n is the number of monomers) pairwise SAPT calculations in the XPS and they are completely independent of one another, we can use the embarrassingly paralleizable version of XPS calculations to distribute each SAPT calculation on each CPU. We name them XPS(KS) (Parallel) and XPS(KS)+D (Parallel) in 6.6. From 6.6, XPol and XPS calculations are apparently less expensive than supersystem calculations as also pointed out before,[65] furthermore, the cost for the XPS calculations can be reduced to $\mathscr{O}(n)$ as XPol if the embarrassingly parallelizable version of XPS is used. In other words, the cost of XPS(KS)+D (Parallel) calculation scales linearly with the number of monomers and such low cost scaling makes XPS(KS)+D (Parallel) flexible for large many-body systems.

In summary, the expensive and problematic sum-over-states dispersion terms in XPS(KS) have been replaced by empirical atom-atom potentials, resulting in a cubic-scaling method that we call XPS(KS)+D. Using a "tuned" LRC functional[227,230] and modest basis sets, the method exhibits an accuracy that is comparable to or better than MP2, SCS-MP2, SCS(MI)-MP2, and CCSD results extrapolated to the basis set limit. Potential energy curves for difficult systems are accurate across a range of intermolecular distances, with the exception of $F^-(H_2O)$ at short distances. XPS(KS)+D provides an alternative to DFT-SAPT and SAPT(DFT) methods[137,161] for computing binding energies of DNA nucleobases, and we have used these approaches together to obtain the best estimates to date of the binding energies for nucleobase tetramers.

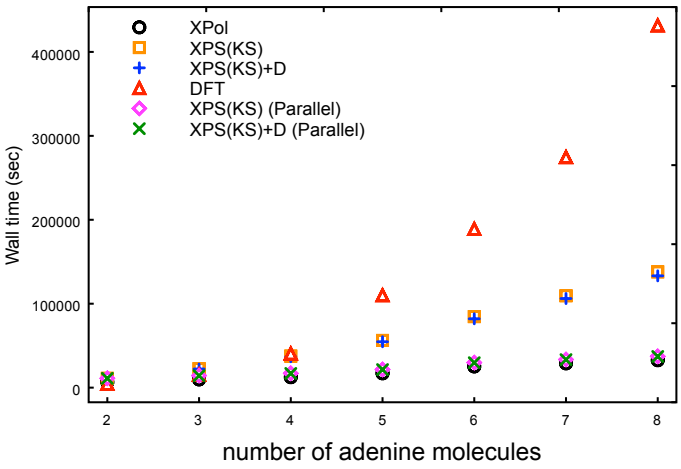

Figure 6.6: Wall time required for XPol, XPS(KS), XPS(KS)+D, and supersystem DFT calculations in stacking $(\text{adenine})_n$. All the calculations use LRC-ωPBEh functional with $C_{\text{HF}} = 0.6$ combined with the aDZ′ basis set. The wall time for the embarrassingly paralleizable version of XPS calculations is also shown.

The low-order scaling of XPS(KS)+D makes this a promising method for use in fragment-based drug design, although the performance for ions suggests that extensions to larger basis sets, and a more extensive parameterization of the dispersion potential, may be in order. Such work is currently underway in our group.

CHAPTER 7

An improved treatment of empirical dispersion and a many-body energy decomposition scheme for the explicit polarization plus symmetry-adapted perturbation theory (XSAPT) method[7.1]

7.1 Introduction

Electronic structure theory has been made a enormous progress over the past several decades, yet the "gold standard" of chemical accuracy, CCSD(T), remains out of reach for most systems. For calculation of non-covalent interactions, which is the topic of the present work, CCSD(T)/aug-cc-pVTZ calculations on $(H_2O)_{20}$ represent the present state-of-the-art,[268] and are only feasible on massively parallel architectures. Second-order Møller-Plesset perturbation theory (MP2) is tractable in larger systems but overestimates binding energies in cases where the binding is dominated by dispersion interactions.[207] Various strategies have been used to correct this deficiency, including spin-component scaled (SCS) MP2 methods,[269,270] sometimes with parameters fit specifically for non-covalent interaction energies,[266] and also MP2.X methods that combine MP2 and MP3 results in an empirical way.[271] These methods achieve a

[7.1] This chapter appeared as a full article in the *Journal of Chemical Physics*, in 2013, volume 139, pages 034107:1–16.

mean accuracy of $\lesssim 0.4$ kcal/mol with respect to complete-basis CCSD(T) binding energies,[272] but scale no better than $\mathcal{O}(N^5)$ with respect to total system size, N.

Density functional theory (DFT) is more affordable but many of the most popular functionals afford a poor description of dispersion interactions.[143,273] One strategy to circumvent this problem is to introduce a rather large number of empirical parameters into the functional,[142,201] then optimize these parameters using data sets that include weak interactions. The M06-2X functional[142] is a popular example of this approach. Alternatively, one might attempt to capture the dispersion interactions directly by means of classical atom–atom potentials, typically with r^{-6} dependence.[11,143,254] Such approaches fall under the moniker "DFT-D", and are often able to reproduce higher-level calculations rather well.[11,13,206,274] The ωB97X-D functional[201] is a popular example. Grimme *et al.*[274] have recently introduced a "third generation" (DFT-D3) correction that further improves the description of non-bonded interactions, in functionals such as ωB97X-D3.[13] Finally, dispersion can be introduced into DFT by means of non-local correlation functionals.[145–147] The LC-VV10 functional[147] is an example of such an approach, which has recently been shown to exhibit outstanding performance for many types of intermolecular interactions.[12] The description of weak interactions in DFT has thus come a long way in the past few years, yet none of these methods scales better than $\mathcal{O}(N^3)$. As such, these DFT methods are still not feasible for applications to molecular liquids or biomolecules.

Fragment-based approaches[43,44,71,275–277] do provide a relatively affordable route to computing binding energies in large systems, by partitioning the supersystem into

subsystems (fragments). In particular, our group has developed a low-cost, monomer-based approach that we call XPol+SAPT (XSAPT).[44,54,55,65] This approach starts from the variational explicit polarization (XPol) method,[252] which is used to capture many-body polarization effects by means of a charge-embedded, monomer-based self-consistent field (SCF) calculation. In a subsequent step, we apply a pairwise-additive form of symmetry-adapted perturbation theory (SAPT).[115,117,149] The resulting XSAPT method generalizes traditional SAPT from two-body to many-body systems. The wall time for an XSAPT calculation scales as $\mathcal{O}(n)$ with respect to the number of monomers, n, assuming that we have ${}_nC_2$ processors to run in "embarrassingly parallel" mode for the second-order SAPT calculations, and is $\mathcal{O}(n^3)$ even in serial mode.[55] When a Hartree-Fock (HF) description of the monomers is used in the XPol calculation, in conjunction with a suitable basis set, errors in dimer binding energies computed by XSAPT lie within 1 kcal/mol of high-level benchmarks.[54,65]

One might try to improve on this accuracy by using instead a Kohn-Sham (KS) description of intramolecular electron correlation, in what we have termed XSAPT(KS).[44] However, this approach suffers from the same problem as SAPT(KS),[162,208] namely, significant overestimation of dispersion energies.[54] Ironically, the dispersion and exchange-dispersion terms in SAPT(KS) are not only the least accurate ones, but also the most expensive to compute, scaling as $\mathcal{O}(N^4)$ and $\mathcal{O}(N^5)$, respectively, with respect to monomer size, N. For this reason, we recently introduced XSAPT(KS)+D,[55] in which the sum-over-states (uncoupled Hartree-Fock[278]) dispersion formula in second-order SAPT(KS) is replaced by empirical atom–atom potentials developed for this

purpose by Hesselmann.[16] In conjunction with a double-ζ basis set, XSAPT(KS)+D exhibits mean errors in binding energies of < 0.5 kcal/mol for the S22A data set[38] and the larger (and more balanced) S66 data set.[10] In this respect, XSAPT(KS)+D is superior to various MP2-type methods extrapolated to the complete basis-set (CBS) limit, and methods containing triple excitations are required in order to do better.[44]

In traditional SAPT, the interaction energy decomposes naturally according to

$$\begin{aligned} E_{\text{int}}^{\text{SAPT}} = E_{\text{elst}}^{(1)} + E_{\text{exch}}^{(1)} + E_{\text{ind}}^{(2)} + E_{\text{exch-ind}}^{(2)} \\ + E_{\text{disp}}^{(2)} + E_{\text{exch-disp}}^{(2)} + \cdots \ . \end{aligned} \tag{7.1}$$

All terms up to second order in the intermolecular interaction are listed explicitly, with subscripts that denote electrostatic (elst), exchange (exch), induction (ind), and dispersion (disp) contributions. This work is mainly focused on developing a similar interaction-energy decomposition scheme for XSAPT, and comparing the various energy components to dimer SAPT results for the S22A.[37] In the course of this analysis, we discovered that the original XSAPT(KS)+D method does not do a good job of reproducing individual energy components, and thus its favorable performance for binding energies is due in no small part to error cancellation. This led us to pursue a "second generation" dispersion correction ("+D2"), using alternative dispersion potentials developed by Podeszwa *et al.*[244]

7.2 Theory

7.2.1 Energy decomposition scheme

The success of XSAPT is based on the fact that the many-body (MB) or non-pairwise–additive contribution to cluster binding energies is dominated by induction

interactions, whereas electrostatic, exchange-repulsion, and dispersion interactions are strictly or nearly pairwise additive.[8,251,279–282] In the context of XPol-SAPT, it makes sense to define the MB contribution to the interaction energy, $E_{\text{int}}^{\text{MB}}$, according to

$$E_{\text{int}}^{\text{MB}} = E_{\text{int}}^{\text{XSAPT}} - \sum_{A} \sum_{B<A} E_{AB}^{\text{XSAPT}} \ , \tag{7.2}$$

where $E_{\text{int}}^{\text{XSAPT}}$ is the overall XSAPT interaction energy and E_{AB}^{XSAPT} is the interaction energy for dimer $A \cdots B$. The SAPT part is pairwise additive by construction (we do not consider the three-body SAPT terms derived by Lotrich *et al.*[249,250]). Thus, the total SAPT interaction energy for a collection of monomers is

$$E_{\text{int}}^{\text{SAPT}} = \sum_{A} \sum_{B<A} E_{AB}^{\text{SAPT}} \ . \tag{7.3}$$

Addition of Eqs. (7.2) and (7.3) affords

$$E_{\text{int}}^{\text{MB}} = E_{\text{int}}^{\text{XSAPT}} - E_{\text{int}}^{\text{SAPT}} - \sum_{A} \sum_{B<A} \left(E_{AB}^{\text{XSAPT}} - E_{AB}^{\text{SAPT}} \right) \ . \tag{7.4}$$

The SAPT interaction energy can be decomposed as in Eq. (7.1) and substituted into Eq. (7.4), with the result rewritten as

$$\begin{aligned} E_{\text{int}}^{\text{XSAPT}} = {} & E_{\text{elst}}^{(1)} + E_{\text{exch}}^{(1)} + E_{\text{ind}}^{(2)} + E_{\text{exch-ind}}^{(2)} \\ & + E_{\text{disp}}^{(2)} + E_{\text{exch-disp}}^{(2)} + \cdots \\ & + \sum_{A} \sum_{B<A} (E_{AB}^{\text{XSAPT}} - E_{AB}^{\text{SAPT}}) + E_{\text{int}}^{\text{MB}} \ . \end{aligned} \tag{7.5}$$

Equation (7.5) is the interaction-energy decomposition scheme for XSAPT. This energy decomposition analysis requires two sets of calculations. First, traditional SAPT calculations are performed on all pairs of monomers, to obtain pairwise electrostatic, exchange, dispersion, and induction terms. Second, and XSAPT calculation

must be performed on the entire system, which provides not only $E_{\text{int}}^{\text{MB}}$ but also the rest of induction terms that are missing in the first set of calculations. These are defined by the double sum in Eq. (7.5).

It is common in SAPT calculations to solve coupled-perturbed Hartree-Fock (CP-HF) equations and thereby include an infinite-order response correction for polarization in the presence of a frozen partner density.[114] In contrast, the XSAPT method treats polarization self-consistently in the XPol step. As such, the infinite-order response correction for induction should be included exactly by the XPol part of the calculation (with further induction corrections vanishing), if the XPol calculation is performed using density embedding,[276] that is, if monomer SCF densities are used to compute the electrostatic interactions between the monomers in the self-consistent XPol calculation. We do not pursue this possibility here, but instead use atom-centered point charges to do the embedding, as in previous work.[54,55,65] This procedure is significantly less expensive (especially if the monomers are large), and implicitly includes some higher-order induction effects into the zeroth-order XPol monomer energies.[65] Therefore, the energy difference between XSAPT and SAPT for all pairs of dimers [the double sum in Eq. (7.5)] partly includes the infinite-order response correction for induction. In addition, some higher-order induction effects are captured by the $\delta E_{\text{int}}^{\text{HF}}$ correction[130,131] that is discussed below.

It is common to truncate the SAPT interaction energy at second order in Eq. (7.5) and incorporate higher-order polarization effects by adding a correction

$$\delta E_{\text{int}}^{\text{HF}} = E_{\text{int}}^{\text{HF}} - \left(E_{\text{elst}}^{(1)} + E_{\text{exch}}^{(1)} + E_{\text{ind,resp}}^{(2)} + E_{\text{exch-ind,resp}}^{(2)}\right) \tag{7.6}$$

to the interaction energy. The "response" (resp) subscripts indicate that the infinite-order response correction for induction is incorporated by solving CPHF equations, and $E_{\rm int}^{\rm HF}$ is the counterpoise-corrected HF binding energy for the dimer. It is recommended to include $\delta E_{\rm int}^{\rm HF}$ term in SAPT calculations involving polar monomers, because induction corrections converge slowly for polar molecules.[130,131] Furthermore, we find that $\delta E_{\rm int}^{\rm HF}$ is necessary in traditional SAPT calculations, not only to obtain quantitative binding energies but also to obtain qualitatively correct potential energy surfaces for induction-dominated systems.[50] In the present work, we assume that this correction term is pairwise additive for many-body XSAPT calculations,

$$\delta E_{\rm int}^{\rm HF} = \sum_{A} \sum_{B<A} \delta E_{AB}^{\rm HF} \, . \tag{7.7}$$

The quality of the results presented herein indicates that this assumption is quite robust.

Adding the $\delta E_{\rm int}^{\rm HF}$ correction term to Eq. (7.5), the XSAPT interaction energy becomes

$$\begin{aligned} E_{\rm int}^{\rm XSAPT} = {} & E_{\rm elst}^{(1)} + E_{\rm exch}^{(1)} + E_{\rm disp}^{(2)} + E_{\text{exch-disp}}^{(2)} \\ & + \Bigg[E_{\rm ind}^{(2)} + E_{\text{exch-ind}}^{(2)} + \sum_{A} \sum_{B<A} \delta E_{AB}^{\rm HF} \\ & + \sum_{A} \sum_{B<A} (E_{AB}^{\rm XSAPT} - E_{AB}^{\rm SAPT}) + E_{\rm int}^{\rm MB} \Bigg] \, . \end{aligned} \tag{7.8}$$

Those terms in square brackets are regarded as the total induction energy, and the total dispersion energy is $E_{\rm disp}^{(2)} + E_{\text{exch-disp}}^{(2)}$. One may argue that this procedure double-counts the higher-order induction terms since both the XPol SCF procedure and $\delta E_{\rm int}^{\rm HF}$ are included those terms. In the charge-embedded, monomer-based SCF calculation for XPol, both infinite-order response and higher-order corrections for induction are

only partly included. Furthermore, we assume that the $\delta E_{\rm int}^{\rm HF}$ correction term is pairwise for the many-body system. Thus, this approach relies partly on cancellation of errors to provide meaningful results. However, thorough comparisons to benchmark calculations will show that this interaction-energy decomposition scheme is accurate and robust.

7.2.2 Second-generation dispersion potential

Our original XSAPT(KS)+D method[55] was based on empirical atom–atom dispersion potentials developed for SAPT(KS) by Hesselmann.[16] Here, we test a second-generation ("+D2") version of XSAPT that employs an alternative dispersion potential developed by Podeszwa *et al.*:[244]

$$E_{\rm disp}^{\rm dlDF} = -\sum_{i\in A}\sum_{\substack{j\in B\\(B\neq A)}}\left[\frac{C_{ij,6}}{r_{ij}^6}\,f_6(\beta_{ij}r_{ij}) + \frac{C_{ij,8}}{r_{ij}^8}\,f_8(\beta_{ij}r_{ij})\right] \tag{7.9}$$

where $C_{ij,6} = (C_{i,6}C_{j,6})^{1/2}$ (similar for $C_{ij,8}$) and $\beta_{ij} = (\beta_i\beta_j)^{1/2}$. This empirical potential has previously been used to correct the results of a "dispersionless" density functional (dlDF),[6,244] hence the notation $E_{\rm disp}^{\rm dlDF}$. The indices i and j in Eq. (7.9) represent nuclei located on different monomers and

$$f_n(r_{ij}) = 1 - \exp(-r_{ij})\sum_{m=0}^{n}\frac{r_{ij}^m}{m!} \tag{7.10}$$

is the Tang-Toennies damping function.[283] The quantities $C_{i,6}$, $C_{i,8}$, and β_i are parameters that are fit to reproduce SAPT(DFT) dispersion energies ($E_{\rm disp}^{(2)} + E_{\rm exch\text{-}disp}^{(2)}$) for a training set of dimers.[244] For hydrogen the values of these parameters depend upon the identity of the nearest-neighbor atom.

The Hesselmann and Podeszwa dispersion potentials were parameterized in completely different ways. Podeszwa's dispersion force field was fitted directly to a set of $E^{(2)}_{\text{disp}} + E^{(2)}_{\text{exch-disp}}$ values computed using SAPT(DFT),[161] for a large training set of dimers. Hesselmann's dispersion potential was fitted to reproduce benchmark intermolecular interaction energies for the S22 data set.[16] Therefore, Podeszwa's potential seems better suited for use in our energy decomposition scheme, since it constitutes an well-defined dispersion component. As in our previous work,[55] we refer to results using Hesselmann's dispersion potential as XSAPT(KS)+D, whereas results using Eq. (7.9), which are presented here for the first time, will be called XSAPT(KS)+D2.

7.2.3 Basis sets and functionals

Our tests indicate that obtaining good results for different energy components requires triple-ζ basis sets augmented with diffuse functions. However, a large number of diffuse functions may cause overpolarization of the monomer-based wave function in the XPol step,[284] especially for anions. As a compromise, all calculations were evaluated using "heavy augmented" basis sets that have diffuse functions only on non-hydrogen atoms. Our basis set of choice is the second-generation Ahlrichs triple-ζ basis set, def2-TZVPP,[197] which exhibits the smallest mean unsigned error for S22A amongst many basis sets.[54,65] We augment def2-TZVPP with diffuse functions taken from Dunning's aug-cc-pVTZ (aTZ) basis set. We will denote this heavy-augmented version of def2-TZVPP as ha-TZVPP (haTZ).

We use long-range-corrected (LRC) density functionals[226–228,240] to obtain correct asymptotic behavior, as in previous work.[55] Specifically, we employ two LRC

functionals based on the short-range ωPBE exchange functional.[222] One of these (LRC-ωPBEh[227]) also contains 20% short-range Hartree-Fock exchange, whereas the other (LRC-ωPBE[240]) does not. The range separation parameter (ω) is determined by system-specific tuning[230] to satisfy the condition

$$\varepsilon_{\mathrm{HOMO}} = -\mathrm{IP} \ , \tag{7.11}$$

where "IP" denotes the lowest ionization potential. For clusters of monomers, we previously took ω for the supersystem to be the one corresponding to the lowest monomer IP,[55] on the assumption that the non-covalent interactions would not significantly affect the IPs (hence the lowest monomer IP equals the cluster IP). As such, the same value of ω was used for all monomers in the cluster calculations reported in Ref. 55. Subsequently, we discovered that the results could be significantly improved in certain cases by using different ω values for different monomers, in order to obtain an exact asymptotic correction (AC) for each monomer unit, as is done in the SAPT(DFT) method[161] and the DFT-SAPT method.[137] In the present work, we compare these two approaches, using the designation "(AC)" whenever different ω values are used for different monomers. Tuned ω values for the monomers considered here are provided in the Supplementary Material.

Our tests indicate that errors for the exchange energy components increase with increasing fraction of short-range HF exchange in the LRC functional. Furthermore, we find that Podeszwa's dispersion potential [Eq. (7.9)] gives lower errors for dispersion components as compared to Hesselmann's potential. We expect that Podeszwa's potential works well with pure functionals in the short range since both exchange

and dispersion components give lower errors. Actually, the LRC-ωPBE functional gives better results when used with Podeszwa's dispersion potential and the LRC-ωPBEh functional is more suitable to use with Hesselmann's dispersion force field, and for this reason we use LRC-ωPBEh for XSAPT(KS)+D and LRC-ωPBE for XSAPT(KS)+D2, with the haTZ basis set in either case. As in our previous work,[55] the parameter s_β in Hesselmann's dispersion potential[16] was optimized in a basis-set-specific way, using the S22A dimer binding energies[38] as benchmarks. The fitted values of s_β for several different basis sets are listed in the Supplementary Material.

In consideration of computational cost, it is not feasible to use a triple-ζ basis set to do a $\delta E_{\rm int}^{\rm HF}$ calculation for large monomers, even with the pairwise-additive approximation used here. We find that the 6-31+G(3d,3pd) basis set affords $\delta E_{\rm int}^{\rm HF}$ estimates comparable to those obtained using triple-ζ basis sets. Comparing this basis set to def2-TZVPP augmented with Pople diffuse sp functions on non-hydrogen atoms, the unsigned difference in the $\delta E_{\rm int}^{\rm HF}$ corrections for S22A data set is only 0.02 kcal/mol. Thus, we use the smaller 6-31+G(3d,3pd) basis set to compute the $\delta E_{\rm int}^{\rm HF}$ corrections throughout this work.

Previously, we showed[55] that XSAPT(KS)+D combined with the partially-augmented aug-cc-pVDZ′ (aDZ′) basis set[157,158] outperforms MP2/CBS and related SCS-MP2/CBS methods, so XSAPT(KS)+D/aDZ′ results are presented here as well, for comparison. The aDZ′ and TZVPP basis sets gave the best results for the XSAPT(HF) method,[54,65] so we also include them in this work.

All XSAPT calculations reported here use smooth ChElPG embedding charges[54]

for the XPol calculations and "projected" (pseudocanonical dimer) basis sets[65] for the SAPT corrections. For supersystem DFT calculations, the def2-QZVP basis set was used. The Euler-Maclaurin-Lebedev (99,590) quadrature grid was used for all semi-local functionals and the (75,302) grid was used for LC-VV10 calculations, except in the case of the $F^-(H_2O)_{10}$ clusters where considerations of cost led us to use a (75,302) grid for the semi-local functionals and a (50,194) grid for LC-VV10.

7.2.4 Computational benchmarks

For dimer benchmarks, we use the CCSD(T)/CBS results in the S22A and S66 data sets.[10,38] (The S22A database revises the energetics of the original S22 database.[186]) Benchmark results for the Ar$\cdots$Ne potential energy curve (dlDF+D_{as}/aTZ level) were taken from Ref. 6. CCSD(T)/CBS potential curves for $(C_6H_6)_2$ were taken from Ref. 5, and for $F^-(H_2O)$ and $Cl^-(H_2O)$ from Ref. 55.

For benchmarks of individual energy *components* for the S22 data set, we use the SAPT2+(3)/aTZ results from Ref. 37, since this is the most accurate SAPT method for S22A binding energies.[117]

For the $F^-(H_2O)_n$ and $Cl^-(H_2O)_n$ clusters ($n \leq 6$), structures were optimized at the MP2/aTZ level of theory and the MP2/CBS correlation energy was estimated using a two-point (aTZ and aQZ) extrapolation. This extrapolated correlation energy was added to the HF/aQZ energy to estimate the MP2/CBS energy. In the $n = 1$ case, a triples correction was added to the MP2/CBS energy, based on the difference between CCSD(T)/aTZ and MP2/aTZ energies, and this corrected result is reported as the CCSD(T)/CBS energy. We also optimized ten isomers of $F^-(H_2O)_{10}$ at the

PM7 level.[285] Resolution-of-identity MP2 (RIMP2) energies in the CBS limit were obtained as indicated above, except that the HF/aTZ and HF/aQZ energies were computed in this case using a dual-basis approximation.[286,287] All of the aforementioned single-point calculations were counterpoise corrected according to the Boys-Bernardi scheme.[74] Coordinates and benchmark energetics for these halide–water clusters are available in the Supplementary Material.

The SAPT2+(3)/aTZ calculations for $F^-(H_2O)$ and $Cl^-(H_2O)$ were performed using the SAPT 2008.2 program[132] with integrals generated by the ATMOL program.[133] The PM7 geometry optimizations were performed using the MOPAC 2012 software.[288] All other calculations were performed using a locally-modified version of Q-Chem.[242,243]

7.3 Results and discussion

7.3.1 S22 and S66 data sets

Interaction energies calculated using XSAPT, traditional SAPT, the effective fragment potential (EFP) method,[275] and several post-HF methods are compared to benchmark CCSD(T)/CBS results for the S22A data set in Table 7.1. The mean unsigned error (MUE) for high-level SAPT2+(3)/aTZ calculations is 0.32 kcal/mol,[37] and is just a little bit worse than the SCS(MI)-MP2/CBS and SCS-CCSD/CBS results.[38] The SAPT2+(3) and XSAPT-based results are all better than the EFP and MP2/CBS results. As noted previously,[37] EFP results are quite poor (MUE = 1.79 kcal/mol) when S22 geometries are used, but the MUE is reduced to 0.91 kcal/mol

Table 7.1: Mean unsigned errors (MUEs), in kcal/mol, and percent errors (in parentheses), with respect to CCSD(T)/CBS benchmarks for the S22A data set along with subsets consisting of the hydrogen-bonded dimers, dispersion-dominated dimers, and dimers of mixed influence. All calculations were performed at S22 geometries except one of the EFP results.

Method	H-Bonded	Disp.-bound	Mixed	All S22
XSAPT(HF)/aDZ′	0.60 (6.55)	0.85 (32.06)	0.30 (9.32)	0.60 (16.70)
XSAPT(HF)/TZVPP	0.40 (3.63)	0.78 (21.99)	0.37 (10.09)	0.53 (12.36)
XSAPT(KS)+D/aDZ′	0.73 (5.93)	0.38 (9.87)	0.52 (12.31)	0.53 (9.39)
XSAPT(KS)+D[a]	0.76 (6.54)	0.30 (7.34)	0.32 (7.26)	0.45 (7.06)
XSAPT(KS)+D (AC)[a]	0.62 (5.91)	0.41 (7.95)	0.40 (9.19)	0.47 (7.70)
XSAPT(KS)+D2[a]	0.88 (5.54)	0.99 (15.17)	0.35 (7.24)	0.75 (9.58)
XSAPT(KS)+D2 (AC)[a]	0.72 (4.57)	1.18 (15.88)	0.52 (11.82)	0.82 (10.99)
SAPT2+(3)[b]	0.42 (4.09)	0.41 (8.79)	0.13 (4.64)	0.32 (5.98)
EFP (S22 geoms.)[c]	2.93 (19.53)	1.70 (42.62)	0.76 (17.43)	1.79 (27.26)
EFP (EFP geoms.)[d]	1.97 (14.51)	0.48 (13.10)	0.34 (7.39)	0.91 (11.73)
MP2[e]	0.24	1.69	0.61	0.88
SCS-MP2[e]	1.54	0.55	0.37	0.80
SCS(MI)-MP2[e]	0.30	0.37	0.17	0.28
SCS-CCSD[e]	0.40	0.23	0.08	0.24

[a]Using the ha-TZVPP basis set and the $\delta E_{\text{int}}^{\text{HF}}$ correction. [b]Using the aTZ basis set, from Ref. 37.
[c]From Ref. 37. [d]Using EFP-optimized geometries, from Ref. 37. [e]CBS limit, from Ref. 38.

at EFP geometries. This suggests that the XSAPT results might also improve if XSAPT-optimized geometries were used (an assertion that is supported by some limited finite-difference optimizations[54,65]), but we have not explored this possibility, as analytic gradients for XSAPT are not available.

The XPol procedure was originally developed as a "next generation" force field,[9,289] where it was combined with pairwise Lennard-Jones (LJ) potentials to account for short-range exchange repulsion and long-range dispersion interactions. LJ parameters

were optimized in Ref. 9 at the B3LYP/6-31G(d) level using 105 hydrogen-bonded dimers. However, we find that this XPol-LJ method affords very poor binding energies for the S22A data set, with a MUE of 3.29 kcal/mol. (The method furthermore predicts that three of the dispersion-dominated dimers are not bound at all.) It was argued in Ref. 9 that the LJ parameters are optimized especially for the hydrogen-bonded systems and that XPol-LJ should provide a good description of hydrogen-bonded interactions. However, for the H-bonded subset of S22, the MUE remains large, 3.16 kcal/mol. This is the reason why we need to use SAPT to capture the rest of interaction following an XPol calculation.

The results for XSAPT(KS)+D are better than those for XSAPT(KS)+D2, but one should recall that in developing the original "+D" dispersion potential, we specifically optimized one parameter (s_β) to minimize errors for S22A binding energies. The main source of error in XSAPT(KS)+D2 comes from dispersion-dominated complexes, especially the π-stacked uracil dimer and π-stacked adenine-thymine complex, where the errors are 3 and 4 kcal/mol, respectively, and the primary source of these errors is error in the dispersion component of the potential. For uracil dimer, the D2 dispersion potential overestimates the SAPT2+(3)/aTZ dispersion energy by 1.8 kcal/mol, and for adenine-thymine by 2.0 kcal/mol. Although we do not understand in detail the origin of these errors, it is noteworthy that of the 79 dimers used as a training set for the D2 dispersion potential, the only π-stacked systems are the sandwich isomer of benzene dimer (for which our method performs very well, as discussed below), and the pyrazine dimer.[244] Thus, some refinement for larger π-stacked

Table 7.2: Mean unsigned errors (MUEs), in kcal/mol, and percent errors (in parentheses), for individual energy components of the S22 data set, with respect to benchmarks computed at the SAPT2+(3)/aTZ level.[37] All calculations were performed at S22 geometries.

Method	Electrostatic	Exchange	Induction	Dispersion
XSAPT(HF)/aDZ′	0.69 (10.74)	2.84 (18.87)	1.92 (61.80)	1.25 (21.95)
XSAPT(HF)/TZVPP	0.41 (7.35)	2.16 (13.43)	1.70 (58.56)	0.60 (9.04)
XSAPT(KS)+D/aDZ′	0.64 (13.37)	3.34 (25.10)	1.86 (60.20)	1.37 (18.08)
XSAPT(KS)+D[a]	0.35 (7.07)	0.81 (6.16)	0.25 (14.22)	0.55 (7.96)
XSAPT(KS)+D (AC)[a]	0.37 (7.54)	0.98 (7.77)	0.20 (11.30)	0.62 (8.52)
XSAPT(KS)+D2[a]	0.32 (6.31)	0.57 (4.81)	0.23 (12.01)	0.38 (6.68)
XSAPT(KS)+D2 (AC)[a]	0.36 (7.16)	0.72 (6.46)	0.23 (11.11)	0.38 (6.68)
EFP[b]	2.03 (34.26)	2.29 (16.53)	1.71 (50.73)	0.80 (12.61)

[a]Using the ha-TZVPP basis and the $\delta E_{\text{int}}^{\text{HF}}$ correction. [b]From Ref. 37.

systems may be in order.

If we eliminate these two π-stacked systems from the data set, then the MUE for XSAPT(KS)+D2/haTZ (AC) is reduced to 0.34 kcal/mol for the dispersion-dominated subset and 0.54 kcal/mol the entire S22A data set. It is not intuitively obvious why the results for XSAPT(KS)+D are much better than the results for XSAPT(KS)+D2 in the dispersion-dominated subset, since the "+D2" dispersion potential is fit directly to dispersion energies ($E_{\text{disp}}^{(2)} + E_{\text{exch-disp}}^{(2)}$) computed using SAPT(DFT).

The individual XSAPT energy components (electrostatics, exchange, induction, and dispersion), listed in Table 7.2, provide a clue as to the main sources of error. The MUE for the dispersion energy obtained using XSAPT(KS)+D2 is actually better than the corresponding MUE for XSAPT(KS)+D. As such, the larger errors in

XSAPT(KS)+D2 binding energies for dispersion-dominated dimers must arise from unsatisfactory error cancellation between the different energy components. This unsatisfactory cancellation of errors makes the XSAPT(KS)+D2 results slightly worse than XSAPT(HF) results, as shown in Table 7.1. For XSAPT(KS)/haTZ with empirical dispersion, using the exact AC for different monomers is a little bit worse than using the same ω value for the whole system as compared to their errors with respect to CCSD(T)/CBS benchmark for the S22A data set. The cancellation of errors does not work so well in S22A data set when using exact AC. However, it seems necessary to use the exact AC for different monomers for hydrogen-bonded systems because XSAPT(KS)+D2 (AC) affords a lower MUE (0.72 kcal/mol) than does XSAPT(KS)+D2 (0.88 kcal/mol) for the H-bonded subset of S22A.

For the XSAPT(HF)/aDZ′, XSAPT(HF)/TZVPP, and XSAPT(KS)+D/aDZ′ methods, MUEs for the various energy components with respect to SAPT2+(3)/aTZ are quite large as shown in Table 7.2. Insofar as these methods provide good results for binding energies, it is clear that cancellation of errors must play a significant role. XSAPT(KS)/haTZ methods with empirical dispersion potentials and $\delta E_{\rm int}^{\rm HF}$ corrections give very good results for individual energy components, especially XSAPT(KS)+D2. The electrostatic and exchange energies predicted by XSAPT(KS)+D and XSAPT(KS)+D2 with exact AC are a bit worse than the corresponding results without using exact AC, or in other words, a bit farther from the SAPT2+(3) results.

SAPT2+(3)/aTZ was selected as a benchmark for the individual energy components because it is the most accurate version of SAPT for the S22A binding energies.[117] It is possible, however, that SAPT2+(3)/aTZ energy *components* do not accurately represent the exact values, owing to truncation of either the basis set or the perturbation series. For example, intramolecular electron correlation contributions to the electrostatic and exchange energies in SAPT2+(3) are truncated at

$$\epsilon^{(1)}_{\mathrm{elst,resp}}(3) = E^{(12)}_{\mathrm{elst,resp}} + E^{(13)}_{\mathrm{elst,resp}} \tag{7.12}$$

and

$$\epsilon^{(1)}_{\mathrm{exch}}(2) = E^{(11)}_{\mathrm{exch}} + E^{(12)}_{\mathrm{exch}} \, , \tag{7.13}$$

respectively. However, the convergence behavior is improved if these are replaced by $\epsilon^{(1)}_{\mathrm{elst,resp}}$(CCSD) and $\epsilon^{(1)}_{\mathrm{exch}}$(CCSD), computed at the CCSD level.[235,290,291] Due to the high computational cost for some of the larger dimers in the S22 data set, we don't consider this possibility in the present work, and treat SAPT2+(3)/aTZ is a good benchmark.

Examining the EFP results in Table 7.2 (which are taken from Ref. 37), we observe that the MUEs for the individual energy components are very large. For two of the dispersion-dominated complexes (π-stacked benzene dimer and π-stacked indole-benzene), the electrostatic interactions are even predicted to be repulsive by EFP.[37] On average, EFP underestimates electrostatic and induction energies in almost all strongly hydrogen-bonded complexes, by several kcal/mol, probably owing to insufficient capture of charge penetration by the screening function that is applied to the multipolar electrostatics.[37] Furthermore, neglect of the charge-transfer term

in the EFP potentials is another source of error for induction energies, especially for hydrogen-bonded complexes.[37] The exchange energies are always underestimated by EFP, which may be caused by neglect of intramolecular correlation effects that are captured by SAPT.[37] Of the various EFP energy components, the dispersion energy best agrees with SAPT2+(3) results, although Podeszwa's (+D2) dispersion potentials provide better agreement.

Another energy decomposition scheme based on the supermolecular method was proposed long ago by Kitaura and Morokuma,[292] and is conventionally known as energy decomposition analysis (EDA). Although the definitions of different energy components in the SAPT and EDA methods are different, they share a common term: electrostatic interaction. Given the same molecular geometry and level of theory, the electrostatic term calculated by SAPT should be similar to the corresponding term computed by EDA. For the S22A data set, the MUE of the electrostatic term calculated by EDA at the BLYP-D3/TZ2P level is 0.32 kcal/mol,[293] quite similar to XSAPT(KS)/haTZ results with empirical dispersion. In short, our interaction-energy decomposition scheme is accurate and can be extended to many-body systems that are not amenable to traditional SAPT energy decomposition.

Table 7.3 presents errors in binding energies, as compared to CCSD(T)/CBS benchmarks, for the S66 data set.[10] This set includes 66 weakly-bound dimers representing binding motifs commonly found in biomolecular structures, and is thought to be more balanced than S22 with respect to different types of interactions. Although the original S66 binding energies[10] have been revised,[35] the revised values are only

Table 7.3: Mean unsigned errors (MUEs), in kcal/mol, and percent errors (in parentheses), with respect to CCSD(T)/CBS benchmarks for the S66 data set along with subsets consisting of the hydrogen-bonded dimers, dispersion-dominated dimers, and dimers of mixed influence. All calculations except the EFP ones were performed at S66 geometries.

Method	H-Bonded	Disp.-bound	Mixed	All S66
XSAPT(HF)/aDZ′	0.26 (3.46)	0.83 (27.66)	0.45 (11.32)	0.54 (15.27)
XSAPT(HF)/TZVPP	0.36 (6.07)	0.44 (14.51)	0.36 (7.48)	0.39 (9.86)
XSAPT(KS)+D/aDZ′	0.28 (4.00)	0.22 (9.06)	0.34 (8.01)	0.27 (7.04)
XSAPT(KS)+D[a]	0.46 (5.13)	0.38 (13.46)	0.35 (7.50)	0.40 (9.11)
XSAPT(KS)+D (AC)[a]	0.37 (3.98)	0.34 (12.25)	0.34 (7.75)	0.35 (8.28)
XSAPT(KS)+D2[a]	0.49 (5.55)	0.34 (11.13)	0.47 (8.33)	0.42 (8.51)
XSAPT(KS)+D2 (AC)[a]	0.34 (3.15)	0.31 (10.06)	0.53 (9.69)	0.38 (7.56)
EFP[b]	0.79	0.65	0.35	0.61
B97-D3[c]	0.26	0.32	0.19	0.26
BLYP-D3[c]	0.40	0.46	0.31	0.39
B2PLYP-D3[c]	0.50	0.13	0.15	0.26
M06-2X[c]	0.24	0.35	0.25	0.28
ωB97X-D[d]	0.16	0.58	0.24	0.33
MP2[e]	—	—	—	0.45
SCS-MP2[e]	—	—	—	0.74
SCS(MI)-MP2[e]	—	—	—	0.28
CCSD[e]	—	—	—	0.62
SCS-CCSD[e]	—	—	—	0.15

[a] Using ha-TZVPP and the $\delta E_{\text{int}}^{\text{HF}}$ correction. [b] Using EFP-optimized geometries, from Ref. 37. [c] Using def2-QZVP, from Ref. 11. [d] Using 6-311++G(3df,3pd), from Ref. 13. [e] CBS limit, from Ref. 10.

marginally different; the MUE between the original and the revised binding energies is 0.08 kcal/mol. We use the original binding energies here, to facilitate comparison to some published *ab initio* and DFT results.[10,11] As shown in previous work,[55] the XSAPT(KS)+D/aDZ′ method affords good cancellation of errors and gives very good results for S66 (MUE = 0.27 kcal/mol). A few supersystem DFT methods have been identified that yield comparable MUEs,[11] including M06-2X, ωB97X-D, BLYP-D3, and B97-D3, and also the double-hybrid B2PLYP-D3 method, and S66 error statistics for these methods are listed in Table 7.3. (Note, however, that a quadruple-ζ basis set was used in most of these DFT benchmarks,[11] so these DFT calculations are considerably more expensive than the XSAPT(KS)+D calculations.) For the -D and -D3 functionals, the classical dispersion potential is crucial to achieving MUEs of $\sim$0.3 kcal/mol. In the case of B2PLYP, for example, the MUE is reduced from 1.60 to 0.26 kcal/mol when the D3 correction is added.[11]

For XSAPT(KS)/haTZ, in both its -D and -D2 form, we obtain MUEs of $\sim$ 0.4 kcal/mol when the $\delta E_{\rm int}^{\rm HF}$ correction is applied. This is superior to MP2, SCS-MP2, and CCSD results in the CBS limit, and superior also to EFP results. For XSAPT(KS)+D2/haTZ (AC), the maximum error is about 3 kcal/mol for the π-stacked uracil dimer. The same system affords the maximum error in the S22 data set, and the MUE for S66 is reduced to 0.33 kcal/mol if we eliminate this problematic complex. Using the exact AC for XSAPT not only reduces the MUE for the H-bonded subset, as in the S22A data set, but also lowers the overall MUE as well. This suggests that using the exact AC (*i.e.*, different ω values for different monomers) is important

in general purpose XSAPT calculations, as is found also in SAPT(DFT) and DFT-SAPT calculations.[162,163]

7.3.2 Potential energy curves

The error statistics for S22A and S66 binding energies demonstrate that XSAPT(KS) methods with empirical dispersion potentials provide a good description of a wide variety of non-covalent interactions at their van der Waals (vdW) minima. It is important also to know how these methods perform across the whole range of intermolecular distances, so in this section we examine some one-dimensional potential energy curves (PECs).

The Ar$\cdots$Ne interaction potential is thought to be a difficult case since several DFT methods, which are recommended for non-covalent interactions and which predict an accurate value of the Ar$\cdots$Ne binding energy at the CCSD(T) vdW minimum, afford qualitatively *incorrect* PECs for this system.[6,55] Figure 7.1 compares the Ar$\cdots$Ne interaction potentials given by various methods. XSAPT(KS)+D provides a good description of the whole range PECs for this system, with XSAPT(KS)/haTZ + $\delta E_{\text{int}}^{\text{HF}}$ affording especially good results, although using the exact AC leads to slight overestimation of the interaction energy in the vicinity of the vdW minimum.

In Fig. 7.1(b), we also compare XSAPT(KS)+D2/haTZ to DFT results obtained using the def2-QZVP basis set. The DFT methods selected for this comparison have been shown to yield accurate binding energies for the dispersion-dominated subset of S66.[11,12] Of these DFT methods, M06-2X overestimates the well depth at the minimum-energy geometry, as demonstrated previously.[6] B2PLYP-D3, which has

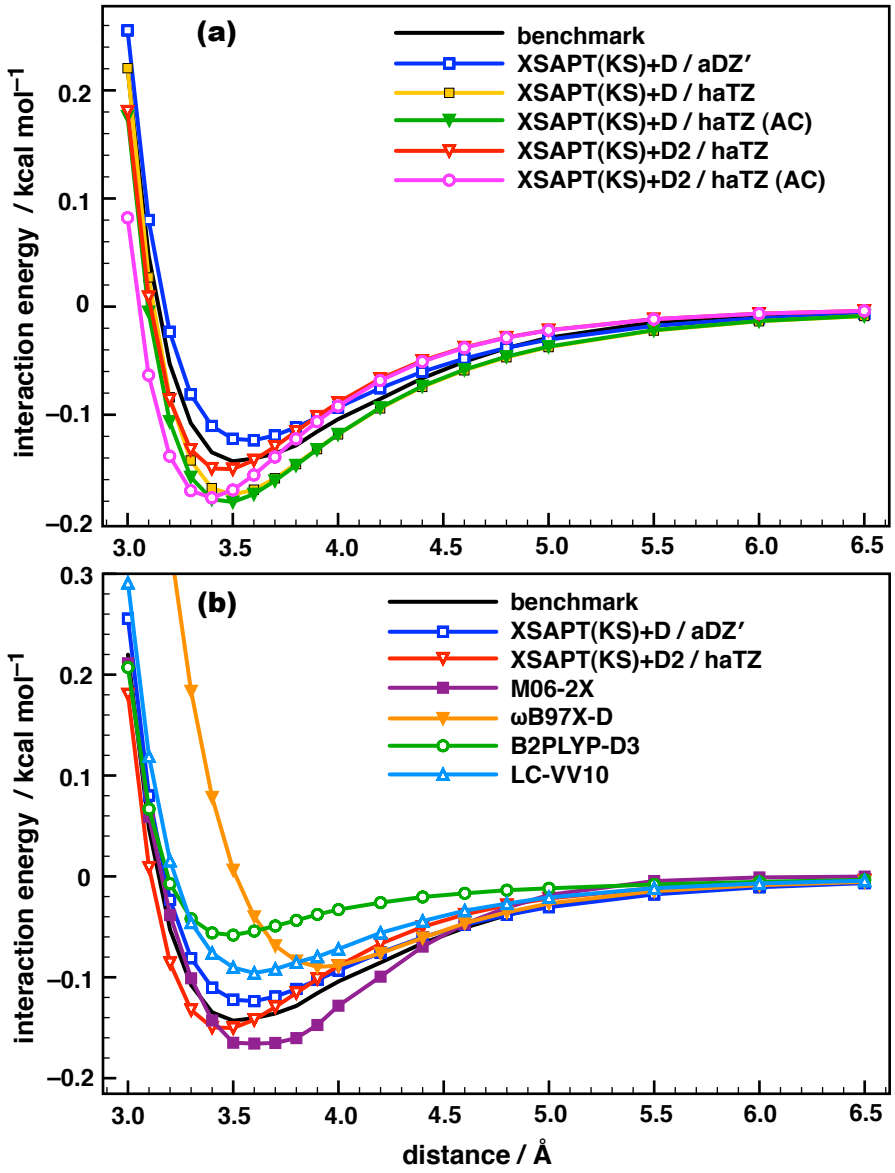

Figure 7.1: Potential energy curves for Ar$\cdots$Ne: (a) comparison of various XSAPT methods, and (b) comparison of XSAPT and DFT methods. Benchmark results were computed using a "dispersionless" density functional (dlDF) augmented with a dispersion correction (the dlDF+D_{as}/aTZ method of Ref. 6). The $\delta E_{\rm int}^{\rm HF}$ correction is included in the XSAPT(KS)/haTZ methods with empirical dispersion potentials. For DFT methods, the def2-QZVP basis set is used Boys-Bernardi counterpoise correction is applied.

the smallest MUE of any DFT-based method for the dispersion-dominated subset of S66 (see Table 7.3), affords an Ar$\cdots$Ne well that is much too shallow. The popular ωB97X-D functional[201] shifts the minimum-energy distance significantly as compared to other methods; the ωB97X-D minimum is located at a distance that is 0.4 Å too large, and the well depth is too shallow as well. The non-local correlation functional LC-VV10 yields a more accurate PEC than either M06-2X or ωB97X-D, consistent with results for the S66 data set, where the MUE for the dispersion-dominated subset is just 0.1 kcal/mol using LC-VV10.[12] However, XSAPT(KS)+D2/haTZ is clearly superior across the entire PEC.

For the remaining homo-monomer systems considered in this work (benzene dimer and water clusters), the XSAPT(KS)+D2/haTZ methods with and without exact AC are the same. Furthermore, the XSAPT(KS)+D/haTZ with and without exact AC affords are very similar results, since these two methods differ only by way of slightly different s_β parameters in the dispersion potential. For brevity, we thus limit our discussion to XSAPT(KS)/haTZ results without exact AC for homo-monomer systems. For the halide–water clusters (hetero-monomer systems), however, we will only show the results *with* exact AC, since this improves the binding energies, consistent with results above for H-bonded systems.

Benzene dimer is considered to be a stringent test of dispersion interactions and we consider both "sandwich" and "T-shaped" isomers, as shown in Fig. 7.2. For the sandwich isomer, XSAPT(KS)+D2/haTZ reproduces the whole CCSD(T)/CBS potential curve almost quantitatively. For the T-shaped isomer, XSAPT(KS)+D2/

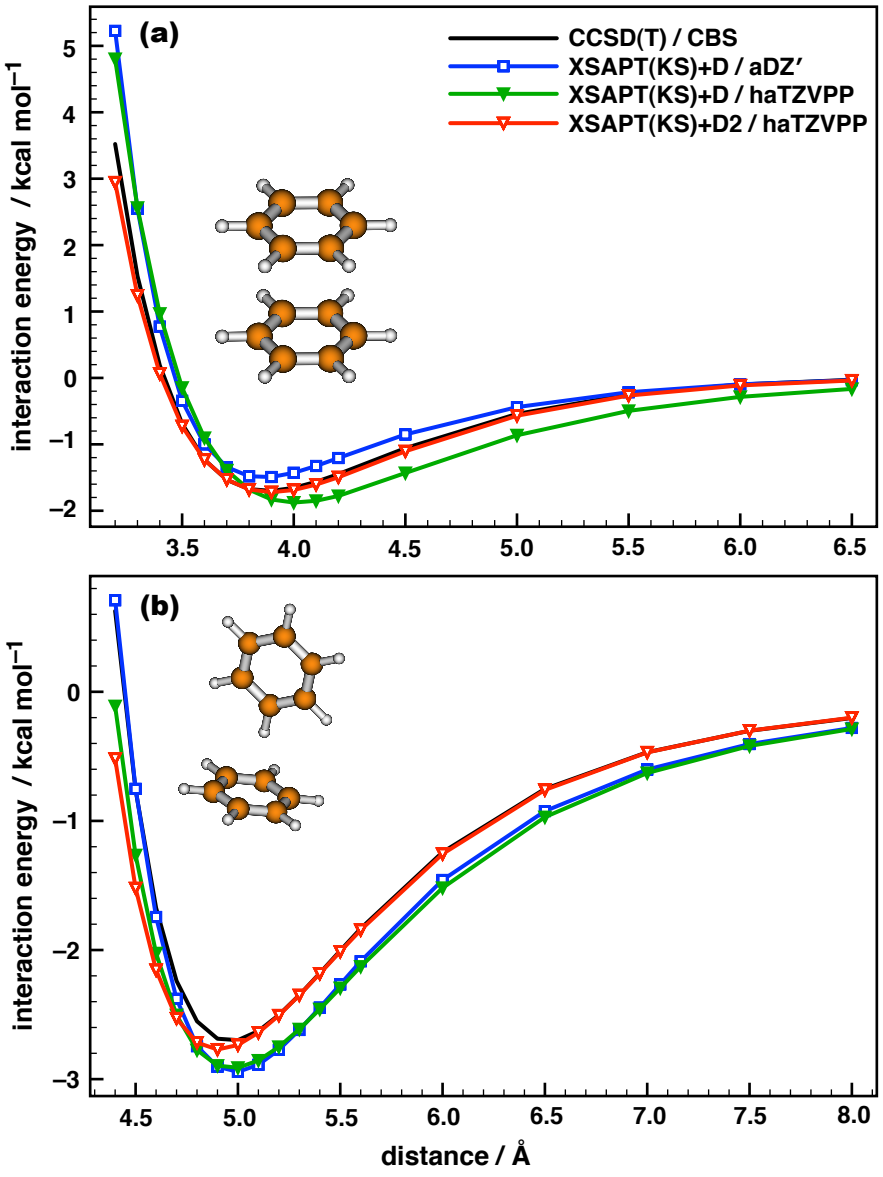

Figure 7.2: Potential energy curves for (a) the "sandwich" and (b) the "T-shaped" isomer of $(C_6H_6)_2$. The distance coordinate in either case is the center-to-center distance between the benzene rings. Benchmark CCSD(T)/CBS results are taken from Ref. 5. The $\delta E_{\rm int}^{\rm HF}$ correction is included in the two XSAPT(KS)/haTZ calculations that include empirical dispersion.

haTZ slightly underestimates the binding energy at short intermolecular distance but is very accurate beyond the vdW minimum. The new +D2 dispersion potentials outperform the old ones for this system.

In Fig. 7.3, we plot PECs for $F^-(H_2O)$ and $Cl^-(H_2O)$, which are known to be challenging cases for both XSAPT as well as traditional SAPT,[50,55] even when third-order corrections are included in the latter.[50] Using the exact AC improves the binding energies by about 0.1 kcal/mol for $F^-(H_2O)$ and 0.3 kcal/mol for $Cl^-(H_2O)$. Values of ω, optimized according to Eq. (7.11), are 0.500, 0.475, and 0.375 bohr^{-1} for H_2O, F^-, and Cl^-, respectively, and the similarity between the H_2O and F^- values explains why use of exact AC has a smaller effect for $F^-(H_2O)$ than for $Cl^-(H_2O)$. This provides further evidence that the exact AC afforded by the IP condition in Eq. (7.11) is necessary, especially for H-bonded systems.

For $F^-(H_2O)$, the XSAPT(KS)+D/aDZ′ and XSAPT(KS)+D/haTZ (AC) methods greatly overestimate the interaction energy at short distance. The second-generation method is significantly better, and in fact is in nearly quantitative agreement with CCSD(T)/CBS results except at monomer separations much smaller than the vdW distance, where it becomes a bit too repulsive. In previous work,[50] we hypothesized that the poorly-shaped PECs for $F^-(H_2O)$ that were obtained using XSAPT(KS)+D might originate in a low-energy $FH\cdots OH^-$ diabatic state[261] that simply cannot be captured by monomer-based methods. While such a state may ultimately lead to other problems, we see in Fig. 7.3 that most of the error in the shape of the PEC is eliminated when the +D dispersion potential is replaced by +D2, and the $\delta E_{\text{int}}^{\text{HF}}$ is

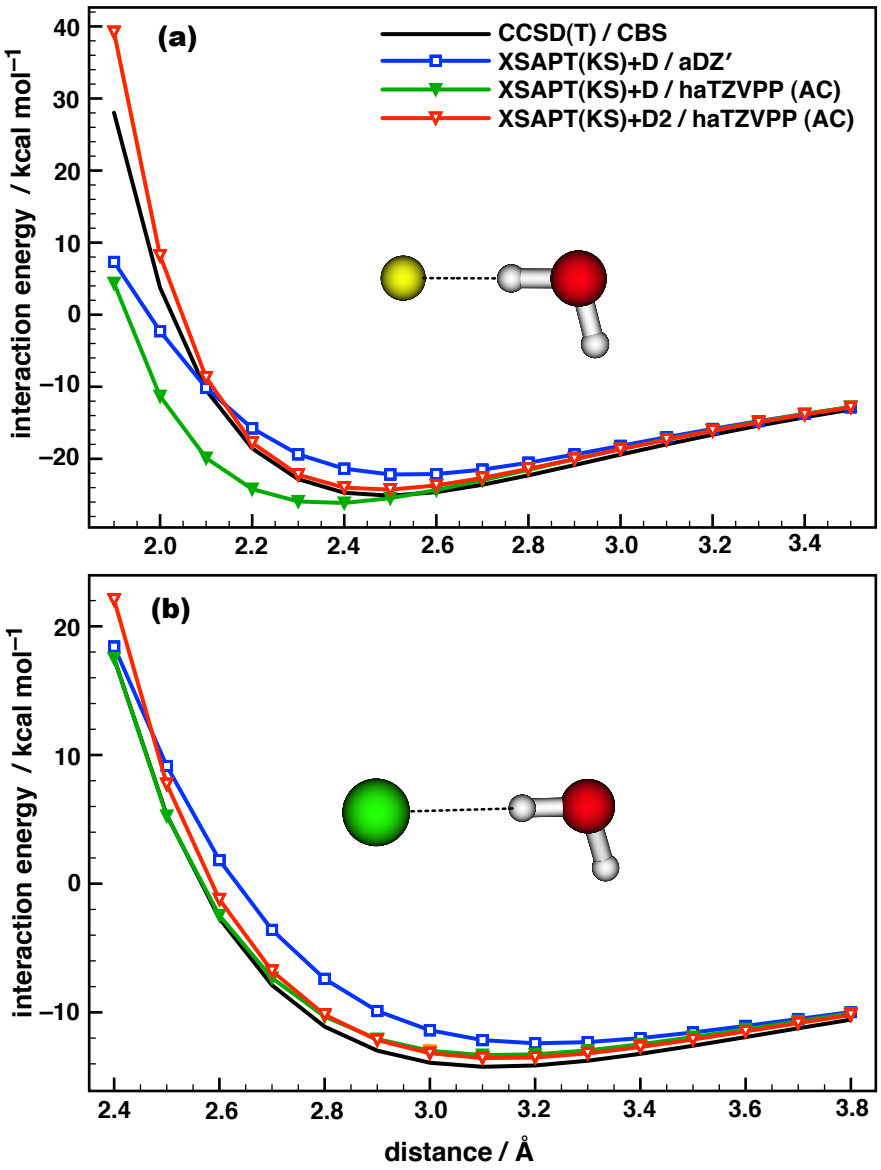

Figure 7.3: Potential energy curves for (a) $F^-(H_2O)$ and (b) $Cl^-(H_2O)$ at a fixed H_2O geometry. The distance coordinate is the halide–oxygen distance and the benchmark is CCSD(T)/CBS. The $\delta E_{\text{int}}^{\text{HF}}$ correction is included in the XSAPT(KS)/haTZ methods with empirical dispersion potentials.

included as well.

7.3.3 Water hexamer

The above results demonstrate the excellent performance of XSAPT(KS)+D2/haTZ (AC) for dimers, including some very challenging ones. We next investigate whether this good performance extends to many-body systems composed of polar monomers, where non-pairwise–additive effects are important. We use $(H_2O)_6$ as our test system since it marks a turning point from two- to three-dimensional hydrogen-bonding networks and has been widely studied. We chose eight low-lying isomers from Ref. 8, where binding energies were computed at the MP2/a5Z-h level with counterpoise corrections. (The basis set is aug-cc-pV5Z with h functions removed.)

Figure 7.4 compares the binding energies of these eight isomers, computed with various methods. Of the XSAPT methods, XSAPT(KS)+D/aDZ′ overestimates the binding energies with respect to the MP2/a5Z-h results, and furthermore makes rather large errors in the relative energies and the energetic ordering of the isomers. We have seen previously that triple-ζ basis sets are often important for H-bonded systems; using XSAPT(KS)+D with the haTZ basis set leads to binding energies that are underestimated rather than overestimated, although the relative energies are improved. In all, however, the +D results cannot be said to be quantitative for water clusters, but the +D2 potential affords good results for both absolute binding energies and relative isomer energies, except for a slight overstabilization of the cyclic-chair isomer. The maximum energy difference between XSAPT(KS)+D2/haTZ and MP2/a5Z-h is just 0.3 kcal/mol. The contribution from $\delta E_{\rm int}^{\rm HF}$ is about 10 kcal/mol and is essential

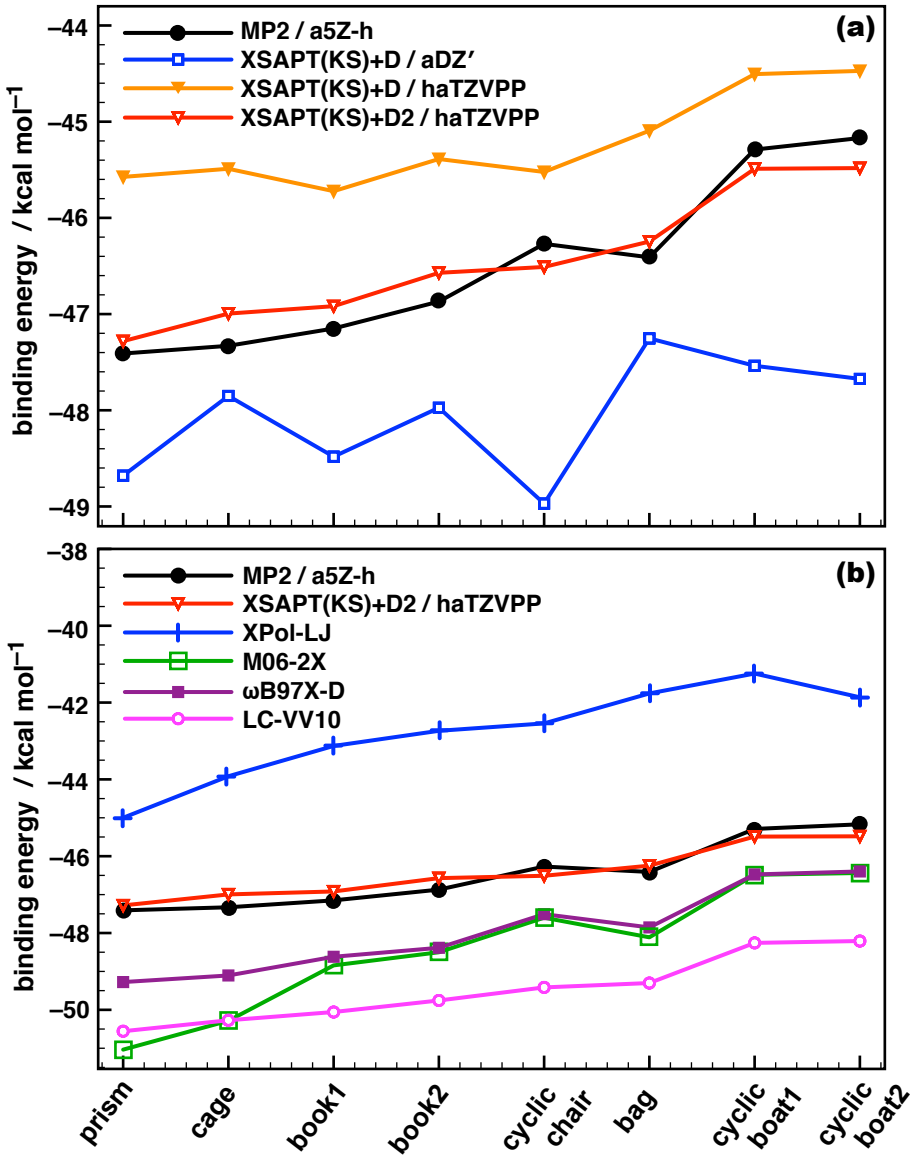

Figure 7.4: Binding energies for eight isomers of $(H_2O)_6$, computed using (a) XSAPT methods and (b) XSAPT, XPol-LJ, and DFT methods. MP2/a5Z-h results are taken from Ref. 8. LJ parameters for XPol-LJ are taken from Ref. 9 where they were optimized for H-bonded systems. The def2-QZVP basis set with counterpoise correction is used for all DFT methods, and the $\delta E_{\text{int}}^{\text{HF}}$ correction is included in the XSAPT(KS)/haTZ methods with empirical dispersion.

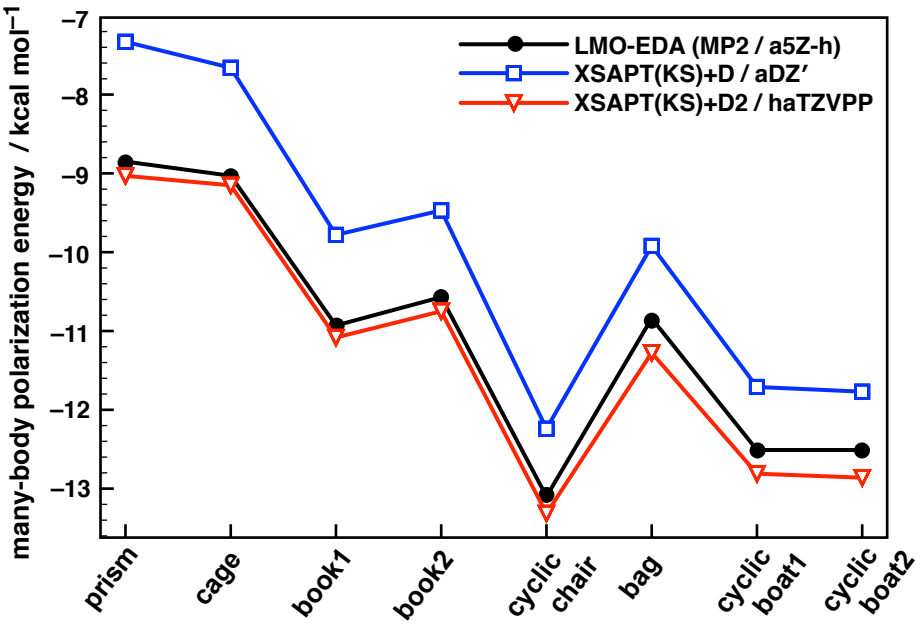

Figure 7.5: Many-body polarization energies for eight $(H_2O)_6$ isomers. MP2/a5Z-h results are taken from Ref. 8. The $\delta E^{\mathrm{HF}}_{\mathrm{int}}$ correction is included in the XSAPT(KS)+D2/haTZ calculations.

for obtaining good results. This is consistent with recommendations to use $\delta E^{\mathrm{HF}}_{\mathrm{int}}$ for SAPT calculations involving polar monomers.[117,130,131] However, the $\delta E^{\mathrm{HF}}_{\mathrm{int}}$ correction is not suitable for use with XSAPT(KS)+D/aDZ′ since this method's success rests on cancellation of errors.

In Fig. 7.4(b) we compare XSAPT to M06-2X and ωB97X-D results, as these two functionals performed best for the H-bonded subset of S66, with MUEs of $\approx$ 0.2 kcal/mol (see Table 7.3). For $(H_2O)_6$, however, the errors are much larger. M06-2X overestimates the binding energies by up to about 4 kcal/mol, and ωB97X-D

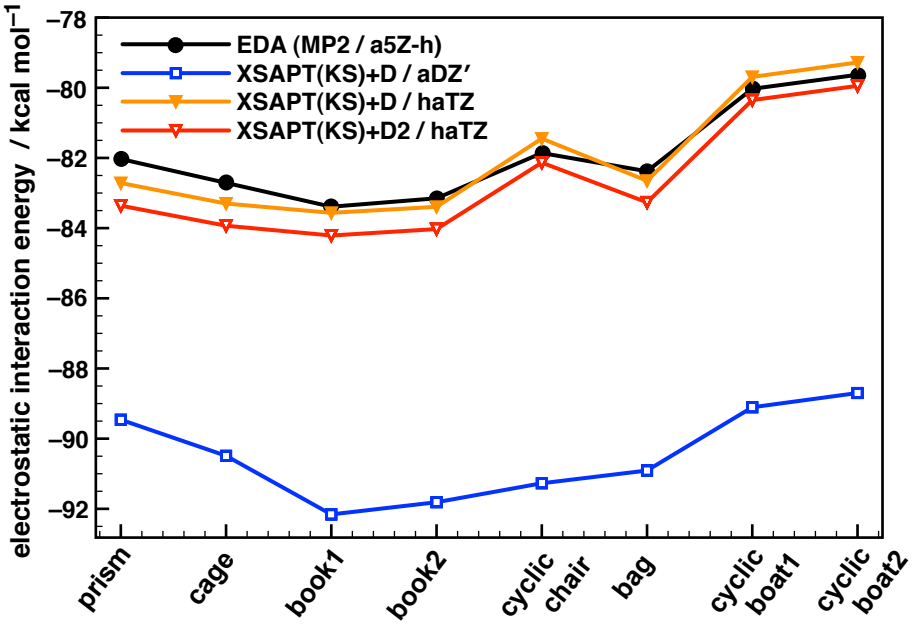

Figure 7.6: Electrostatic interaction energy for eight $(H_2O)_6$. MP2/a5Z-h results are taken from Ref. 8.

by up to about 2 kcal/mol, consistent with another recent study.[294] The LC-VV10 functional also consistently overestimates the binding energies, by about 3 kcal/mol, although the relative energies are quite good. Despite being parameterized for H-bonded complexes,[9] the XPol-LJ results are quite poor, with binding energies that are underestimated by up to about 5 kcal/mol. Unlike these other methods, the performance of XSAPT(KS)+D2/haTZ for dimers does extend to this larger system.

The importance of many-body effects in $(H_2O)_6$ has been quantified recently by Chen and Li,[8] at the MP2/a5Z-h level, using localized molecular orbital energy

decomposition analysis (LMO-EDA), which is a simplified version of the Kitaura-Morokuma EDA scheme.[292] These authors point out that the many-body effects are dominated by polarization interactions, whereas the other energy components are strictly or nearly pairwise additive.[8] Figure 7.5 compares the many-body polarization energies for the $(H_2O)_6$ isomers considered here, computed using LMO-EDA at the MP2/a5Z-h level[8] or using our XSAPT-based energy-decomposition scheme, the latter of which assumes that the many-body energy arises exclusively from polarization. The +D and +D2 versions of XSAPT(KS)/haTZ give very similar results for the many-body contribution to the energy, so only the latter are shown, for clarity. These results are in good agreement with the LMO-EDA results, whereas the XSAPT(KS)+D/aDZ′ method consistently underestimates the many-body energy.

Electrostatic energies from LMO-EDA and from XSAPT(KS)-based methods are shown in Fig. 7.6. The XSAPT(KS)/aDZ′ method greatly overestimates the electrostatic energies, as compared to values extracted from MP2/a5Z-h calculations, whereas XSAPT(KS)/haTZ methods with empirical dispersion potentials show good agreement with LMO-EDA for the electrostatic energy. This is consistent with our previous conclusions that triple-ζ basis sets are necessary for hydrogen-bonded systems.[54,55]

7.3.4 Halide–water clusters

The SAPT method does not afford chemical accuracy for strongly-interacting systems, especially ions at short intermolecular distances, where the perturbation series converges slowly or may even diverge.[50,170] As demonstrated in Fig. 7.3 and also Ref.

Table 7.4: Binding energies (in kcal/mol) for $F^-(H_2O)$ and $Cl^-(H_2O)$.

Method	$F^-(H_2O)$	$Cl^-(H_2O)$
CCSD(T)/CBS	−32.364	−15.540
MP2/CBS	−32.248	−15.748
M06-2X[a]	−35.673	−16.298
ωB97X-D[a]	−33.729	−15.482
LC-VV10[a]	−35.010	−16.492
SAPT0/aDZ′	−33.926	−14.671
SAPT0/aTZ	−37.571	−16.524
SAPT2+(3)/aTZ	−34.072	−15.533
XSAPT(KS)+D/aDZ′	−27.437	−13.516
XSAPT(KS)+D[b]	−36.217	−14.627
XSAPT(KS)+D (AC)[b]	−36.204	−14.787
XSAPT(KS)+D2[b]	−33.010	−14.686
XSAPT(KS)+D2 (AC)[b]	−33.150	−15.046

[a]Using the def2-QZVP basis set with counterpoise correction. [b]Using the ha-TZVPP basis set and the $\delta E_{\text{int}}^{\text{HF}}$ correction.

55, the first-generation XSAPT(KS)+D/aDZ′ method does not afford good results for $F^-(H_2O)$ or $Cl^-(H_2O)$, as PECs for these systems exhibit qualitatively incorrect shapes at short distance. This problem is resolved by the D2 dispersion potential, however, and the halide–water PECs have reasonable shapes and the binding energies are accurate, as demonstrated above. In this section, we extend these anion systems from two-body to many-body systems, investigating binding energies for $X^-(H_2O)_n$ up to $n = 6$. All cluster geometries were optimized at the MP2/aTZ level of theory. Since traditional SAPT is limited to dimers, we first compare the binding energies for $X^-(H_2O)$ dimers (X = F, Cl) calculated using various methods, in Table 7.4.

The difference between CCSD(T)/CBS and MP2/CBS binding energies is only

0.1–0.2 kcal/mol for these dimers, and as such we will use MP2/CBS as a benchmark for larger $X^-(H_2O)_n$ clusters. Examining the DFT results in Table 7.4, we see that both M06-2X and LC-VV10 overestimate the binding energies, by $\sim$3 kcal/mol for $F^-(H_2O)$ and $\sim$1 kcal/mol for $Cl^-(H_2O)$. Addition of the D3 correction proposed by Grimme[274] (M06-2X-D3) leads to even worse results. The ωB97X-D binding energy for $Cl^-(H_2O)$ is quite accurate but that for $F^-(H_2O)$ is overestimated by about 1 kcal/mol. SAPT0/aDZ$'$ binding energies are reasonable (errors $\sim$1 kcal/mol) and are significantly better than SAPT0/aTZ results. This is consistent with SAPT0 results for other systems, where the aDZ$'$ basis set leads to favorable error cancellation,[117,157] which is why this basis set was proposed in the first place.

The high-level SAPT2+(3)/aTZ method affords almost the same binding energy as CCSD(T)/CBS for $Cl^-(H_2O)$ but overestimates the $F^-(H_2O)$ binding energy by 1.7 kcal/mol, consistent with previous results where methods beyond SAPT0 were employed.[50] Coupled to the additional observation that supersystem DFT errors are consistently larger for $F^-(H_2O)$ than they are for $Cl^-(H_2O)$, these results suggest that fluoride–water is an especially challenging test of monomer-based quantum chemistry.

Examining the XSAPT results in Table 7.4, we observe that XSAPT(KS)+D/aDZ$'$ underestimates the $X^-(H_2O)$ binding energy, especially for X = F, but the new XSAPT(KS)+D2/haTZ method performs much better, with errors $<$ 1 kcal/mol when exact AC is used. Once again, this is consistent with the need for triple-ζ basis sets for H-bonded systems.[54,55] In terms of the exact AC (versus using the same ω value for all monomers), results with and without exact AC differ by only

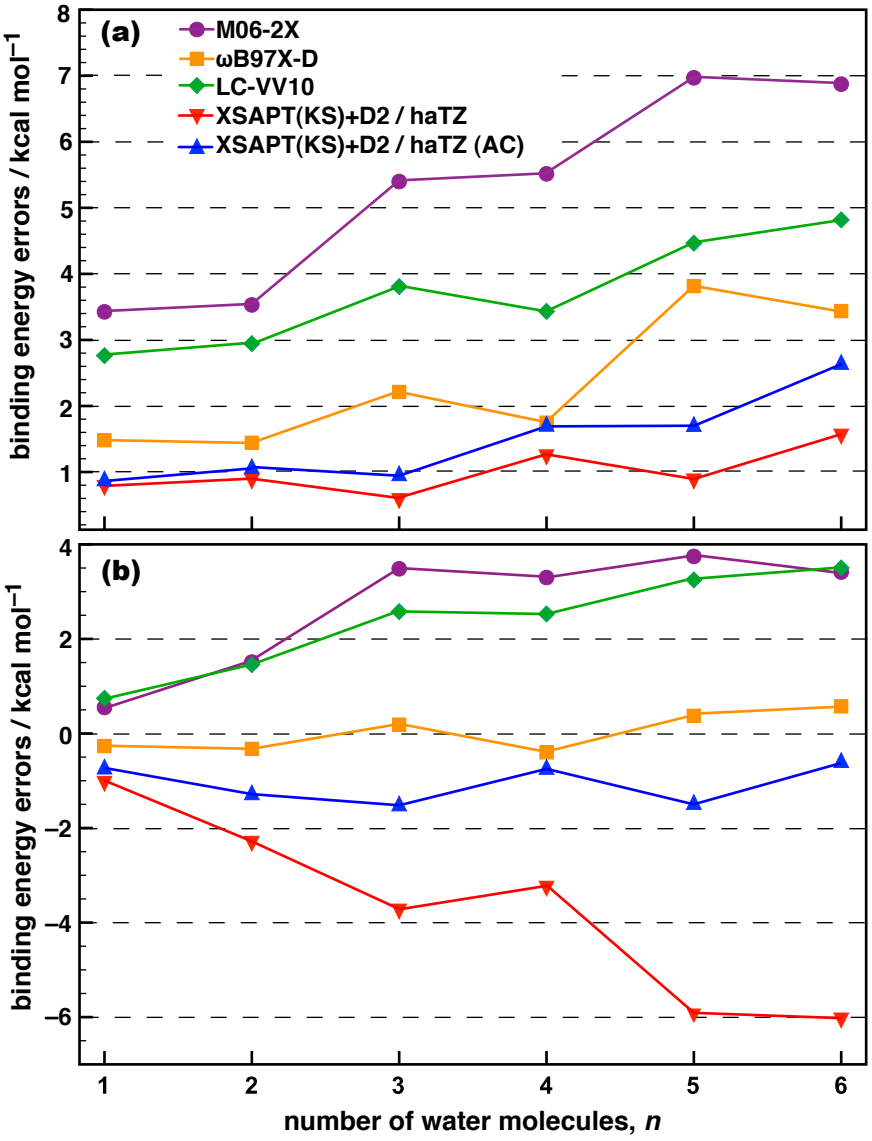

Figure 7.7: Errors in binding energies with respect to MP2/CBS benchmarks for (a) $F^-(H_2O)_n$ and (b) $Cl^-(H_2O)_n$, up to $n = 6$. The def2-QZVP basis set and counterpoise correction was used for all DFT calculations. The $\delta E_{\text{int}}^{\text{HF}}$ correction is included in the XSAPT(KS)+D2/haTZ calculations. Positive and negative errors imply that the binding energies are over- and underestimated, respectively.

Table 7.5: Mean unsigned errors, with respect to RIMP2/CBS benchmarks, for ten isomers of $F^-(H_2O)_{10}$.

Method	MUE / kcal mol^{-1}	
	binding energy	relative energy
M06-2X[a]	9.81	1.340
ωB97X-D[a]	3.71	0.483
LC-VV10[a]	7.91	0.778
XSAPT(KS)+D2[b]	4.03	0.989
XSAPT(KS)+D2 (AC)[b]	1.68	0.751

[a]Using the def2-QZVP basis set with counterpoise correction. [b]Using the haTZ basis set and the $\delta E_{\text{int}}^{\text{HF}}$ correction.

0.14 kcal/mol for $F^-(H_2O)$, but for $Cl^-(H_2O)$ exact AC improves the binding energy by 0.36 kcal/mol.

XSAPT(KS)+D2/haTZ results for $X^-(H_2O)_n$ up to $n = 6$, both with and without exact AC, are compared to MP2/CBS benchmarks and to three DFT methods in Fig. 7.7. Oddly, for $F^-(H_2O)_n$ the XSAPT(KS)+D2/haTZ results are slightly better when the same ω value is used for both F^- and H_2O; differences between the two approaches range up to about 1 kcal/mol and appear to grow slightly larger with cluster size. These discrepancies may be coincidence (or some cancellation of errors), because for $Cl^-(H_2O)_n$, where the ω values optimized for Cl^- and H_2O are quite different, use of the exact AC is crucial. Exact AC in this case reduces the errors from ≈ 6 kcal/mol for $Cl^-(H_2O)_5$ and $Cl^-(H_2O)_6$ down to ~ 1 kcal/mol for all cluster sizes.

Beran[83] has suggested that errors in fragment-based calculations of molecular clusters ought to be extensive, *i.e.*, proportional to the number of monomer units. In applications to water clusters, our group has indeed observed some numerical evidence of a roughly-constant error per hydrogen bond, for clusters with $\gtrsim 5$ hydrogen bonds.[65] The slight uptick in the errors for $F^-(H_2O)_n$ clusters, as a function of increasing cluster size [Fig. 7.7(a)] reminds us of these observations. To investigate this further, we performed calculations on a set of ten isomers of $F^-(H_2O)_{10}$, whose structures were taken from Ref. 71 and then optimized at the PM7 level. MUEs for these $F^-(H_2O)_{10}$ binding energies, with respect to RIMP2/CBS benchmarks, are listed in Table 7.5. Results from several supersystem DFT methods are included in Table 7.5, and in these cases the def2-QZVP basis set and Boys-Bernardi counterpoise correction were used, consistent with other DFT results presented here.

All three DFT methods (M06-2X, ωB97X-D, and LC-VV10) afford very large errors for these $F^-(H_2O)_{10}$ clusters, although errors for relative isomer energies are much smaller. M06-2X is the worst of the bunch, with a MUE of almost 10 kcal/mol and a maximum error of about 12 kcal/mol. None of these methods comes close to achieving "chemical accuracy" in the absolute binding energies of these clusters, although ωB97X-D and LC-VV10 exhibit errors < 0.8 kcal/mol in the relative energies. Errors in the absolute binding energies are much larger than those observed in a recent study of DFT methods applied to $SO_4^{2-}(H_2O)_6$ clusters,[203] confirming the challenging nature of fluoride–water clusters.

In contrast to these supersystem DFT results, the XSAPT(KS)+D2/haTZ (AC)

method exhibits a much smaller MUE (1.7 kcal/mol) for the absolute binding energies, which is much smaller than the corresponding error without exact AC. This MUE is in line with the errors observed in smaller $F^-(H_2O)_n$ clusters, suggesting that the error does not grow appreciably with cluster size, at least in the size regime $n \leq 10$. Admittedly, errors of ≈ 1.7 kcal/mol in the binding energies are larger than what we observe in other systems, and fluoride–water clusters remain challenging even for the improved XSAPT(KS)+D2 method. On the other hand, these systems are challenging for supersystem DFT as well, and probably deserve to be considered more routinely in evaluating the performance of various methods for non-covalent interactions. Our benchmark MP2/CBS results are available in the Supplementary Material.

Although ωB97X-D performs well for $Cl^-(H_2O)_n$, with errors $\lesssim 0.5$ kcal/mol (superior to XSAPT-based methods), the XSAPT-based methods are far superior for $F^-(H_2O)_n$ clusters, where ωB97X-D exhibits errors > 3 kcal/mol for the $n = 5$, $n = 6$, and $n = 10$ cases. The M06-2X and LC-VV10 functionals yield very poor results, even for the comparatively benign chloride–water clusters.

7.4 Conclusions

A second-generation ("D2") version of our XPol+SAPT(KS) method with empirical dispersion has been introduced and tested. An exact asymptotic correction (AC) scheme, in which the AC is optimized separately for each monomer according to the IP criterion, is found to be necessary in general to obtain good binding energies, especially for hydrogen-bonded complexes. The XSAPT(KS)+D2/haTZ method with

exact AC exhibits MUEs of 0.82 and 0.38 kcal/mol for S22A and S66 binding energies, respectively, although two outliers (the π-stacked adenine-thymine and uracil-uracil dimers) suggest that some further refinement of the empirical dispersion potentials may be in order, using data sets that contain additional π-stacked dimers.

A variety of other challenging systems have been considered as well, including $\mathrm{Ar}\cdots\mathrm{Ne}$, $(C_6H_6)_2$, $(H_2O)_6$, $Cl^-(H_2O)_n$, and $F^-(H_2O)_n$. The XSAPT(KS)+D2/haTZ method affords accurate one-dimensional potential energy scans for the dimers and accurate relative energies for the larger clusters. In particular, this method corrects certain qualitative problems in the short-range description of halide–water interaction potentials that were observed in the "first generation" version of the method.[55]

The accuracy of XSAPT(KS)+D2/haTZ (AC) is superior to that of popular DFT approaches for non-covalent interactions, including M06-2X, ωB97X-D, and LC-VV10; since XSAPT is a monomer-based approach, the cost is also much lower than the cost of supersystem DFT.[55] For the very challenging halide–water clusters, XSAPT(KS)+D2/haTZ (AC) binding energies are of moderate accuracy (1–2 kcal/mol errors with respect to MP2/CBS results), but M06-2X and LC-VV10 yield even worse results (errors of 8–12 kcal/mol), as does ωB97X-D for $F^-(H_2O)_n$. We recommend fluoride–water clusters as tests cases for any method intended to capture non-covalent interactions, and our benchmark results for these systems are available in the Supplementary Material.

Finally, we have introduced an interaction-energy decomposition scheme for XSAPT that extends SAPT energy decomposition analysis to many-body systems. The

different energy components (electrostatic, exchange, induction, and dispersion) for the S22A data set are in very good agreement with benchmark SAPT2+(3)/aTZ results, demonstrating that our energy decomposition scheme is robust. Using this energy-decomposition scheme in conjunction with XSAPT(KS)+D2, the many-body contributions to the binding energies of $(H_2O)_6$ isomers are reproduced almost perfectly as compared to benchmark calculations. Therefore, XSAPT(KS)+D2 (AC) not only yields good binding energies for different non-covalent systems, but furthermore we can decompose these binding energy into physical meaningful energy components for many-body systems, which is not possible in traditional SAPT.

CHAPTER 8

Accurate and efficient quantum chemistry calculations for noncovalent interactions in many-body systems: The XSAPT family of methods[8.1]

8.1 Background

8.1.1 Quantum chemistry for non-covalent interactions

Non-covalent or "non-bonded" interactions are responsible for the properties of a variety of complex systems ranging from the structures of both single- and double-stranded DNA,[248] drug binding to both proteins and DNA,[295,296] and also crystal engineering and crystal structure prediction.[297] Electronic structure calculations of non-covalent interactions have seen much progress in recent years, due to improvements in both algorithms and computer power. In particular, symmetry-adapted perturbation theory[114–117,149,150] (SAPT) provides a natural decomposition of non-covalent interactions into physical meaningful components (electrostatics, induction, and dispersion), along with a corresponding exchange term for each. The dispersion

[8.1]This chapter appeared as a full article in the *Journal of Physical Chemistry A*, in 2015, volume 119, pages 235–252.

(van der Waals) interaction is particularly interesting as it is a purely quantum-mechanical effect arising solely from intermolecular electron correlation. Dispersion is therefore absent at the level of Hartree-Fock molecular orbital (MO) theory, and has historically been difficult to describe with density functional theory (DFT) as well, because popular semilocal functionals fail to account for long-range electron correlation.

Various strategies have been devised to incorporate dispersion into DFT, including highly-parameterized meta-GGA[298] functionals where non-bonded interactions are included in the fitting set. The "Minnesota" family of functionals are prime examples of this approach.[142] Alternatively, explicit r^{-2k} dependence ($k = 3, 4, \ldots$) can be added to DFT *a posteriori*, via classical atom-atom potentials, in a "DFT+D" approach popularized by Grimme.[143,254,274] The ωB97X-D functional[201] is one of the best-performing examples of a DFT+D functional. Finally, "double hybrid" functionals that mix second-order Møller-Plesset (MP2) correlation with DFT,[299] and other non-local correlation functionals,[105,147] also do a better job of describing non-covalent interactions as compared to traditional semilocal GGAs. However, the cost of these methods scales no better than $\mathcal{O}(N^3)$ with respect to total system size, N, which limits their routine application to systems with $\lesssim 100$ atoms.

Prior to the advent of these newer DFT-based approaches, the $\mathcal{O}(N^5)$ MP2

method was considered the simplest way to incorporate dispersion in electronic structure calculations. MP2 performs well for hydrogen-bonded systems, and thus continues to play a vital role in the study of water clusters[300] (and recently, bulk liquid water[301]). However, MP2 significantly overestimates π-stacking interactions and other dispersion-dominated interactions.[186,207] This behavior stems from poor effective C_6 coefficients,[302,303] which at the MP2 level correspond to an uncoupled Hartree-Fock (UCHF) description of the frequency-dependent polarizabilities for the monomers. Moreover, slow convergence of the MP2 correlation energy to the complete basis set (CBS) limit requires costly counterpoise correction[80] to eliminate basis set superposition error (BSSE). A self-consistent treatment of double excitations, *i.e.*, the $\mathcal{O}(N^6)$ coupled-cluster singles and doubles (CCSD) method, also underestimates $\pi\cdots\pi$ interactions,[304] and on average represents only a modest improvement upon MP2, with errors of 0.7–1.0 kcal/mol relative to converged CCSD(T) values.[272]

In short, the $\mathcal{O}(N^7)$ CCSD(T) method remains the "gold standard" for noncovalent quantum chemistry, though there is some recent effort to explore quantum Monte Carlo techniques as an alternative.[40,305] CBS extrapolation is required to obtain converged CCSD(T) results, but higher-order electron correlation effects are consistently < 0.1 kcal/mol.[33,141,306] The CCSD(T)/CBS limit can more affordably be obtained by adding a correction

$$\delta_{\mathrm{CCSD(T)}} = E_{\mathrm{CCSD(T)}} - E_{\mathrm{MP2}} \tag{8.1}$$

to the MP2/CBS binding energy, as this correction is generally converged in triple-ζ

basis sets,[34] whereas the MP2/CBS extrapolation requires a basis of at least aug-cc-pVQZ (aQZ) quality. Nevertheless, this approach remains prohibitively expensive except for small systems. For example, a recent CCSD(T)/aTZ calculation on $(H_2O)_{17}$ required 3.3 hours on 120,000 processors simply for the "(T)" part of the calculation.[307]

The performance of various electronic structure methods that have been suggested for non-covalent interactions, and which scale better than $\mathcal{O}(N^7)$ is summarized in Fig. 8.1 for the S66 data set of non-covalent dimers.[10] Among these methods, the MP2/CBS results are actually the worst, and this is a direct result of severe overestimation of π-stacking interactions. The best-performing method is SCS-MI-CCSD (spin-component scaled CCSD for molecular interactions[308]), but its sixth-order scaling is also severely limiting. The MP2C approach[309–312] (MP2 with coupled dispersion) also affords very small errors, with only fifth-order scaling, but this method is formulated exclusively for dimers. Finally, Fig. 8.1 shows selected DFT results using functionals that afford good results for S66; however, this good performance is not transferred to anionic systems such as halide–water clusters, $X^-(H_2O)_n$,[56] except the ωB97X-V function as discussed below.

In view of these remarks, it is clear that quantum chemistry calculations with sub-kcal/mol accuracy remain out of reach for large non-covalent assemblies, such as the HIV protease + inhibitor system that is shown in Fig. 8.2. With a binding pocket consisting of 16 nearby amino acids plus two crystallographic waters, the total system size for a reasonable quantum chemistry model system amounts to 323 atoms,

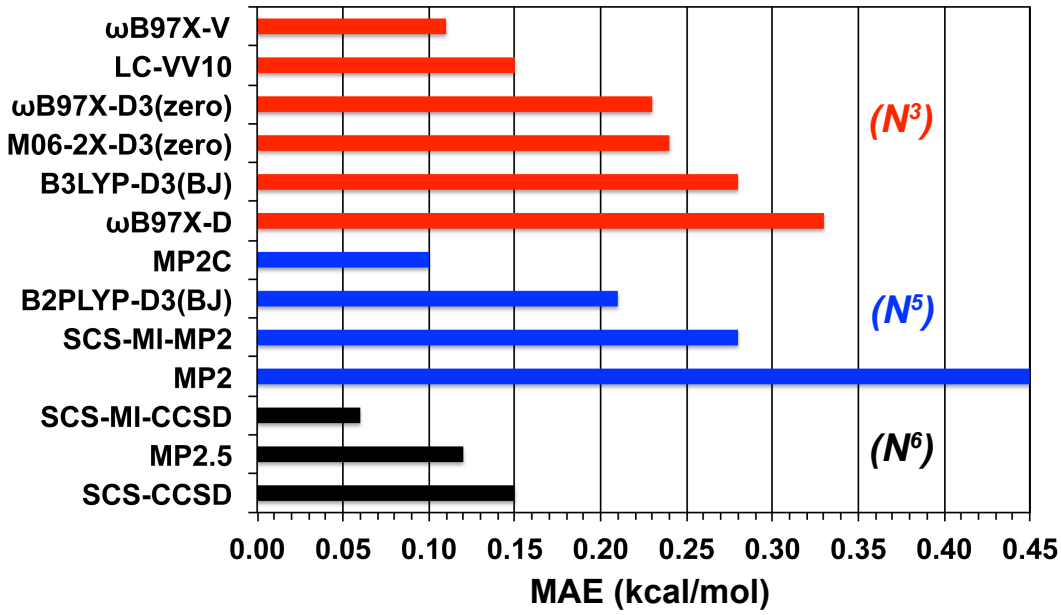

Figure 8.1: Mean absolute error (MAE) with respect to CCSD(T)/CBS benchmarks, for binding energies in the S66 database[10] of non-covalently-bound dimers. The various methods are color-coded according to how their cost scales as a function of system size. All MP2- and CCSD-based results[10] are evaluated in the CBS limit. M06-2X-D3(zero), B3LYP-D3(BJ), and B2PLYP-D3(BJ) calculations employ the def2-QZVP basis.[11] ωB97X-V and LC-VV10 calculations use aug-cc-pVTZ.[12]. The ωB97X-D and ωB97X-D3(zero) calculations use the 6-311++G(3df, 3pd) basis.[13] [Here, "zero" indicates the "zero-damping" function of Ref. 14, which damps empirical dispersion to zero as $R \to 0$, whereas Becke-Johnson (BJ) damping damps it to a finite value.] The counterpoise correction is employed for all methods except M06-2X-D3(zero), B3LYP-D3(BJ), and B2PLYP-D3(BJ).

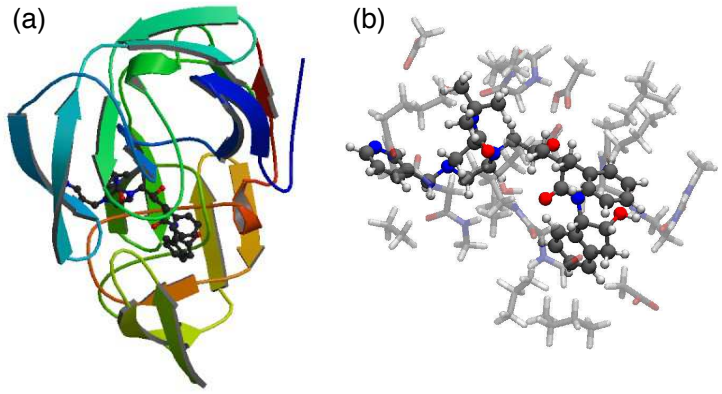

Figure 8.2: (a) Structure of the protease inhibitor indinavir bound to HIV protease, as obtained from PDB crystal structure 1HSG.[1] (b) An enlarged view of the binding pocket, consisting of indinavir (opaque ball-and-stick model) along with 16 amino acids and 2 crystallographic waters (translucent tubular model). [Panel (b) is reproduced from Ref. 2; copyright 2011 American Institute of Physics.]

or 10,626 basis functions using aug-cc-pVTZ. Fragment-based quantum chemistry methods offer a way to surmount this predicament.[43,44] This article describes a family of fragmentation methods that we have developed in an attempt to achieve sub-kcal/mol accuracy for non-covalent interactions, yet remain affordable enough to be applied to systems such as the one in Fig. 8.2(b), where the monomers naturally form fragments but the protease inhibitor molecule is 92 atoms by itself. As such, any method that aims to describe this system must be efficient both for large fragments as well as systems comprised of a large number of fragments.

The XSAPT family of methods that is described in this article has been developed by our group over the past several years.[44,51,54–56,65] These methods employ the variational explicit polarization (XPol) method of Xie *et al.*[252] to generate monomer

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

8.1.2 Symmetry-adapted perturbation theory (SAPT)

It is useful to review the original SAPT methodology.[114–117,149] SAPT is a direct, perturbative expansion of the intermolecular (dimer) interaction energy based on noninteracting monomer wave functions, and BSSE is avoided because subtraction of monomer energies is not required. A double perturbation expansion is employed, in which *intra*molecular electron correlation (Møller-Plesset fluctuation potentials[114,151] or a cluster *ansatz*[154–156]) is one perturbation, and the intermolecular Coulomb operators comprise the other. Methods that include intramolecular correlation, such as SAPT2+, SAPT2+(3), SAPT2+3, and SAPT(CCSD), are generally quite accurate[117,172] but scale no better than $\mathcal{O}(N^7)$, the same as CCSD(T). However, an accurate SAPT calculation may be able to use a smaller basis set as compared to CCSD(T), owing to SAPT's intrinsic lack of BSSE, and furthermore SAPT comes with an informative energy decomposition.

A comparatively low-cost means to introduce intramolecular electron correlation is to swap Kohn-Sham (KS) MOs into the SAPT formalism, an approach that has been called SAPT(KS).[208] This approach was considered and rejected more than a decade ago, however, because it fails to yield accurate intermolecular interaction energies due to inaccurate asymptotic behavior of the exchange-correlation (XC) potentials,

$v_{\rm xc}(r)$, that are obtained from standard functionals.[162,208,209] The asymptotic (large r) behavior should be[210]

$$v_{\rm xc}(r) \sim -\frac{1}{r} + \Delta_\infty \ , \tag{8.2}$$

where the limiting ($r \to \infty$) value is[210,211]

$$\Delta_\infty = {\rm IP} + \varepsilon_{\rm HOMO} \ . \tag{8.3}$$

In Eq. (8.3), IP denotes the lowest ionization potential and $\varepsilon_{\rm HOMO}$ is the KS eigenvalue for the highest occupied molecular orbital (HOMO). The failure of standard density-functional approximations to satisfy Eq. (8.2) leads to large errors in SAPT(KS) dispersion energies,[54] even while the energetics of strongly hydrogen-bonded systems are somewhat improved as compared to calculations that use HF wave functions for the monomers.[54]

This failure of SAPT(KS) is partially ameliorated by using an asymptotically correct $v_{\rm xc}$ to compute KS orbitals for the monomers,[51,162–164] though dispersion energies remain quite poor[51,209,213,214] for the same reason that MP2 dispersion energies are poor. It is possible to solve coupled KS equations to obtain frequency-dependent density susceptibilities for the monomers, and this improved approach was developed independently by Heßelmann and Jansen,[118,137,163,164,213] who named the method DFT-SAPT, and by Misquitta *et al.*,[161,162,205,214,235,313] who called it SAPT(DFT). This approach, which is closely related to MP2C, scales as $\mathcal{O}(N^6)$ but can be reduced to $\mathcal{O}(N^5)$ using resolution-of-identity techniques.[137,161,205,313]

8.1.3 "Extended" SAPT (XSAPT)

Most electronic structure methods for intermolecular interactions function either as supersystem calculations, or else (like SAPT) are designed to compute pairwise interactions only. Benchmark non-covalent data sets have also largely focused on dimers. However, many-body contributions to the interaction energy are often quite significant, especially in polar systems where non-additive interactions are dominated by induction (*i.e.*, polarization).[8,251] For example, the many-body contribution to polarization in isomers of $(H_2O)_6$ is about 10 kcal/mol, whereas electrostatic, exchange-repulsion, and dispersion interactions are nearly pairwise additive.[8,56] Likewise, electron correlation effects have often been found to be largely pairwise-additive in small clusters of small molecules, provided that many-body induction effects are incorporated self-consistently,[84] but many-body dispersion is more important in systems with a large number of monomers.[314]

Although the equations for three-body SAPT have been written down,[249,250,315] their cost scales as $\mathcal{O}(N^7)$ with respect to the size of the largest trimer, whereas the present work will focus on methods that scale as $\mathcal{O}(N^3)$–$\mathcal{O}(N^5)$ with respect to the size of the largest *dimer*. In recognition of the qualitative observations above, our group has recently developed an "extended" version of SAPT in which many-body induction and polarization effects are incorporated into the zeroth-order wave functions by means of the XPol method.[252] Other components of the intermolecular interaction are included via pairwise SAPT. The result is a monomer-based method

that we call XSAPT,[44,54–56,65] which is aimed at doing fast calculations for non-covalent assemblies, including molecular and ionic clusters but also protein–ligand binding systems such as the one depicted in Fig. 8.2. The XPol procedure starts with a charged-embedded self-consistent field (SCF) calculation on each monomer, whose cost is therefore $\mathcal{O}(n)$ with respect the number of monomers, n. This is followed by an embarrassingly-parallelizable $\mathcal{O}(n^2)$ pairwise SAPT calculation. SAPT-style energy decomposition analysis is available, including a term that directly measures the many-body contribution to the interaction energy.[56]

8.2 Overview of XSAPT

We briefly review the theory behind XSAPT; see Ref. 54 for a detailed derivation. We assume that covalent bonds remain intact in the fragmentation of the system, as appropriate for non-covalent assemblies.

8.2.1 Many-body polarization: XPol

XPol is a fragment-based MO method that has been put forward as a means to obtain the polarization term in a "next-generation" force field.[252,289,316–318] Upon partitioning the system into fragments, the XPol wave function is written as a direct product of fragment wave functions, $|\Psi\rangle = |\Psi_A\rangle|\Psi_B\rangle|\Psi_C\rangle\cdots$. The XPol energy is[65]

$$E_{\mathrm{XPol}} = \sum_{A=1}^{n} \left(2 \sum_{a \in A} \mathbf{c}_a^\dagger \mathbf{f}^A \mathbf{c}_a + E_{\mathrm{nuc}}^A \right) + E_{\mathrm{embed}} \,, \tag{8.4}$$

where the term in parenthesis is the ordinary HF energy expression for monomer A, whose Fock matrix is $\mathbf{f}^A = \mathbf{h}^A + \mathbf{J}^A - \frac{1}{2}\mathbf{K}^A$. Crucially, the MOs $\mathbf{c}_a$ for fragment A

are represented using only those atomic orbitals (AOs) that are centered on atoms in fragment A. This partitioning of the basis leads to $\mathcal{O}(n)$ scaling and furthermore excludes BSSE, by construction. (Charge transfer between fragments is also excluded, at least in small basis sets.) The quantity E_{embed} in Eq. (8.4) is an electrostatic embedding potential that could in principle be obtained from the monomer electron densities, $\rho_A = |\Psi_A|^2$, but more often consists of the charge–density interactions that arise once each ρ_A is collapsed onto some set of atom-centered point charges.[54,65]

Upon variational minimization of Eq. (8.4) with respect to the $\mathbf{c}_a$, one obtains a set of monomer SCF equations that involve a modified Fock matrix $\mathbf{F}^A$ for monomer A:[54,65]

$$F^A_{\mu\nu} = f^A_{\mu\nu} - \frac{1}{2}\sum_{J\notin A} q_J (\mathbf{I}_J)_{\mu\nu} + \sum_{J\in A} M_J \, (\mathbf{\Lambda}_J)_{\mu\nu} \; . \tag{8.5}$$

The second term on the right is the interaction of fragment A with the embedding charge q_J, where $(\mathbf{I}_J)_{\mu\nu}$ is a charge–density Coulomb integral involving shell pair $\mu\nu \in A$. In the final term, $M_J = \partial E_{\text{embed}}/\partial q_J$ is easy to compute (see Ref. 54), and

$$(\mathbf{\Lambda}_J)_{\mu\nu} = \frac{\partial q_J}{\partial P_{\mu\nu}} \; . \tag{8.6}$$

Calculation of the $\mathbf{\Lambda}$ tensor does require some non-trivial overhead when the embedding charges q_J are "CHELPG" charges[95] that are fit to reproduce the monomer electrostatic potentials.[54,319] This is our preferred choice, as it seems physically sound and moreover the use of Mulliken or Löwdin charges often leads to convergence failure in the XPol procedure, when non-minimal basis sets are employed.[65,319] In any case, the monomer XPol SCF equations ($\mathbf{F}^A\mathbf{C}^A = \mathbf{S}^A\mathbf{C}^A\boldsymbol{\varepsilon}^A$) are iterated to self-consistency using a "dual SCF" procedure consisting of an outer loop over monomers and an inner

loop over SCF cycles for a particular monomer, updating the embedding charges as each $|\Psi_A\rangle$ is converged.

The final term in Eq. (8.5) is equal to $\partial E_{\text{embed}}/\partial P_{\mu\nu}$, and ensures that the converged XPol energy is fully variational.[44,65,252] In contrast, the fragment Fock matrix that is traditionally used in the fragment molecular orbital (FMO) method,[320,321] and also in the electrostatically-embedded many-body expansion,[45,64] is

$$F^A_{\mu\nu} = f^A_{\mu\nu} - \sum_{J \notin A} q_J (\mathbf{I}_J)_{\mu\nu} \; . \tag{8.7}$$

This Fock matrix does *not* afford a variational method, because it omits the response of the embedding charges to changes in the fragment wave functions. As a result, analytic gradients for FMO and other methods that use Eq. (8.7) with self-consistent embedding charges *should*, in principle, require solution of coupled-perturbed equations,[322,323] although these response terms have often been neglected.[79,324,325] In addition, Eq. (8.7) omits the polarization work that diminishes the middle term in Eq. (8.5) by a factor of two.[326,327]

8.2.2 Symmetry-adapted perturbation theory

In the original XPol method of Xie *et al.*,[252,289] intermolecular dispersion (van der Waals) and exchange (Pauli repulsion) interactions are incorporated using empirical Lennard-Jones or Buckingham potentials. We find, however, that this does not afford benchmark-quality results for non-covalent interactions,[56] so we instead choose to use the XPol fragment wave functions as zeroth-order states for application of second-order SAPT.

In SAPT, the Hamiltonian for the dimer $A \cdots B$ is partitioned according to[114,115]

$$\hat{H} = \hat{F}^A + \hat{F}^B + \xi \hat{W}^A + \eta \hat{W}^B + \zeta \hat{V} \ , \tag{8.8}$$

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

$$E_{\rm int}^{\rm SAPT} = \sum_{\zeta=1}^{\infty} \sum_{\kappa=0}^{\infty} \left(E_{\rm pol}^{(\zeta\kappa)} + E_{\rm exch}^{(\zeta\kappa)} \right) \tag{8.9}$$

where $\kappa = \xi + \eta$. The terms $E_{\rm pol}^{(\zeta\kappa)}$ constitute the *polarization expansion*,[328,329] which neglects exchange of electrons between monomers. (The term "polarization expansion" is historical, and should not be confused with the way in which we use the term "polarization" below, namely, to mean induction.) To correct this, each term in Eq. (8.9) has a corresponding exchange term $E_{\rm exch}^{(\zeta\kappa)}$ arising from the antisymmetrizer that is used to project out the Pauli-forbidden components of the interaction energy. Of these exchange terms, it has historically only been possible to evaluate $E_{\rm exch}^{(10)}$ exactly,[127] whereas other exchange terms are evaluated within the *single-exchange approximation*, in which permutations involving more than one pair of electrons are neglected. (The resulting formulas involve the square of the overlap matrix, $\mathbf{S}^2$, and for this reason the single-exchange approximation is often called the "S^2 approximation".[114,115,250]) Recently, an analytic form for $E_{\rm exch}^{(20)}$ has been reported,[174,175] but its implementation is not yet widely available. The single-exchange approximation is expected to be accurate at or beyond the van der Waals contact distance,[115] although

problems for anionic systems necessitate some rescaling of the higher-order exchange interactions.[50,52]

Neglecting *intra*molecular electron correlation but treating $\hat{V}$ to second order (the so-called SAPT0 approximation[117]), we have

$$E_{\text{int}}^{\text{SAPT0}} = E_{\text{elst}}^{(1)} + E_{\text{exch}}^{(1)} + E_{\text{ind}}^{(2)} + E_{\text{exch-ind}}^{(2)} + E_{\text{disp}}^{(2)} + E_{\text{exch-disp}}^{(2)} \,. \tag{8.10}$$

(We have dropped the index κ, since $\kappa = 0$.) Explicit expressions for these terms can be found in Ref. 114 (MO basis) or Ref. 137 (AO basis). The dispersion and exchange-dispersion terms are MP2-like in both cost and accuracy.

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

$$\delta E_{\text{int}}^{\text{HF}} = E_{\text{int}}^{\text{HF}} - \left(E_{\text{elst}}^{(1)} + E_{\text{exch}}^{(1)} + E_{\text{ind,resp}}^{(2)} + E_{\text{exch-ind,resp}}^{(2)} \right) \tag{8.11}$$

to the SAPT interaction energy.[117] Here, $E_{\text{int}}^{\text{HF}}$ is the counterpoise-corrected HF binding energy for the dimer.

8.2.3 Combining XPol with SAPT

The partition of the dimer Hamiltonian in Eq. (8.8) can be generalized to an arbitrary number of monomers,[54]

$$\hat{H} = \sum_A \left(\hat{F}^A + \xi_A \hat{W}^A \right) + \sum_A \sum_{B>A} \zeta_{AB} \hat{V}_{AB} \,. \tag{8.12}$$

with zeroth-order wave functions taken to be direct products of XPol monomer wave functions. Modification of the SAPT perturbation is required in order to avoid double-counting, because some part of electrostatics and polarization is already included at

the XPol level, but this modification is straightforward.[54,65] The resulting XSAPT energy expression, including all terms through second order in the intermolecular interactions, is[44,54]

$$\begin{aligned} E_{\mathrm{XSAPT}} = \sum_{A=1}^{n} \left[\sum_{a \in A} \left(2\varepsilon_a^A - \mathbf{c}_a^\dagger \mathbf{f}^A \mathbf{c}_a \right) + E_{\mathrm{nuc}}^A \right] \\ + \sum_A \sum_{B>A} \Big(E_{\mathrm{RSPT}}^{[0;1_{AB}]} + E_{\mathrm{exch}}^{[0;1_{AB}]} \\ + E_{\mathrm{RSPT}}^{[0;2_{AB}]} + E_{\mathrm{exch}}^{[0;2_{AB}]} \Big) + E_{\mathrm{3B}} \; . \end{aligned} \tag{8.13}$$

The superscript $[0; n_{AB}]$ indicates a term that is zeroth-order in the monomer fluctuation potentials but nth order in the intermolecular perturbation, V_{AB}.[44,54] The connection to traditional two-body SAPT is that

$$E_{\mathrm{RSPT}}^{[0;1_{AB}]} = E_{\mathrm{elst},A}^{(1)} + E_{\mathrm{elst},B}^{(1)} \tag{8.14a}$$

$$E_{\mathrm{exch}}^{[0;1_{AB}]} = E_{\mathrm{exch},A}^{(1)} + E_{\mathrm{exch},B}^{(1)} \tag{8.14b}$$

and

$$E_{\mathrm{RSPT}}^{[0;2_{AB}]} = E_{\mathrm{ind},A}^{(2)} + E_{\mathrm{ind},B}^{(2)} + E_{\mathrm{disp},AB}^{(2)} \tag{8.15a}$$

$$E_{\mathrm{exch}}^{[0;2_{AB}]} = E_{\mathrm{exch\text{-}ind},A}^{(2)} + E_{\mathrm{exch\text{-}ind},B}^{(2)} + E_{\mathrm{exch\text{-}disp},AB}^{(2)} \; . \tag{8.15b}$$

Two perturbations can couple three monomers, and second-order XSAPT thus contains three-body induction couplings that have no analogues in dimer SAPT.[54] This is the meaning of the E_{3B} term in Eq. (8.13):

$$E_{\mathrm{3B}} = \sum_{A,C} \sum_{B>A}{}' \sum_{D>C}{}' \Big(E_{\mathrm{RSPT}}^{[0;1_{AB},1_{CD}]} + E_{\mathrm{exch}}^{[0;1_{AB},1_{CD}]} \Big) \; . \tag{8.16}$$

The primed summations indicate that these terms vanish unless no more than three of the indices A, B, C, and D are distinct. Except for some exploratory calculations

in Ref. 54, these terms have been neglected in our previous work on XSAPT, and we will neglect them here unless stated otherwise. Results for water clusters and halide–water clusters, however, will demonstrate that these terms are important in large clusters of polar monomers. In order to compute the E_{3B} term, one must calculate all $n-2$ fragments for each unique dimer pair, which increases the computational cost from $\mathcal{O}(n^2)$ to $\mathcal{O}(n^3)$. Alternatively, $\mathcal{O}(n^2)$ scaling can be recovered (with a large prefactor) by storing induction amplitudes on disk.[54]

One of the attractive features of traditional SAPT is its energy decomposition analysis, and XSAPT extends this to many-body systems in a largely analogous way. We include a $\delta E_{\text{int}}^{\text{HF}}$ correction of the form given in Eq. (8.11), whose goal is to incorporate higher-order induction effects, and for many-body systems we assume that this correction is pairwise additive:

$$\delta E_{\text{int}}^{\text{HF}} = \sum_{A} \sum_{B>A} \delta E_{AB}^{\text{HF}} \, . \tag{8.17}$$

This assumption appears to be robust.[56] The SAPT interaction energy can be decomposed as in Eq. (8.10), and the resulting XSAPT energy decomposition is[56]

$$\begin{aligned} E_{\text{int}}^{\text{XSAPT}} = {} & E_{\text{int}}^{\text{SAPT}} + \delta E_{\text{int}}^{\text{HF}} + E_{\text{int}}^{\text{MB}} \\ & + \sum_{A} \sum_{B>A} \left(E_{AB}^{\text{XSAPT}} - E_{AB}^{\text{SAPT}} \right) \, , \end{aligned} \tag{8.18}$$

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

$$E_{\text{int}}^{\text{SAPT}} = \sum_{A} \sum_{B>A} E_{AB}^{\text{SAPT}} \, . \tag{8.19}$$

The XSAPT interaction energy can be rewritten as

$$\begin{aligned} E_{\text{int}}^{\text{XSAPT}} = & E_{\text{elst}}^{(1)} + E_{\text{exch}}^{(1)} + E_{\text{disp}}^{(2)} + E_{\text{exch-disp}}^{(2)} \\ & + \Bigg[E_{\text{ind}}^{(2)} + E_{\text{exch-ind}}^{(2)} + \sum_A \sum_{B>A} \delta E_{AB}^{\text{HF}} \\ & + \sum_A \sum_{B>A} \left(E_{AB}^{\text{XSAPT}} - E_{AB}^{\text{SAPT}} \right) + E_{\text{int}}^{\text{MB}} \Bigg] . \end{aligned} \tag{8.20}$$

Here, the terms $E_{\text{elst}}^{(1)}$, $E_{\text{exch}}^{(1)}$, etc., represent the sum of these energy components over all pairs of dimers, and the many-body contribution to the interaction energy[56]

$$E_{\text{int}}^{\text{MB}} = E_{\text{int}}^{\text{XSAPT}} - \sum_A \sum_{B>A} E_{AB}^{\text{XSAPT}} . \tag{8.21}$$

The term in square brackets in Eq. (8.18) is regarded as the total induction energy, which includes a many-body contribution.

In dimer SAPT calculations, an infinite-order polarization correction (in the presence of a frozen partner density) can be included by solving coupled-perturbed equations.[114] However, XSAPT treats polarization self-consistently and the infinite-order response correction for induction should be included exactly, via the XPol procedure, if density embedding is used.[276] We prefer CHELPG embedding for reasons of cost, however. The pairwise difference between XSAPT and SAPT in Eq. (8.18) partly includes the infinite-order response correction for induction. An infinite-order polarization correction is still included in $\delta E_{\text{int}}^{\text{HF}}$ by solving coupled-perturbed equations,[179,180] which is the meaning of the "resp" (response) subscripts in Eq. (8.11).

In the language of traditional dimer SAPT, our XPol monomer wave functions are computed in a monomer-centered basis set, which largely excludes the description of

charge transfer between monomers and is typically less accurate than if the zeroth-order wave functions are computed using the dimer-centered basis set.[153] The correct choice of a dimer basis is ambiguous in a many-body system, however, so we choose instead to converge the XPol wave functions in the monomer-centered basis, then compute the pairwise SAPT corrections in a "projected"[65] (pseudocanonicalized[286,330] monomer-centered) basis set.

8.2.4 XSAPT(KS)+D

Especially for strongly hydrogen-bonded systems, inclusion of intermolecular electron correlation effects may be important,[54,158] but involves methods whose cost scales as $\mathcal{O}(N^7)$ within the wave function-based SAPT formalism.[117] SAPT(KS) represents a low-cost way to include such effects, though asymptotic correction of the XC potential is required.[162,208,209] Various "splicing" schemes have been used in this context,[199,215,331] but result in "stray" XC potentials[216] that do not correspond to any well-defined energy functional, $v_{\mathrm{xc}}^{\mathrm{AC}} \neq \delta E_{\mathrm{xc}}/\delta\rho$. This is potentially problematic in the context of geometry optimizations,[216] and fatal to any attempt to derive analytic energy gradients.

We sidestep this problem using long-range corrected (LRC) density functionals,[219,222,224,227] which correctly reproduce the asymptotic $\sim r^{-1}$ behavior of v_{xc}. To achieve the proper limiting value, $\Delta_\infty = 0$ in Eq. (8.2), we apply a monomer-specific

"tuning" procedure, as suggested by Baer and co-workers,[229,230] in which the range-separation parameter, ω, is adjusted in order to satisfy the condition

$$\varepsilon_{\mathrm{HOMO}}(\omega) = -\mathrm{IP}(\omega) \; . \tag{8.22}$$

In the context of SAPT(KS), this significantly improves the quality of the various interaction energy components as compared to benchmark results,[51] and represents a promising alternative to other AC schemes. Dispersion energies are still not of benchmark quality, however, owing to problems with the uncoupled KS description of dispersion.[51]

To correct the latter problem, we replace the second-order dispersion and exchange-dispersion terms in SAPT with empirical atom–atom dispersion potentials,[55,56] following along the lines of the "SAPT(KS)+D" method introduced by Heßelmann.[16] This has the added benefit of reducing the scaling from $\mathcal{O}(N^5)$ to $\mathcal{O}(N^3)$.[16] At first glance, this approach seems similar in spirit to dispersion-corrected DFT,[143] but the separation of dispersion from other parts of the energy is much cleaner in SAPT, whereas in DFT+D there is a potential double-counting problem for mid-range intermolecular distances, where the short-range DFT correlation may not have decayed completely to zero as the (damped) long-range dispersion potentials are turning on. Indeed, Grimme[253] suggests that dispersion in DFT+D is a model-dependent quantity with no real physical meaning.

Our original version[55] of XSAPT(KS)+D used Heßelmann's SAPT(KS)+D dispersion potential,[16] which was fit to reproduce S22 benchmark binding energies, and that "first generation" (+D1) approach affords an impressive MAE of only

0.3 kcal/mol for the S66 data set.[55] However, XSAPT(KS)+D1 benefits from favorable error cancellation and does *not* accurately reproduce individual energy components.[56] In subsequent work, we avoided fitting directly to binding energies and instead pursued a second-generation (+D2) method using alternative dispersion potentials developed by Podeszwa *et al.*,[244] which were fit to reproduce distance-dependent dispersion potentials, $E_{\text{disp}} = E^{(2)}_{\text{disp}} + E^{(2)}_{\text{exch-disp}}$, obtained from SAPT(DFT) calculations. XSAPT(KS)+D2 accurately reproduces not only total binding energies but also individual energy components.[56]

Tests on the S22 and S66 data sets reveal that the primary source of errors in XSAPT(KS)+D2 calculations comes from π-stacked complexes, where in some cases the dispersion energy is overestimated by $\sim$2 kcal/mol as compared to SAPT2+(3)/aTZ results.[56] Such systems are underrepresented in the training set used to parameterize the D2 potentials,[244] and here we report for the first time a third-generation dispersion potential for XSAPT. (It should be stressed that our "D3" dispersion potential is unrelated to Grimme's "D3" correction[274] for DFT.)

The new D3 dispersion potential uses the same functional form as the D2 potential,[244]

$$E^{\text{D3}}_{\text{disp}} = -\sum_{i\in A}\sum_{\substack{j\in B\\(B\neq A)}}\left[\frac{C_{ij,6}}{r^6_{ij}}f_6(\beta_{ij}r_{ij}) + \frac{C_{ij,8}}{r^8_{ij}}f_8(\beta_{ij}r_{ij})\right], \tag{8.23}$$

where

$$f_n(r_{ij}) = 1 - \exp(-r_{ij})\sum_{m=0}^{n}\frac{r^m_{ij}}{m!} \tag{8.24}$$

is the Tang-Toennies damping function,[283] and i and j represent nuclei located on different monomers. We take $C_{ij,6} = (C_{i,6}C_{j,6})^{1/2}$, $C_{ij,8} = (C_{i,8}C_{j,8})^{1/2}$, and $\beta_{ij} =$

$(\beta_i\beta_j)^{1/2}$, where $C_{i,6}$, $C_{i,8}$, and β_i are parameters fit to reproduce SAPT2+(3)/aTZ dispersion energies,

$$\begin{aligned} E_{\text{disp}}^{\text{SAPT2+(3)}} = E_{\text{exch-disp}}^{(20)} + E_{\text{disp}}^{(20)} + E_{\text{disp}}^{(21)} \\ + E_{\text{disp}}^{(22)}(\text{SDQ}) + \text{E}_{\text{disp}}^{(22)}(\text{T}) + \text{E}_{\text{disp}}^{(30)} \, . \end{aligned} \tag{8.25}$$

For hydrogen, these parameters depend upon the identity of the nearest-neighbor atom. This is the similar procedure used in Ref. 244 to obtain the D2 potential, but we have expanded the training set to include additional π-stacked systems as well as the ionic systems $F^-(H_2O)$ and $Cl^-(H_2O)$. (The list of systems can be found in the Supporting Information.) For the latter two systems, the halide–water distance is short and dispersion is especially important. In these two cases, we use benchmarks from the highest level SAPT theory, SAPT2+3(CCD)/aTZ:[183,184]

$$\begin{aligned} E_{\text{disp}}^{\text{SAPT2+3(CCD)}} = E_{\text{exch-disp}}^{(20)} + E_{\text{disp}}^{(2)}[\text{CCD}] \\ + E_{\text{disp}}^{(22)}[\text{S(CCD)}] + E_{\text{disp}}^{(22)}[\text{T(CCD)}] \\ + E_{\text{disp}}^{(30)} + E_{\text{exch-disp}}^{(30)} \\ + E_{\text{ind-disp}}^{(30)} + E_{\text{exch-ind-disp}}^{(30)} \, . \end{aligned} \tag{8.26}$$

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

The LRC-ωPBE functional[240] with monomer-specific, tuned values of ω is used in all of our XSAPT+D3 calculations, as we have previously observed that errors in the exchange energy components increase as short-range HF exchange is added to LRC

functionals.[56] Tuned values of ω are listed in the SI and differ, in some cases, from values reported previously.[56] In the case of thymine, for example, the optimally-tuned value changes from $\omega = 0.625\ a_0^{-1}$ to $0.275\ a_0^{-1}$. The large discrepancy comes from the fact that the condition in Eq. (8.22) cannot be satisfied for some monomers and basis sets. In such cases, we selected the closest point of approach between the $\varepsilon_{\mathrm{HOMO}}(\omega)$ and $-\mathrm{IP}(\omega)$ curves. The binding energy errors for the π-stacked uracil dimer and the adenine–thymine dimer were 3 and 4 kcal/mol, respectively, at the XSAPT(KS)+D2 level,[56] and these were the outliers amongst the S22 dimers. Simply using these newly-tuned values of ω determined in this study, the XSAPT(KS)+D2 errors for these two systems are reduced to 2.0 and 2.5 kcal/mol, respectively. We therefore recommend these new ω values.

8.2.5 sd-XSAPT(KS)

Introduction of empirical dispersion into XSAPT calculations reduces the scaling from fifth-order to third-order with respect to dimer size, and is further motivated by the fact that second-order, uncoupled dispersion energies are not nearly of benchmark quality. An alternative approach, which scales as $\mathcal{O}(N^4)$, is to omit the fifth-order exchange-dispersion term in SAPT0, then scale the fourth-order dispersion term by an empirical factor. This method was introduced recently by Ochsenfeld and co-workers,[332] who called it sd-SAPT0. We have implemented the corresponding sd-XSAPT(KS) method, and following Ref. 332, we neglect the $\delta E_{\mathrm{int}}^{\mathrm{HF}}$ correction in this approach.

Using the S22B binding energies[34] to fit the dispersion scaling parameter, the best-performing combination was found to be the LRC-ωPBEh functional[227] (20% short-range HF exchange), 6-31G(d,2p) basis set, and a scaling parameter $c_{\text{disp}} = 0.657$, in which case the root-mean-square deviation (RMSD) for S22 binding energies is 0.366 kcal/mol. That $c_{\text{disp}} < 1$ can be understood in terms of the neglect of the repulsive exchange-dispersion interaction and the fact that second-order perturbation theory tends to overestimate dispersion in the first place. (The double-ζ basis set also helps in this regard, as dispersion interactions converge slowly to the basis-set limit, and scaling back the basis set is a well-established way to reduce errors in second-order dispersion energies,[117] albeit by error cancellation.) Since the sd-XSAPT(KS) method is based on fitting to obtain accurate total binding energies, it is not recommended as a means to do energy decomposition analysis.

8.3 Performance benchmarks

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

8.3.1 Timings

So as not to obfuscate the fact that the primary purpose of XSAPT is *efficient* calculation of intermolecular interactions, we lead off with data illustrating the efficiency of the method. Figure 8.3 plots timings for XSAPT(KS)+D calculations on π-stacked $(\text{adenine})_n$. Serial timings represent the total wall time required, which scales as $\mathcal{O}(n^2)$, whereas parallel timings represent the wall time required when the calculation

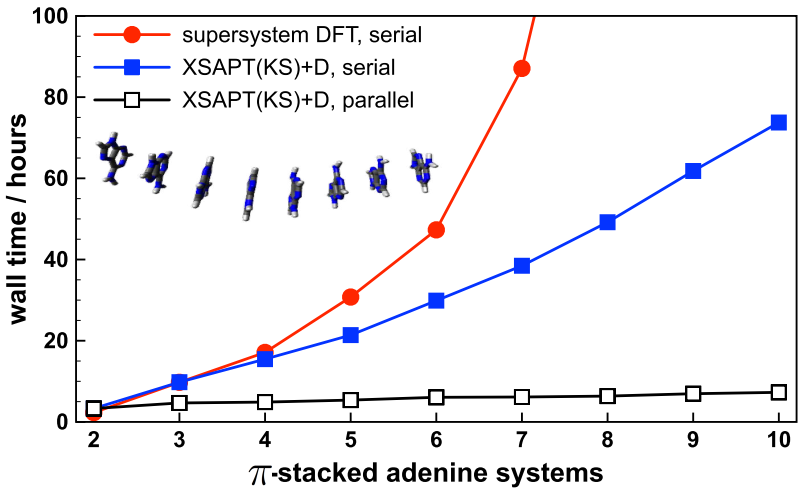

Figure 8.3: Timings for XSAPT(KS) and supersystem DFT calculations for π-stacked $(\text{adenine})_n$ systems. All calculations use the LRC-ωPBE functional and the hpTZVPP basis set.

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

8.3.2 Validation of the D3 dispersion potential

To test the performance of the new D3 dispersion potential, we have used the D2 and D3 potentials to compute the stacking interaction between DNA base pairs in 10 different nucleobase tetramers,[15] as compared to the dispersion interaction (=

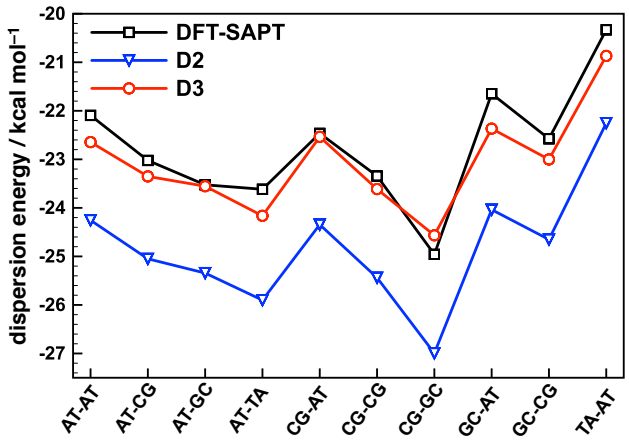

Figure 8.4: Comparison of E_{disp}(D2), E_{disp}(D3), and E_{disp}(DFT-SAPT) for the stacking interaction in nucleobase tetramers. The DFT-SAPT benchmarks come from Ref. 15, corrected by a factor of 1.1 as suggested in Ref. 16.

$E^{(2)}_{\text{disp}} + E^{(2)}_{\text{exch-disp}}$) obtained from DFT-SAPT calculations. DFT-SAPT dispersion energies from Ref. 15 were multiplied by a factor of 1.1 as an approximate correction for basis-set incompleteness at the aTZ level, as suggested by Heßelmann.[16] The results, shown in Fig. 8.4, indicate good agreement between E_{disp}(D3) and E_{disp}(DFT-SAPT), and the D3 dispersion potential is much closer to the benchmark as compared to D2 results.

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

$$E^{\text{MP2+D3}}_{\text{int}} = E^{\text{MP2}}_{\text{int}} - E^{(2)}_{\text{disp}} - E^{(2)}_{\text{exch-disp}} + E_{\text{disp}}(\text{D3}) \;, \tag{8.27}$$

in which we use MP2 to incorporate intramolecular correlation, then subtract out

the second-order dispersion and replace it with the D3 empirical potential. This method is similar in spirit to the MP2(CCD) method.[52,173] We plan to investigate this "MP2+D3" approach in the future.

8.3.3 Basis set convergence

The $E^{(1)}_{\text{elst}}$ and $E^{(1)}_{\text{exch}}$ energy components are common to both SAPT(KS) and XSAPT(KS), and in the former case we can perform benchmark SAPT2+(3)/aTZ calculations for dimers, which we have done for the representative systems $F^-(H_2O)$, $(H_2O)_2$, and both the T-shaped and parallel-displaced isomers of $(C_6H_6)_2$. SAPT(KS) results for the same systems, using either the dimer-centered or the "projected" (pseudocanonicalized monomer-centered) approach, were used to select an AO basis set. [The SAPT2+(3)/aTZ benchmarks employ the dimer-centered basis.] Mean errors in $E^{(1)}_{\text{elst}}$ and $E^{(1)}_{\text{exch}}$, with respect to the SAPT2+(3)/aTZ benchmarks, are provided in the SI for 21 different basis sets ranging from double-ζ to quadruple-ζ quality. A brief summary is presented here.

Using the dimer-centered approach, we find that the errors are quite small for both Dunning (aug-cc-pVXZ) and Ahlrichs (def2) basis sets, provided that diffuse functions are included. In that case, calculations of triple-ζ quality, or possibly even double-ζ quality, appear to be converged to the basis-set limit. Unfortunately, however, the dimer-centered construction is ill-defined for a many-body system, which is why we turn to the pseudocanonicalization approach.[65] The data in the SI show that diffuse functions are also essential in this approach, although in this case there remains significant discrepancy between triple- and quadruple-ζ results. These differences are

smaller when Ahlrichs basis sets are employed, in which case it matters little whether the diffuse functions are drawn from the Ahlrichs, Pople, or Dunning basis sets.

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

8.4 Illustrative applications

The remainder of this paper is dedicated to illustrating the power and utility of XSAPT-based methods. Details about how the benchmark calculations were obtained, along with an explanation of non-standard basis set nomenclature, can be found in the SI.

8.4.1 Biologically-relevant dimers

Since the S22 data set[186] was used to fit both the D3 dispersion potential and the scaling factor for sd-XSAPT(KS), the S66 data set[10] will be used to evaluate the accuracy of these methods. S66 consists of CCSD(T)/CBS binding energies for 66 weakly-bound dimers related to biomolecular structures, and we use the recently-revised S66 binding energies.[333] Augmented triple-ζ basis sets are essential in order

Table 8.1: Errors[a] in S66 binding energies for XSAPT(KS)+D3.

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

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

to obtain accurate results for individual energy components, and error statistics for XSAPT(KS)+D3 in a variety of triple-ζ basis sets are shown in Table 8.1. The best-performing basis sets are pTZVPP and hpTZVPP.

Figure 8.5 shows S66 error statistics for a variety of methods that exhibit reasonably small MAEs. The new XSAPT(KS)+D3 method slightly outperforms the previous two generations (D1 and D2), and in particular reduces the errors in the π-stacked outliers. It is worth mentioning that the MP2C[309–311] and SAPT2+(3) methods,[117] which exhibit excellent performance for S66, are only formulated for dimers, and that the Coulomb-attenuated MP2 method[17,176,334] (att-MP2) contains a parameter that was optimized using this very data set. It also bears mention that all of the methods that outperform XSAPT(KS)+D3 exhibit at least fifth-order scaling with respect to the size of the *supersystem*, whereas XSAPT(KS)+D scales as $\mathcal{O}(N^3)$ with respect to dimer size and $\mathcal{O}(n^2)$ with respect to the number of monomers. XSAPT(KS)+D also affords an energy decomposition analysis that is discussed in Section 8.4.8.

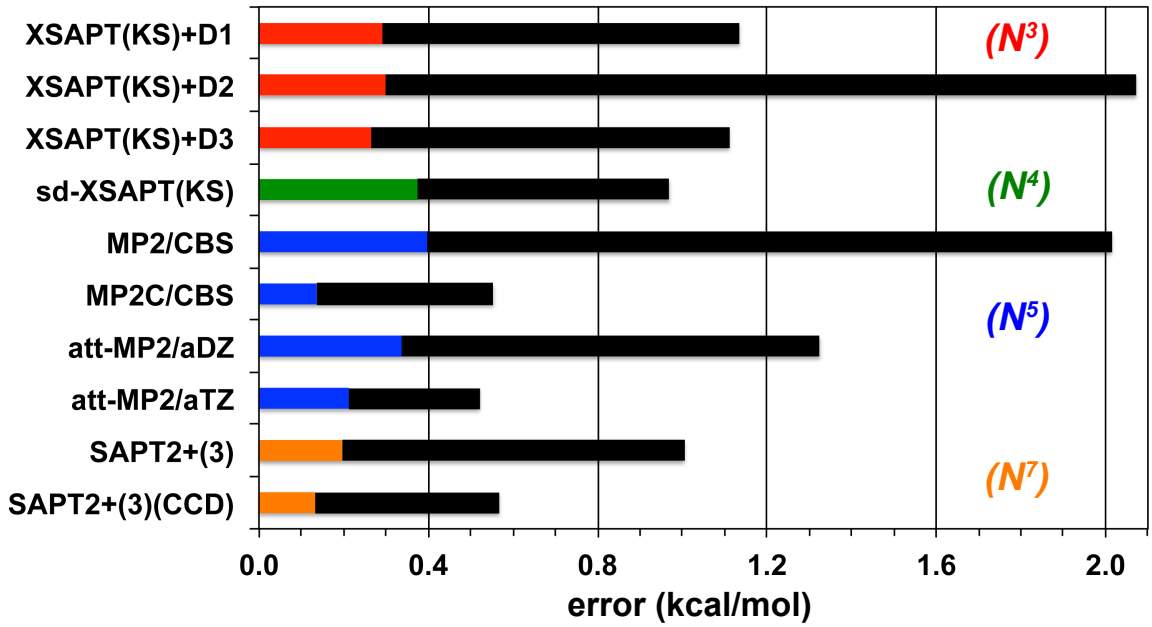

Figure 8.5: MAEs (colored bars) and maximum errors (in black) computed for S66 binding energies with respect to CCSD(T)/CBS benchmarks. These methods are color-coded according to how their cost scales with system size. The jun-cc-pVDZ basis set is used for XSAPT(KS)+D1 calculations and the hpTZVPP basis set for XSAPT(KS)+D2 and +D3 calculations. The 6-31G(d,2p) basis set is used for sd-XSAPT(KS) and aug-cc-pVTZ basis setfor SAPT2+(3) and SAPT2+(3)(CCD). The att-MP2 data were obtained from 17 and data for the other supermolecular methods were obtained from Ref. 10.

8.4.2 Potential energy curves

Comparison of the "sandwich" and "T-shaped" isomers of $(C_6H_6)_2$ represents a stringent test for theoretical models,[5] since the two are stabilized by very different types of interactions (dispersion competes with quadrupolar electrostatics). In Figure 8.6, the XSAPT(KS)+D3 method and the ωB97X-V density functional[105] reproduce CCSD(T)/CBS potential curves for both isomers nearly quantitatively. In contrast, the att-MP2 method shifts the minimum to shorter distances for both isomers. The sd-XSAPT(KS) method significantly shortens the van der Waals contact distance for the π-stacked isomer, while significantly underestimating the binding energy for the T-shaped isomer.

In Figure 8.7, we plot the potential energy curves for $F^-(H_2O)$ and $Cl^-(H_2O)$, which are known to be challenging cases for SAPT.[50,55] CCSD(T)/CBS correlation energies were evaluated using a two-point (aTZ, aQZ) extrapolation,[73] then added to the HF/aQZ energy to obtain the CCSD(T)/CBS results. The XSAPT(KS)+D3, att-MP2, and ωB97X-V methods reproduce the CCSD(T)/CBS potential curves for $F^-(H_2O)$ nearly quantitatively, whereas sd-XSAPT(KS) significantly overestimates the binding energy. These methods perform even better for $Cl^-(H_2O)$, with only a slight underestimation of the binding energy on the part of sd-SAPT(KS).

The geometry of H_2O is held fixed in the calculation shown in Fig. 8.7. Upon relaxation of the geometry, the $F^-(H_2O)$ binding energy increases from 25 to 32 kcal/mol at the minimum-energy structure, and the XSAPT(KS)+D3 method somewhat overestimates the binding energy. The relaxed geometry is somewhat problematic for

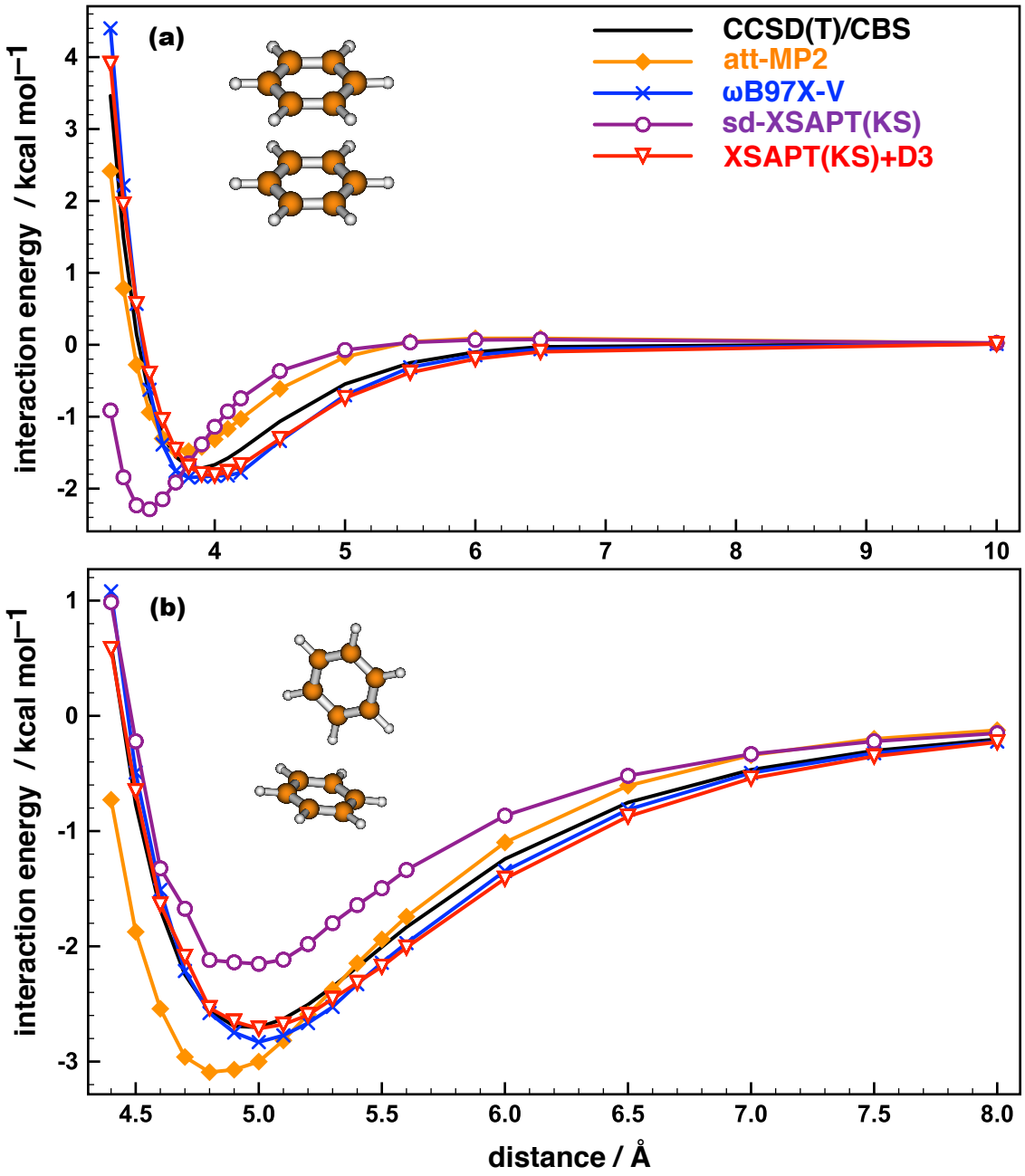

Figure 8.6: Potential energy curves for the (a) sandwich and (b) T-shaped isomers of $(C_6H_6)_2$. The distance coordinate is the center-to-center distance between the benzene rings. Benchmark CCSD(T)/CBS results are taken from Ref. 5. The aug-cc-pVTZ basis set is used for the att-MP2 and ωB97X-V calculations. The hpTZVPP basis set is used for the XSAPT(KS)+D3 calculations.

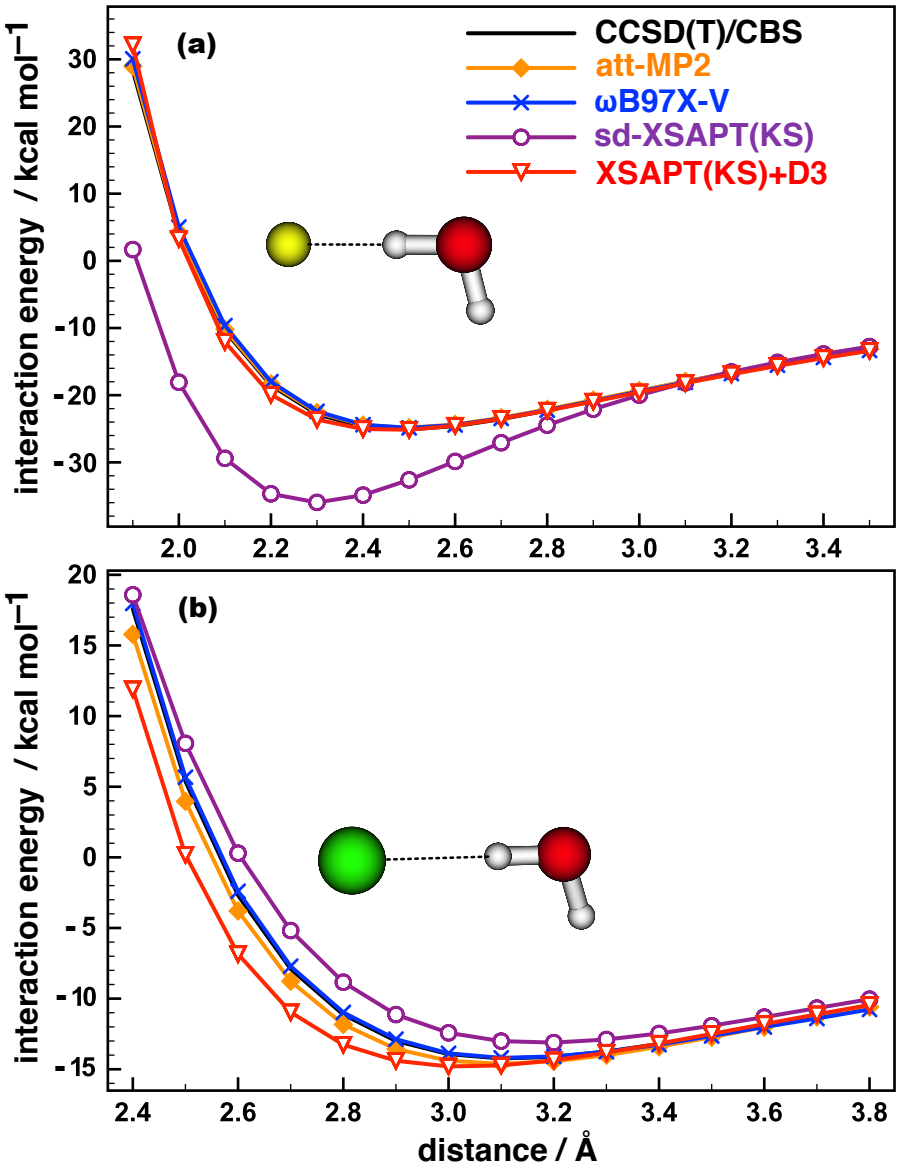

Figure 8.7: Potential energy curves for (a) $F^-(H_2O)$ and (b) $Cl^-(H_2O)$ at a fixed H_2O geometry. The distance coordinate is the halide-oxygen distance and the benchmark is CCSD(T)/CBS. The aug-cc-pVTZ basis set is used for the att-MP2 and ωB97X-V calculations. The hpTZVPP basis set is used for the XSAPT(KS)+D3 calculations.

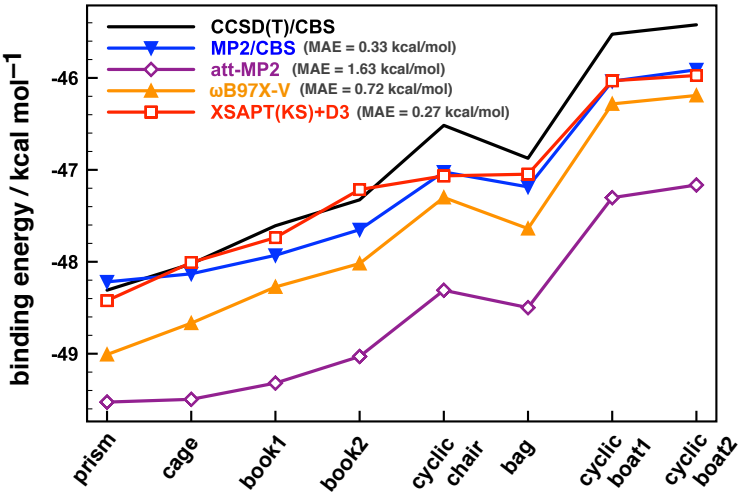

Figure 8.8: Binding energies for eight isomers of $(H_2O)_6$. The aTZ basis set is used for the att-MP2 and ωB97X-V calculations, whereas the hpTZVPP basis set is used for the XSAPT(KS)+D3 calculations. (MAEs for the whole data set, with respect to the CCSD(T)/CBS benchmarks, are also listed for each method).

fragment-based methods, as the proton affinity of F^- leads to an unusually long O–H bond length of 1.06 Å versus 0.96 Å when the H_2O geometry is optimized separately. [For $Cl^-(H_2O)$, the O–H bond length increases only to 0.98 Å upon relaxation.] We note that other fragment based methods, such as the effective fragment potential (EFP) method,[335] must also use rigid, EFP-optimized geometries; otherwise, large errors in non-covalent binding energies are obtained.[56]

8.4.3 Many-body system: $(H_2O)_6$

We next consider a cluster of polar monomers that exhibits significant many-body polarization effects.[8,251] Eight low-lying structures of $(H_2O)_6$ are considered,[8] and their CCSD(T)/CBS binding energies are evaluated using a two-point extrapolation[73]

of CCSD(T)-F12 correlation energies (cc-pVDZ-F12 and cc-pVTZ-F12 basis sets, using the corresponding near-complete auxiliary basis sets cc-pVDZ-F12-CABS and cc-pVTZ-F12-CABS).[191,192] The HF energy is evaluated using the cc-pVTZ-F12 basis set. For comparison, the MP2 correlation energy in the CBS limit was evaluated using a two-point (aTZ, aQZ) extrapolation,[73] then added to the HF/aQZ energy to obtain the MP2/CBS result.

Figure 8.8 compares the binding energies for isomers of $(H_2O)_6$ computed using various methods, but sd-XSAPT(KS) results are not shown in Fig. 8.8 because this method overestimates the binding energies by an average of 5.46 kcal/mol. This is consistent with the fact that sd-XSAPT(KS) overestimates the binding energy of water dimer already by 0.56 kcal/mol, hence sd-XSAPT(KS) cannot be recommended for water clusters. The att-MP2,[17,176] ωB97X-V,[105] and MP2/CBS methods all afford accurate relative energies, as does XSAPT(KS)+D3 except for a slight overstabilization of the cyclic chair isomer, but the most accurate absolute binding energies are obtained using XSAPT(KS)+D3. Total binding energies predicted using XSAPT(KS)+D3 are more accurate than those obtained using ωB97X-V which is one of the best DFT approaches for non-covalent interactions,[105] although it should be noted that *relative* energies are slightly better with ωB97X-V, which does not over stabilize the cyclic chair.

One factor that influences the accuracy of XSAPT(KS) results is the tuning of ω. Tuned values listed in the SI were obtained using a step size $\Delta\omega = 0.025\ a_0^{-1}$ to scan $\varepsilon_{\mathrm{HOMO}}(\omega)$ and $-\mathrm{IP}(\omega)$, and $\omega = 0.500\ a_0^{-1}$ is thus determined to be the

optimal value for H_2O. However, if the step size is decreased to 0.005 a_0^{-1}, then the optimal value changes to 0.485 a_0^{-1}. For water dimer, the binding energy changes by only 0.07 kcal/mol between these two values, but errors accumulate as the number of fragments increases. For isomers of $(H_2O)_6$, the difference in binding energies between $\omega = 0.485$ and 0.500 a_0^{-1} is 0.68 kcal/mol on average, but is 3.13 kcal/mol for the isomers of $(H_2O)_{20}$ that are discussed below. As such, we use the more finely-tuned value ($\omega = 0.485\ a_0^{-1}$) for water clusters.

8.4.4 Larger clusters: $(H_2O)_{20}$

Medium-sized water clusters have long attracted interest from the quantum chemistry community; for example, $(H_2O)_{16,17}$ are considered to be transition structures from "all-surface" to "internally solvated" arrangements of the hydrogen-bonding network,[336,337] and $(H_2O)_{20,24}$ are the building blocks of ice clathrates.[338] Here, we use ten low-energy isomers of $(H_2O)_{20}$, obtained using the TIP4P force field,[339] to benchmark the methods introduced above, and—by comparison to $(H_2O)_6$ results—to understand whether errors increase with system size. MP2/CBS results for these ten isomers were estimated as described in Section 8.4.3.

To estimate the CCSD(T)/CBS binding energies, we use explicitly-correlated CCSD(T) calculations reported recently using the generalized energy-based fragmentation (GEBF) method.[18] We take

$$E_{\mathrm{CBS}}^{\mathrm{CCSD(T)}} \approx E_{\mathrm{CBS}}^{\mathrm{MP2}} + \delta_{\mathrm{MP2\text{-}F12}}^{\mathrm{CCSD(T)\text{-}F12a}}\ , \tag{8.28}$$

where the correction to the MP2/CBS result is equal to the difference between the

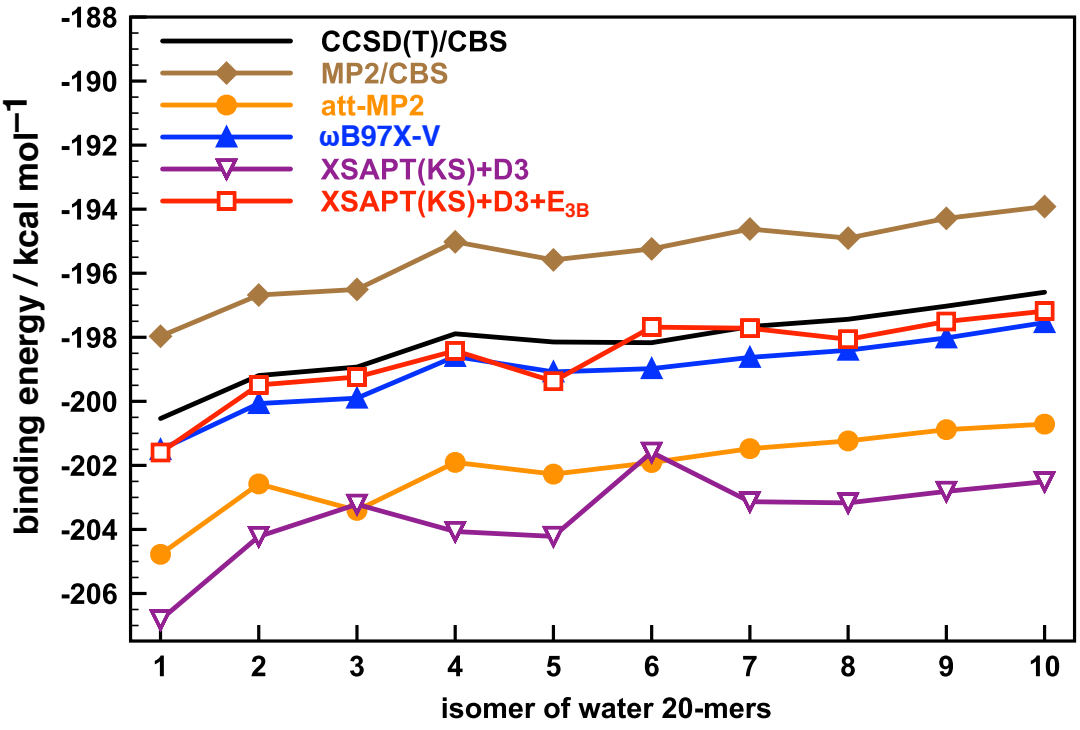

Figure 8.9: Binding energies for ten low-energy isomers of $(H_2O)_{20}$. The aTZ basis set is used for the att-MP2 and ωB97X-V calculations and the hpTZVPP basis set is used for XSAPT methods. The isomers are number as in Ref. 18.

CCSD(T)-F12a/aDZ and MP2-F12/aDZ binding energies that were reported in Ref. 18 using the GEBF approximation. Unlike the calculations reported in Ref. 18, which used the GEBF approximation at both the MP2 and CCSD(T) levels of theory, we evaluate the full MP2/CBS energy. This changes the binding energies by an average of 1.75 kcal/mol relative to the benchmarks reported in Ref. 18, and we believe that our CCSD(T)/CBS results for $(H_2O)_{20}$ are one of the most accurate binding benchmarks for large water clusters in the literature.

Table 8.2: MAEs[a] for ten low-energy isomers of $(H_2O)_{20}$.

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

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

Results for total binding energies are shown in Fig. 8.9, with error statistics listed in Table 8.2. The correction to MP2/CBS in Eq. (8.28) is negative for these clusters, and CCSD(T)/CBS total binding energies are 2.7 kcal/mol larger, on average, than MP2/CBS binding energies. Unfortunately, XSAPT(KS)+D3 results for total binding energies are not significantly better than MP2/CBS results in this case, and can be worse, depending on the basis set that is used.

In an attempt to understand this loss of accuracy relative to the $(H_2O)_6$ results, we investigated the neglected three-body induction couplings, E_{3B} in Eq. (8.16). For isomers of $(H_2O)_6$, $E_{3B} = 0.13$ kcal/mol (on average) at the XSAPT(KS)+D3/hpTZVPP level. However, $E_{3B} = 4.95$ kcal/mol (on average) for the $(H_2O)_{20}$ isomers considered here, and the XSAPT(KS)+D3 total binding energies are significantly improved by the addition of this three-body correction term.

It is evident both from the plot in Fig. 8.9 and from the error statistics in Table 8.2 that the ωB97X-V and MP2/CBS methods afford very good relative energies, as does att-MP2. XSAPT(KS)+D3+E_{3B} relative energies are somewhat worse, even if the absolute binding energies are slightly more accurate, at least when the hpTZVPP basis set is employed. The reasons for this are unclear, although in fitting the D3 potential we considered a variety of intermolecular distances but only one intermolecular orientation per monomer. Also unclear is why isomer **6** (see Fig. 8.9) poses such a problem for XSAPT, although this same isomer has been noted to pose problems for DFT methods as well.[18] It is the most highly-coordinated of all $(H_2O)_{20}$ isomers considered here, so the problem may again be orientational dependence, although Wang *et al.*[18] suggest that both basis set and electron correlation effects must be considered in order to obtain an accurate relative energy for this isomer. Extension of XSAPT to larger basis sets, by means of an AO rather than an MO implementation, is currently underway in our group and may help in this capacity.

8.4.5 Halide–water clusters

Halide–water clusters are difficult cases for popular DFT methods.[56,105] Table 8.3 shows binding-energy errors, with respect to CCSD(T)/CBS benchmarks, for various methods applied to the minimum-energy structures of $X^-(H_2O)_{n=1-6}$, for X = F, Cl. (Benchmarks were obtained as described for water hexamer in Section 8.4.3.)

The MP2/CBS method cannot be considered a benchmark, sub-kcal/mol level of theory for $F^-(H_2O)_n$, with a MAE of 1.0 kcal/mol and a maximum error of 1.6 kcal/mol. Errors are smaller for $Cl^-(H_2O)_n$ (MAE = 0.2 kcal/mol, maximum =

Table 8.3: MAEs[a] for binding energies of $X^-(H_2O)_{n=1-6}$.

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

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

0.3 kcal/mol). Interestingly, the att-MP2 method performs in the opposite way: highly accurate results for $F^-(H_2O)_n$, but a MAE of 1.0 kcal/mol for $Cl^-(H_2O)_n$. The ωB97X-V functional performs well for both, as shown previously.[105]

For XSAPT calculations of $X^-(H_2O)_n$, we find that the three-body induction couplings are important, reducing the errors by approximately a factor of two, although the MAE for $F^-(H_2O)_n$ remains $>$ 1 kcal/mol. Parker *et al.*[173] have proposed a "δMP2" correction,

$$\delta E_{\mathrm{MP2}} = E_{\mathrm{int}}^{\mathrm{MP2}} - E_{\mathrm{int}}^{\mathrm{SAPT2}} \,, \tag{8.29}$$

to account for missing terms such as high-order coupling between induction and dispersion. This correction, which we find is especially important in ionic systems,[52] is equal to the difference between the counterpoise-corrected MP2 binding energy

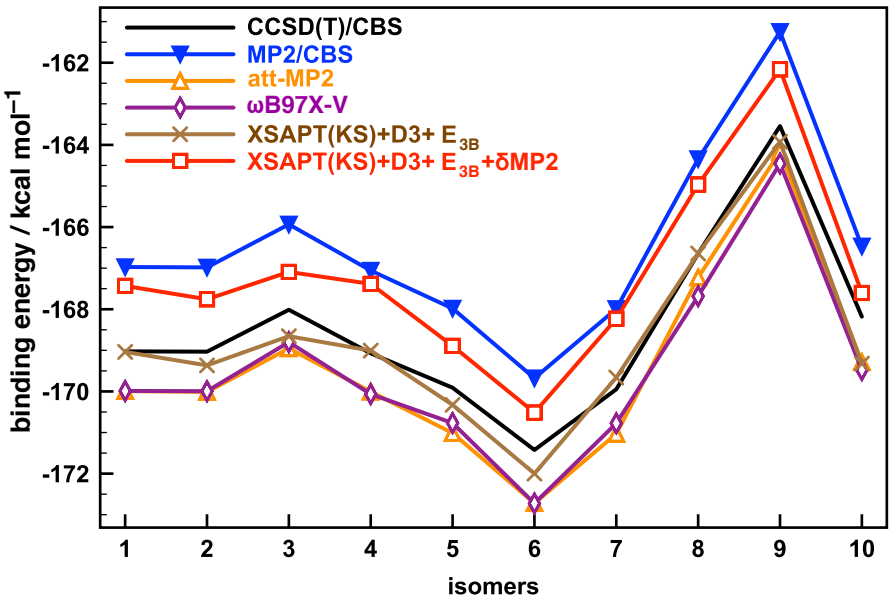

Figure 8.10: Binding energies for ten isomers of $F^-(H_2O)_{10}$. The att-MP2 and ωB97X-V calculations use the aTZ basis set and XSAPT calculations use the hpTZVPP basis set, except that the aTZ basis set was used to evaluate the δMP2 correction.

for the dimer and the SAPT2 binding energy. In XSAPT calculations, we apply the δE_{MP2} correction in a pairwise way, for dimers that include X^-, and this further reduces the errors, especially for X = F. [The SAPT2 part of δE_{MP2} also makes this correction $\mathcal{O}(N^6)$ with respect to dimer size, although this can be reduced to $\mathcal{O}(N^5)$ using density-fitting techniques.[158]] The hpTZVPP basis set works better than the haTZVPP basis set, which may be an overpolarization problem[340] in the latter case, due to the larger number of diffuse functions haTZVPP. In any case, MAEs of < 1 kcal/mol in total binding energies are achievable for the difficult case of $X^-(H_2O)_n$ clusters, if both E_{3B} and δE_{MP2} are included.

Table 8.4: MAEs[a] in binding energies and relative energies of ten isomers of $F^-(H_2O)_{10}$.

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

[a]With respect to CCSD(T)/CBS benchmarks. [b]Using the "zero-damping" function of Ref. 14. [c]Using the hpTZVPP basis set for XSAPT. [d]Using the aTZ basis set for δMP2. [e]Using the haTZVPP basis set for XSAPT.

We next examine relative energies for ten isomers of $F^-(H_2O)_{10}$ that were considered in Ref. 56. The following scheme is used to obtain CCSD(T)/CBS benchmarks:

$$E_{\mathrm{CBS}}^{\mathrm{CCSD(T)}} \approx E_{\mathrm{CBS}}^{\mathrm{MP2}} + \delta_{\mathrm{MP2\text{-}F12}}^{\mathrm{CCSD(T)\text{-}F12}} \,. \tag{8.30}$$

Here, the correction to the MP2/CBS result is equal to the difference between CCSD(T)-F12/aDZ and MP2-F12/aDZ energies. Binding energies for a variety of XSAPT and DFT approaches are shown in Figure 8.10, and error statistics are listed in Table 8.4. Amongst supersystem methods, ωB97X-V and att-MP2 perform the best.

In the case of XSAPT calculations, we find that the E_{3B} contribution is 5–6 kcal/mol, much larger than its contribution in neutral systems, which makes sense given that E_{3B} is an induction correction. The δMP2 term contributes 1.6 kcal/mol on average, but in contrast to its effect in the smaller halide–water clusters, here the δMP2 term has a deleterious effect on the accuracy of total binding energies. This discrepancy may arise from the manner in which we obtain the CCSD(T)/CBS benchmarks. Specifically, we used the CCSD(T)-F12 results to extrapolate directly to CBS limit in small halide–water clusters but the additive scheme in Eq. (8.30) is used for $F^-(H_2O)_{10}$. Although this additive scheme seems to work well in neutral systems, it is not well tested for anionic systems and may not be appropriate in such cases. We note that the deviation between XSAPT results and these putative CCSD(T)/CBS benchmarks is typically comparable to, or smaller than, the $\sim$2 kcal/mol magnitude of the $\delta_{\mathrm{MP2\text{-}F12}}^{\mathrm{CCSD(T)\text{-}F12}}$ correction in Eq. (8.30) although XSAPT(KS)+D3+E_{3B}/hpTZVPP offers the best binding energies among all methods considered in Table 8.4.

Even if the deviation persists, and is indeed a problem with XSAPT, we note that the ωB97X-D, LC-VV10, M06-2X, M06-2X-D3(zero), B2PLYP, and B2PLYP-D3(zero) supersystem methods all afford errors that are unacceptably large, and cannot be recommended for binding energies of halide–water clusters. (While accurate relative energies are sufficient for structure determination, accurate *total* binding energies in order to compute, *e.g.*, the binding affinity of a drug molecule to a protein.) XSAPT works better with the hpTZVPP basis set than with haTZVPP, which may again be an overpolarization effect. Such effects are well known in QM/MM calculations, where a simple solution is a Gaussian "blurring" of point charges nearby to the QM region.[340] In future work, we plan to test a Gaussian-blurred version of the XPol procedure.

8.4.6 CH_4 in a dodecahedral $(H_2O)_{20}$ cage

The isolated $CH_4@(H_2O)_{20}$ gas-phase cluster has been used as a model system to study the interaction between methane and clathrate hydrates.[342–344] A recent quantum Monte Carlo (QMC) benchmark affords a binding energy of -5.3 ± 0.5 kcal/mol,[341] for $CH_4@(H_2O)_{20} \rightarrow CH_4 + (H_2O)_{20}$, while various other electronic structure methods predict binding energies ranging from -4 to -7 kcal/mol (see Table 8.5). DFT methods generally overestimate the binding energy by about 1 kcal/mol while MP2 and MP2C afford accurate binding energies. It is therefore curious that the double-hybrid B2PLYP functional requires an empirical dispersion correction to get anywhere close to the benchmark binding energy. The nonlocal LC-VV10 functional also severely underestimates the binding energy despite its very good performance for S66. The

Table 8.5: Binding energy of CH_4 to $(H_2O)_{20}$.

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

[a]From Ref. 341. [b]CBS extrapolation.

[b]Using the hpTZVPP basis set.

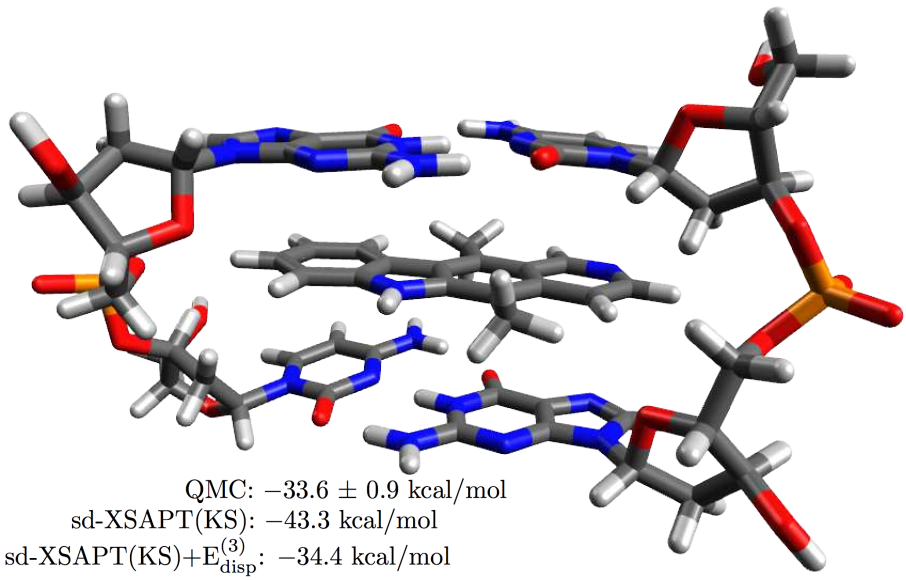

Figure 8.11: Ellipticine molecule intercalated into a GC:GC segment of DNA; binding energies computed with various methods are shown. XSAPT calculations used three fragments: neutral ellipticine and two single-stranded GC complexes, each with with a −1 charge.

att-MP2, DFT-SAPT, and XSAPT(KS)+D3 methods each underestimate the binding energy somewhat. It should be noted that the DFT-SAPT calculation requires the use of $(H_2O)_{20}$ as one monomer unit, whereas in XSAPT the monomers are CH_4 and H_2O.

8.4.7 Anti-cancer drug intercalated into DNA

Predicting accurate non-covalent interaction between biomolecules and drug candidates (*i.e.*, protein–ligand interactions) is a crucial component in drug discovery and design, where docking[345,346] (with force fields or empirical scoring functions) and *ab initio* screening[347,348] (with low-level quantum-chemical methods) are theoretical

mainstays. Here, we consider intercalation of the anti-cancer agent ellipticine[349] into DNA, which involves insertion between two Watson-Crick CG base pairs, linked by their respective phosphate sugar puckers as depicted in Fig. 8.11. The structure depicted in the figure consists of 157 atoms, and a benchmark binding energy is available from QMC calculations.[350] The sd-XSAPT(KS) method overestimates the binding energy by about 10 kcal/mol with respect to this benchmark.

The sd-XSAPT(KS) method ignores the exchange-dispersion component and scales the dispersion component by a factor optimized against CCSD(T)/CBS benchmarks for small dimers; this method may therefore overestimate binding energies in systems with very large monomers. Dispersion-corrected DFT is also known to overestimate binding energies in such systems,[351] and can be improved in such cases by a three-body interatomic dispersion energy ($E^{(3)}_{\rm disp}$) based on the Axilrod-Teller-Muto three-body dispersion formula.[351] For ellipticine intercalated into DNA, we obtain $E^{(3)}_{\rm disp} = 8.90$ kcal/mol, and the corrected sd-XSAPT(KS)+$E^{(3)}_{\rm disp}$ binding energy is -34.4 kcal/mol, which lies within the statistical error bars of the QMC benchmark. The PBE+MBD* method,[40] where "MBD*" is a many-body dispersion correction, yields a binding energy of -35.4 kcal/mol,[350] which is also within the QMC error bars. PBE+MBD* and sd-XSAPT(KS)+$E^{(3)}_{\rm disp}$ are the methods that come closest to the QMC result so far.

8.4.8 Energy decomposition

Table 8.6 shows a statistical summary of the energy components for the dimers in the S22 and S66 data sets, as computed by XSAPT methods and also by EFP.[335] The

Table 8.6: MAEs (in kcal/mol) and percent errors (in parentheses) for individual energy components of the S22 and S66 data sets.

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

[a]Errors with respect to SAPT2+(3)/aTZ energy components. [b]Error with respect to CCSD(T)/CBS binding energy. [c]Using LRC-ωPBEh/jun-cc-pVDZ with 60% short-range HF exchange. [d]Using LRC-ωPBE/haTZVPP. [e]Using LRC-ωPBE/hpTZVPP.

benchmark is SAPT2+(3)/aTZ, results of which were reported for S22 in Ref. 51 and are reported here for S66, for the first time.

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

In contrast to the $X^-(H_2O)_n$ systems, for which the hpTZVPP basis set works slightly better than haTZVPP (possibly owing to overpolarization), here the haTZVPP basis set affords slightly better results for induction energies. We note that charge-transfer interactions show up in the induction energy within the SAPT formalism, but only if the basis set is diffuse enough so that basis functions centered on monomer A extend significantly over monomer B. This may be why the more diffuse haTZVPP performs better here for induction energies. In neutral systems, the overpolarization caused by diffuse basis functions is not large, and it is better to use a large basis set in order to capture charge transfer, whereas overpolarization is more significant in anionic systems, and the more diffuse basis leads to larger errors in binding energies.[52]

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

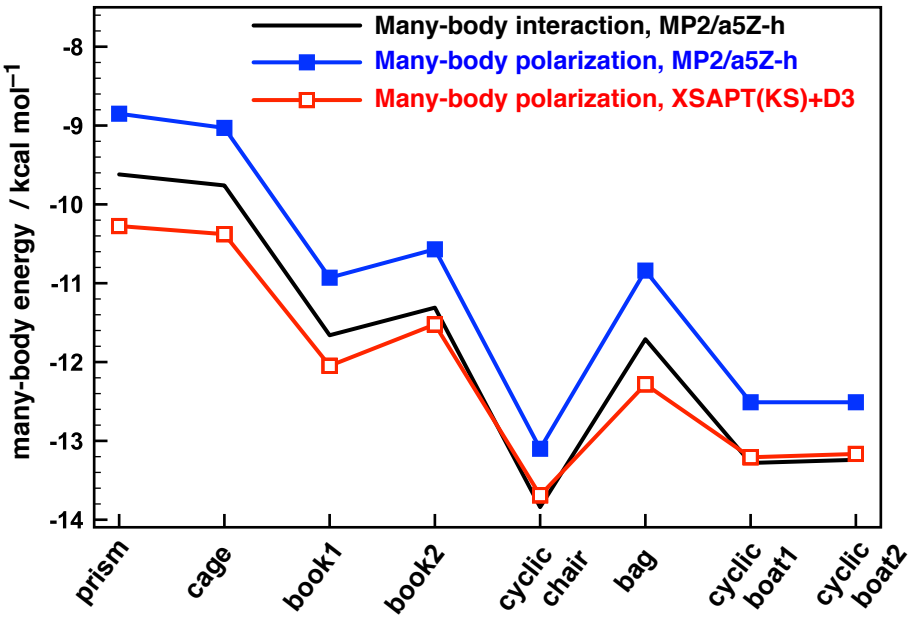

Figure 8.12: Many-body interactions for isomers of $(H_2O)_6$. MP2/a5Z-h results are taken from Ref. 8 and the hpTZVPP basis is used for XSAPT(KS)+D3 calculations.

For the eight $(H_2O)_6$ isomers discussed in Section 8.4.3, energy components have been quantified by Chen and Li[8] at the MP2/a5Z-h level, using a localized molecular orbital energy decomposition analysis. Many-body effects in $(H_2O)_6$ are dominated by polarization interactions, whereas the other energy components are strictly or nearly pairwise additive.[8] Figure 8.12 compares the many-body polarization and total many-body energies for these $(H_2O)_6$ isomers, as reported by Chen and Li, to XSAPT(KS)+D3 results. The latter method inherently assumes that the many-body part of the interaction arises exclusively from polarization. The many-body polarization energies using XSAPT(KS)+D3 are consistently overestimated as compared to the many-body polarization energies evaluated at the MP2/a5Z-h level, but are much closer to the total many-body energies using MP2/a5Z-h.

8.5 Summary

Two new XSAPT-based methods based on a modified dispersion interaction, XSAPT(KS)+D3 and sd-XSAPT(KS), are reported in this article. It has been demonstrated that XSAPT(KS)+D3 is very successful in predicting binding energies for a wide range of challenging systems ranging from benzene dimer to large water- and halide–water clusters. The sd-XSAPT(KS) method performs well for large, dispersion-bound systems, such as a ligand–DNA intercalation complex considered here and (based on preliminary calculations) the L7 database[19] of large organic dimers. However, this method performs less well for water clusters, where the double-ζ basis set that is used in fitting the dispersion scaling parameter cannot adequately

describe electrostatic and induction interactions.

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

There is still room for progress with XSAPT, including the formulation and implementation of analytic energy gradients for geometry optimizations and simulations, and the combination of XSAPT and TDDFT response theory for dispersion energy along the lines of SAPT(DFT). Extension of XSAPT to include intramolecular correlation based on either a Møller-Plesset[114,151] or coupled-cluster formalism[154–156] is possible, as are improvements to the empirical dispersion potential and the charge embedding scheme. Several of these lines of development are currently underway in our group.

CHAPTER 9

Atomic orbital implementation of extended symmetry-adapted perturbation theory (XSAPT) and benchmark calculations for large supramolecular complexes

9.1 Introduction

Efficient and accurate modeling of non-covalent interactions—and in particular, dispersion (van der Waals) interactions—is an active topic of research in computational chemistry, due to the importance of such interactions in the supramolecular chemistry of host/guest complexes, crystal packing, protein folding, and the conformational energies of large molecules in general.[117,138,352–354] Density functional theory (DFT) based on popular semilocal functions is fundamentally incapable of modeling the long-range electron correlation effects that give rise to dispersion, which have inherently quantum-mechanical, many-body, and non-local properties. To do better requires either a non-local correlation functional,[144,145,147] or else the addition of *a posteriori* corrections. Examples of the latter approach include empirical atom–atom dispersion potentials,[143] as well as similar potentials wherein the C_6 coefficients are

determined on-the-fly based on the electron density[355–358] or further include the electrodynamic response effects due to the surrounding polarizable atoms by solving the self-consistent screening equation of classical electrodynamics.[359–362]

On the other hand, second-order Møller-Plesset perturbation theory (MP2), while effective at describing intermolecular interactions in systems dominated by electrostatics and polarization,[52] usually overestimates the dispersion energy, especially for π-stacking.[57,363] This is due to the fact that dispersion at the MP2 level operates within the uncoupled Hartree-Fock (HF) approximation.[278] Various strategies have been used to address this deficiency, including spin-component scaled MP2 (SCS-MP2),[263] long-range attenuated MP2 (att-MP2),[176] and a "coupled" MP2 approach (MP2C).[309,310] At present, the $\mathcal{O}(N^7)$ coupled-cluster method with single, double, and perturbative triple excitations [CCSD(T)], extrapolated to the complete basis-set (CBS) limit, remains the gold standard for non-covalent interactions, with corrections to binding energies beyond CCSD(T) being < 0.1 kcal/mol.[33,141]

A perturbative view of intermolecular interactions results in an alternative method called symmetry-adapted perturbation theory (SAPT),[114–117,149,150] which provides a solid theoretical basis to resolve one of the principle shortcomings of the supermolecular approach to non-covalent interactions, namely, basis-set superposition error (BSSE). This is bypassed in SAPT because the perturbation series computes the binding energy directly and energy subtraction is not required. Moreover, the intermolecular energy in SAPT is naturally decomposable into physically-meaningful

components. In order to obtain accurate interaction energies, intramolecular electron correlation effects must be included, which can be accomplished using either a Møller-Plesset formalism[114,151] or a coupled-cluster formulation.[154–156,172,239,364] Unfortunately these high-level SAPT methods such SAPT2+3 or SAPT(CCSD), while generally quite accurate,[117,172] exhibit the same $\mathcal{O}(N^7)$ scaling as CCSD(T). On the other end of the spectrum, the SAPT0 method describes the monomers at the HF level and uses second-order perturbation theory for the intermolecular interactions. SAPT0 has been applied to very large systems including a proflavine/DNA complex,[160] protein/ligand complexes,[332] and π-stacked complexes of graphene with other aromatic molecules.[365,366]

Alternatively, intramolecular electron correlation can be incorporated into SAPT in a low-cost way using Kohn-Sham (KS) orbitals for the monomers, with the remaining long-range intermolecular correlation handled by the perturbative expansion. This approach is known as SAPT(KS).[208] Proper asymptotic behavior of the monomer DFT exchange-correlation potentials is crucial to the success of such calculations,[162,208,209] and we have demonstrated that this can be achieved using long-range corrected (LRC) density functionals.[51] A DFT-based alternative to SAPT(KS) is DFT-SAPT,[149,150] in which DFT response theory is used to compute frequency-dependent polarizabilities for the monomers, which are then used to compute the "coupled Kohn-Sham" (CKS) dispersion interaction, via a Casimir-Polder-type formalism. Proper asymptotic behavior of the exchange-correlation potential remains important, and is accomplished using various "splicing" schemes.[199,215,331]

Traditional SAPT is a well-defined method for dimers, and while three-body non-additivity corrections have been derived,[249,250,367] they are expensive to evaluate. For this reason, our group has developed an "extended" version of SAPT (XSAPT),[54,57,65] in which the monomer wave functions incorporate many-body polarization effects via the variational explicit polarization (XPol) method.[252] These monomer wave functions are then used in subsequent SAPT calculations, thereby extending traditional SAPT to many-body systems.

The computational scaling of XSAPT depends on the treatment of the dispersion and exchange-dispersion energies.[57] The SAPT0-style treatment via second-order perturbation theory scales as $\mathcal{O}(N^5)$ and affords a poor description of dispersion.[54] An alternative is to omit the exchange-dispersion term and then to scale the direct dispersion term by an empirical factor, an $\mathcal{O}(N^4)$ strategy originally introduced by Ochsenfeld and co-workers[332] in the context of SAPT0 and later implemented by the present authors for use with XSAPT.[57] A further approximation is to neglect the dispersion term also, replacing the sum of dispersion and exchange-dispersion energies with empirical atom–atom potentials.[55–57] (This idea was originally introduced in the context of traditional SAPT by Heßelmann.[16]) This "XSAPT(KS)+D" method exhibits $\mathcal{O}(N^3)$ scaling and avoids any double-counting of dispersion interactions at short and medium range. (The same cannot be said of dispersion-corrected DFT,[253,254,368] except in the context of the "dispersionless" density functional developed by Szalewicz and co-workers.[6,369]) Three generations of empirical dispersion potentials were developed for use in XSAPT,[55–57] with the latest ("+D3") affording

accurate results for both total binding energies and individual energy components for many different systems.[57]

For small molecules, the performance of XSAPT(KS)+D3 is typically superior to alternative supermolecular methods with similar scaling,[57] but results for systems with large monomers have been limited. There is reason to question the validity of atomic-pairwise dispersion potentials in systems composed of a large number of highly polarizable centers,[370] such as molecular crystals or supramolecular complexes assembled from highly conjugated monomers.[371–373] The present work therefore considers examples from the latter category. Such examples often involve monomers that are rather large, so to facilitate XSAPT(KS)+D3 calculations in these cases, we have reformulated XSAPT in the atomic orbital (AO) basis. This formulation avoids the four-index integral transformation that is required in the original, molecular orbital (MO) version of the method, and exhibits $\mathcal{O}(N^3)$ scaling with respect to dimer size.

9.2 Methods

9.2.1 Atomic orbital implementation of XSAPT

In XSAPT(KS), we consider intermolecular correlation through second order and incorporate intramolecular correlation via monomer DFT calculations. The interaction energy can be written as

$$\begin{aligned} E_{\text{int}}^{\text{SAPT0}} = E_{\text{elst}}^{(1)} + E_{\text{exch}}^{(1)} + E_{\text{ind}}^{(2)} + E_{\text{exch-ind}}^{(2)} \\ + E_{\text{disp}}^{(2)} + E_{\text{exch-disp}}^{(2)} + \delta E_{\text{int}}^{\text{HF}}. \end{aligned} \tag{9.1}$$

The various terms in Eq. (9.1) are discussed below, but we note here that the final term,

$$\delta E_{\rm int}^{\rm HF} = E_{\rm int}^{\rm HF} - \Big[E_{\rm elst}^{(10)}({\rm HF}) + {\rm E}_{\rm exch}^{(10)}({\rm HF}) + E_{\rm ind,resp}^{(20)}({\rm HF}) + {\rm E}_{\rm exch\text{-}ind,resp}^{(20)}({\rm HF}) \Big] \ , \tag{9.2}$$

is often incorporated to capture polarization effects beyond second order, by means of a counterpoise-corrected HF calculation of the binding energy ($E_{\rm int}^{\rm HF}$). The energy components in Eq. (9.2) are described at the HF level, neglecting intramolecular correlation, but a response ("resp") correction, as described below, is included in the second-order induction and exchange-induction terms.

The non-dispersion terms in Eq. (9.1) are

$$E_{\rm non\text{-}disp}^{\rm SAPT0} = E_{\rm elst}^{(1)} + E_{\rm exch}^{(1)} + E_{\rm ind}^{(2)} + E_{\rm exch\text{-}ind}^{(2)} \ , \tag{9.3}$$

By evaluating these terms in the AO basis, one sidesteps the need for a costly four-index integral transformation. The AO formulation also circumvents problems with linear dependencies (which we have encountered in large systems when using a resolution-of-identity implementation of XSAPT), and naturally facilitates use of linear-scaling algorithms for construction of the Coulomb (**J**) and exchange (**K**) matrices. Closed-shell formulas to implement Eq. (9.1) in the AO basis were first provided by Heßelmann *et al.*,[137,374] and a more efficient formulation of $E_{\rm exch}^{(1)}(S^2)$ and $E_{\rm exch\text{-}ind}^{(2)}$ was introduced later by Beer.[375] Open-shell formulas have been presented by Hapka *et al.*[376] Below, we provide both the closed-shell and open-shell AO expressions, using Beer's formulation since it requires construction of one fewer exchange matrix for $E_{\rm exch}^{(1)}(S^2)$ and one fewer Coulomb matrix for $E_{\rm exch\text{-}ind}^{(2)}$, as compared to Heßelmann's

formulation. The closed-shell AO-expression of an exact form for $E^{(1)}_{\rm exch}$ shown below is based on its open-shell expression in Ref. 376. Two minor typographical errors in Ref. 375 and 376 are corrected here.

In the AO basis, the closed-shell (CS) first-order electrostatic energy can be expressed as

$$[E^{(1)}_{\rm elst}]_{\rm CS} = {\rm tr}\big(2\,\mathbf{P}^A\mathbf{V}^B + 2\,\mathbf{P}^B\mathbf{V}^A + 4\,\mathbf{P}^B\mathbf{J}^A\big) + V_0\ , \tag{9.4}$$

where $\mathbf{P}^{A,B}$ are the one-particle density matrices for monomers A and B, $\mathbf{V}^{A,B}$ are the corresponding electrostatic potentials, $\mathbf{J}^{A,B} \equiv \mathbf{J}[\mathbf{P}^{A,B}]$ are the Coulomb matrices for monomers A and B, and V_0 is the internuclear repulsion energy. Note that ${\rm tr}(\mathbf{P}^B\mathbf{J}^A) = {\rm tr}(\mathbf{P}^A\mathbf{J}^B)$, which is why Eq. (9.4) does not at first appear to be symmetric with respect to interchange of A and B.

The corresponding open-shell (OS) version is

$$\begin{aligned}[E^{(1)}_{\rm elst}]_{\rm OS} = {\rm tr}\Big[&(\mathbf{P}^A_\alpha + \mathbf{P}^A_\beta)\mathbf{V}^B \\ &+ (\mathbf{P}^B_\alpha + \mathbf{P}^B_\beta)(\mathbf{V}^A + \mathbf{J}^A_\alpha + \mathbf{J}^A_\beta)\Big] + V_0\ ,\end{aligned} \tag{9.5}$$

where σ (with $\sigma = \alpha$ or β) in the subscript indicates the σ-spin component for the corresponding matrix.

Regarding the exchange terms in Eq. (9.1), an exact form for $E^{(1)}_{\rm exch}$ was written down almost 40 years ago,[127] but analytic forms of $E^{(2)}_{\rm exch\text{-}ind}$ and $E^{(2)}_{\rm exch\text{-}disp}$ were only published recently,[174,175] so the second-order exchange energies have historically always been evaluated using the "single exchange" (or "S^2") approximation.[115,116] This approximation is accurate beyond the van der Waals (vdW) contact distance,[115,174,175] except when one of the monomers is an anion, in which case the $E^{(1)}_{\rm exch}(S^2)$ and

$E^{(2)}_{\text{exch-ind}}(S^2)$ terms must be scaled in order to obtain accurate results.[50,52] (This scaling is based on the ratio of the exact and S^2 results for first-order exchange.)

Defining

$$\mathbf{\Omega}^{A,B} = \mathbf{V}^{A,B} + 2\ \mathbf{J}^{A,B} \tag{9.6}$$

$$\mathbf{h}^{A,B} = \mathbf{\Omega}^{A,B} - \mathbf{K}^{A,B}\ , \tag{9.7}$$

the closed-shell exact first-order exchange energy can be expressed as

$$\begin{aligned}
[E^{(1)}_{\text{exch}}]_{\text{CS}} = 2\,\text{tr}\Big(&-\mathbf{P}^A\mathbf{K}^B + \mathbf{T}^A\mathbf{h}^B + \mathbf{T}^B\mathbf{h}^A \\
&+ \mathbf{T}^{AB}\mathbf{h}^A + \mathbf{T}^{AB}\mathbf{h}^B + \mathbf{T}^A(2\ \mathbf{J}[\mathbf{T}^B] - \mathbf{K}[\mathbf{T}^B]) \\
&+ \mathbf{T}^{AB}(2\ \mathbf{J}[\mathbf{T}^A] - \mathbf{K}[\mathbf{T}^A]) \\
&+ \mathbf{T}^{AB}(2\ \mathbf{J}[\mathbf{T}^B] - \mathbf{K}[\mathbf{T}^B]) \\
&+ \mathbf{T}^{AB}(2\ \mathbf{J}[\mathbf{T}^{BA}] - \mathbf{K}[\mathbf{T}^{BA}])\Big)\ .
\end{aligned} \tag{9.8}$$

The matrices $\mathbf{T}^{A,B,AB}$ can be obtained via back-transformed the block matrix of a matrix $\mathbf{D}$ to the AO basis by the MO coefficient matrices $\mathbf{C}^A$ and $\mathbf{C}^B$ of monomer A and B, respectively. They are defined as

$$\mathbf{T}^A = \mathbf{C}^A\mathbf{D}_{aa}(\mathbf{C}^A)^\dagger\ , \tag{9.9}$$

$$\mathbf{T}^B = \mathbf{C}^B\mathbf{D}_{bb}(\mathbf{C}^B)^\dagger\ , \tag{9.10}$$

$$\mathbf{T}^{AB} = \mathbf{C}^A\mathbf{D}_{ab}(\mathbf{C}^B)^\dagger\ , \tag{9.11}$$

$$\mathbf{T}^{BA} = \mathbf{C}^B\mathbf{D}_{ba}(\mathbf{C}^A)^\dagger\ . \tag{9.12}$$

The matrix $\mathbf{D} = \begin{pmatrix} \mathbf{D}_{aa} & \mathbf{D}_{ab} \\ \mathbf{D}_{ba} & \mathbf{D}_{bb} \end{pmatrix}$ can be calculated as

$$\mathbf{D} = [\mathbf{1} + \mathbf{S}^{AB}]^{-1} - \mathbf{1} = -\mathbf{S}^{AB} + (\mathbf{S}^{AB})^2 - (\mathbf{S}^{AB})^3 + \cdots\ , \tag{9.13}$$

where $\mathbf{S}^{AB}$ is the dimer overlap matrix between the occupied molecular orbitals of both fragments. The matrix $\mathbf{S}^{AB}$ can be written as $\mathbf{S}^{AB} = \begin{pmatrix} \mathbf{0} & \mathbf{S}_{ab} \\ \mathbf{S}_{ba} & \mathbf{0} \end{pmatrix}$ where $S_{ab} = \langle \phi_a | \phi_b \rangle$ with $a \in A$ and $b \in B$ used to label occupied MOs belonging to fragments A and B, respectively.

The matrices $\mathbf{K}^{A,B} = \mathbf{K}[\mathbf{P}^{A,B}]$ are the usual Hartree-Fock exchange matrices. The $\mathbf{J}[\mathbf{X}]$ and $\mathbf{K}[\mathbf{X}]$ are Coulomb and exchange matrices formed from the generalized density matrix $\mathbf{X}$, respectively:

$$(\mathbf{J}[\mathbf{X}])_{\mu\nu} = \sum_{\lambda\sigma} X_{\lambda\sigma} (\mu\nu|\lambda\sigma) \; , \tag{9.14}$$

$$(\mathbf{K}[\mathbf{X}])_{\mu\nu} = \sum_{\lambda\sigma} X_{\lambda\sigma} (\mu\lambda|\sigma\nu) \; . \tag{9.15}$$

The corresponding open-shell version is

$$\begin{aligned}
[E^{(1)}_{\text{exch}}]_{\text{OS}} = \sum_{\sigma} \text{tr} \Big\{ & -\mathbf{P}^A_\sigma \, \mathbf{K}^B_\sigma + \mathbf{T}^A_\sigma \, \mathbf{h}^B_\sigma + \mathbf{T}^B_\sigma \, \mathbf{h}^A_\sigma \\
& + \mathbf{T}^{AB}_\sigma \, \mathbf{h}^A_\sigma + \mathbf{T}^{AB}_\sigma \, \mathbf{h}^B_\sigma + \mathbf{T}^A_\sigma \Big(\mathbf{J}[\mathbf{T}^B_\sigma] - \mathbf{K}[\mathbf{T}^B_\sigma] \Big) \\
& + \mathbf{T}^{AB}_\sigma \Big(\mathbf{J}[\mathbf{T}^A_\sigma] - \mathbf{K}[\mathbf{T}^A_\sigma] \Big) \\
& + \mathbf{T}^{AB}_\sigma \Big(\mathbf{J}[\mathbf{T}^B_\sigma] - \mathbf{K}[\mathbf{T}^B_\sigma] \Big) \\
& + \mathbf{T}^{AB}_\sigma \Big(\mathbf{J}[\mathbf{T}^{BA}_\sigma] - \mathbf{K}[\mathbf{T}^{BA}_\sigma] \Big) \Big\} \\
& + \mathbf{T}^A_\alpha \, \mathbf{J}[\mathbf{T}^B_\beta] + \mathbf{T}^A_\beta \, \mathbf{J}[\mathbf{T}^B_\alpha] \\
& + \mathbf{T}^{AB}_\alpha \, \mathbf{J}[\mathbf{T}^A_\beta] + \mathbf{T}^{AB}_\beta \, \mathbf{J}[\mathbf{T}^A_\alpha] \\
& + \mathbf{T}^{AB}_\alpha \, \mathbf{J}[\mathbf{T}^B_\beta] + \mathbf{T}^{AB}_\beta \, \mathbf{J}[\mathbf{T}^B_\alpha] \\
& + \mathbf{T}^{AB}_\alpha \, \mathbf{J}[\mathbf{T}^{BA}_\beta] + \mathbf{T}^{AB}_\beta \, \mathbf{J}[\mathbf{T}^{BA}_\alpha] \; .
\end{aligned} \tag{9.16}$$

Errors introduced by the S^2 approximation in $E^{(1)}_{\text{exch}}$ and $E^{(2)}_{\text{exch-ind}}$ are somewhat cancelled by the errors arising from the S^2 approximation used in the $\delta E^{\text{HF}}_{\text{int}}$ term, and

the S^2 approximation to the first-order exchange energy avoids the need to invert the dimer overlap matrix. The S^2 approximation will be used for all calculations reported in this work. The first-order exchange energy within the S^2 approximation can be expressed as

$$\begin{aligned}[E^{(1)}_{\rm exch}(S^2)]_{\rm CS} = -2\,{\rm tr}\Big(&\mathbf{P}^A\mathbf{K}^B + \mathbf{O}^\dagger\mathbf{h}^A + \mathbf{O}\mathbf{h}^B \\ &- \mathbf{P}^B\mathbf{S}\mathbf{O}\mathbf{\Omega}^A - \mathbf{O}\mathbf{S}\mathbf{P}^A\mathbf{\Omega}^B + \mathbf{O}\mathbf{K}^\dagger[\mathbf{O}]\Big)\,.\end{aligned} \tag{9.17}$$

The quantity

$$\mathbf{O} = \mathbf{P}^A\mathbf{S}\mathbf{P}^B \tag{9.18}$$

is a generalized density matrix, with $\mathbf{S}$ and $\mathbf{P}$ denoting the overlap and density matrices, respectively.

The corresponding open-shell version is:

$$\begin{aligned}[E^{(1)}_{\rm exch}(S^2)]_{\rm OS} = -\sum_\sigma {\rm tr}\Big(&\mathbf{P}^A_\sigma\,\mathbf{K}^B_\sigma + \mathbf{O}^\dagger_\sigma\,\mathbf{h}^A_\sigma \\ &+ \mathbf{O}_\sigma\,\mathbf{h}^B_\sigma - \mathbf{P}^B_\sigma\mathbf{S}\mathbf{O}_\sigma\mathbf{\Omega}^A \\ &- \mathbf{O}_\sigma\mathbf{S}\mathbf{P}^A_\sigma\mathbf{\Omega}^B + \mathbf{O}_\sigma\mathbf{K}^\dagger[\mathbf{O}_\sigma]\Big)\,,\end{aligned} \tag{9.19}$$

Construction of non-symmetric $\mathbf{K}[\mathbf{O}]$ and $\mathbf{K}[\mathbf{T}^{BA}]$ is the rate-determining step in evaluating the non-dispersion terms in SAPT0 for first-order exchange term with and without involving S^2 approximation, respectively. The cost of open-shell AO-SAPT is twice that of the closed-shell versions, since the rate-determining step amongst the non-dispersion terms in SAPT0 must be done for each spin density.

The second-order induction energy can be written as

$$E^{(2)}_{\rm ind} = E^{(2)}_{\rm ind}(A \leftarrow B) + E^{(2)}_{\rm ind}(B \leftarrow A), \tag{9.20}$$

where $E^{(2)}_{\rm ind}(A \leftarrow B)$ indicates the induction energy for monomer A due to the perturbing field of a frozen charge density from B. In the AO basis,

$$[E^{(2)}_{\rm ind}]_{\rm CS} = 2\,{\rm tr}\big(\mathbf{X}^A\,\mathbf{\Omega}^B + \mathbf{X}^B\,\mathbf{\Omega}^A\big) \tag{9.21}$$

where

$$\begin{aligned} X^A_{\mu\nu} &= \sum_{ar} c^A_{\nu r}\; U_{ra}\; (c^A_{\mu a})^\dagger \\ X^B_{\mu\nu} &= \sum_{bs}\; c^B_{\nu s}\; U_{sb}\; (c^B_{\mu b})^\dagger\;. \end{aligned} \tag{9.22}$$

The quantity U_{ra} is defined as

$$U_{ra} = \frac{\Omega_{ra}}{\varepsilon_a - \varepsilon_r} \tag{9.23}$$

and U_{sb} is defined similarly, with $r \in A$ and $s \in B$ used to label virtual MOs belonging to fragments A and B, respectively.

The corresponding open-shell version is

$$[E^{(2)}_{\rm ind}]_{\rm OS} = {\rm tr}\big(\mathbf{X}^A_\alpha\,\mathbf{\Omega}^B + \mathbf{X}^A_\beta\,\mathbf{\Omega}^B + \mathbf{X}^B_\alpha\,\mathbf{\Omega}^A + \mathbf{X}^B_\beta\,\mathbf{\Omega}^A\big) \tag{9.24}$$

Within the S^2 approximation, the second-order exchange-induction energy for monomer A is

$$\begin{aligned} &[E^{(2)}_{\text{exch-ind}}(S^2)(A \leftarrow B)]_{\rm CS} \\ &\quad = -2\,{\rm tr}\Big(\mathbf{X}^A\mathbf{K}^B + \mathbf{X}^A\mathbf{S}\mathbf{P}^B\mathbf{h}^A + \mathbf{P}^B\mathbf{S}\mathbf{X}^A\mathbf{h}^B \\ &\qquad - \mathbf{P}^B\mathbf{S}\mathbf{X}^A\mathbf{S}\mathbf{P}^B\mathbf{\Omega}^A - \mathbf{X}^A\mathbf{S}\mathbf{O}^\dagger\mathbf{\Omega}^B - \mathbf{O}\mathbf{S}\mathbf{X}^A\mathbf{\Omega}^B \\ &\qquad + 2\;\mathbf{O}^\dagger\mathbf{J}[\mathbf{X}^A] - 2\;\mathbf{P}^B\mathbf{S}\mathbf{O}\mathbf{J}[\mathbf{X}^A] - \mathbf{X}^A\mathbf{K}[\mathbf{O}] \\ &\qquad + \mathbf{X}^A\mathbf{S}\mathbf{P}^B\mathbf{K}^\dagger[\mathbf{O}] + \mathbf{P}^B\mathbf{S}\mathbf{X}^A\mathbf{K}[\mathbf{O}]\Big)\;. \end{aligned} \tag{9.25}$$

The corresponding term $E^{(2)}_{\rm exch-ind}(S^2)(B \leftarrow A)$ is evaluated analogously.

The $E^{(2)}_{\text{ind}}$ and $E^{(2)}_{\text{exch-ind}}$ terms are commonly replaced by their "response" analogues, $E^{(2)}_{\text{ind,resp}}$ and $E^{(2)}_{\text{exch-ind,resp}}$, in which coupled-perturbed Hartree-Fock (CPHF) equations are solved in order to compute the infinite-order correction for induction arising from a frozen partner density.[179,180] The CPHF coefficients can be back-transformed to the AO basis as $\mathbf{U}$ in Eq. (9.22) and substituted into the $\mathbf{X}^A$ and $\mathbf{X}^B$ matrices in Eq. (9.21) and (9.25), in which case they afford the response-corrected energies $E^{(2)}_{\text{ind,resp}}$ and $E^{(2)}_{\text{exch-ind,resp}}$, respectively.

Subject to the S^2 approximation, the corresponding open-shell version of the second-order exchange-induction energy for monomer A due to the perturbing field of a frozen charge density on monomer B is

$$\begin{aligned}
&[E^{(2)}_{\text{exch-ind}}(S^2)(A \leftarrow B)]_{\text{OS}} \\
&\quad = -\sum_{\sigma} \text{tr}\Big(\mathbf{X}^A_\sigma \mathbf{K}^B_\sigma + \mathbf{X}^A_\sigma \mathbf{S} \mathbf{P}^B_\sigma \mathbf{h}^A_\sigma \\
&\qquad\quad + \mathbf{P}^B_\sigma \mathbf{S} \mathbf{X}^A_\sigma \mathbf{h}^B_\sigma - \mathbf{P}^B_\sigma \mathbf{S} \mathbf{X}^A_\sigma \mathbf{S} \mathbf{P}^B_\sigma \mathbf{\Omega}^A \\
&\qquad\quad - \mathbf{X}^A_\sigma \mathbf{S} \mathbf{O}^\dagger_\sigma \mathbf{\Omega}^B - \mathbf{O}_\sigma \mathbf{S} \mathbf{X}^A_\sigma \mathbf{\Omega}^B \\
&\qquad\quad + \mathbf{O}^\dagger_\sigma \mathbf{J}[\mathbf{X}^A_\sigma] - \mathbf{P}^B_\sigma \mathbf{S} \mathbf{O}_\sigma \mathbf{J}[\mathbf{X}^A_\sigma] \\
&\qquad\quad - \mathbf{X}^A_\sigma \mathbf{K}[\mathbf{O}_\sigma] + \mathbf{X}^A_\sigma \mathbf{S} \mathbf{P}^B_\sigma \mathbf{K}^\dagger[\mathbf{O}_\sigma] \\
&\qquad\quad + \mathbf{P}^B_\sigma \mathbf{S} \mathbf{X}^A_\sigma \mathbf{K}[\mathbf{O}_\sigma] \Big) \\
&\qquad + \text{tr}\Big(\mathbf{O}^\dagger_\alpha \mathbf{J}[\mathbf{X}^A_\beta] + \mathbf{O}^\dagger_\beta \mathbf{J}[\mathbf{X}^A_\alpha] \\
&\qquad\quad - \mathbf{P}^B_\alpha \mathbf{S} \mathbf{O}_\alpha \mathbf{J}[\mathbf{X}^A_\beta] - \mathbf{P}^B_\beta \mathbf{S} \mathbf{O}_\beta \mathbf{J}[\mathbf{X}^A_\alpha] \Big) \, .
\end{aligned} \tag{9.26}$$

9.2.2 Intramolecular correlation using tuned LRC-DFT

The asymptotic (large r) behavior of the exchange-correlation potential is[210]

$$v_{\rm xc}(r) \sim -\frac{1}{r} + \Delta_\infty \tag{9.27}$$

with $\Delta_\infty \to {\rm IP} + \varepsilon_{\rm HOMO}$ as $r \to \infty$.[210,211] Here, "IP" denotes the lowest ionization potential and $\varepsilon_{\rm HOMO}$ is the KS eigenvalue for the highest occupied molecular orbital (HOMO). To achieve proper asymptotic behavior $v_{\rm xc}(r) \sim -1/r$, Baer and co-workers[229,230] propose to tune the range-separation parameter (ω) in LRC-DFT in order to satisfy the condition $\varepsilon_{\rm HOMO}(\omega) = -{\rm IP}(\omega)$, or in other words $\Delta_\infty = 0$.

Correct asymptotic behavior is crucial for obtaining accurate energy components in DFT-based SAPT,[162,208,209] and the non-empirical tuning procedure of Baer and co-workers, applied separately to each monomer, affords such behavior and provides accurate energy components.[51,56,57] At the same time, tuned LRC-DFT retains the relationship $v_{\rm xc} = \delta E_{\rm xc}/\delta\rho$ that is sacrificed when using the asymptotic "splicing" schemes[199,215,331] that have traditionally been employed in DFT-based SAPT. Tuned LRC-DFT functionals also provide a better description of properties such as polarizabilities and isotropic C_6 coefficients, and afford smaller delocalization errors, as compared to splicing methods.[377,378] These quantities are directly related to the SAPT energy components, and Hapka *et al.*[378] have demonstrated that SAPT based on tuned LRC-DFT provides better or comparable results as compared to the traditional splicing approach.[378] We will denote by "$\omega_{\rm IP}$" the value that satisfies the tuning condition

$$\varepsilon_{\rm HOMO}(\omega_{\rm IP}) = -{\rm IP}(\omega_{\rm IP})\ . \tag{9.28}$$

Tuned values of ω_{IP} are available in the Supporting Information.

It should be noted that the value ω_{IP} exhibits a troublesome dependence on system size.[379,380] An alternative method to select ω is the global density-dependent (GDD) tuning procedure,[26] in which the optimal value

$$\omega_{\mathrm{GDD}} = C\langle d_{\mathrm{x}}^2 \rangle^{-1/2} \tag{9.29}$$

is related to the average distance d_{x} between an electron in the outer regions of a molecule and the exchange hole in the region of localized orbitals, and C is an empirical constant for a given LRC functional. Following the procedure in Ref. 26, we determined $C = 0.885$ for the LRC-ωPBE functional,[227] with $\omega = 0.3\ a_0^{-1}$. (Details can be found in the Supporting Information.) The basis set in these calculations is def2-TZVPP augmented with diffuse functions on non-hydrogen atoms that are taken from Dunning's aug-cc-pVTZ (aTZ) basis set. (Henceforth, we refer to this basis as "haTZVPP".) Our optimized parameter C is almost the same as that determined in Ref. 26 ($C = 0.90$) for LRC-ωPBE with $\omega = 0.4\ a_0^{-1}$ and the def2-TZVPP basis set. Since LRC-ωPBE(ω_{GDD}) provides a better description of polarizabilities in polyacetylene as compared to ω_{IP},[378] it is anticipated that using ω_{GDD} in place of ω_{IP} may afford more accurate energy components, especially in conjugated systems. (Many of the supramolecular complexes considered here fall into this category.)

9.2.3 Dispersion corrections

The expensive and inaccurate SAPT0-style description of dispersion ($E_{\mathrm{disp}}^{(2)} + E_{\mathrm{exch\text{-}disp}}^{(2)}$) can be replaced by a third-generation dispersion potential ("+D3") that we have

recently developed.[57] It should be emphasized that our D3 potential is unrelated to Grimme's D3 correction for DFT.[274] Whereas the latter is a model-dependent quantity with no real physical meaning,[253] our D3 correction is a true dispersion potential, and exhibits errors of $\lesssim 0.2$ kcal/mol with respect to dispersion energies obtained from high-level SAPT calculations.[57]

It should be stressed that the fitting set for D3 consists of monomers containing no more than 20 non-hydrogen atoms, whereas the molecules considered here are much larger. It has been demonstrated that pairwise dispersion methods overestimate the dispersion in certain supramolecular complexes and molecular crystals,[371–373] in what Dobson has classified as "type-B" non-additivity.[370] Even MP2 (and thus SAPT0) fails to capture type B effects,[370] and various methods have been proposed to include type-B non-additivity within a pairwise scheme. These include the many-body dispersion (MBD) method,[359–362] an empirical Axilrod-Teller-Muto (ATM) three-body dispersion term $E_{\rm disp,3B}^{\rm ATM(Grimme)}$ developed by Grimme,[274] and another ATM-type correction that is derived "on the fly" from the electron density.[351] The latter is based on the Tkatchenko-Scheffler (TS) formalism,[358] and will be denoted here as $E_{\rm disp,3B}^{\rm ATM(TS)}$. Preliminary results for a DNA intercalation complex suggest that these corrections, in conjunction with sd-XSAPT(KS), afford accurate interaction energies for large supramolecular complexes.[57]

In detail, the three-body ATM triple-dipole term is

$$E_{\rm disp,3B}^{\rm ATM} = \sum_{A<B<C} C_9^{ABC} \left(\frac{3\cos\theta_a \cos\theta_b \cos\theta_c + 1}{R_{AB}^3 R_{AC}^3 R_{BC}^3} \right) f_{\rm damp}^{ABC} \tag{9.30}$$

where the θ_x are the internal angles in the atomic ABC triangle. The coefficients

C_9^{ABC} are functionals of the electron density that depend on the atom-in-molecule C_9^X coefficients ($X = A, B, C$) and static polarizabilities α_0^X, which in turn depend on the free-atom quantities $C_{9,\text{free}}^X$ and $\alpha_{0,\text{free}}^X$. Hirshfeld volume partitioning[356] is used to determine the atomic volume ratio relative to that of the free atom:

$$v^X = \frac{V^X}{V_{\text{free}}^X} \ . \tag{9.31}$$

The C_9 coefficient and polarizability for atom X are then taken to be

$$\begin{aligned} C_9^X &= (v^X)^3 \ C_{9,\text{free}}^X \\ \alpha_0^X &= v^X \ \alpha_{0,\text{free}}^X \end{aligned} \ . \tag{9.32}$$

The damping function f_{damp}^{ABC} in Eq. (9.30) is a product of two-body Tang-Toennies damping functions,

$$f_{\text{damp}}^{ABC} = f_6^{AB} f_6^{AC} f_6^{BC} \ , \tag{9.33}$$

with

$$f_6^{AB}(R_{AB}) = 1 - e^{-b_{AB} R_{AB}} \sum_{k=0}^{6} \frac{b_{AB}^k R_{AB}^k}{k!} \tag{9.34}$$

and a range parameter $b_{AB} = 0.31 \times R_{\text{vdW}}^{AB} + 3.43 \ a_0^{-1}$. Here, $R_{\text{vdW}}^{AB} = R_{\text{vdW}}^A + R_{\text{vdW}}^B$ is a sum of atom-in-molecule vdW radii that are again obtained from the corresponding free-atom quantities using Hirshfeld volume partitioning:

$$R_{\text{vdW}}^X = (v^X)^{1/3} \ R_{\text{vdW,free}}^X \ . \tag{9.35}$$

To obtain total electron densities for calculating $E_{\text{disp,3B}}^{\text{ATM(TS)}}$, one must perform separate calculations on the dimer as well as both monomers. Then

$$\begin{aligned} E_{\text{disp,3B}}^{\text{ATM(TS)}}(\text{int}) = E_{\text{disp,3B}}^{\text{ATM(TS)}}(\text{dimer}) - E_{\text{disp,3B}}^{\text{ATM(TS)}}(\text{mono1}) \\ - E_{\text{disp,3B}}^{\text{ATM(TS)}}(\text{mono2}) \ . \end{aligned} \tag{9.36}$$

To save time for large dimers, average hybridized values of α_0, C_9, and R_{vdW} are sometimes determined in advance and used to estimate $E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$.[351] We denote this as $E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$(param), the use of which sidesteps the need for a self-consistent field calculation on the dimer.

It has been argued that inclusion of a three-body dispersion term may overestimate the extent to which the two-body dispersion should be reduced, and that contributions up to sixth order are necessary for a converged dispersion interaction.[40] However, the C_8R^{-8} term is also included in our D3 dispersion potential, whereas this term is omitted in the MBD* approach of Tkatchenko *et al.*[40] The magnitude of the C_8R^{-8} term is roughly 1/3 that of the C_6R^{-6} term,[255] and we expect that the C_8R^{-8} term reduces the overcompensation caused by $E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$.

As a "control experiment", we tested the combination of $E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$ with our D2 and D3 dispersion potentials,[51,57] for the S66 data set, and we find that the addition of $E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$ to XSAPT(KS)+D changes the mean absolute errors (MAEs) in S66 interaction energies by no more than 0.12 kcal/mol for these dimers, This is consistent with the high-quality results that we obtain using XSAPT(KS)+D2 and +D3 in small molecules, where atom-wise non-pairwise-additive dispersion is not important.

On the other hand, our empirical dispersion potentials do not explicitly include "type-A" non-additive effects, in which crowding by neighboring atoms results in a reduction in atomic polarizabilities and thus dispersion energies.[370] Type A effects *are* accounted for by methods such as the Tkatchenko-Scheffler correction[358] and the Becke-Johnson exchange-dipole model (XDM),[355,357] because these corrections

depend directly on the electron density. Grimme's D3 correction[274] also includes type A effects because its atomic dispersion parameters depend upon the number of bonds. Within XSAPT, type-A non-additivity is included *implicitly* via the XPol charge-embedding step, wherein the electron density of each fragment is squeezed by the neighboring charges, thereby reducing the dispersion interaction.

9.2.4 Data sets

The L7 (Ref. 19) and S12L (Ref. 20) data sets are used here to assess the performance of various methods on large organic complexes. Structures and nomenclature for these complexes are shown in Fig. 9.1. The L7 binding energies range from 2–27 kcal/mol, and were revised recently by Hansen *et al.*[21] using the domain-based local pair natural orbital (DPLNO) approximation to CCSD(T),[381–384] extrapolated to the CBS limit. These new benchmarks, which are the ones used here, come with estimated uncertainties of approximately ±1 kcal/mol.[21,353] Although Calbo *et al.*[385] have also recently revised the L7 benchmarks, differences as large as 4.9 kcal/mol are observed with respect to the Hansen benchmarks. The most like reason for these discrepancies are differences in the truncation thresholds used in the DLPNO procedure,[384] as discussed in Ref. 21.

The S12L data set consists of 12 host/guest complexes assembled from six host molecules and two guests each. Whereas six of seven L7 complexes are dispersion-dominated (see Fig. 9.1), the S12L complexes incorporate hydrogen bonding, dispersion, π-stacking, and cation-dipole binding. Benchmark gas-phase binding energies

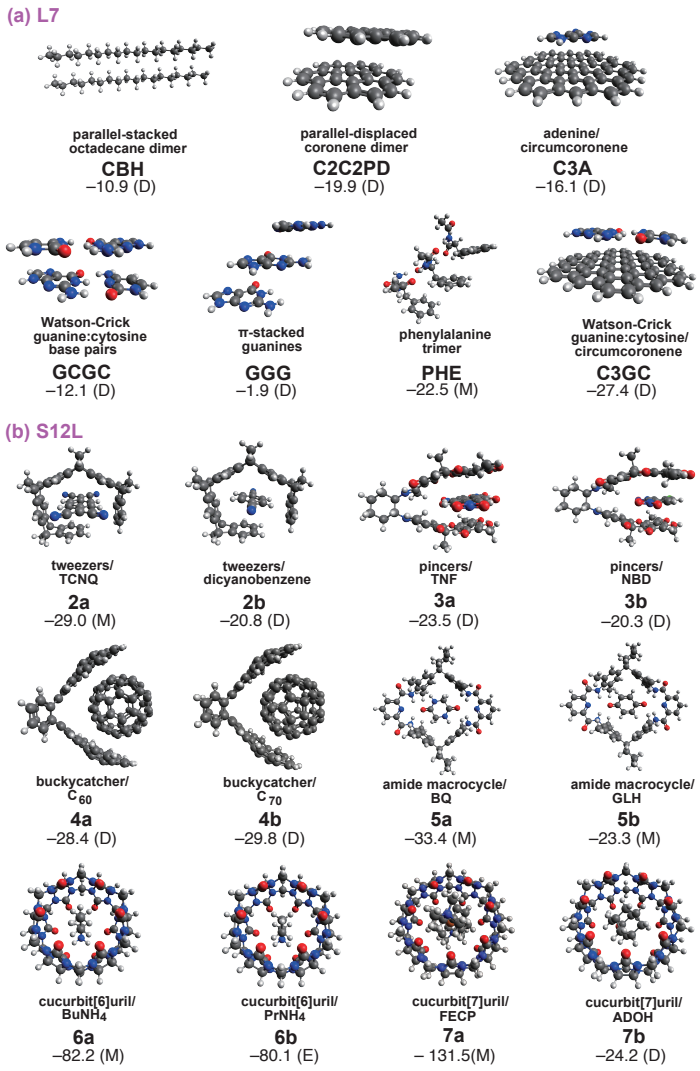

Figure 9.1: Structures of the complexes in the (a) L7 and (b) S12L data sets,[19,20] with benchmark binding energies (in kcal/mol) taken from Ref. 21 (for L7) and Ref. 22 (for S12L, except that the benchmark from **7a** is taken from Ref. 23). The coloring system is as follows: white spheres (hydrogen), gray (carbon), dark blue (nitrogen), red (oxygen), green (chlorine), and orange (iron). The designations "D", "E", and "M" indicate, respectively, cases where the intermolecular interactions are dispersion-dominated, electrostatics-dominated, or of mixed character. (Case D means $|E_{\rm elst}/E_{\rm disp}| > 2$, case E means $|E_{\rm elst}/E_{\rm elst}| < 1/2$, and case M is the intermediate regime.[7]) These classifications are based on energetics computed at the XSAPT(KS)+D3+$E_{\rm disp,3B}^{\rm ATM(TS)}(\omega_{\rm GDD})$/hpTZVPP level.

for S12L have been obtained by back-correcting experimental free energies of association,[20,23] using semi-empirical harmonic vibrational frequencies and COSMO-RS solvation energies.[386] The back-corrected binding energies range from 20–132 kcal/mol with estimated uncertainties of $\sim$ 2 kcal/mol,[20,23] which is thought mainly to arise from the solvation correction, which is 1.4 kcal/mol on average.[22] Quantum Monte Carlo (QMC) binding energies with narrow statistical error bars have been determined for six of the S12L systems (**2a**, **2b**, **4a**, **5a**, **6a**, and **7b**).[40] The MAE between these six QMC benchmarks and the latest back-corrected experimental benchmarks is 1.6 kcal/mol,[22] versus 2.4 kcal/mol when compared to older, back-corrected experimental benchmarks,[20] which provides another suggestion as to the level of accuracy that can be anticipated from the benchmarks. The newer benchmarks from Ref. 22 are used for all S12L complexes except **7a**, for which only the older benchmark from Ref. 20 is available.

9.2.5 Computational methods

The L7 and S12L data set will be examined using both XSAPT(KS)+D3 and sd-XSAPT(KS). We will also examine dispersion energies obtained from DFT-SAPT calculations.[42] The non-dispersion terms in these calculations were performed in the cc-pVTZ basis set (abbreviated TZ hereafter), whereas the dispersion terms were extrapolated to the CBS limit using a two-point cc-pVDZ $\rightarrow$ cc-pVTZ scheme (DZ,TZ), then multiplying the CBS limit by a factor of 1.08 to account for basis-set incompleteness.[24]

Regarding basis sets, this dependence was tested at the SAPT(KS)/LRC-ωPBE

Table 9.1: Mean absolute percentage errors[a] for $F^-(H_2O)$, $(H_2O)_2$, and the T-shaped and parallel-displaced isomers of $(C_6H_6)_2$.

Basis Set	MAPE (%)	
	$E^{(1)}_{\text{elst}}$	$E^{(1)}_{\text{exch}}$
dimer-centered TZ	10.0	4.0
pseudocanonical[b] hpTZVPP[c]	5.1	4.7
pseudocanonical[b] aQZVPP[d]	3.4	3.9

[a]With respect to SAPT2+(3)/aTZ benchmarks.

[b]Equivalent to the "projected" basis of Ref. 65.

[c]Basis set is def2-TZVPP augmented with diffuse functions on non-hydrogen atoms, taken from Pople basis sets.

[d]Basis is def2-QZVPP augmented with diffuse functions taken from aug-cc-pVQZ.

level using a few representative systems. Error with respect to SAPT2+(3)/aTZ benchmarks, for the energy components $E^{(1)}_{\text{elst}}$ and $E^{(1)}_{\text{exch}}$, are shown in Table 9.1 for several different basis sets. Although the pseudocanonicalized[65] quadruple-ζ basis set performs the best, the hpTZVPP basis is only slightly inferior and for reasons of cost will be used here for all XSAPT(KS)+D3 calculations. (This basis also affords good results in small-molecule XSAPT calculations.[56,57]) The def2-TZVPP basis set is used for Fe in complex **7a**.

A scaling factor of $c_{\text{disp}} = 0.657$ was used in previous sd-XSAPT(KS)/6-31G(d,2p) calculations using IP tuning, and essentially the same value ($c_{\text{disp}} = 0.661$) was optimized here in the ω_{GDD} case using the S22 data set, with essentially the same RMSD (0.37 kcal/mol versus 0.34 kcal/mol, respectively). We will use the former value for sd-XSAPT(KS) calculations based on IP tuning, and the latter value for

sd-XSAPT(KS) calculations based on ω_{GDD} tuning.

For the $\delta E_{\mathrm{int}}^{\mathrm{HF}}$ correction in Eq. (9.2), the 6-31+G(3d,3pd) basis set is used for the L7 complexes, as in previous work,[56] whereas for S12L we use the $\delta E_{\mathrm{int}}^{\mathrm{HF}}$ corrections computed in the TZ basis and reported in Ref. 42. For the two smallest systems in S12L (**2a** and **2b**), we have verified that the $\delta E_{\mathrm{int}}^{\mathrm{HF}}$ correction computed using 6-31+G(3d,3pd) differs by at most 0.1 kcal/mol from the TZ value in Ref. 42.

The LRC-ωPBE functional[227] (no short-range HF exchange) and the LRC-ωPBEh functional[227,240] (20% short-range HF exchange) are used for the XSAPT(KS)+D3 and sd-XSAPT(KS) calculations, respectively, with monomer-specific tuned ω values in either case. Atom-centered ChElPG charges,[54,319] computed on-the-fly from the monomer wave functions, are used to do the XPol embedding.[54,65] Then, the converged XPol wave functions are used to compute the pairwise SAPT corrections in a projected (pseudocanonicalized) basis set, which captures some intermolecular charge-transfer interactions.[65] The supersystem calculation needed to evaluate $E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$(int) in Eq. (9.36) was performed at the PBE/hpTZVPP level, as in previous work.[351] The D2 and D3 dispersion potentials are not parameterized for iron, nor is $E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$(int), so in complex **7a** involving ferrocene the pairwise dispersion potentials involving Fe are set to zero.

Benchmark binding energies for S12L include monomer deformation (or relaxation) energies, which are not included in XSAPT calculations. In principle we could compute them using the same density functionals that we employ for the monomers in the XSAPT calculations, however these functionals are not dispersion corrected

so the results are suspect in cases where the monomers are large. Instead, we turn to results from Ref. 42 where the S12L deformation energies were computed using a nonlocal DFT functional (NLDFT),[148] and also using MP2 and spin-component scaled (SCS) MP2. These methods deviate from one another, and from additional B97M-V/aTZ results computed here, by as much as 11 kcal/mol (see the Supporting Information). For reasons of cost, the B97M-V calculations for complexes **4a**, **4b**, **7a**, and **7b** use the "heavy augmented" haTZ basis set, which differs from aTZ by removal of the diffuse functions on hydrogen. MAEs for S12L binding energies, with respect to the back-corrected experimental benchmarks of Ref. 22, are 2.4 (NLDFT), 6.4 (SCS-MP2), and 15.9 kcal/mol (MP2). Hence we select NLDFT to compute the the deformation energies, and note that the B97M-V/aTZ deformation energies reported here differ from the NLDFT results in Ref. 42 by no more than 2.8 kcal/mol.

Finally, results for several DFT and MP2-based methods will also be shown for comparison. The ωB97X-V and B97M-V functionals are among the best-performing DFT methods for non-covalent interactions,[57,105,202] and we will report ωB97X-V/ and B97M-V/aTZ results without counterpoise correction. An Euler-Maclaurin-Lebedev quadrature grid ($N_r = 75, N_\Omega = 302$) is used to evaluate the semi-local part of these functionals and the SG-1 grid[91] is used for the VV10 nonlocal correlation part.

Closely related to MP2 is the SAPT0 method[117] and the empirically-scaled sSAPT0 method.[173] (The latter has been called the "bronze standard" of SAPT.[173]) These two methods will be tested for L7 using the jun-cc-pVDZ (jaDZ) basis set, also

known as aug-cc-pVDZ′ and which removes some diffuse functions relative to aug-cc-pVDZ.[157,158] This basis is selected because it often affords favorable error cancellation for MP2-type methods.[157,158] Lastly, the low-cost HF-3c method[387] will be tested for both L7 and S12L. HF-3c consists of a minimal-basis Hartree-Fock calculation with three empirical corrections: for dispersion, for BSSE, and for basis-set incompleteness.

All calculations except SAPT0, *s*SAPT0, and HF-3c were performed using a locally-modified version of Q-Chem.[97] [An MO implementation of XSAPT(KS)+D3 is available in Q-Chem v. 4.3 and the AO implementation reported here will be available in v. 4.4.] The SAPT0 and *s*SAPT0 calculations performed using the beta5 version of the Psi4 program,[187] and the HF-3c calculations were performed using the Orca program,[188] v. 3.0.3.

A few final notes concerning supramolecular DFT, MP2, and XSAPT calculations warrant mention. First, we set the self-consistent field (SCF) convergence criterion to $\tau_{\mathrm{SCF}} = 10^{-7}$ a.u., two orders of magnitude tighter than the Q-Chem default, because we have observed that the binding energies of the buckycatcher complexes (**4a** and **4b** in S12L) change by several kcal/mol as the threshold is tightened from the Q-Chem default of $\tau_{\mathrm{SCF}} = 10^{-5}$ a.u. Furthermore, and at great cost, we set the integral screening threshold to $\tau_{\mathrm{ints}} = 10^{-12}$ a.u. to avoid negative eigenvalues in the overlap matrix. These could be circumvented, in principle, and a looser threshold used for integral screening, if the overlap matrix were computed with a separate threshold, but that capability is not yet implemented.

ChElPG charges were computed based on Lebedev grid using a "head space"

of 3.0 Å (see Ref. 319) with radial shells spaced $\Delta r = 0.25$ Å apart for L7 and $\Delta r = 0.50$ Å for S12L. For L7, each Lebedev shell is a (110,590) grid and for S12L it is (86,434).

9.3 Results and discussion

9.3.1 L7 Data set

Wave Function Methods

MAEs, mean deviations (MDs), and maximum errors in binding energies for the L7 data set, using a variety of supersystem methods, are listed in Table 9.2. (Many of these are taken from the literature[19,177,385,388–390] but summarized here for ease of comparison; the HF-3c, ωB97X-V, and B97M-V calculations are new.) As expected, MP2/CBS is unreliable for these dispersion-bound complexes, with a MAE of 8.9 kcal/mol and a MD of −8.9 kcal/mol, indicating that this method grossly overestimates dispersion energies. This is especially true for the π-stacked parallel-displaced coronene dimer (C2C2PD in Fig. 9.1), whose binding energy is overestimated by a factor of two, an error of 19.1 kcal/mol. This problem is mitigated by including half of the MP3 correlation energy: the MP2.5/CBS method reduces the MAE to 1.8 kcal/mol with a maximum error of 3.0 kcal/mol, albeit at $\mathcal{O}(N^6)$ cost.

Alternatively, spin-component scaling tends to improve the performance of MP2 for non-covalent interactions. In the original SCS-MP2 method,[263] the same- and opposite-spin scaling parameters were optimized using high-quality reaction energies, although they have also been optimized for molecular interaction (MI) using S22

Table 9.2: Mean absolute errors, mean deviations, and maximum absolute errors in L7 binding energies,[a] computed using various supersystem methods.

Method	error (kcal/mol)		
	MAE	MD	max
—Wave Function Methods—			
MP2/CBS[b]	8.94	−8.94	19.08
MP2.5/CBS[b]	1.77	−1.76	3.00
SCS-MP2/CBS[b]	2.80	−1.79	7.63
SCS(MI)-MP2/CBS[b]	5.19	−4.83	11.81
MP2C/CBS[b]	0.99	−0.99	2.32
att-MP2/aTZ[c]	1.71	−0.99	2.53
att-SCS-MP2/aTZ[c]	1.32	0.13	2.96
HF-3c	1.35	−1.10	2.37
HF-3c+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$	1.02	−0.02	1.78
—DFT Methods—			
PBE-TS[d,e,f]	3.80	−3.80	5.55
PBE-TS/HI[d,e,g]	3.27	−3.27	7.43
TPSS+D3BJ/def2-QZVP[b]	1.08	−1.07	1.73
B3LYP+D3BJ/def2-QZVP[b]	2.61	−2.61	3.89
M06-2X+D30/def2-QZVP[b]	1.51	−0.65	3.13
M06-2X/def2-QZVP[b]	2.69	2.44	6.19
PW6B95+D3BJ/def2-QZVP[b]	0.84	−0.53	2.31
PW6B95+D3BJ+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$/ def2-QZVP[b]	0.90	0.56	1.69
B2PLYP+D3BJ/def2-TZVP[h,i]	0.98	−0.28	2.15
B2PLYP+NL/def2-TZVP[h,i]	1.72	−1.42	2.86
PBE-XDM/pc-2-spd[j]	2.33	1.34	5.49
LC-ωPBE-XDM/pc-2-spd[j]	1.42	−1.41	2.57
ωB97X-D/aTZ (CP)[k]	2.49	−2.49	3.44
ωB97X-V/aTZ	2.43	−2.43	3.76
B97M-V/aTZ	2.30	−2.30	3.65
B97M-V/aTZ (CP)	1.42	−1.42	2.71

[a] With respect to CCSD(T)/CBS benchmarks.[21,353]

[b] From Ref. 19.

[c] From Ref. 177.

[d] From Ref. 388.

[e] Using a plane-wave basis set.

[f] With TS dispersion (Hirshfeld).

[g] With TS dispersion (iterative Hirshfeld).

[h] From Ref. 385

[i] With $E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$.

[j] From Ref. 390.

[k] From Ref. 389.

interaction energies, in a method called SCS(MI)-MP2.[?] For the L7 data set, SCS-MP2 actually outperforms SCS(MI)-MP2, with an MAE of 2.8 kcal/mol for the former and 5.2 kcal/mol for the latter. This suggests that SCS-MP2 is more suitable for general applications as compared to SCS(MI)-MP2, or perhaps that the SCS(MI)-MP2 scaling parameters should be re-optimized using larger molecules. Nevertheless, the SCS-MP2/CBS method still overestimates the binding energy of C2C2PD by 7.63 kcal/mol.

MP2's deficiency for dispersion-bound complexes is the result of its uncoupled HF treatment of dispersion. In the MP2C method,[309,310] the dispersion component of an MP2 calculation is replaced by a more accurate dispersion energy calculated using DFT response theory. [The cost of evaluating the dispersion energy scales as $\mathcal{O}(N^4)$ if resolution-of-identity techniques are used,[310] although the overall cost remains $\mathcal{O}(N^5)$.] MP2C exhibits a MAE of 1.0 kcal/mol for L7, with a maximum error of 2.3 kcal/mol.

We have focused on CBS results in this discussion owing to the slow convergence of MP2 interaction energies with respect to the one-particle basis set. This convergence problem is ameliorated by the attenuated MP2 (att-MP2) method,[176] which attenuates (at long range) the Coulomb operator used in the correlation energy calculation. This method not only improves the strongly overbound π-stacking energies, but also removes significant BSSE so that the method is accurate in a finite basis set. Promising results are obtained using the aTZ basis set,[17] and the att-MP2/aTZ

method affords a MAE of 1.7 kcal/mol for L7, comparable to MP2.5/CBS but without the $\mathcal{O}(N^6)$ terms or the need for quadruple-ζ results for CBS extrapolation. The SCS concept can be incorporated into attenuated MP2 (att-SCS-MP2),[177] and for L7 this method affords a MAE of 1.3 kcal/mol with a maximum error of 3.0 kcal/mol.

At the other end of the spectrum of computational cost, we find that the minimal-basis HF-3c method affords a MAE of only 1.4 kcal/mol, which is further reduced to just 1.0 kcal/mol by adding the $E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$ three-body dispersion correction. (The dispersion correction included in HF-3c is Grimme's D3 correction,[387] hence the three-body correction is anticipated to be necessary in large complexes, even given the parameterized nature of HF-3c.) The mean deviation for HF-3c+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$ is practically zero, indicating no significant bias. Thus, HF-3c seems like a promising method for screening of large structures, *e.g.*, in computational drug design.

In summary, the best MP2-based methods for the L7 data set are MP2C/CBS and att-SCS-MP2/aTZ, and obviously the latter is far more efficient. In fact, the favorable performance of the MP2C/CBS approach may be a coincidence, since the calculations reported in Ref. 19 used only double- and triple-ζ basis sets for the CBS extrapolation, and performed the DFT response calculations (to evaluate the coupled Kohn-Sham dispersion correction) in the small 6-31G*(0.25) basis set. Especially promising is the performance of the low-cost HF-3c+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$ method, which is comparable to that of the best wave function methods discussed above, with an MAE of only 1.0 kcal/mol.

Table 9.3: Atomic three-body dispersion energies (in kcal/mol) for L7 complexes, using two versions of the Axilrod-Teller-Muto triple-dipole scheme.

System	$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$	$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$
CBH	2.53	0.73
GGG	1.09	0.26
C3A	5.08	1.23
C3GC	9.04	2.31
C2C2PD	7.72	1.72
GCGC	3.73	1.02
PHE	1.16	0.39

DFT Methods

Traditional semi-local DFT does not describe the nonlocal dispersion interaction, but the pairwise Tkatchenko-Scheffler (TS) dispersion model[358] can be directly combined with semi-local functionals to capture dispersion interactions. For example, the PBE functional predicts that five of the L7 dimers are unbound (with the octadecane dimer and phenylalanine trimer being the exceptions), but the PBE-TS binds these species and affords a MAE of 3.8 kcal/mol for the binding energies. The mean deviation (-3.8 kcal/mol) demonstrates that the TS model overcorrects in these cases. The MAE is reduced, but only to 3.3 kcal/mol, using a TS dispersion model based on the iterative Hirshfeld partitioning scheme (PBE-TS/HI in Table 9.2).[391] This suggests that a dispersion correction beyond the pairwise TS approach is needed to improve the description of intermolecular interactions in large supramolecular complexes.

The $E_{\mathrm{disp,3B}}^{\mathrm{ATM}}$ correction accounts for atomic three-body dispersion effects, and the size of this term for each of the L7 dimers is listed in Table 9.3. The larger π-stacked

systems (C3A, C2C2PD, and C3GC) afford large, repulsive values of $E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$, of up to 9 kcal/mol for C3GC, whereas the $E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$ corrections are considerably smaller. We will show below that $E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$ is more suitable to combine with our pairwise D3 potential to reproduce dispersion energies calculated using DFT response theory. Adding the $E_{\mathrm{disp,3B}}^{\mathrm{ATM}}$ correction to PBE-TS or PBE-TS/HI reduces the MAEs for L7 to 2.2 and 3.1 kcal/mol, respectively (see Table 9.2).

Among DFT methods that incorporate Grimme's D3 dispersion correction,[274] the PW6B95 hybrid meta-GGA functional[392] performs the best, with a MAE of 0.8 kcal/mol (Table 9.2). This is consistent with the good performance of PW6B95 + D3BJ for the S66 data set.[11] [Here, "D3BJ" indicates use of the Becke-Johnson (BJ) damping function for the D3 dispersion correction.[14]] PW6B95+D3BJ has also been recommended for use in calculations involving water clusters,[393] biomolecules,[394] transition metal catalysts,[395] large host/guest complexes,[20,22,396] and geometry optimizations.[397] Addition of the $E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$ term to PW6B95+D3BJ slightly increases the mean errors for L7 but decreases the maximum error. We conclude that PW6B95+D3BJ should be reasonably accurate for a wide range of applications.

With the exception of the phenylalanine trimer, all of the L7 systems are unbound at the B3LYP level,[389] but Grimme's D3 overcorrects and the B3LYP+D3BJ method overestimates the L7 binding energies with a MAE of 2.6 kcal/mol. A non-empirical meta-GGA functional, TPSS, affords accurate results when combined with Grimme's D3 (MAE = 1.1 kcal/mol). The highly parametrized M06-2X functional only successfully captures middle-range (but not long-range) non-local correlation, and affords a

MAE of 2.7 kcal/mol. In these large complexes, the long-range correlations are important and long-range correlation must be added to M06-2X via an empirical dispersion correction. The M06-2X+D30 method (where "0" indicates the "zero damping" function[274]) gives reasonable results for L7, with a MAE of 1.5 kcal/mol. The Becke-Johnson XDM dispersion correction, with two different density functionals, affords MAEs of 1.4 and 2.3 kcal/mol that are comparable to—but not significantly better than—the performance of the best DFT-D3 methods.

The ωB97X-D/aTZ[201] with counterpoise correction overestimates the binding energies of the L7 complexes with a MAE of 2.5 kcal/mol.[389] Considering a more modern collection of non-local dispersion-corrected functionals, B97M-V[202] and ωB97X-V[105] functionals combined with aTZ basis set each overestimate the binding energies of the L7 complexes, with MAEs of 2.3, 2.4 kcal/mol, respectively, even though B97M-V and ωB97X-V afford quite accurate results for many other non-covalent systems.[57,105,202] The overestimation of non-local dispersion-corrected functional in large complexes partly comes from the BSSE. By using the counterpoise correction in B97M-V/aTZ calculations, its MAE reduces to 1.4 kcal/mol. For the C2C2PD system, B97M-V with and without counterpoise correction yield binding energies of -21.5 and -22.3 kcal/mol, as compared to the benchmark value of -20.0 kcal/mol. The BSSE in C2C2PD system contributes 0.8 kcal/mol error in B97M-V calculations. Given that these functionals accurately reproduce the binding energy of the π-stacked benzene dimer[105,202] (which is, admittedly, about ten times smaller than the binding energy of C2C2PD), the coronene dimer appears to be a challenging test

for non-covalent interactions with non-local vdW functionals. Note also that for the largest of the L7 complexes (C3GC), which contains ≈4,000 aTZ basis functions, the B97M-V calculation is about six times faster than the ωB97X-V calculation, due to the absence of HF exchange.

The VV10 non-local correlation functional (included in ωB97X-V and B97M-V) is also rather expensive to evaluate for large systems, and less-expensive alternatives are of interest. Among the DFT methods tested here, the PW6B95+D3BJ approach is unique in the fact that it is able to reproduce the benchmark binding energies for both benzene dimer and C2C2PD. Moreover, according to Hujo and Grimme,[398,399] self-consistent evaluation of the non-local part of the functional is rarely necessary, and can accurately be approximated in the final step using the converged electron density from the semi-local part of the functional. Such a strategy has been used in the context of double-hybrid density functionals, replacing the D3BJ dispersion potential in B2PLYP+D3BJ+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$ by non-self-consistent VV10. However, this replacement slightly increases the MAE, from 1.0 kcal/mol to 1.7 kcal/mol.[385]

(X)SAPT Methods

Next, we turn from supersystem methods to SAPT-based methods, for which L7 error statistics are listed in Table 9.4. SAPT and XSAPT afford similar results (which, for dimers, is by design[65]), so we mainly focus XSAPT results in the following discussion.

We first note the importance of dispersion in these systems, as evidenced by the poor performance of SAPT0 and its scaled-dispersion counterpart, *s*SAPT0, which

Table 9.4: Mean absolute errors, mean deviations, and maximum absolute errors in L7 binding energies,[a] computing using (X)SAPT-based approaches.[b,c]

Method	error (kcal/mol)		
	MAE	MD	max
—SAPT-Based Methods—			
SAPT0/jun-cc-pVDZ	5.47	−4.51	11.68
*s*SAPT0/jun-cc-pVDZ	5.47	−4.51	11.74
—XSAPT-Based Methods—			
XSAPT(KS)+D3(ω_{IP})	3.18	−2.77	6.16
XSAPT(KS)+D3(ω_{GDD})	3.50	−3.11	6.49
sd-XSAPT(KS)(ω_{IP})	2.47	−1.84	4.17
sd-XSAPT(KS)(ω_{GDD})	1.96	−1.23	2.73
XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$($\omega_{\mathrm{IP}}$)	1.57	1.57	3.99
XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$($\omega_{\mathrm{GDD}}$)	1.38	1.23	3.88
sd-XSAPT(KS)+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$($\omega_{\mathrm{IP}}$)	2.73	2.50	5.27
sd-XSAPT(KS)+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$($\omega_{\mathrm{GDD}}$)	3.33	3.10	6.68
XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$($\omega_{\mathrm{IP}}$)	2.30	−1.67	4.43
XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$($\omega_{\mathrm{GDD}}$)	2.61	−2.01	4.76
sd-XSAPT(KS)+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$($\omega_{\mathrm{IP}}$)	1.66	−0.75	2.51
sd-XSAPT(KS)+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$($\omega_{\mathrm{GDD}}$)	1.15	−0.14	2.72

[a] With respect to CCSD(T)/CBS benchmarks.[21,353]

[b] All complexes treated as dimers.

[c] The jun-cc-pVDZ basis set is used for (*s*)SAPT0 calculations, hpTZVPP for (X)SAPT(KS)+D2/D3, and 6-31G(d,2p) for sd-XSAPT(KS).

exhibit errors as large as 12 kcal/mol. Even the MAEs (5.5 kcal/mol for both methods) are larger than those documented in previous SAPT0/jaDZ benchmarks for smaller molecules,[173] despite the favorable error cancellation in MP2-type dispersion that comes from using the jaDZ basis.[157,158]

Moving on to the XSAPT results, we note that both XSAPT(KS)+D3 and sd-XSAPT(KS) binding energies are improved, in nearly all cases, by including a three-body dispersion term, either $E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$ or $E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$. The TS version of this three-body correction works better with XSAPT(KS)+D3 while the Grimme version performs better with sd-XSAPT(KS). As shown below for S12L, by comparison to dispersion energies computed using DFT response theory, the TS version of the three-body ATM correction affords superior results (as compared to $E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$) when combined with our D3 two-body dispersion potential. As such, the better performance of Grimme's three-body correction in the context of sd-XSAPT(KS) may arise from error cancellation.

Examining the best of these three-body, dispersion-corrected XSAPT methods—namely XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$ and sd-XSAPT(KS)+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$—we find that better results are obtained using ω_{GDD} rather than ω_{IP} to tune the range-separation parameter. This observation is consistent with the expectation that ω_{GDD} should provide a better description of SAPT energy components, an expectation that is in turn based on the observation that LRC-ωPBE(ω_{GDD}) affords a better description of polarizabilities as compared to the corresponding ω_{IP} approach.[378] The largest difference in tuned ω values (ω_{GDD} versus ω_{IP}) for XSAPT(KS)+D3 occur for the largest

L7 monomer, circumcoronene, which is interesting in light of the documented dependence of ω_{IP} on system size.[379,380] For circumcoronene, the difference between ω_{IP} and ω_{GDD} is 0.071 a_0^{-1}.

The best XSAPT methods in Table 9.4 are XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}(\omega_{\mathrm{GDD}})$ and sd-XSAPT(KS)+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}(\omega_{\mathrm{GDD}})$, with MAEs of 1.4 and 1.2 kcal/mol, respectively. These two methods are comparable to the best supersystem methods in Table 9.2. Although we also applied XSAPT(KS)+D2 to the L7 data set, the resulting MAEs were found to be at least 0.2 kcal/mol larger than the corresponding D3 results, consistent with our previous observation that XSAPT(KS)+D3 is generally superior to +D2 for both total binding energies and dispersion energies.[57] In Fig. 9.2, we summarize the performance of various electronic structure methods for L7.

The ratios $-E_{\mathrm{comp}}/E_{\mathrm{int}}$, where E_{comp} is an energy component obtained from XSAPT energy decomposition analysis and E_{int} is the total XSAPT interaction energy, can be used to categorize the nature of the association interaction.[56] These ratios, computed at the XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}(\omega_{\mathrm{GDD}})$ level, are shown in Fig. 9.3 and the individual energy components can be found in the Supporting Information. According to this analysis, all of the L7 complexes are dispersion-dominated except for the hydrogen-bonded phenylalanine trimer (PHE), where the electrostatic interaction is slightly larger than the dispersion interaction. This is consistent with Hobza's classification[7] (used in Fig. 9.1) that classifies PHE as having "mixed" interactions (dispersion and electrostatics of comparable importance).

In the (X)SAPT calculations discussed thus far, each system is treated as a dimer

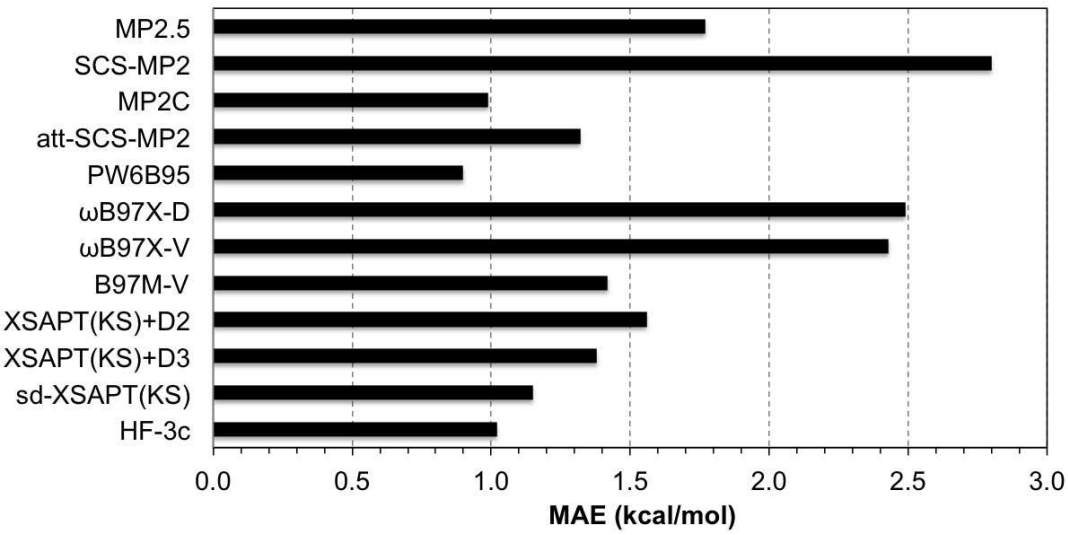

Figure 9.2: Mean absolute error (MAE), with respect to CCSD(T)/CBS benchmarks, for L7 binding energies. The MP2.5, SCS-MP2, and MP2C energies are extrapolated to the basis set limit whereas the att-SCS-MP2 and ωB97X-V calculations use the aTZ basis set without counterpoise correction. The ωB97X-D and B97M-V energies using the aTZ basis set are corrected with counterpoise correction. The D3BJ and $E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$ dispersion corrections are applied to PW6B95/def2-QZVP calculations. The $E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$ correction is added to the XSAPT(KS)+D2/D3 calculations, which employ the hpTZVPP basis set; for the HF-3c calculations the $E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$ correction is also included but is evaluated using a minimal basis set. The $E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$ is used in the sd-XSAPT(KS) calculations, along with the 6-31G(d,2p) basis set. The ω_{GDD} tuning procedure is used in all XSAPT(KS) variants.

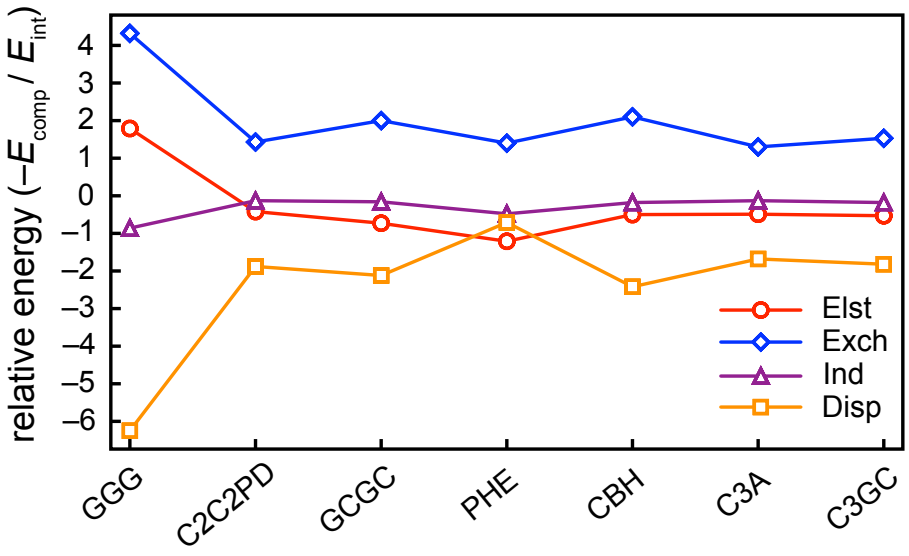

Figure 9.3: Ratio of each energy contribution relative to the total interaction energy, computed at the XSAPT(KS)+D3+$E^{\mathrm{ATM(TS)}}_{\mathrm{disp,3B}}(\omega_{\mathrm{GDD}})$/hpTZVPP level, for the L7 data set.

Table 9.5: Binding energies (in kcal/mol) for the trimers and tetramers from L7, computed at the XSAPT(KS)+D3+$E^{\mathrm{ATM(TS)}}_{\mathrm{disp,3B}}(\omega_{\mathrm{GDD}})$/hpTZVPP level.

System	Two-body[a]	Many-body
GGG	−1.51	−1.59
C3GC	−24.51	−25.87
PHE	−22.79	−22.78
GCGC	−12.33	−13.21

[a] Treated as dimers, as in Table 9.4.

even though the GGG, C3GC, and PHE complexes contain three monomers and GCGC contains four. In these calculations, we have used the π-stacked guanine dimer as one monomer in GGG, the hydrogen-bonded GC base pair as one monomer in GCGC and in G3GC (where it interacts with circumcoronene, in the latter example), and phenylalanine dimer as one monomer in PHE. Using XSAPT, however, we can treat each fragment as a monomer, and in Table 9.5 the best XSAPT method from Table 9.4, which is XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}(\omega_{\mathrm{GDD}})$, is used to compute the binding energies of the four complexes that consist of more than one monomer, using a many-body treatment.

For the GGG and PHE complexes, the binding energies are essentially the same in either treatment. The monomers in these two systems are arranged into layers (with each layer being a monomer), and the interaction between the second and third layer is only slightly disturbed by the presence of the first layer. For GCGC, the difference between the two-body and the many-body approach is 0.9 kcal/mol and for C3GC it is 1.4 kcal/mol, differences that may stem from the assumption that the $\delta E_{\mathrm{int}}^{\mathrm{HF}}$ correction is pairwise-additive in many-body systems, and the related fact that the $\delta E_{\mathrm{int}}^{\mathrm{HF}}$ correction used in XSAPT is obtained from two-body SAPT without electrostatic embedding using a dimer-centered basis.[56] Furthermore, the infinite-order response correction for induction is assumed to be implicitly include in the XPol step with negligible double-counting of higher-order corrections for induction.[56] These assumptions appear to be robust, since XSAPT affords very good binding energies for

systems including water clusters and halide–water clusters,[57] and moreover the many-body XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}(\omega_{\mathrm{GDD}})$ treatment slightly reduces the MAE for L7 as compared to the strictly pairwise treatment. Hence, the XSAPT can be used as an accurate and efficient quantum-mechanical method for non-covalent interactions in many-body systems.

9.3.2 S12L Data set

Error statistics for S12L, using a variety of supersystem and (X)SAPT-based methods, are compiled in Table 9.6 and will be discussed below. Note that the mean deviations in Table 9.6 are uniformly negative, indicating that these methods consistently overbind the S12L dimers with respect to both QMC and back-corrected experimental benchmarks. The latter benchmarks are consistently larger than the QMC ones, by an average of 1.6 kcal/mol, which explains why the errors in Table 9.6 are slightly larger when the comparison is to QMC benchmarks.

Although ∼1 kcal/mol is often used as the standard for "chemical accuracy", this is too stringent a standard for binding energies of large supramolecular complexes, where deviations of 2–3 kcal/mol are to be expected.[400] Estimated uncertainties for back-corrected experimental binding energies in S12L are 2 kcal/mol,[20,23] while the average statistical error in the QMC benchmarks is 1.2 kcal/mol.[40] In view of errors of this magnitude, Heßelmann's NLDFT approach,[148] the PW6B95+D3BJ+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$ method, DFT-SAPT, MP2C, and XSAPT(KS)+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}(\omega_{\mathrm{GDD}})$ with dispersion calculated by D3 or DFT response theory [disp(CKS)] can be considered as acceptable

Table 9.6: Mean absolute errors, mean deviations, and maximum absolute errors for S12L binding energies with respect to back-corrected experimental results[20,22] and QMC benchmarks.[40]

Method	error (kcal/mol)					
	Experiment			QMC		
	MAE	MD	max	MAE	MD	max
—Supersystem Methods—						
MP2/CBS[a]	15.90	−15.90	53.28	17.87	−17.87	52.98
SCS-MP2/CBS[a]	6.36	−2.96	25.23	7.19	−5.05	27.65
MP2C/CBS [(DZ,TZ) extrapolation][a,b]	1.78	−0.98	6.87	3.21	−3.21	9.47
MP2C/CBS [1.08×(DZ,TZ)][a,b]	4.36	−4.36	12.61	6.40	−6.40	15.21
NLDFT/def2-QZVP[a]	2.35	−1.08	5.48	3.05	−2.93	8.08
PW6B95+D3BJ/def2-QZVP(-g,-f)[c,d]	2.75	−2.35	6.60	4.57	−4.40	9.20
PW6B95+D3BJ+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$/def2-QZVP(-g,-f)[c,d]	1.48	−0.18	3.40	2.78	−2.28	6.00
B2PLYP+D3BJ+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$/def2-TZVP[e]	3.62	−0.92	12.33	4.00	−3.89	14.78
B2PLYP+NL+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$/def2-TZVP[e]	5.22	−5.22	17.91	6.74	−6.74	20.00
PBE-XDM/pc-2-spd[f]	2.28	2.00	4.40	1.22	−0.05	2.70
LC-ωPBE-XDM/pc-2-spd[f]	5.61	−5.61	11.70	7.65	−7.65	14.20
ωB97X-D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$/def2-QZVP(-g,-f)[g]	2.16	−1.29	5.04	4.08	−4.08	5.82
B97M-V/aTZ	6.07	−6.07	12.78	7.51	−7.51	14.74
B97M-V/aTZ (CP)	4.20	−4.17	9.58	5.79	−5.79	11.80
HF-3c	5.39	−3.92	11.38	6.51	−4.90	11.57
HF-3c+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$	3.86	−1.74	8.01	4.73	−2.78	9.21
—(X)SAPT-Based Methods—						
DFT-SAPT [(DZ,TZ) extrapolation][a,b,h,i]	2.86	2.13	6.29	1.99	−0.43	4.89
DFT-SAPT [1.08×(DZ,TZ)][a,b,h,i]	2.10	−1.56	7.64	3.87	−3.87	10.13
XSAPT(KS)+D2(ω_{IP})[b,j]	9.63	−9.623	18.50	12.19	−12.19	21.10
XSAPT(KS)+D2(ω_{GDD})[b,j]	8.37	−8.37	18.72	10.70	−10.70	21.10
XSAPT(KS)+D2+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$($\omega_{\mathrm{IP}}$)[b,j]	2.71	−1.60	8.99	4.59	−4.59	8.99
XSAPT(KS)+D2+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$($\omega_{\mathrm{GDD}}$)[b,j]	1.73	−0.76	5.32	3.13	−3.13	7.92
XSAPT(KS)+D3(ω_{IP})[b,j]	10.92	−10.92	21.56	14.02	−14.02	24.16
XSAPT(KS)+D3(ω_{GDD})[b,j]	9.67	−9.67	21.86	12.57	−12.57	24.38
XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$($\omega_{\mathrm{IP}}$)[b,j]	4.80	−2.89	8.16	6.42	−6.42	10.76
XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$($\omega_{\mathrm{GDD}}$)[b,j]	3.33	−1.64	8.39	4.97	−4.97	10.99
XSAPT(KS)+disp[CKS,(DZ,TZ)](ω_{IP})[b,h,j]	2.24	0.52	7.08	2.50	−2.45	7.08
XSAPT(KS)+disp[CKS,(DZ,TZ)](ω_{GDD})[b,h,j]	2.26	1.77	5.53	1.74	−0.99	3.64
XSAPT(KS)+disp[CKS,1.08×(DZ,TZ)](ω_{IP})[b,h,j]	3.65	−3.17	10.34	5.89	−5.89	10.34
XSAPT(KS)+disp[CKS,1.08×(DZ,TZ)](ω_{GDD})[b,h,j]	2.27	−1.91	6.29	4.43	−4.43	8.89
sd-XSAPT(KS)(ω_{GDD})[b,k]	6.28	−6.04	25.61	7.40	−7.40	26.34
sd-XSAPT(KS)+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$($\omega_{\mathrm{GDD}}$)[b,k]	5.38	−3.87	22.11	6.23	−5.29	23.14
sd-XSAPT(KS)+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$($\omega_{\mathrm{GDD}}$)[b,k]	5.56	1.98	11.13	4.51	0.20	12.94

[a] From Ref. 42.
[b] Using the NLDFT deformation energies from Ref. 42.
[c] The f functions on hydrogen and g functions on other atoms are removed.
[d] From Ref. 20.
[e] From Ref. 385.
[f] From Ref. 390.
[g] From Ref. 22.
[h] DFT-SAPT dispersion energies ["disp(CKS)"] were extrapolated to the CBS limit using a (DZ,TZ) scheme.[42]
[i] Using cc-pVTZ for the non-dispersion terms.
[j] Using the hpTZVPP basis set.
[k] Using the 6-31G(d,2p) basis set.

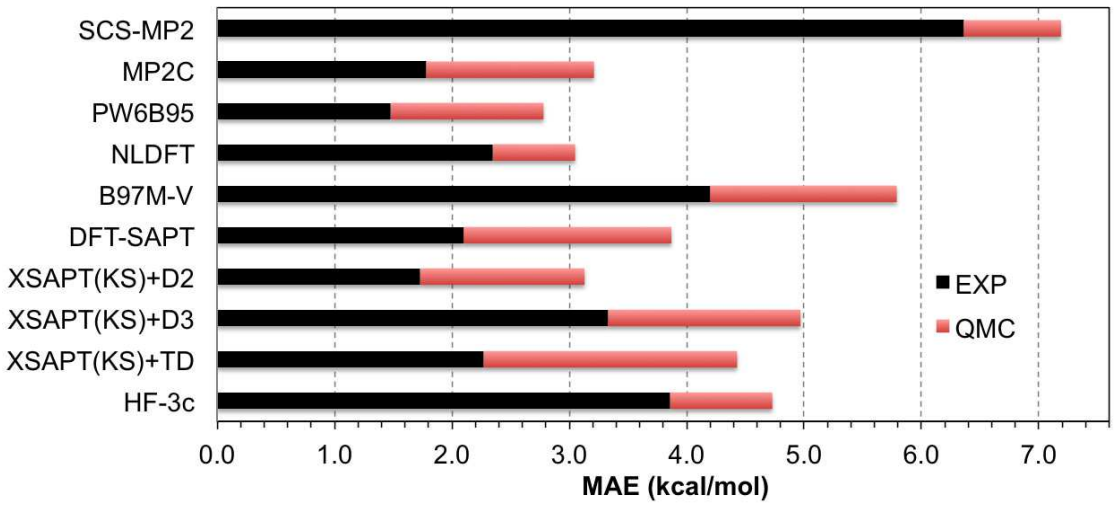

Figure 9.4: Mean absolute errors (MAE) in S12L binding energies, with respect to QMC and back-corrected experimental benchmarks. The SCS-MP2 and MP2C energies have been extrapolated to the CBS limit whereas the NLDFT and B97M-V calculations used the def2-QZVP and aTZ basis sets, respectively, except that the B97M-V calculations use the haTZ basis set for **4a**, **4b**, **7a**, and **7b**. The counterpoise correction is used to remove the BSSE in B97M-V calcualtions. The PB6B95 calculations use the def2-QZVP(-g-f) basis, which omits g and f functions, and include the $E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$ correction. This correction is included in the HF-3c calculations as well, but is evaluated using a minimal basis set, consistent with HF-3c itself. XSAPT(KS) calculations use the hpTZVPP basis set and the ω_{GDD} tuning procedure, and include the $E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}(\omega_{\mathrm{GDD}})$ correction. Here, "+TD" means dispersion energies are calculated by DFT-SAPT at the CBS limit using the 1.08×(DZ,TZ) extrapolation scheme of Ref. 24.

methods for very large supramolecular calculations, as summarized in Fig. 9.4. A detailed analysis can be found below.

Atomic-Pairwise Dispersion Potentials

To address the validity of atomic-pairwise dispersion potentials in the much larger S12L complexes, we compare the D2 and D3 dispersion energies to those computed using DFT response theory, with results shown in Table 9.7. In general, both the TS and Grimme variants of the three-body dispersion terms greatly reduce the errors in the pairwise dispersion potentials, though the $E_{\text{disp,3B}}^{\text{ATM(TS)}}$ correction performs better in both the D2 and the D3 case. Surprisingly, D2-based potentials give smaller errors than the corresponding D3-based potentials, in contrast to previous results for S22 and S66 and for stacking complexes of nucleobase tetramers.[57] [In Ref. 57, the atomic-pairwise dispersion potentials were benchmarked against SAPT2+(3) and DFT-SAPT results.] Two possible explanations come to mind. First, the D2 dispersion potential is known to slightly overestimate the dispersion energy, especially for π-stacked systems, as compared to dispersion benchmarks,[57] whereas the D3 potential resolves this overestimation. Hence, the repulsive $E_{\text{disp,3B}}^{\text{ATM(TS)}}$ term is more suitable to combine with D2 than D3. Alternatively, the DFT-SAPT dispersion benchmarks in Table 9.7, which were taken from Ref. 42, may not be accurate enough.

According to Ref. 42, two factors influence the quality of DFT-SAPT dispersion benchmarks. First, the coupled value of $E_{\text{exch-disp}}^{(2)}$ is *not* obtained from DFT response theory but rather via empirical scaling of the uncoupled value. A scaling factor of 0.686 was obtained by fitting $E_{\text{exch-disp}}^{(2)}$ for molecules in the S22 data set,

Table 9.7: Mean and maximum absolute errors for the S12L data set,[a] comparing D2 and D3 dispersion potentials.

Dispersion potentials	error (kcal/mol)	
	MAE	max
—D2 Potential—		
D2	6.46	12.43
D2+$E_{\text{disp,3B}}^{\text{ATM(TS)}}$	1.66	3.87
D2+$E_{\text{disp,3B}}^{\text{ATM(TS)}}$(param)	0.75	2.36
D2+$E_{\text{disp,3B}}^{\text{ATM(Grimme)}}$	4.39	9.23
—D3 Potentials—		
D3	7.75	15.69
D3+$E_{\text{disp,3B}}^{\text{ATM(TS)}}$	2.18	4.81
D3+$E_{\text{disp,3B}}^{\text{ATM(TS)}}$(param)	2.97	5.02
D3+$E_{\text{disp,3B}}^{\text{ATM(Grimme)}}$	5.66	12.30

[a] With respect to DFT-SAPT/CBS dispersion energies extrapolated using the 1.08 × (DZ,TZ) scheme of Ref. 42.

with $R^2 = 0.9993$ for the fit. It is possible, however, that the relationship between the coupled and uncoupled values of $E^{(2)}_{\text{exch-disp}}$ is different in large molecules where dispersion plays a more prominent role. Second, DFT-SAPT dispersion energies were extrapolated to CBS limit using only double- and triple-ζ basis sets, then multiplied by an empirical factor of 1.08 (again, obtained from S22 benchmarks[24]) to account for residual basis-set incompleteness. Thus, the superior performance of D2 in comparison to the DFT-SAPT dispersion energies for the S12L data set may come from error cancellation. The $E^{\text{ATM(TS)}}_{\text{disp,3B}}$(param) correction, combined with D3 dispersion, slightly degrades the dispersion energies as compared to D3+$E^{\text{ATM(TS)}}_{\text{disp,3B}}$. On the other hand, D2+$E^{\text{ATM(TS)}}_{\text{disp,3B}}$(param) is superior to D2+$E^{\text{ATM(TS)}}_{\text{disp,3B}}$, which may indicate error cancellation in the D2 results. The performance of both D2 and D3 combined with XSAPT(KS) will be further discussed below.

Supersystem Approaches

Let us now discuss the errors for the supersystem methods, which are listed in Table 9.6 and summarized in Fig. 9.4. The MP2/CBS method significantly overestimates the binding energies, as expected, with errors as large as 53 kcal/mol, whereas SCS-MP2/CBS reduces the MAE dramatically, to 6–7 kcal/mol depending on which set of benchmarks is used. MP2C/CBS can achieve a MAE as low as 2–3 kcal/mol, using a (DZ,TZ) extrapolation.[42] However, application of the empirical scaling factor of 1.08 that is suggested in Ref. 42 to account for basis-set incompleteness considerably degrades the quality of the results, increasing the MAEs to 4–6 kcal/mol depending on the choice of benchmarks. Note that the S12L benchmarks have been revised since

the study in Ref. 42 that suggested this scaling factor, and results obtained with the new benchmarks thus call this factor into question. Simply using the (DZ,TZ) extrapolation, with the small 6-31G*(0.25) basis set for the coupled Kohn-Sham response equations, affords a MAE of 1.0 kcal/mol for L7 and 2–3 kcal/mol for S12L.

Heßelmann's non-local functional[148] ("NLDFT") and the PW6B95+D3BJ approach both perform well for S12L, with MAEs of 2–3 kcal/mol in the former case and no larger than 4.6 kcal/mol in the latter case, again depending on whether the QMC or the back-corrected experimental benchmarks are selected. Addition of the three-body $E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$ correction to PW6B95+D3BJ reduces the MAE by more than 1 kcal/mol, and PW6B95+D3BJ+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$ appears to be a promising DFT approach for use in large complexes. The B2PLYP+NL+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$ method, introduced quite recently,[385] overestimates the S12L binding energies by 5–7 kcal/mol.

Other non-local density functionals perform less well. B97M-V/aTZ with counterpoise correction calculations overbind most of the S12L complexes, with MAEs of 4.2–5.8 kcal/mol depending on which set of benchmarks is used for comparison, and either value is larger than the MAE observed when B97M-V/aTZ is applied to the L7 data set. The maximum deviation is for complex **4b**, and is 9.6 kcal/mol respect to back-corrected experimental benchmark; this maximum error is also larger than the maximum error for L7. It is important to correct BSSE in these large complexes, and the maximum counterpoise correction in S12L complexes comes from complex **4b** with 3.2 kcal/mol BSSE. Without counterpoise correction, the MAE of B97M-V/aTZ calculations increases to 6.1–7.5 kcal/mol depending on which set of benchmarks is used

for comparison, and the maximum deviation increases to 12.8 kcal/mol for complex **4b**. The non-local ωB97X-V functional was similar in its performance to B97M-V for the L7 data set, but is considerably more expensive due to the inclusion of Hartree-Fock exchange and as such we only consider B97M-V calculations in S12L complexes. We note, however, that ωB97X-D3 with a quadruple-ζ basis set affords a MAE of 2.2 kcal/mol for S12L,[22] much smaller (for the same set of benchmarks) than the 4.2 kcal/mol MAE of B97M-V/aTZ.

Finally, considering XDM-based methods we find that the favorable performance of the LC-ωPBE-XDM for the L7 data set does not carry over to S12L, and in the latter case this approach exhibits a MAE of 5.6–7.6 kcal/mol. PBE-XDM performs reasonable well in both L7 and S12L data sets with MAEs of $\sim$ 2 kcal/mol in both data sets, and in particular exhibits the smallest MAE with respect to the QMC benchmarks for S12L, of all methods compared in Table 9.6. However, the strong dependence of the XDM correction on the choice of underlying functional, which has been noted previously,[390] is somewhat bothersome.

(X)SAPT-Based Methods

Turning to the SAPT- and XSAPT-based results, we note first that the maximum difference between the ω_{GDD} and ω_{IP} tuning procedures (= 0.218 a_0^{-1}, for for the benzoquinone guest molecule in complex **5a**) is much larger than for the L7 data set. Error statistics are again listed in Table 9.6 and Fig. 9.4. Some of these results are labeled "XSAPT(KS)+disp(CKS)", by which we mean that dispersion energies are computed at the coupled Kohn-Sham (CKS) level based on DFT response theory

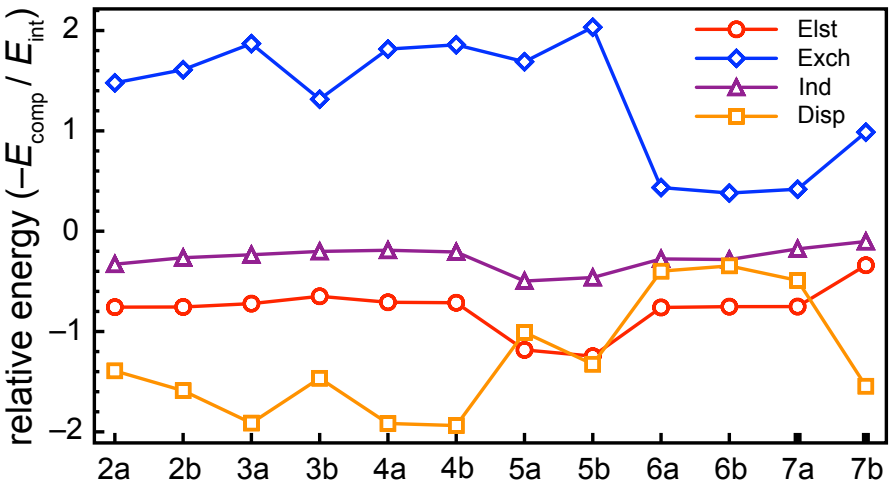

Figure 9.5: Ratio of each energy contribution relative to the total interaction energy for the S12L data set, computed at the XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$($\omega_{\mathrm{GDD}}$)/hpTZVPP level.

(as in DFT-SAPT), but are then combined with an XSAPT(KS) description of the remaining energy components.

The ratios $-E_{\mathrm{comp}}/E_{\mathrm{int}}$, for each energy component E_{comp} according to the XSAPT decomposition analysis,[56] are plotted in Fig. 9.5 as computed at the XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$($\omega_{\mathrm{GDD}}$) level. (The raw data can be found in the Supporting Information.) Almost all of the systems are dominated by dispersion except complexes **5a**, **6a**, **6b** and **7a**, where the dispersion remains large but is not dominant. The electrostatic and induction interactions have about the same pattern across all complexes although the mixed-type dimers **5a** and **5b** have slightly larger electrostatic and induction contributions. On the other hand, most complexes have large exchange-repulsion and

attractive dispersion, with no clear pattern. The non-dispersion part of the interaction energy is positive for all complexes except **6a**, **6b**, and **7b**, whose non-dispersion energies are −54.0, −57.2, and −70.5 kcal/mol, respectively. Clearly, accurate description of dispersion interactions is very important for structure determination and stabilization in these complexes involving large monomers.

DFT-SAPT affords a good description of binding energies for S12L (see Table 9.6), with a MAEs of 2–3 kcal/mol, depending on the benchmarks. For reasons of cost, these calculations[42] use a (DZ,TZ) extrapolation to the CBS limit, and in Ref. 42 it is suggested to multiply the extrapolated energies by an empirical factor of 1.08, to account for basis-set incompleteness, and this correction *slightly* reduces the MAE with respect to the back-corrected experimental benchmarks, by 0.8 kcal/mol, although the maximum error increases by 1.4 kcal/mol. (Note that the benchmarks have been updated since the ones that were used in Ref. 42 to determine the empirical correction factor.[21,22]) With respect to the QMC benchmarks, however, empirical scaling seriously degrades the quality of the DFT-SAPT/CBS results, increasing the MAE from 2.0 to 3.9 kcal/mol and increasing the maximum error from 4.9 to 10.1 kcal/mol. As such, this empirical scaling factor appears to be ill-advised.

For reasons of cost, we have performed fifth-order scaling SAPT0/jaDZ calculations only the two smallest S12L complexes, **2a** and **2b**, obtaining binding energies of −43.3 and −27.2 kcal/mol, respectively. These are too large by 14.3 and 6.4 kcal/mol, respectively, as compared to the back-corrected experimental results. This is consistent with the overstabilization of the L7 complexes at the SAPT0/jaDZ level (see

Table 9.8: Atomic three-body dispersion energies for S12L.

System	energy (kcal/mol)		
	$E_{\text{disp,3B}}^{\text{ATM(TS)}}$	$E_{\text{disp,3B}}^{\text{ATM(TS)}}$(parm)[a]	$E_{\text{disp,3B}}^{\text{ATM(Grimme)}}$[b]
2a	8.04	6.53	1.80
2b	5.65	4.62	1.20
3a	10.46	8.58	1.80
3b	6.42	5.33	0.70
4a	13.40	11.28	3.20
4b	14.48	12.14	3.50
5a	3.80	2.93	1.00
5b	3.46	2.72	1.00
6a	5.59	4.04	2.20
6b	4.48	3.28	1.80
7a	11.43	7.90	4.60
7b	9.13	6.32	3.30

[a] Values are estimated using the average hybridized values of α, C_9, and R_{vdW} without requiring any supersystem calculations.

[b] From Ref. 23.

Table 9.2), and also consistent with L7 results is the fact that the "bronze standard" (empirically scaled) *s*SAPT0/jaDZ approach affords essentially the same binding energies (−43.2 and −27.1 kcal/mol). The error cancellation facilitated by the use of the jaDZ basis in small systems[117,157,158] does not extend to larger ones.

The three-body dispersion correction $E_{\text{disp,3B}}^{\text{ATM(TS)}}$, which is as large as 9.0 kcal/mol for the C3GC complex in L7 (Table 9.3), is even larger for S12L complexes, up to 14.5 kcal/mol for complex **4b** (Table 9.8). Incorporation of $E_{\text{disp,3B}}^{\text{ATM(TS)}}$ into XSAPT(KS) + D3(ω_{GDD}) reduces the MAE by one third, from 9.7 to 3.3 kcal/mol. As was found for the L7 data set, better XSAPT(KS) results are obtained using ω_{GDD} as

Table 9.9: Mean absolute deviations and maximum deviations in DFT-SAPT energy components for the S12L data set, versus those obtained using XSAPT(KS).

Component	deviation (kcal/mol)			
	ω_{IP} tuning		ω_{GDD} tuning	
	MAE	max	MAE	max
$E^{(1)}_{\mathrm{elst}}$	0.70	1.87	0.67	2.07
$E^{(1)}_{\mathrm{exch}}$	2.23	6.13	0.77	1.82

compared to ω_{IP}. To investigate the latter point in more detail, Table 9.9 compares how the first-order electrostatic and exchange energy components in XSAPT(KS) deviate from the corresponding DFT-SAPT energy components,[42] when either ω_{IP} or ω_{GDD} tuning is employed. For $E^{(1)}_{\mathrm{exch}}$, the latter tuning scheme affords results that are much more consistent with DFT-SAPT, and is hence recommended for (X)SAPT calculations that employ LRC functionals to describe the monomers.

The XSAPT(KS)+disp(CKS) results in Table 9.6 use CKS dispersion energies computed using asymptotically-corrected density functionals.[42] The CKS dispersion energies are extrapolated to the CBS limit using a (DZ,TZ) scheme, both with and without the scaling factor of 1.08 suggested in Ref. 42. Using ω_{GDD} tuning for the density functional, we obtain a MAE of 2.3 kcal/mol for S12L as compared to the back-corrected experimental benchmarks. This makes the XSAPT(KS)+disp(CKS) approach superior to XSAPT(KS)+D3, even when three-body dispersion terms are included in the latter. The accuracy of XSAPT(KS)+disp(CKS) is similar to that of DFT-SAPT itself (see Table 9.6). These results support the notion that CKS

dispersion energies should be more accurate than atom–atom dispersion potentials.

Regarding the empirical scaling factor in the CBS extrapolation, we find that XSAPT(KS)+disp(CKS) performs better *without* this factor (MAE = 1.7 kcal/mol) than with it (MAE = 4.4 kcal/mol), as compared to the QMC benchmarks. For both this method and the closely-related DFT-SAPT approach, the mean deviations are negative when the scaling factor is included by positive when it is omitted. This suggests that the factor of 1.08 that was fitted using small systems leads to overcorrection in larger systems; an optimal scaling factor for S12L probably lies between 1.00 and 1.08.

Interestingly, all XSAPT(KS)+D2 methods perform better than the corresponding +D3 approaches, with the best version of the former being XSAPT(KS) + D2 + $E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}(\omega_{\mathrm{GDD}})$, with a MAE of 1.7 kcal/mol. In fact, the +D2+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$ treatment of dispersion outperforms the +disp(CKS) approach, which should in principle be more accurate, suggesting that the +D2+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$ correction may benefit from some error cancellation. The best of the sd-XSAPT methods, sd-XSAPT(KS)(ω_{GDD}), affords MAEs of 6–7 kcal/mol that remain as large as ∼5 kcal/mol when three-body dispersion corrections are included.

A Closer Look at a Few Examples

Within the S12L data set, the largest $E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$ contribution is from dimers **4a** (13.4 kcal/mol) and **4b** (14.5 kcal/mol), which are the buckycatcher complexes with C_{60} and C_{70}, and the smallest contributions occur in the amide macrocycle complexes **5a** (3.8 kcal/mol) and **5b** (3.5 kcal/mol). Whether atomic non-pairwise-additive

dispersion effects are important depends upon the shape, topology, and conjugation of a given host/guest complex. Let us analyze the two extreme cases: the buckycatcher/C_{60} (C_{60}@$C_{60}H_{28}$) complex, **4a**, which has been widely used as a stringent test of theoretical methods for describing dispersion,[20,23,40,42,246,274,373,385,401,402] and the amide macrocycle/benzoquinone complex, **5a**, whose intermolecular interactions are classified as being of mixed type. Binding energies for these two complexes, computed at various levels of theory, are provided in Table 9.10.

Methods listed in Table 9.10 overestimate the binding energy of the buckycatcher/C_{60} complex by anywhere from 2–53 kcal/mol as compared to QMC benchmarks. MP2 performs exceptionally poorly, as expected, with an MP2/CBS binding energy that is 53 kcal/mol too large, whereas the SCS-MP2/CBS and MP2C/CBS approaches afford binding energies that are overestimated by "only" 28 and 10 kcal/mol, respectively. These errors may be due in part to residual BSSE arising from the fact that a (DZ,TZ) extrapolation is used in these cases; for the C2C2PD complex, (aTZ,aQZ) extrapolation has been shown to reduce MP2 binding energies as compared to the CBS limit estimated using (DZ,TZ) extrapolation.[202]

The B2PLYP+D3BJ+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$ and B2PLYP+NL+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$ methods[385] overestimate the binding energy of **4a** by 15 and 20 kcal/mol, respectively, which is likely an artifact of the uncoupled MP2 treatment of dispersion in these double hybrids, which is especially problematic in π-stacked systems. This can be remedied by instead using attenuated MP2 correlation in the double hybrid functionals,[177] although we have not attempted such calculations here.

Table 9.10: Binding energies of complexes **4a** and **5a** from the S12L data set.

Method	binding energy (kcal/mol)	
	4a	**5a**
Experiment[a]	−28.4	−33.4
QMC[b]	−25.8±1.5	−33.4±1.0
HF-3c+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$	−34.2	−27.6
MP2/CBS[c]	−78.8	−39.5
SCS-MP2/CBS[c]	−53.5	−30.1
MP2C/CBS[c,d]	−35.3	−35.4
B2PLYP+D3BJ+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$/ def2-TZVP[e]	−40.6	−33.1
B2PLYP+NL+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$/ def2-TZVP[e]	−45.8	−35.2
NLDFT/def2-QZVP[c]	−33.9	−33.0
PW6B95+D3BJ+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$/ def2-QZVP(-g,-f)[a]	−31.8	−31.9
PBE+MBD*[b]	−28.3	−33.8
DFT-SAPT[c,d,f,g]	−35.9	−34.5
XSAPT(KS)+D3(ω_{GDD})[d,h]	−50.2	−38.2
XSAPT(KS)+D3 +$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$($\omega_{\mathrm{GDD}}$)[d,h]	−36.8	−34.4
XSAPT(KS) +disp(CKS)(ω_{GDD})[d,f,h]	−34.7	−37.1
PBE-XDM/pc-2-spd[i]	−27.5	−30.7
B97M-V/aTZ (CP)	−37.6	−35.2

[a] From Ref. 22.

[b] From Ref. 40.

[c] From Ref. 42.

[d] Using NLDFT deformation energies from Ref. 42.

[e] From Ref. 385.

[f] DFT-SAPT dispersion energies [+disp(CKS)] were extrapolated using the 1.08×(DZ,TZ) scheme of Ref. 42.

[g] Using the cc-pVTZ for the non-dispersion terms.

[h] Using the hpTZVPP basis set.

[i] From Ref. 390.

Heßelmann's NLDFT functional along with the PW6B95+D3BJ+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$ approach each perform pretty well, overestimating the binding energy of **4a** by 8 and 6 kcal/mol, respectively, as compared to QMC benchmark. B97M-V/aTZ with counterpoise correction overestimates the binding energy of **4a** by 12 kcal/mol. The best methods for **4a** in Table 9.10 are PBE+MBD* and PBE-XDM, which only overestimate the binding energy by 2.5 and 1.7 kcal/mol, respectively. The DFT-SAPT and XSAPT(KS)+disp(CKS)(ω_{GDD}) methods overestimate the binding energy by 10 and 9 kcal/mol, respectively, which is probably at least partially due to uncertainty in the DFT-SAPT dispersion energies at the CBS limit, owing to use of the aforementioned 1.08×(DZ,TZ) extrapolation procedure.

The $E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$ correction is very important in **4a**, reducing the error in the XSAPT(KS)+D3 binding energy from 24 to 11 kcal/mol. The remaining 11 kcal/mol may come from the lack of dynamical dielectric screening effects in the dispersion coefficients of the pairwise D3 dispersion potential.[373] Such dynamical screening effects are very important in large supramolecular and solid-state calculations; for example, the C_6 dispersion coefficients are reduced by a factor of 1.6−1.8 in diamond and silicon relative to free atoms.[373,403] The dynamical screening effect in complex **4a**, as computed using the PBE+TS-vdW method, is +9 kcal/mol,[373] which comes from the reduction of C_6 coefficients in C_{60} by a factor of 1.1−1.3 after forming the host-guest complex **4a**.[373] Combining the 9 kcal/mol dynamical screening effect in **4a** with the XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$($\omega_{\mathrm{GDD}}$) binding energy affords −27.8 kcal/mol, which is close to both the QMC and the experimental benchmarks.

For the mixed-type complex **5a**, there are hydrogen bonds between host and guest molecules although dispersion still plays a prominent role. All of the methods in Table 9.10 agree well with the QMC benchmark, and the maximum deviation is only 6 kcal/mol, for MP2/CBS. The XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}(\omega_{\mathrm{GDD}})$ binding energy lies within the statistical error bars of the QMC benchmark. The aforementioned dynamical screening effect for **5a**, computed using the PBE+TS-vdW method, is only 0.1 kcal/mol,[373] which is probably the reason for the good performance of XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}(\omega_{\mathrm{GDD}})$.

Finally, let us discuss complexes **6a** and **6b** whose binding energies are consistently overestimated by the methods considered here. The guest molecule is a cation in these two examples, and the cation/dipolar binding leads to a large errors in continuum solvation energies (up to 6 kcal/mol, according to the estimates in Ref. 22), which are used to back-correct the solution-phase experiments to obtain gas-phase binding energies. This is likely the reason why **6a** has the largest deviation (at 3.6 kcal/mol) between the QMC and the back-corrected experimental benchmark. In particular, it has been demonstrated that the inclusion of counter-ions, leading to overall neutral charge, generally improves the results as compared to the experimental free energies.[396,404–406] COSMO-RS solvation energies for multiply-charged species are also improved by including counter-ions,[405] and the most recent S12L benchmarks (used herein) do use counter-ions in some cases.[22]

The XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}(\omega_{\mathrm{GDD}})$ method overestimates the binding energies of **6a** and **6b** by 2.6 and 1.9 kcal/mol, respectively. We have recently observed

some problems with XPol charge embedding for systems involving ions, and preliminary results suggest that significantly improved results for ion–molecule binding energies are obtained using Gaussian blurring of the embedding point charges. (This has been seen in other contexts as well.[407]) Here, however, XSAPT(KS) + D3 + $E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}(\omega_{\mathrm{GDD}})$ with Gaussian blurring slightly increases the overestimation of the **6a** and **6b** binding energies, so the point-charge embedding is not the culprit. As compared to the more accurate QMC benchmark, XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}(\omega_{\mathrm{GDD}})$ overestimates the binding energy for **6a** by 3.8 kcal/mol, and replacing its dispersion potential by disp(CKS) reduces the error to 1.2 kcal/mol. Even in this electrostatically-dominated system, and accurate description of dispersion interactions is important if quantitative results are required.

One final remark bears mention. In XSAPT calculations, the polarized wave function for each fragment is used for the subsequent SAPT calculations, and typically the dipole moment of this polarized wave function is larger than that of the unpolarized one, often significantly. For example, upon polarizaton by the neighboring charges, the dipole moment of circumcoronene in C3GC increases from 0.09 D to 2.48 D, and the dipole moment of C_{70} in $C_{70}@C_{60}H_{28}$ increases from 0.01 D to 1.39 D. For the cation complexes **6a** and **6b**, the dipole moment for the host molecule increases from approximately zero to almost 7 D, because the guest is a cation. (Dipole moments of the polarized and unpolarized wave functions can be found in the Supporting Information.) The polarized wave functions generated in the XPol step should capture the environmental effects created by the rest of fragments and can in principle be used

density-based schemes for describing dispersion, such as the TS-vdW model.[358] This should capture some many-body dispersion effects without the need for supersystem calculations. We are currently investigating such an approach.

9.4 Conclusions

Accurate description of dispersion interactions is important for modeling large, supramolecular assemblies, and here we have examined two data sets consisting mostly of dispersion-dominated complexes. Using an AO implementation of XSAPT, we have examined the performance of XSAPT(KS)+D3 based on a LRC-DFT description of the monomers. As in previous work,[51,56,57] monomer-specific tuning of the range separation parameter is essential, and here we find that the GDD tuning procedure[26] is superior to the more traditional IP tuning scheme.[230] Optimally-tuned values of ω differ by as much as 0.218 a_0^{-1} between these two methods. We therefore recommend the use of LRC-ωPBE based on ω_{GDD} for both SAPT and XSAPT calculations.

The Axilrod-Teller-Muto three-body dispersion correction, $E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$, provides a significant improvement for binding energies computed using XSAPT(KS)+D3, reducing MAEs from 3.5 kcal/mol for L7 and 9.7 kcal/mol for S12L to 1.4 and 3.3 kcal/mol, respectively. Remaining errors arise primarily from the absence of dynamical dielectric screening effects[370] in the D3 pairwise dispersion coefficients, which become very important when the monomers are large. For example, the binding energy of the buckycatcher/C_{60} complex [**4a** in Figure 9.1(b)] computed at the XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}(\omega_{\mathrm{GDD}})$ level is -36.8 kcal/mol as compared to a QMC

benchmark of -25.8 ± 1.5 kcal/mol.[40] Addition of the +9 kcal/mol dynamical screening effect predicted by PBE+TS-vdW method[373] reduces the aforementioned XSAPT result to -27.8 kcal/mol, which is close to the QMC benchmark.

Concerning the application of supersystem methods to L7 and S12L, we find that the newly-developed B97M-V functional,[202] with an aTZ basis set and counterpoise correction scheme, affords a MAE of 1.4 and 4.2 kcal/mol, respectively, for L7 and S12L complexes. Its maximum error in S12L is up to 9.6 kcal/mol, despite the fact that this functional performs well for many other non-covalent systems,[52,57,105] including difficult cases involving ions.[52] Modeling of large complexes must therefore still be considered difficult even with the latest DFT functionals. However, the best supersystem method that we have found for these large complexes is PW6B95+D3BJ+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$, with MAEs of 0.9 kcal/mol for L7 and 1.5 kcal/mol for S12L. The low-cost, empirically-corrected HF-3c+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(Grimme)}}$ method performs reasonable well, yielding MAEs of 1.0 kcal/mol for L7 and 3.9 kcal/mol for S12L.

CHAPTER 10

A Simple Correction for Non-Additive Dispersion within Extended Symmetry-Adapted Perturbation Theory

10.1 Introduction

The developments of accurate and efficient dispersion models have mitigated the failure of semi-local and hybrid density functionals in dispersion descriptions,[408,409] which have become popular in this decade.[143,273] Density functional theory added with dispersion potential (DFT+D) methods provide an efficient way to understand complex molecular systems, in systems such as biomolecules, supramolecules, and condensed matter.[352,410] The simplest way to account for the dispersion is to use atomic pairwise-additive contributions from localized multipoles:

$$E_{\text{disp}} = -\sum_{i>j}\left(\frac{C_6^{ij}}{R_{ij}^6} + \frac{C_8^{ij}}{R_{ij}^8} + \frac{C_{10}^{ij}}{R_{ij}^{10}} + \cdots\right) \tag{10.1}$$

where R_{ij} is the interatomic distance for atoms i and j. For each pair of atoms, such as atoms i and j, C_6^{ij} is dipole−dipole dispersion coefficient, C_8^{ij} represents dipole−quadrupole dispersion coefficient, C_{10}^{ij} expresses quadrupole−quadrupole and dipole−octupole dispersion coefficients, and so on. The dispersion coefficients can be determined by appealing to *ab initio* calculations[274,358,411–413] or at least partly

based on fitting to a benchmark training set.[6,16,57,244,253,254,368,414,415] Eq. (10.1) is well-defined for well-separated atoms at long ranges but diverges to $-\infty$ at small interatomic separation. Thus, a damping function has been used to avoid the divergence at short range and to avoid double-counting of correlation effects which have already been partly captured by the main body of density functionals at short range.[143] These damping functions often include several functional-dependent parameters to control their domain for the specific density functional.[143]

In the early stage of DFT+D methods, the dispersion potentials did not explicitly include non-additive effects which measure the departure of the benchmark dispersion from the dispersion in Eq. (10.1) based on gas-phase C_n^{ij} coefficients.[370] In other words, the gas-phase C_n^{ij} dispersion coefficients are predetermined and constant quantities which are independent of molecular environment during molecular simulations. The first step to fix this departure is to introduce the environment-dependent C_n^{ij} coefficients. The environment-dependent C_n^{ij} coefficients are identified as "type-A" non-additive effects,[370] in which crowding or squeezing by neighboring atoms results in a reduction in atomic polarizabilities and thus dispersion energies. Type-A effects *are* accounted for by methods such as the Tkatchenko-Scheffler (TS) correction[358] and the Becke-Johnson exchange-hole dipole model (XDM)[412] because these corrections depend directly on the local electron density. Grimme's D3 correction[274] also includes type-A effects because its atomic dispersion parameters depend upon the number of bonds, i.e. the hybridization of atoms.

An alternative way to obtain dispersion interactions is based on symmetry-adapted

perturbation theory (SAPT)[115] where its dispersion is genuine at all distances and different from those add-on dispersion models used in DFT. The SAPT calculation can be performed economically by using Kohn-Sham (KS) orbitals for the monomers to incorporate intramolecular electron correlation. The approach is known as SAPT(KS).[208] By using asymptotically corrected DFT, such as long-range corrected (LRC) DFT, the accurate predication of the non-dispersion components of the energy can be achieved;[51] however, errors in SAPT(KS) dispersion energies remain unacceptably large. Accurate dispersion can be computed by an alternative DFT-based SAPT method, DFT-SAPT[149,150] where the "coupled" Kohn-Sham" (CKS) dispersion interaction is calculated via a Casimir-Polder-type formalism based on frequency-dependent polarizabilities for the monomers. The computational scaling of SAPT is still high since the second-order dispersion and the corresponding exchange-dispersion scale at minimum $\mathcal{O}(N_f^4)$ and $\mathcal{O}(N_f^5)$, respectively, with respect to the fragment size N_f. An approximation is to omit those dispersion and exchange-dispersion terms by replacing the sum of them with empirical atom−atom potentials where the parameters are fitted to reproduce accurate dispersion interactions calculated by DFT-SAPT[6,244] and high-level SAPT2+(3).[57] Those SAPT-based dispersion potentials ("+D") avoid any double-counting of dispersion interactions at short and medium range, and represent genuine dispersion interactions in all range, whereas it is not the case of the add-on dispersion potentials in dispersion-corrected DFT.[253,254,368] This SAPT(KS)+D method based on LRC-DFT for monomers exhibits $\mathcal{O}(N_f^3)$ scaling and affords good results for small data sets as well as the whole potential energy curve for the benzene

dimer.[51] The latest version of SAPT-based empirical dispersion potentials ("+D3") was developed by our group.[57] Our SAPT-based "+D3" dispersion potential is unrelated to Grimme's +D3 correction for DFT[274] except that each is the third-generation of such a correction.

The traditional SAPT method was generalized to many-body systems by our group via the combination of explicit polarization (XPol) method[252] and SAPT. This many-body SAPT method is called "extended" version of SAPT (XSAPT).[54,57,65] The combination of XSAPT based on the monomer description by LRC-DFT and empirical dispersion potentials, XSAPT(KS)+D, affords accurate binding energies and individual energy components for many different systems.[55–57] The drawback of SAPT-based dispersion potentials are that their parameters are predetermined and independent of molecular environment in many-body systems similar to early stage DFT+D methods.[253,254,368]

In this work, the environment-dependent dispersion contraction (DC) effect is introduced within the XSAPT scheme. In the following section, the prescription for estimating the magnitude of DC within the XSAPT scheme is presented. Then, we evaluate the significance of DC using XSAPT in many-body systems. A brief conclusion can be found in the last section.

10.2 Computational details

Within XSAPT, type-A non-additivity is included *implicitly* via the XPol charge-embedding step, wherein the electron density of each fragment is squeezed by the

neighboring charges, thereby reducing the dispersion interaction. The dispersion energy difference based on XSAPT (with charge embedding) and SAPT (without charge embedding),

$$E_{\rm disp}^{\rm DC-full} = [E_{\rm disp}^{(2)} + E_{\rm exch\text{-}disp}^{(2)}]_{\rm XSAPT} - [E_{\rm disp}^{(2)} + E_{\rm exch\text{-}disp}^{(2)}]_{\rm SAPT}, \tag{10.2}$$

should account for the squeezing effects (dispersion contraction, DC) by the neighboring charges where $E_{\rm disp}^{(2)}$ and $E_{\rm exch\text{-}disp}^{(2)}$ are second-order dispersion and exchange-dispersion terms, repsectively. The calculation of $E_{\rm exch\text{-}disp}^{(2)}$ with $\mathcal{O}(N_f^5)$ scaling is much more expensive than the calculation of $E_{\rm disp}^{(2)}$ with $\mathcal{O}(N_f^4)$ scaling, and it is not practical to use XSAPT in systems with large size of monomers if the $E_{\rm exch\text{-}disp}^{(2)}$ term is explicitly included. In practice, $E_{\rm exch\text{-}disp}^{(2)}$ is generally several times smaller than $E_{\rm disp}^{(2)}$ and decays exponentially with respect to the separation of monomers. Thus, it is expected that $[E_{\rm exch\text{-}disp}^{(2)}]_{\rm XSAPT}$ and $[E_{\rm exch\text{-}disp}^{(2)}]_{\rm SAPT}$ terms cancel out (at least partially), and leads to a computationally favorable, $\mathcal{O}(N_f^4)$ scaling, way to estimate DC effect,

$$E_{\rm disp}^{\rm DC-part} = [E_{\rm disp}^{(2)}]_{\rm XSAPT} - [E_{\rm disp}^{(2)}]_{\rm SAPT}. \tag{10.3}$$

$E_{\rm disp}^{\rm DC-full}$ and $E_{\rm disp}^{\rm DC-part}$ are used to distinguish the estimation of DC effect by using Eq. (10.2) and (10.3), respectively. Although the CKS method and a very large basis set have to used to obtain accurate intermolecular dispersion energy,[153] the dispersion energy *difference* in Eq. (10.2) and (10.3) should converge faster and is less sensitive to the method involved. A similar idea using the rapid convergence in energy difference

between the CKS and uncoupled Hartree-Fock dispersion energies has been successfully used in MP2C method.[311] In this work, the DC energies in Eq. (10.2) and (10.3) are calculated by uncoupled SAPT(KS) method, and the basis-set dependence for DC energies will be tested in section 10.3.2. Thorough comparisons to benchmark calculations, as presented in section 10.3.2, support the accuracy and robustness of using uncoupled SAPT(KS) method and double-ζ basis set to estimate DC energies according to Eq. (10.2) or (10.3). Then the scaling of XSAPT(KS)+D3+DC method is $\mathcal{O}(N_f^3)$ using a triple-ζ basis set for the non-dispersion terms, and $\mathcal{O}(N_f^5)$ or $\mathcal{O}(N_f^4)$ using a double-ζ basis set for the DC terms which depend on using Eq. (10.2) or (10.3), respectively.

The widely studied $(H_2O)_6$ cluster with eight isomers,[41] anionic $X(H_2O)_{n=1-6}$ clusters ($X = F^-$, Cl^-, and SO_4^{2-}), and cationic $M(H_2O)_{n=1-6}$ clusters ($M = Li^+$, Na^+, and K^+) are employed in this work to study the importance of DC in different sizes and charges of clusters. Our "+D3" parameters were extended to include Li, Na, and K. The values for those parameters and the details of fitting procedures are shown in the Supporting Information for this paper. Larger clusters, ten isomers of $F^-(H_2O)_{10}$ and ten isomers of $(H_2O)_{20}$, are used in the analysis to see the contributions of DC in such large clusters.

The geometries of $F^-(H_2O)_{n=1-6}$, $Cl^-(H_2O)_{n=1-6}$, and $F^-(H_2O)_{10}$ are taken from Ref. 57. The geometries of $(H_2O)_6$ and $(H_2O)_{20}$ are taken from Ref. 8 and 18, respectively. The starting global minima of $SO_4^{2-}(H_2O)_{n=3-6}$ optimized using basin-hopping

Monte Carlo are taken from Ref. 416, and subsequently those structures were optimized at the resolution-of-identity (RI) MP2/aug-cc-pVDZ level. In the rest of the paper, aug-cc-pVXZ basis sets are abbreviated as aXZ. For completeness, the global minima of $SO_4^{2-}(H_2O)_{n=1-2}$ were obtained manually at the same level of method although $SO_4^{2-}(H_2O)_{n=1-2}$ structures are unstable in the gas phase with respect to electron autodetachment.[417,418] The geometries of cation $M(H_2O)_{n=1-6}$ clusters were optimized to obtain global minima manually at the method of TPSS functional[419] with Grimme's D3 dispersion correction[274] using Becke−Johnson damping[14] combined with def2-TZVPP basis set.[197]

The starting geometries of building blocks for clathrate hydrates, $(H_2O)_{20}$, $(H_2O)_{24}$, and $(H_2O)_{28}$, are taken from Ref. 25. Then, small guest molecules, CH_4 at $(H_2O)_{20}$, CH_4 at $(H_2O)_{24}$, and tetrahydrofuran (THF) at $(H_2O)_{28}$, are put into water cages for doing geometry optimizations at the level of B97M-V/6-31+G(d) with freezing oxygen atoms on cages.

All binding energies in this work are based on the monomer geometries in clusters, and the monomer relaxation energies are not considered. The binding energies at the level of CCSD(T) (coupled cluster singles and doubles with perturbative triples) extrapolated to complete basis-set (CBS) limit for $(H_2O)_6$, $(H_2O)_{20}$, $F^-(H_2O)_{n=1-6}$, $Cl^-(H_2O)_{n=1-6}$, and $F^-(H_2O)_{10}$ are taken from Ref. 57. For the $SO_4^{2-}(H_2O)_{n=1-6}$

clusters, the MP2 correlation energies in the CBS limit were estimated using a two-point (aTZ/aQZ) extrapolation, and this correlation energy was added to the Hartree-Fock/aQZ energy to estimate the MP2/CBS energy. Then, a triples correction,

$$\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}} = E_{\mathrm{CCSD(T)}} - E_{\mathrm{MP2}} \ , \tag{10.4}$$

based on def2-TZVPP basis set was added to MP2/CBS energy to compute the CCSD(T)/CBS energy.

For the three alkali−water clusters, the MP2/CBS energies were estimated at a similar way but using def2-TZVPP and def2-QZVPP basis sets to alkali atoms for triple- and quadruple-ζ basis sets, respectively, in the corresponding procedures. The $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ correction is evaluated using def2-TZVPP basis set for alkali atoms and heavy-aug-cc-pVTZ basis set for the remaining atoms, in which the diffuse functions on hydrogen in aug-cc-pVTZ are removed.

For the three cages of clathrate hydrates with small guest molecules, the MP2/CBS were also estimated at a similar way as $SO_4^{2-}(H_2O)_{n=1-6}$ clusters but using def2-TZVPP and def2-QZVPP basis sets for triple- and quadruple-ζ basis sets, respectively, in the corresponding procedures. The $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ correction is evaluated using def2-TZVP basis set, but CCSD(T) energies are calculated at the domain based local pair-natural orbital (DLPNO) scheme[382,383] with tight PNO of truncation threshold[384] which approaches the canonical CCSD(T) results within 1 kJ/mol.[420] All geometries in Cartesian coordinates and the corresponding CCSD(T)/CBS binding energies are available in Supporting Information.

Some modern popular density-functional methods including PW6B95-D3 (zero damping),[274,392] M11,[421] ωB97X-D3,[13] ωB97X-V,[105] and B97M-V[202] will be tested because they have shown good performance for non-covalent interactions.[11,105,202,203,422] A series of DFT methods recommended for clathrate hydrates[344] will also be tested here for the interactions between water cages and guest molecules. The def2-TZVPPD basis set[197,423] was used in those DFT calculations.

The LRC-ωPBE functional,[227,240] which has no short-range HF exchange, is used in all XSAPT(KS) calculations. Two ways to tune the range-separation parameter (ω) in LRC-ωPBE are used in this work. They are ω_{IP} tuning[230] which adjusts ω to make the lowest ionization potential (IP) equal the minus HOMO (highest occupied molecular orbital) energy, and ω_{GDD} tuning[26,60] where ω is global density-dependent (GDD) and is related to the average distance between an electron in the outer regions of a molecule and the exchange hole in the region of localized orbitals. The advantage of using LRC functional in XSAPT is that the exchange-correlation (XC) potential is the functional derivation of the XC energy.[51] On the other hand, Chai *et al.*[217] developed an asymptotically-corrected (AC) model potential called a strictly localized Fermi−Amaldi (LFAs) scheme which can be directly added to any semi-local density functional, and the corrected XC potential is still the functional derivation of the XC energy. The LFAs scheme combined with PBE functional is tested here for the first time in the context of XSAPT(KS).

The individual energy components for dimers in S22 and S66 data sets[10,186] calculated by XSAPT(KS) energy decomposition scheme[56] based on different density functionals are compared with the benchmark SAPT2+(3)/aTZ energy components[117] to test the performance of those functionals. The SAPT2+(3)/aTZ individual energy components for dimers in S22 data set are taken from Ref. 51, and the individual energy components for dimers in S66 data set are shown here for the first time.

In this work, XSAPT(KS)+D3 calculations were employed hpTZVPP basis set (def2-TZVPP augmented with diffuse functions on non-hydrogen atoms that are taken from Pople basis sets) which was recommended in our previous works.[56,57] Although the tuning ω values are dependent on monomer's geometry, those tuning ω values used in this work are based on the monomers optimized at the level of RIMP2/aug-cc-pVDZ. Those tuned values of ω_{IP} and ω_{GDD} are available in the Supporting Information.

The TPSS geometry optimizations, DLPNO-CCSD(T) and MP2 calculations for clathrate hydrates were performed using the ORCA program,[188] v. 3.0.3, with TightSCF convergence. In TPSS calculations, the RI approximation[424] for the Coulomb integrals matching default auxiliary basis sets[425] was applied to speed up the computations, and the numerical quadrature grid in ORCA denoted as Grid4 was used. The SAPT2+(3), and the single-point MP2 and CCSD(T) calculations for $SO_4^{2-}(H_2O)_{n=1-6}$ and alkali−water clusters were performed using the BETA5 version of the PSI4 program.[187] The RI-JK algorithm[426] for SCF parts and RI approximation for correlation parts are employed in all MP2 and CCSD(T) calculations in this work.

For those DFT and XSAPT(KS) calculations, the SCF convergence criterion is set to $\tau_{\mathrm{SCF}} = 10^{-7}$ a.u. and the integral screening threshold is set to $\tau_{\mathrm{ints}} = 10^{-12}$ a.u. A fairly dense Euler-Maclaurin-Lebedev quadrature grid ($N_{\mathrm{radial}} = 75, N_{\mathrm{angular}} = 302$) has been employed in those supersystem DFT calcualtions. All supersystem single-point calculations are eliminated basis-set superposition error (BSSE) by using counterpoise (CP) correction.[80] All XSAPT(KS) calculations reported here use smooth atom-centered ChElPG embedding charges[54] for the XPol calculations and "projected"[65] (pseudocanonicalized[286,330] monomer-centered) basis sets for the SAPT corrections. ChElPG charges were computed based on Lebedev grid using a "head space" of 3.0 Å (see Ref. 319) with radial shells spaced $\Delta r = 0.25$ Å apart where each Lebedev shell consists of a (110,590) grid. The non-dispersion terms in XSAPT(KS) calculations are efficiently evaluated in the atomic-orbital AO basis, in a way that scales no worse than cubically with respect to system size,[60] and the dispersion energy is evaluated by our third-generation, SAPT-based dispersion potential ("+D3").[57] All DFT and XSAPT(KS) calculations were performed using a locally-modified version of Q-Chem.[97]

10.3 Results and discussion

10.3.1 Asymptotically-corrected density functionals for XSAPT(KS)

First, we compare the performance of a series of density functionals in XSAPT energy decomposition.[56] Fig. 10.1 shows the mean absolute errors (MAEs) of the non-dispersion energy components for the dimers in S22 and S66 data sets, as computed by

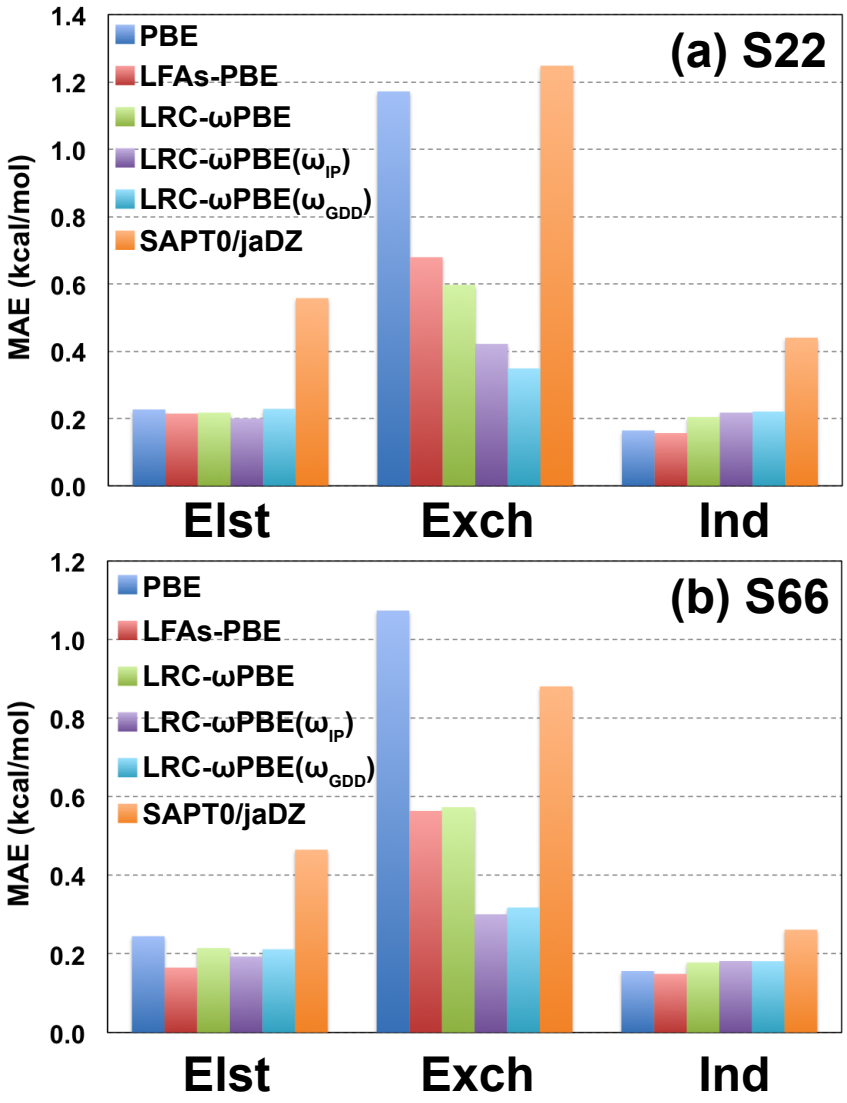

Figure 10.1: Mean absolute errors (MAEs) in kcal/mol for individual energy components, which are electrostatics (Elst), exchange (Exch), and Induction (Ind), of the (a) S22 and (b) S66 data sets using XSAPT(KS) based on a series of density functionals and SAPT0 methods with respect to SAPT2+(3)/aTZ energy components. The hpTZVPP basis set is used for XSAPT(KS) calculations and the jaDZ basis set is used for SAPT0 calculations.

XSAPT(KS)/hpTZVPP methods and also by SAPT0/jun-cc-pVDZ (jaDZ) method with respect to the $\mathcal{O}(N^7)$ scaling SAPT2+(3)/aTZ method.

For S22, MAEs for the electrostatics (Elst) and induction (Ind) energy components calculated using XSAPT(KS) are similar in all density functionals tested here. For the exchange (Exch) energy component calculated using XSAPT(KS), MAEs decrease in the order, PBE > LFAs-PBE/LRC-ωPBE > LRC-ωPBE($\omega_{\mathrm{IP}}/\omega_{\mathrm{GDD}}$), due to an improvement of asymptotical XC potentials. Thus, tuning is necessary in LRC-DFT to obtain accurate energy components in (X)SAPT(KS).[51] MAEs for both tuning methods, LRC-ωPBE($\omega_{\mathrm{IP}}/\omega_{\mathrm{GDD}}$), are about 0.2 kcal/mol for Elst and Ind, and 0.4 kcal/mol for Exch. Although the LFAs-PBE method is computational efficient,[217] its MAE for Exch energy component is about 1.5 times worse than the MAE using tuned LRC-ωPBE. The SAPT0 combined with jaDZ basis set has been suggested to use in large systems[160] since it provides fortuitous error cancellation between the overestimation of the uncoupled Hartree-Fock dispersion and the underestimation due to basis set incompleteness error for providing good total binding energies.[157] MAEs of SAPT0/jaDZ individual energy components are all worse than the corresponding energy components using XSAPT(KS)/hpTZVPP with all tested density functionals. Although SAPT0/jaDZ provides poor performance for energy decomposition, it still has been used to do energy decomposition in different kinds of systems.[160,427] MAEs of dispersion energy components for the dimers in the S22 data set are 0.1 and 1.5 kcal/mol for SAPT-based D3 dispersion potential and SAPT0/jaDZ method, respectively.

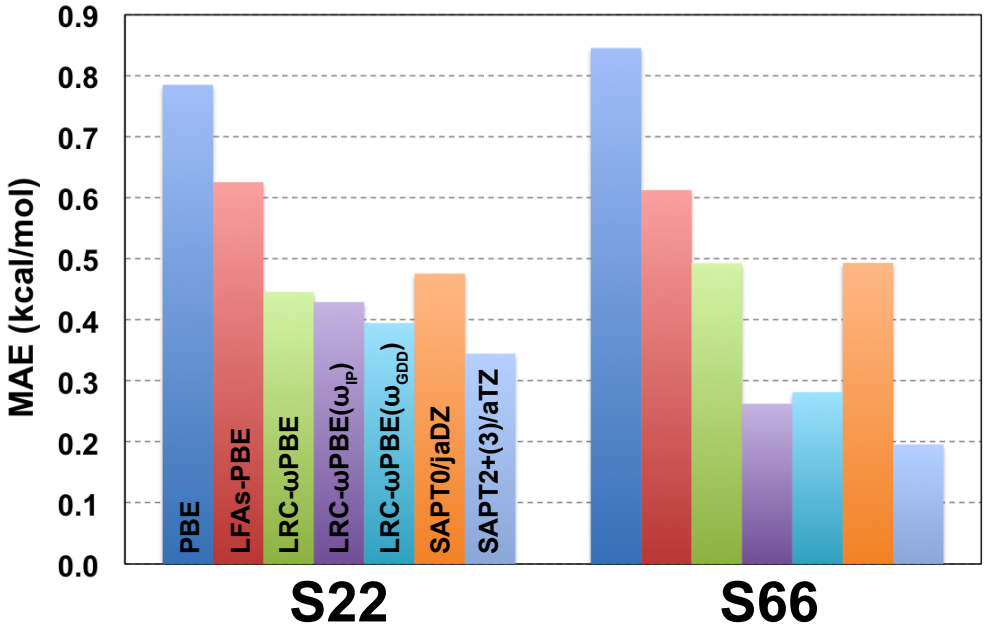

Figure 10.2: Mean absolute errors (MAEs) in kcal/mol with respect to CCSD(T)/CBS benchmarks, for binding energies in the S22 and S66 data sets using XSAPT(KS)+D3, SAPT0, and SAPT2+(3) methods. The hpTZVPP basis set is used for XSAPT(KS)+D3 calculations. The jaDZ and aTZ basis sets are used for SAPT0 and SAPT2+(3) calculations, respectively.

For S66, XSAPT(KS)/hpTZVPP and SAPT0/jaDZ show similar performance as in S22 except the MAE of Exch energy components using XSAPT(PBE)/hpTZVPP is larger than the corresponding MAE using SAPT0/jaDZ. MAEs of dispersion energy components for the dimers in S66 are 0.2 and 1.4 kcal/mol for SAPT-based D3 dispersion potential and SAPT0/jaDZ method, respectively.

MAEs of total binding energies using XSAPT(KS)+D3, SAPT0/jaDZ, and SAPT2+(3)/aTZ with respect to CCSD(T)/CBS for the dimers in S22 and S66 data sets are shown in Fig. 10.2. In XSAPT(KS), the tuned LRC-ωPBE(ω_{IP}/ω_{GDD}) methods still

outperform PBE, LFAs-PBE, and untuned LRC-ωPBE methods. Both SAPT2+(3) and CCSD(T) methods have seventh-order scaling but BSSE-free SAPT2+(3) method is not required to use a very large basis set to remove BSSE where it is required for all supersystem methods. MAEs of SAPT2+(3)/aTZ with respect to CCSD(T)/CBS are 0.3 and 0.2 kcal/mol, respectively, for S22 and S66 data sets. The $\mathcal{O}(N_f^3)$ scaling XSAPT(KS)+D3/hpTZVPP method using tuned LRC-ωPBE($\omega_{\mathrm{IP}}/\omega_{\mathrm{GDD}}$) is only slightly worse than the expensive $\mathcal{O}(N_f^7)$ scaling SAPT2+(3)/aTZ method in total energy computations. SAPT0/jaDZ provides good error cancellation for total binding energies, and is comparable to the performance of XSAPT(KS) using untuned LRC-ωPBE. Furthermore, SAPT0/jaDZ provides total binding energies with smaller MAEs than XSAPT(KS) using PBE or LFAs-PBE functionals. Thus, XSAPT(KS)+D3 method using tuned LRC-ωPBE($\omega_{\mathrm{IP}}/\omega_{\mathrm{GDD}}$) for monomer description provides accurate results for both total binding energies and individual energy components. In the rest of the work, all XSAPT(KS)+D3 results are based on LRC-ωPBE($\omega_{\mathrm{IP}}/\omega_{\mathrm{GDD}}$) for monomer description.

10.3.2 Basis set dependence of dispersion contraction

To find a small basis set for computational efficiency in dispersion contraction which includes $\mathcal{O}(N_f^4)$ scaling dispersion and $\mathcal{O}(N_f^5)$ exchange-dispersion terms, the dispersion contraction was evaluated using an extensive series of double-ζ basis sets. We have selected three anionic clusters, which are $F^-(H_2O)_6$, $Cl^-(H_2O)_6$, and $SO_4^{2-}(H_2O)_6$, with large DCs for this test, and the corresponding DCs are 2.66, 1.60, and 3.08 kcal/mol, respectively, computed by XSAPT(KS)/hpTZVPP using tuned LRC-ωPBE

(ω_{IP}). Since LRC-ωPBE using ω_{IP} and ω_{GDD} tunings show similar results in this test, only the ω_{IP} data are shown here. Fig. 10.3 shows the mean absolute percentage error (MAPE) for XSAPT(KS) using a series of double-ζ basis set using both DC-full [Eq. (10.2)] and DC-part [Eq. (10.3)] schemes as compared to XSAPT(KS)/hpTZVPP using DC-full scheme. The basis sets are ordered by increasing the numbers of basis functions, which is a easy way to see the convergence of the calculation with respect to the basis set size.

Using DC-full [Eq. (10.2)] scheme, the results converge smoothly by increasing the size of double-ζ basis set. The MAPE for aug-cc-pVDZ basis set is only 2.3%. Using DC-part [Eq. (10.3)] scheme, the smallest MAPE is at 6-31+G(d) basis set with 5.3%, then the MAPE keeps increasing by increasing the size of basis set. In the DC-full scheme, the *good* performance relies on the convergence of both dispersion energy difference and exchange-dispersion energy difference in Eq. (10.2). Thus, a large basis set has to use in DC-full scheme to hold this delicate balance for dispersion and exchange-dispersion terms. Omitting the part from exchange-dispersion energy difference in large basis set damages this balance and generates large MAPE in DC-part scheme. Luckily, the 6-31+G(d) basis set gives pretty good results in DC-part scheme based on fortuitous error cancellation, and we will subsequently use it in the rest of the work to test its performance.

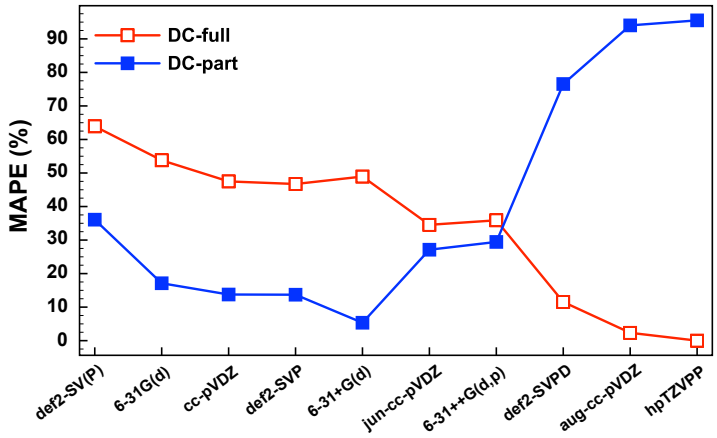

Figure 10.3: Mean absolute percentage errors (MAPEs) in the test set including $F^-(H_2O)_6$, $Cl^-(H_2O)_6$, and $SO_4^{2-}(H_2O)_6$ complexes obtained using DC-full [Eq. (10.2)] and DC-part [Eq. (10.3)] in a series of basis sets as compared to DC-full using hpTZVPP basis set. All calculations are used XSAPT(KS) with tuned LRC-ωPBE(ω_{IP}).

10.3.3 Anionic $\mathbf{X(H_2O)}_{n=1-6}$ ($\mathbf{X = F^-}$, $\mathbf{Cl^-}$, and $\mathbf{SO_4^{2-}}$) and cationic $\mathbf{M(H_2O)}_{n=1-6}$ ($\mathbf{M = Li^+}$, $\mathbf{Na^+}$, and $\mathbf{K^+}$)

Fig. 10.4 shows MAEs of binding energies for XSAPT(KS)+D3 and supersystem methods. In XSAPT(KS)+D3 method based on tuned LRC-ωPBE, ω_{GDD} tuning works slightly better (0.89 kcal/mol on average) than ω_{IP} tuning in anionic clusters. On the other hand, ω_{IP} tuning works slightly better (0.38 kcal/mol on average) than ω_{GDD} tuning in cationic clusters. For the whole set of clusters, ω_{IP} and ω_{GDD} tunings offer MAEs with 1.90 and 1.64 kcal/mol, respectively.

The performance of XSAPT(KS)+D3 is improved by including the DC correction except ω_{IP} tuning in $Na^+(H_2O)_n$ and $K^+(H_2O)_n$, and ω_{GDD} tuning in $K^+(H_2O)_n$.

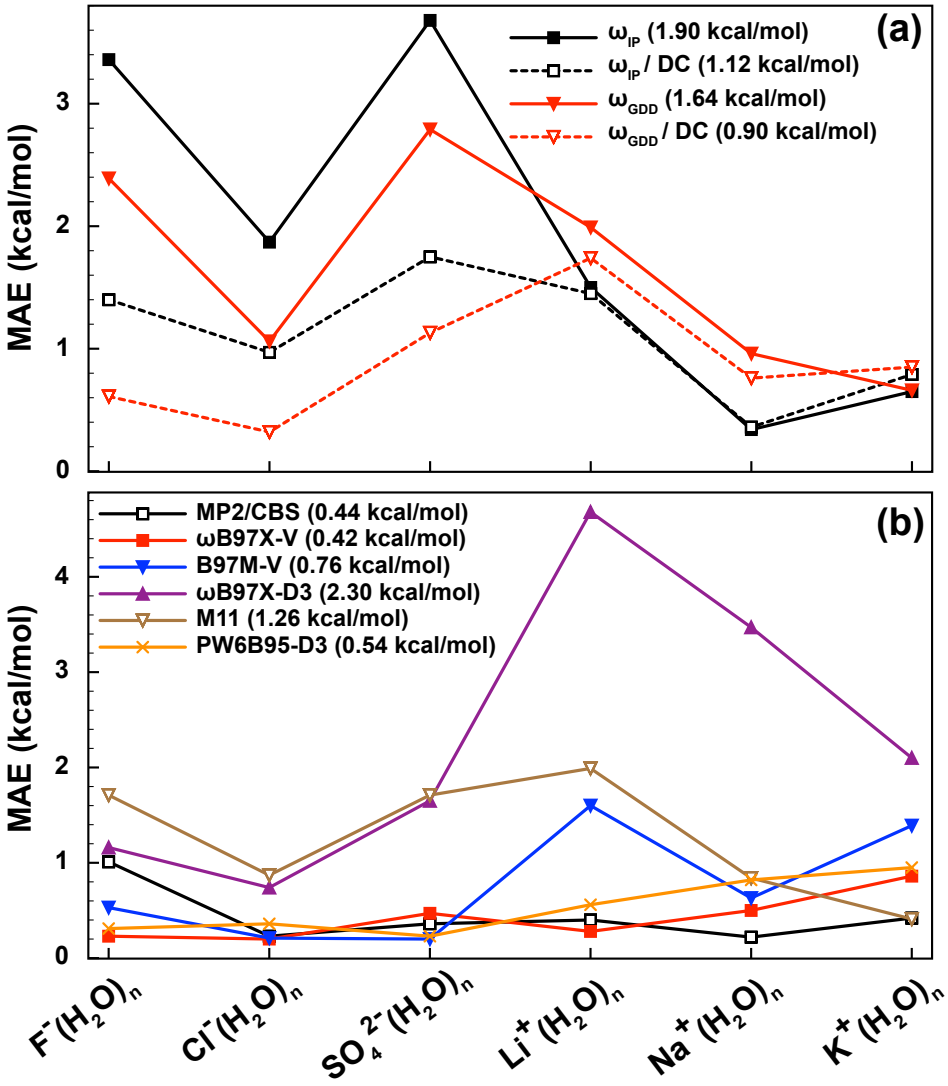

Figure 10.4: Mean absolute errors (MAEs) in kcal/mol for binding energies of ionic anionic $X(H_2O)_{n=1-6}$ (X = F^-, Cl^-, and SO_4^{2-}) and cationic $M(H_2O)_{n=1-6}$ (M = Li^+, Na^+, and K^+) clusters with respect to CCSD(T)/CBS benchmarks using (a) XSAPT(KS)+D3 and (b) supersystem methods. The hpTZVPP basis set is used for XSAPT(KS)+D3 and DC calculations, and the def2-TZVPPD basis set is used for DFT calculations with CP corrections. MAEs for the whole data set are also listed for each method.

Especially, MAEs in anionic clusters are largely reduced by including DC corrections. For example, MAEs in $F^-(H_2O)_n$ clusters are reduced from 3.36 and 2.39 kcal/mol to 1.40 and 0.61 kcal/mol by including DC corrections in XSAPT(KS) calculations using ω_{IP} and ω_{GDD} tunings, respectively. In Ref. 57, we blamed the large errors of XSAPT(KS)+D3 in halide-water clusters because of missing the three-body induction couplings and "δMP2" term[52,173] which accounts for the high-order coupling between induction and dispersion. Including those coupling corrections in anionic systems makes it expensive to use XSAPT(KS) methods in systems with large size of monomers. Here we show that the large errors in anionic clusters actually come from the missing squeezing effects of dispersion in many-body systems and its contribution is up to 3 kcal/mol in $SO_4^{2-}(H_2O)_6$ cluster. Hence, it is essential to include DC correction for capturing dispersion squeezing effect in many-body systems. Then, the XSAPT(KS)+D3+DC method with $\mathcal{O}(N_f^5)$ scaling affords MAEs with 1.12 and 0.90 kcal/mol by using ω_{IP} and ω_{GDD} tunings, respectively, for the whole ionic data set. The $\mathcal{O}(N_f^5)$ scaling of DC-full correction can be replaced by DC-part correction with only $\mathcal{O}(N_f^4)$ scaling in a 6-31+G(d) double-ζ basis set as shown in Sec. 10.3.2. By using DC-part correction with 6-31+G(d) basis set in XSAPT(KS)+D3+DC method, the MAEs change marginally and the corresponding MAEs are 1.10 and 0.95 kcal/mol. Thus, the scaling of XSAPT(KS)+D3+DC is $\mathcal{O}(N_f^3)$ for the non-dispersion terms using a triple-ζ basis set and $\mathcal{O}(N_f^4)$ for the DC-part correction using a double-ζ basis set.

The MP2/CBS method affords the MAE with < 1 kcal/mol for all ionic clusters

except $F^-(H_2O)_n$ with a MAE of 1.01 kcal/mol, and its MAE for the whole ionic data set is 0.44 kcal/mol. The ωB97X-V and PW6B95-D3 are the only two methods which afford MAEs with < 1 kcal/mol for all ionic clusters, and the corresponding MAEs for the whole ionic data set are 0.42 and 0.54 kcal/mol, respectively. The ωB97X-D3 method does not work well in ionic clusters especially in cationic clusters where the MAE is up to 4.68 kcal/mol in $Li^+(H_2O)_n$. The M11 method shows similar performance as ωB97X-D3 in anionic clusters but performs better in cationic clusters. The highly parametrized M11 and dispersion-corrected ωB97X-D3 functionals are long-range-corrected, and affords MAEs of 1.26 and 2.30 kcal/mol, respectively. The non-local dispersion B97M-V calculation is much faster than the ωB97X-V calculation, due to the absence of HF exchange, however, the neglect of HF exchange degrades its performance in cationic clusters. For the whole set of ionic clusters, B97M-V performs about two times worse than ωB97X-V. Hence, it is important to incorporate both long-range correction and non-local dispersion in DFT simultaneously for accurate binding descriptions in different kinds of systems. This may be the reason why ωB97X-V shows good performance in both cationic and anionic clusters, and its good performance can extend in large size of clusters also as shown below. The good performance of PW6B95-D3 is only coincidence here since it shows pretty large errors in $(H_2O)_{20}$ clusters as shown below.

10.3.4 $(H_2O)_6$, $(H_2O)_{20}$, and $F^-(H_2O)_{10}$

Eight low-lying structures[8] of $(H_2O)_6$ are considered here. Fig. 10.5 compares the binding energies for isomers of $(H_2O)_6$ computed using various methods. XSAPT(KS)

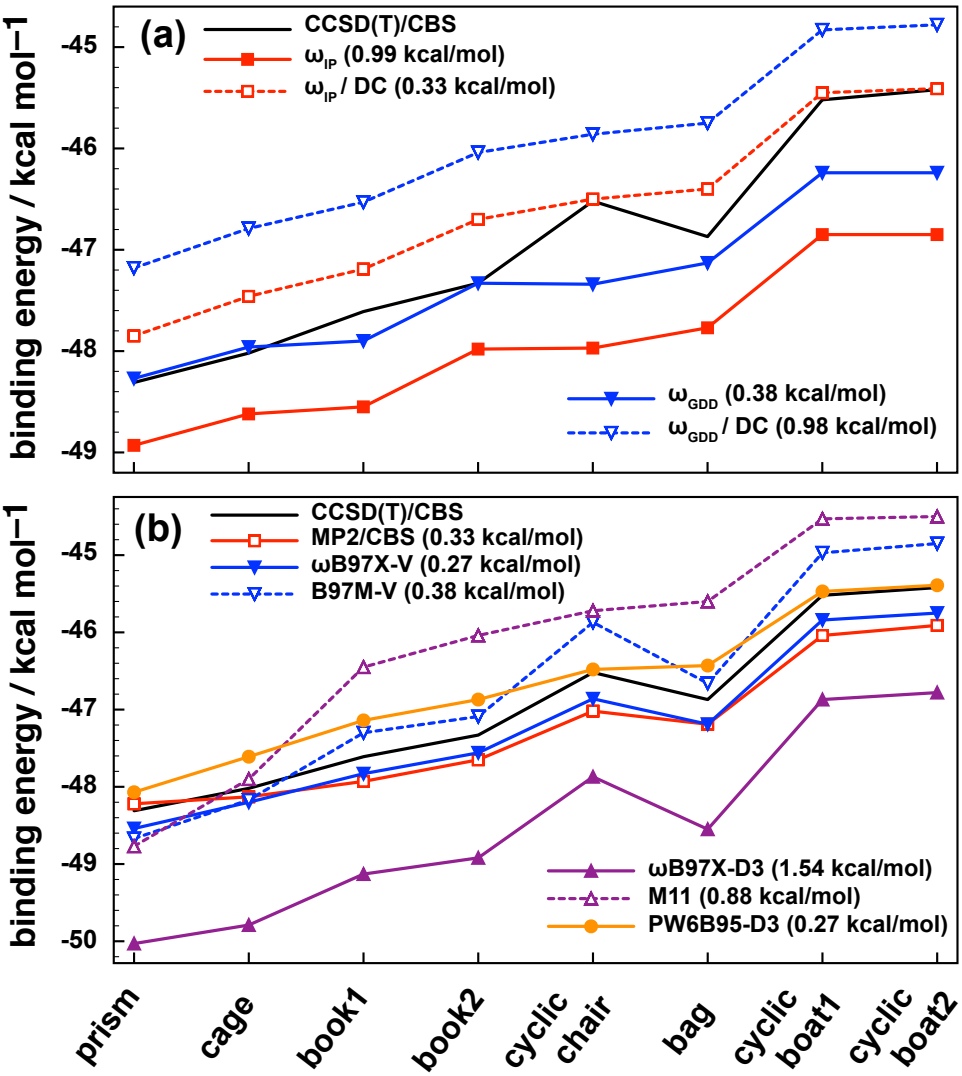

Figure 10.5: Binding energies for eight isomers of $(H_2O)_6$ using (a) XSAPT(KS)+D3 and (b) supersystem methods. The hpTZVPP basis set is used for XSAPT(KS)+D3 and DC calculations, and the def2-TZVPPD basis set is used for DFT calculations with CP corrections. MAEs for the whole data set, with respect to the CCSD(T)/CBS benchmarks, are also listed for each method.

+ D3 shows MAEs of 0.99 and 0.38 kcal/mol by using ω_{IP} and ω_{GDD} tunings, respectively. The corresponding MAEs move to 0.33 and 0.98 kcal/mol by including DC corrections. Using the DC-part correction with 6-31+G(d) basis set, XSAPT(KS)+D3+DC slightly degrades to 0.45 and 1.10 kcal/mol in ω_{IP} and ω_{GDD} tunings, respectively.

For the relative energies calculated by CCSD(T)/CBS, isomer bag is 0.35 kcal/mol more stable than isomer cyclic chair, and isomer cyclic boat1 is 0.10 kcal/mol more stable than isomer cyclic boat2. XSAPT(KS)+D3 (ω_{IP}/ω_{GDD}) gives the reverse prediction that isomer cyclic chair is 0.20 kcal/mol more stable than isomer bag, and isomer cyclic boat1 and isomer cyclic boat2 have the same binding energies. XSAPT(KS)+D3+DC (ω_{IP}/ω_{GDD}) still indicates that isomer cyclic chair is more stable than isomer bag by 0.10 kcal/mol, but isomer cyclic boat1 is 0.04 kcal/mol more stable than isomer cyclic boat2 where it matches the binding order computed by CCSD(T)/CBS.

Those supersystem methods shown in Fig. 10.5 give MAEs < 1 kcal/mol except ωB97X-D3 where its MAE is 1.54 kcal/mol. ωB97X-V and PW6B95-D3 are the two best methods in $(H_2O)_6$ binding predictions, and their binding MAEs are only 0.27 kcal/mol. However, PW6B95-D3 predicts that isomer cyclic chair is 0.05 kcal/mol more stable than isomer bag where the binding energy order is reverse as compared to CCSD(T)/CBS results. M11 also predicts the wrong binding energy order between isomers cyclic chair and bag, furthermore, isomers prism and cage are too stable in M11 prediction. MP2/CBS, ωB97X-V, B97M-V, and ωB97X-D3 give the correct

Table 10.1: MAEs in binding energies and relative energies for ten low-energy isomers of $(H_2O)_{20}$ with respect to CCSD(T)/CBS benchmarks. The hpTZVPP basis set is used for XSAPT(KS)+D3 and DC calculations, and the def2-TZVPPD basis set is used for DFT calculations with CP corrections. The relative energy for each isomer is relative to the lowest energy isomer **1**.

Method	MAE / kcal mol^{-1}	
	binding energy	relative energy
—supersystem methods—		
MP2/CBS	2.68	0.19
ωB97X-V	1.68	0.07
B97M-V	2.86	0.33
ωB97X-D3	6.12	0.33
M11	10.20	1.06
PW6B95-D3	6.31	0.29
—XSAPT methods—		
XSAPT(KS)+D3 (ω_{IP})	6.42	1.02
XSAPT(KS)+D3+DC (ω_{IP})	1.21	0.87
XSAPT(KS)+D3 (ω_{GDD})	3.48	1.02
XSAPT(KS)+D3+DC (ω_{GDD})	1.88	0.88

binding order for all $(H_2O)_6$ isomers. ωB97X-V is the best method to accurately predict both absolute and relative binding energies.

Ten low-energy isomers of $(H_2O)_{20}$,[18,339] obtained using the TIP4P force field,[428] are used here to see whether errors increases with system size. Results for total binding energies are shown in Fig. 10.6, with error statistics listed in Table 10.1. All supersystem methods underbind all isomers of $(H_2O)_{20}$ except ωB97X-D3 which largely overbinds all isomers. MP2/CBS, ωB97X-V, and B97M-V methods perform reasonable well with MAEs of 2.68, 1.68, and 2.86 kcal/mol, respectively. ωB97X-D3, M11, and PW6B95-D3 methods perform poor in large $(H_2O)_{20}$ clusters with MAEs of 6.12,

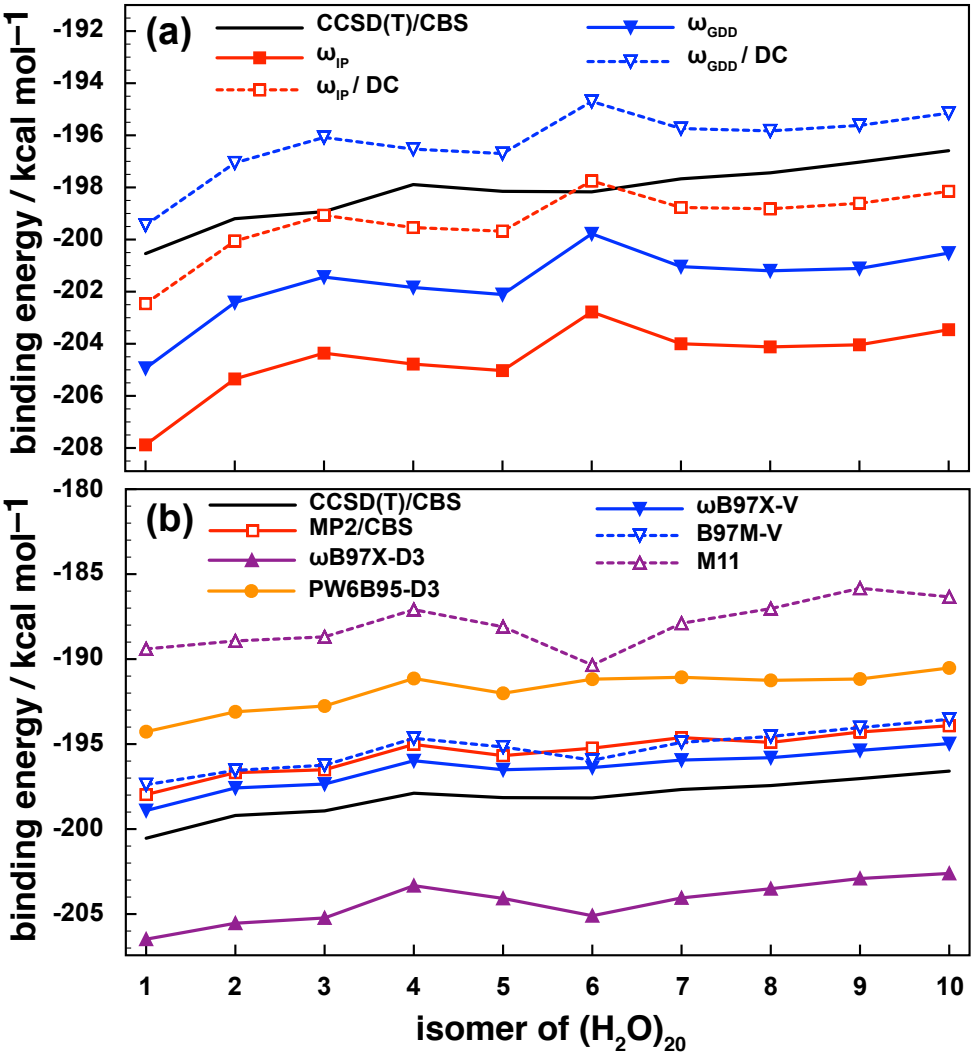

Figure 10.6: Binding energies for ten low-energy isomers of $(H_2O)_{20}$ using (a) XSAPT(KS)+D3 and (b) supersystem methods. The hpTZVPP basis set is used for XSAPT(KS)+D3 and DC calculations, and the def2-TZVPPD basis set is used for DFT calculations with CP corrections. MAEs for the whole data set, with respect to the CCSD(T)/CBS benchmarks, are also listed for each method. The isomers are number as in Ref. 18.

Table 10.2: The contributions of MAEs with respect to CCSD(T)/CBS average binding energies in $(H_2O)_6$ and $(H_2O)_{20}$ clusters. The hpTZVPP basis set is used for XSAPT(KS)+D3 and DC calculations, and the def2-TZVPPD basis set is used for DFT calculations with CP corrections.

	percentage error	
Method	$(H_2O)_6$	$(H_2O)_{20}$
—supersystem methods—		
MP2/CBS	0.7%	1.4%
ωB97X-V	0.6%	0.8%
B97M-V	0.8%	1.4%
ωB97X-D3	3.3%	3.1%
M11	1.9%	5.1%
PW6B95-D3	0.6%	3.2%
—XSAPT methods—		
XSAPT(KS)+D3 (ω_{IP})	2.1%	3.2%
XSAPT(KS)+D3+DC (ω_{IP})	0.7%	0.6%
XSAPT(KS)+D3 (ω_{GDD})	0.8%	1.8%
XSAPT(KS)+D3+DC (ω_{GDD})	2.1%	0.9%

10.20, and 6.31 kcal/mol, respectively. The performance of PW6B95-D3 is very good in $(H_2O)_6$ with a MAE of 0.27 kcal/mol, but decays very fast in $(H_2O)_{20}$ with a MAE of 6.31 kcal/mol. As in $(H_2O)_6$, XSAPT(KS)+D3 (ω_{GDD}) performs better in $(H_2O)_{20}$ than XSAPT(KS)+D3 (ω_{IP}). The XSAPT(KS)+D3 (ω_{IP}/ω_{GDD}) errors can be significantly reduced by using the DC correction. XSAPT(KS)+D3+DC (ω_{IP}) and XSAPT(KS)+D3+DC (ω_{GDD}) give MAEs of 1.21 and 1.88 kcal/mol, respectively. Using the DC-part correction with 6-31+G(d) basis set, the MAE of XSAPT(KS)+D3+DC (ω_{GDD}) slightly increases to 2.23, but the MAE of XSAPT(KS)+D3+DC (ω_{IP}) slightly decreases to 0.93 kcal/mol.

The average binding energies at the level of CCSD(T)/CBS are -46.95 and -198.16 for $(H_2O)_6$ and $(H_2O)_{20}$, respectively. The contributions of MAEs with respect to CCSD(T)/CBS average binding energies in $(H_2O)_6$ and $(H_2O)_{20}$ clusters are shown in Table 10.2. The binding percentage errors from $(H_2O)_6$ to $(H_2O)_{20}$ clusters are increased at the level of MP2/CBS, B97M-V, M11, PW6B95-D3, XSAPT(KS)+D3 (ω_{IP}), and XSAPT(KS)+D3 (ω_{GDD}), but are decreased at the level of XSAPT(KS)+D3+DC (ω_{GDD}). ωB97X-V, ωB97X-D3, and XSAPT(KS)+D3+DC (ω_{IP}) show similar binding percentage errors in both $(H_2O)_6$ and $(H_2O)_{20}$ clusters. M11 and PW6B95-D3 methods are not recommended to use in water clusters since their binding errors increase about three and five times, respectively, from $(H_2O)_6$ to $(H_2O)_{20}$ clusters. ωB97X-V and XSAPT(KS)+D3+DC (ω_{IP}) are the only two methods which show small binding percentage errors ($< 1\%$) in both $(H_2O)_6$ to $(H_2O)_{20}$ clusters.

All supersystem methods afford very good relative energies except M11. The relative energy for each isomer is relative to the lowest energy isomer **1**. The relative energies of XSAPT(KS)+D3 ($\omega_{\mathrm{IP}}/\omega_{\mathrm{GDD}}$) are somewhat worse. MAEs of relative energies for M11 and XSAPT(KS)+D3 ($\omega_{\mathrm{IP}}/\omega_{\mathrm{GDD}}$) are about 1 kcal/mol. Including the DC corrections to XSAPT(KS)+D3 slightly decreases the MAE to about 0.9 kcal/mol although the absolute binding energies of XSAPT are significantly improved by including the DC corrections. In Fig 10.6(a), isomer **6** has the largest relative error in XSAPT(KS)+D3 ($\omega_{\mathrm{IP}}/\omega_{\mathrm{GDD}}$) (about 2.7 kcal/mol), and can be slightly reduced to about 2.3 kcal/mol by employing DC corrections. Isomer **6** is the most highly coordinated of all $(H_2O)_{20}$ isomers considered here, and it has been noted that isomer **6** also has the largest relative error for some DFT methods as indicated by Wang *et al.*[18] Such errors may come form the D3 dispersion potential, and CKS dispersion for XSAPT may correct this problem. This work is currently underway in our group.

Finally, large anionic clusters $F^-(H_2O)_{10}$ with ten isomers are considered here. Results for total binding energies are shown in Fig. 10.7, with error statistics listed in Table 10.3. Among supersystem methods, ωB97X-V and B97M-V perform the best with MAEs of 0.24 and 0.15 kcal/mol, respectively. MAEs of the rest of supersystem methods are in the range of 1 to 5 kcal/mol. The relative energies of supersystem methods are quite accurate with MAEs of 0.3 kcal/mol except ωB97X-D3 and M11 which show MAEs of 0.46 and 0.77 kcal/mol, respectively. The relative energy for each isomer is relative to the lowest energy isomer **6**. Both MAEs for

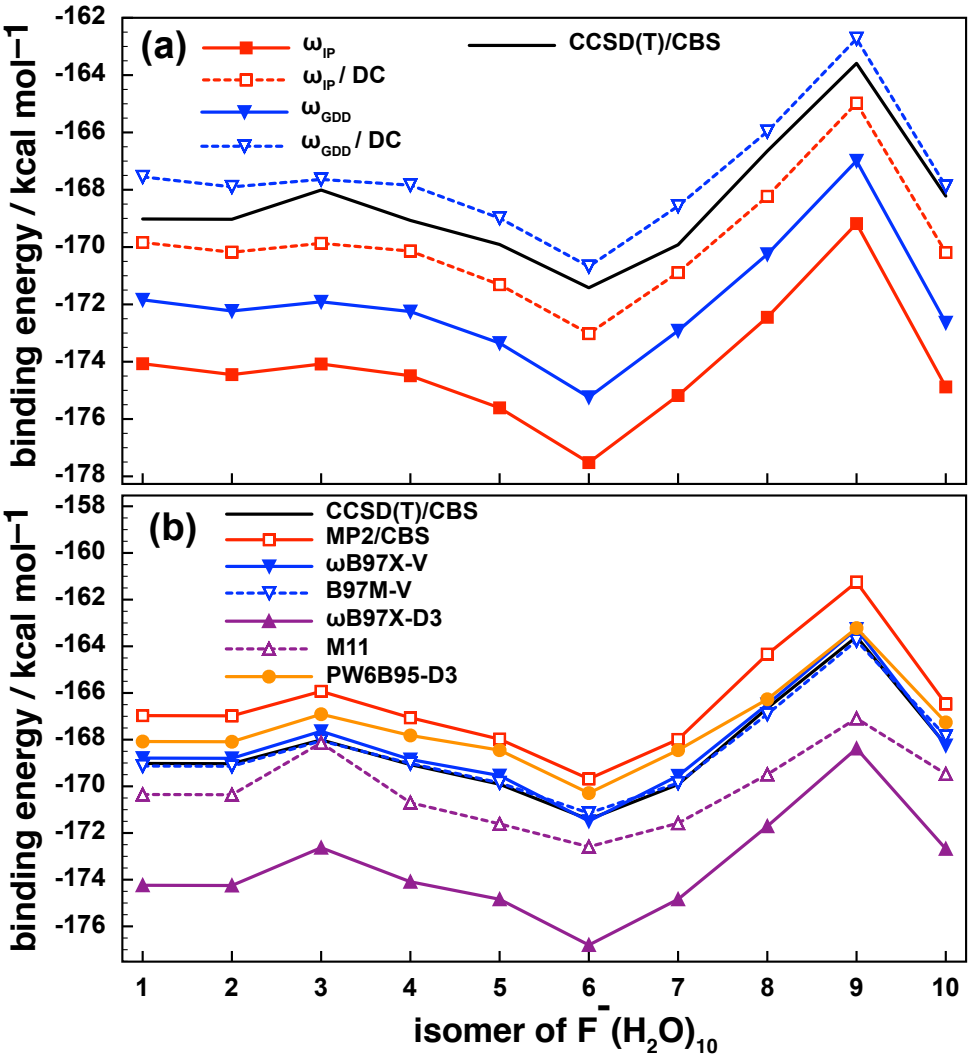

Figure 10.7: Binding energies for ten isomers of $F^-(H_2O)_{10}$ using (a) XSAPT(KS)+D3 and (b) supersystem methods. The hpTZVPP basis set is used for XSAPT(KS)+D3 and DC calculations, and the def2-TZVPPD basis set is used for DFT calculations with CP corrections. MAEs for the whole data set, with respect to the CCSD(T)/CBS benchmarks, are also listed for each method.

Table 10.3: MAEs in binding energies and relative energies for ten isomers of $F^-(H_2O)_{10}$ with respect to CCSD(T)/CBS benchmarks. The hpTZVPP basis set is used for XSAPT(KS)+D3 and DC calculations, and the def2-TZVPPD basis set is used for DFT calculations with CP corrections. The relative energy for each isomer is relative to the lowest energy isomer **6**.

Method	MAE / kcal mol^{-1}	
	binding energy	relative energy
—supersystem methods—		
MP2/CBS	2.02	0.31
ωB97X-V	0.24	0.32
B97M-V	0.15	0.31
ωB97X-D3	4.96	0.46
M11	1.65	0.77
PW6B95-D3	1.00	0.32
—XSAPT methods—		
XSAPT(KS)+D3 (ω_{IP})	5.71	0.56
XSAPT(KS)+D3+DC (ω_{IP})	1.38	0.38
XSAPT(KS)+D3 (ω_{GDD})	3.48	0.53
XSAPT(KS)+D3+DC (ω_{GDD})	0.91	0.37

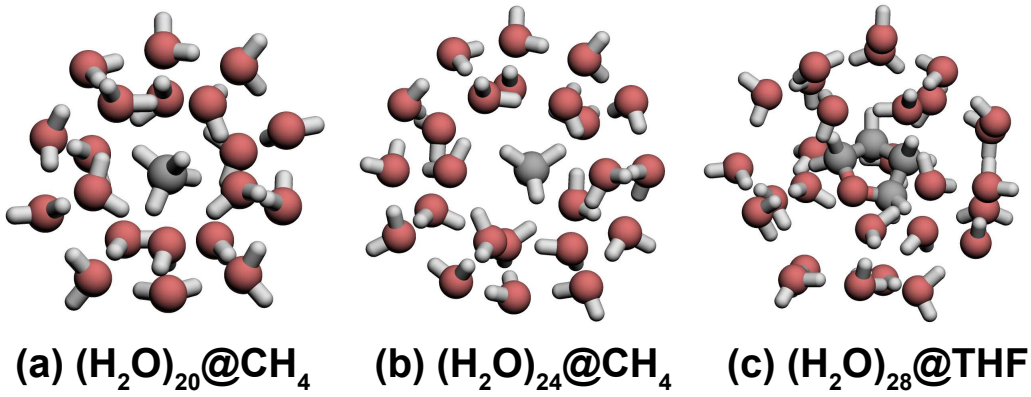

Figure 10.8: Lowest-energy structures for three classes of building blocks in clathrate hydrates with small guest molecules: (a) $(H_2O)_{20}$ with CH_4, (b) $(H_2O)_{24}$ with CH_4, and (c) $(H_2O)_{28}$ with THF. The initial geometries of $(H_2O)_{20}$, $(H_2O)_{24}$, and $(H_2O)_{28}$ are taken from Ref. 25. The host-guest complexes were subsequently optimized at the level of B97M-V/6-31+G(d) with freezing oxygen atoms on cages.

absolute and relative binding energies of XSAPT(KS)+D3 ($\omega_{\mathrm{IP}}/\omega_{\mathrm{GDD}}$) can be effectively reduced by using DC corrections. MAEs for absolute binding energies of XSAPT(KS)+D3+DC (ω_{IP}) and XSAPT(KS)+D3+DC (ω_{GDD}) are 1.38 and 0.91 kcal/mol, respectively, and the corresponding MAEs for relative binding energies are 0.38 and 0.37 kcal/mol. Using the DC-part correction with 6-31+G(d) basis set, the MAE of XSAPT(KS)+D3+DC (ω_{GDD}) slightly increases to 0.98, but the MAE of XSAPT(KS)+D3+DC (ω_{IP}) slightly decreases to 1.32 kcal/mol. Thus, DC correction is essential to include in XSAPT(KS)+D3 for improving both absolute and relative binding energies.

Table 10.4: The absolute binding energies, and their MAEs and MAPEs with respect to CCSD(T)/CBS benchmarks for $(H_2O)_{20}$@CH_4, $(H_2O)_{24}$@CH_4, and $(H_2O)_{28}$@THF. The hpTZVPP basis set is used for XSAPT(KS)+D2/D3 and DC calculations, and the def2-TZVPPD basis set is used for DFT calculations with CP corrections.

Method	binding energy / kcal mol^{-1}			MAE (kcal mol^{-1})	MAPE
	$(H_2O)_{20}$@CH_4	$(H_2O)_{24}$@CH_4	$(H_2O)_{28}$@THF		
—supersystem methods—					
CCSD(T)/CBS	−4.88	−4.32	−12.98	–	–
MP2/CBS	−5.01	−4.42	−13.32	0.19	2.60%
ωB97X-V	−6.47	−5.69	−15.92	1.97	29.02%
B97M-V	−5.78	−4.77	−14.39	0.92	13.29%
ωB97X-D3	−7.68	−6.43	−16.51	2.82	44.55%
M11	−5.97	−3.23	−9.61	1.85	24.53%
PW6B95-D3	−6.35	−5.65	−14.87	1.57	25.24%
—DFT recommended in previous studies—					
revPBE	6.60	3.50	6.10	12.79	187.82%
B3LYP	3.35	2.11	1.73	9.79	143.67%
B3LYP-D	−5.90	−4.32	−14.10	0.72	9.91%
PBE-D	−5.73	−4.80	−13.95	0.77	12.01%
B97-D	−4.61	−4.35	−12.99	0.10	2.01%
BLYP-D3	−7.29	−5.35	−15.98	2.15	32.22%
M05-2X	−5.86	−4.83	−12.56	0.64	11.72%
M06-2X	−6.27	−4.10	−11.24	1.12	15.69%
ωB97X-D	−6.52	−5.10	−14.55	1.33	21.32%
—XSAPT(KS)+D2 methods—					
XSAPT(KS)+D2 (ω_{IP})	−5.14	−4.85	−13.85	0.55	8.11%
XSAPT(KS)+D2+DC (ω_{IP})	−4.91	−4.68	−13.38	0.26	3.99%
XSAPT(KS)+D2 (ω_{GDD})	−5.04	−4.80	−13.79	0.49	6.97%
XSAPT(KS)+D2+DC (ω_{GDD})	−4.81	−4.63	−13.31	0.24	3.72%
—XSAPT(KS)+D3 methods—					
XSAPT(KS)+D3 (ω_{IP})	−3.53	−3.78	−12.53	0.78	14.55%
XSAPT(KS)+D3+DC (ω_{IP})	−3.30	−3.60	−12.06	1.07	18.67%
XSAPT(KS)+D3 (ω_{GDD})	−3.44	−3.73	−12.48	0.84	15.69%
XSAPT(KS)+D3+DC (ω_{GDD})	−3.20	−3.56	−12.00	1.14	19.85%

10.3.5 Small guest molecules in clathrate hydrates

Clathrate hydrate can be used to trap gas (guest) molecules by its hydrogen-bonded polyhedral cage-like water framework (host), and such host-guest complex is considered as a potential fuel storage material.[429–431] Two cubic structures, type I and type II, of clathrate hydrates are generally formed, and the hexagonal cubic structure as type H is rarely formed. In this work, only water cages in type I and type II are considered. Type I is formed by two kinds of water cages which are $(H_2O)_{20}$ (12 pentagonal faces; $\mathbf{5^{12}}$) and $(H_2O)_{24}$ (12 pentagonal and 2 hexagonal faces; $\mathbf{5^{12}6^2}$). Type II is also formed by two kinds of water cages which are $\mathbf{5^{12}}$ and $(H_2O)_{28}$ (12 pentagonal and 4 hexagonal faces; $\mathbf{5^{12}6^4}$).[430,432] The type of guest molecules inside the host cages of clathrate hydrates depends upon the cavity sizes and shapes. Furthermore, the lattice of host cage in clathrate hydrates can be maintained by the interactions between the host cages and guest molecules under suitable temperature-pressure conditions.[432]

There are lots of theoretical studies related to the binding interactions between host water cages and guest molecules. revPBE,[433] B3LYP, B3LYP-D, M05-2X,[434] B97-D,[254] ωB97X-D,[201] and vdW-DF[144] functionals were used without benchmarking the binding interactions with respect to accurate high-level of methods at CBS limit.[343,435–446] HF and MP2 with small basis sets were also favorable in those calculations.[443,447–450] The intermolecular interactions between $\mathbf{5^{12}}$ and methane have been benchmarked for various DFT methods with respect to MP2 using only double-ζ to triple-ζ basis set extrapolation scheme,[342,344] which is known to be inadequate.[451] Among various DFT methods, B97-D,[254] ωB97X-D,[201] M06-2X,[434] and PBE-D were

recommended to be suitable methods for interactions between $\mathbf{5^{12}}$ and methane,[344,452] and they subsequently used to study interactions for different combinations of water cages and guest molecules.[452–455] BLYP-D3 was also suggested to use in cage-guest interaction in clathrate hydrates by studying a variety of guest species in $\mathbf{5^{12}}$ with respect to only CCSD/CBS benchmark.[456] To the best of our knowledge, the accurate high-level benchmark based on quantum Monte Carlo (QMC) is only available for intermolecular interaction between $\mathbf{5^{12}}$ and methane.[341] Thus, it is essential to have accurate high-level benchmarks for intermolecular interactions between various sizes of cages and guest molecules.

The water cages $\mathbf{5^{12}}$ and $\mathbf{5^{12}6^2}$ are used here as methane hosts in various MP2, DFT, and XSAPT computations with respect to CCSD(T)/CBS benchmark. The hydrogen storage can be achieved under industrially attainable conditions by encapsulating THF in $\mathbf{5^{12}6^4}$ to stabilize the type II clathrate hydrate.[457] The water cage $\mathbf{5^{12}6^4}$ is used here as THF host to study the interactions between them. The geometries of $(H_2O)_{20}$@CH_4, $(H_2O)_{24}$@CH_4, and $(H_2O)_{28}$@THF optimized at the level of B97M-V/6-31+G(d) are shown in Fig. 10.8. Results for total binding energies using various methods with error statistics are listed in Table 10.4.

MP2/CBS gives binding energies of −5.01, −4.42, and −13.32 kcal/mol, respectively, for $(H_2O)_{20}$@CH_4, $(H_2O)_{24}$@CH_4, and $(H_2O)_{28}$@THF. The $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ correction is equal to the binding energy difference of DLPNO-CCSD(T)/def2-TZVP and MP2/def2-TZVP methods, and the corresponding values are 0.14, 0.10, and 0.34 kcal/mol. The repulsive correction of $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ makes CCSD(T)/CBS have slightly

smaller binding energy than MP2/CBS in all three cases. Although the geometries of $(H_2O)_{20}@CH_4$ are different in Ref. 341 and this work, we still compare the high-level binding results used in the two studies here. In Ref. 341, MP2 binding energy at CBS limit was obtained by extrapolation energies from aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ basis sets to CBS limit, and results a binding energy of -5.04 kcal/mol. MP2/CBS binding energies in both works are almost the same. The $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ correction in Ref. 341 was estimated by the two- and three-body binding energy difference of the CCSD(T)-F12b[193] and MP2-F12[458] levels with the VTZ-F12[459] basis set. The corresponding $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ corrections truncated at the two- and three-body of many-body expansion (MBE) were -0.90 and -0.14 kcal/mol, respectively. The $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ corrections are attractive in Ref. 341 based on MBE and repulsive in this work based on full calculations. The sign difference in both $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ corrections means that two- and three-body truncation are not enough and higher-body contributions of $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ in Ref. 341 are important. The QMC calculations in Ref. 341 gave a binding energy of -5.3 ± 0.5 kcal/mol for $(H_2O)_{20}@CH_4$. Our corresponding CCSD(T)/CBS result is -4.88 kcal/mol which is within statistical uncertainty of QMC.[341] Thus, the CCSD(T)/CBS binding energies in this work for the three cage-guest complexes are expected to give reliable binding benchmarks. To the best of our knowledge, it is the first time to have supersystem CCSD(T)/CBS benchmarks for bindings between various sizes of water cages and guest molecules.

The small $\delta_{\mathrm{MP2}}^{\mathrm{CCSD(T)}}$ corrections make MP2/CBS as a reliable method to study interactions between water cages and guest molecules. The MAE and MAPE for

Table 10.5: MAEs and MAPEs with respect to CCSD(T)/CBS benchmarks using B97-D combined with 6-311++G(2d,2p) and def2-TZVPPD basis sets for $(H_2O)_{20}$@CH_4, $(H_2O)_{24}$@CH_4, and $(H_2O)_{28}$@THF. The calculations with and without CP corrections are included in each level of theory.

basis set	CP	MAE/kcal mol^{-1}	MAPE
6-311++G(2d,2p)	NO	1.69	21.66%
6-311++G(2d,2p)	YES	0.17	3.29%
def2-TZVPPD	NO	0.42	5.15%
def2-TZVPPD	YES	0.10	2.01%

MP2/CBS are 0.19 kcal/mol and 2.60%, respectively. The revPBE and B3LYP predict unbound in all three complexes, and it is not suitable to use them in modeling of clathrate hydrates. Including the add-on dispersion correction[254] greatly improves the binding results. The MAEs for B3LYP-D and PBE-D are 0.72 and 0.77 kcal/mol, respectively. The functionals with Grimme's D3 dispersion correction,[274] BLYP-D3, ωB97X-D3, and PW6B95-D3, give slightly large MAEs of 2.15, 2.82, and 1.57 kcal/mol, respectively. The three functionals, ωB97X-D, M06-2X, and B97X-D, recommended[344] in binding between $(H_2O)_{20}$ and CH_4, show MAEs of 1.33, 1.12, and 0.10 kcal/mol, respectively. Only B97-D functional is reliable in binding between water cages and guest molecules with MAPE of only 2%. However, B97-D functional in Ref. 344 was recommended to use with 6-311++G(2d,2p) basis set without CP correction. The results of B97-D with 6-311++G(2d,2p) and def2-TZVPPD basis sets are shown in Table 10.5 including both with and without CP corrections. B97-D/6-311++G(2d,2p) without CP correction gives large MAE and MAPE of 1.69 kcal/mol and 21.66%, respectively. B97-D combined with Ahlrichs's def2-TZVPPD

gives small MAE and MAPE of 0.42 kcal/mol and 5.15%, respectively, even without CP correction. Errors can be largely reduced by including CP corrections in both methods, especially in 6-311++G(2d,2p) basis set, and def2-TZVPPD basis set still works better with B97-D than 6-311++G(2d,2p) basis set. Thus, the CP correction is highly suggested to use with B97-D to remove BSSE.

It is interesting that the MAE is larger in the newer generation of Minnesota-class functional. Among M05-2X, M06-2X and M11, M05-2X works the best with a MAE of 0.64 kcal/mol, and M11 works the worst with a MAE of 1.85 kcal/mol. ωB97X-D and ωB97X-D3 have similar trends in which the old generation of ωB97X-D works better than the new generation of ωB97X-D3. The two non-local dispersion functionals, ωB97X-V and B97M-V, does not work better than some of the old generation dispersion-corrected functionals, such as M05-2X, B97X-D, and B3LYP-D, and the corresponding MAEs are 1.97 and 0.92 kcal/mol, respectively. Although ωB97X-V works well in many different system in the previous sections, it does show large errors in some systems, such as the binding interactions between water cages and guest molecules which provide target systems for future functional development.

XSAPT(KS)+D3 ($\omega_{\mathrm{IP}}/\omega_{\mathrm{GDD}}$) gives a MAE of 0.8 kcal/mol. Including the DC correction slightly degrades the performance with a MAE of 1.1 kcal/mol. On the other hand, XSAPT(KS)+D2 ($\omega_{\mathrm{IP}}/\omega_{\mathrm{GDD}}$) works surprisingly well with a MAE of 0.5 kcal/mol. Including the DC correction further improves the performance with a MAE of 0.25 kcal/mol and a MAPE of 4%. Using the DC-part correction with 6-31+G(d) basis set, the MAE of XSAPT(KS)+D2+DC slightly increases to 0.46 and

Table 10.6: Individual energy components of $(H_2O)_{20}$@CH_4, $(H_2O)_{24}$@CH_4, and $(H_2O)_{28}$@THF complexes using XSAPT(KS)+D2+DC (ω_{IP})/hpTZVPP.

XSAPT	energy component / kcal mol^{-1}		
	$(H_2O)_{20}$ @CH_4	$(H_2O)_{24}$ @CH_4	$(H_2O)_{28}$ @THF
Electrostatics	−2.87	−1.13	−5.65
Exchange	8.92	3.69	11.52
Induction	−1.09	−0.37	−2.04
Dispersion	−9.86	−6.87	−17.20
Total	−4.91	−4.68	−13.38

0.39 kcal/mol, respectively, by using ω_{IP} and ω_{GDD} tunings.

All supersystem methods and XSAPT(KS)+D2 in Table 10.4 show the encapsulation of methane in the $(H_2O)_{20}$ cage to be more favorable than in the $(H_2O)_{24}$ cage. The trend has been shown before by using M05-2X and B97X-D.[343] The reverse trend shown by XSAPT(KS)+D3 makes it unsuitable to use in clathrate hydrate systems. Due to the small MAE and MAPE, XSAPT(KS)+D2+DC (ω_{IP}/ω_{GDD}), and supersystem B97-D with CP correction are recommended to be use in clathrate hydrate systems. The computational cost of XSAPT is much lower than those supersystem methods since the maximal sizes of calculations in XSAPT are only for two-body H_2O-H_2O or H_2O-guest systems.

The XSAPT(KS)+D2+DC (ω_{IP}) analysis[56] of the contributions to the binding energy for the three cage-guest systems is reported in Table 10.6. As compared

the binding components in $(H_2O)_{20}$@CH_4 and $(H_2O)_{24}$@CH_4, electrostatics and induction are similar in both systems. CH_4 is at the center of $(H_2O)_{20}$ and is inclined to one side of $(H_2O)_{24}$; therefore, CH_4 has larger exchange and dispersion interactions with $(H_2O)_{20}$ than with $(H_2O)_{24}$. The net binding energy is larger in $(H_2O)_{20}$@CH_4 than in $(H_2O)_{24}$@CH_4. The total binding energy between $(H_2O)_{28}$ and THF is larger than binding interactions between CH_4 and the two water cages since the host-guest interaction is stronger for polar guest molecules than for non-polar molecules. $(H_2O)_{28}$@THF has larger energy components in all species as compared to CH_4 in $(H_2O)_{20}$ and $(H_2O)_{24}$. The larger electrostatics and induction may come from the hydrogen-bonded interaction between THF and H_2O. Due to the large size of THF, it has large exchange and dispersion interactions with $(H_2O)_{28}$. The net total binding energy is -13.38 kcal/mol in $(H_2O)_{28}$@THF, and is much larger than CH_4 in $(H_2O)_{20}$ and $(H_2O)_{24}$. Such large binding energy between THF and $(H_2O)_{28}$ has been used to stabilize the type II clathrate hydrate.[457] It is expected that different kinds of auxiliary guest molecules for stabilizing water cages can be found in the future through theoretical design to achieve gas storage in clathrate hydrates under industrially attainable conditions.

10.4 Conclusions

A dispersion contraction (DC) has been introduced to include non-additive dispersion effects in XSAPT(KS)+D where the dispersion potentials are environment-independent and the squeezing effect by neighboring atoms is not captured. The

DC effect can be obtained by the squeezing effects from the neighboring charges, and is calculated by dispersion energy difference based on XSAPT (with charge embedding) and SAPT (without charge embedding). However, the second-order dispersion and exchange-dispersion terms in DC correction scales as $\mathcal{O}(N_f^4)$ and $\mathcal{O}(N_f^5)$, respectively. The full DC scheme is denoted as DC-full with $\mathcal{O}(N_f^5)$ scaling, and the aug-cc-pVDZ basis set is accurate enough to reproduce DC-full at a triple-ζ basis set. If only second-order dispersion is considered in DC correction, such scheme is denoted as DC-part with $\mathcal{O}(N_f^4)$ scaling. The 6-31+G(d) basis set is accurate enough for DC-part to capture the DC effect based on fortuitous error cancellation. The scaling of XSAPT(KS)+D+DC method is $\mathcal{O}(N_f^3)$ using a triple-ζ basis set for the non-dispersion terms, and $\mathcal{O}(N_f^4)$ using a double-ζ basis set for the DC corrections. The DC correction is especially important in clusters with polar molecules, for example, the DC contributions are 3, 4, and 5 kcal/mol in $SO_4^{2-}(H_2O)_6$, $F^-(H_2O)_{10}$, and $(H_2O)_{20}$ clusters, respectively. Thus, it is important to include DC correction in XSAPT(KS)+D3 to achieve chemical accuracy (energies within about 1 kcal/mol). For example, XSAPT(KS)+D3+DC with ω_{IP} tuning gives MAEs of 1.1, 0.3, 1.2, 1.4 kcal/mol for ionic $X/M(H_2O)_{n=1-6}$ ($X = F^-$, Cl^-, and SO_4^{2-} and $M = Li^+$, Na^+, and K^+), $(H_2O)_6$, $(H_2O)_{20}$, and $F^-(H_2O)_{10}$ clusters, respectively. A series of DFT methods recommended for non-covalent interactions have also been tested in this work. ωB97X-V is the only DFT method recommended to be used, and its corresponding MAEs are 0.4, 0.3, 1.7, and 0.2 kcal/mol. However, XSAPT(KS)+D3+DC with ω_{IP}

tuning and ωB97X-V do not work well in clathrate hydrates, and the corresponding MAEs are 1.1 and 2.0 kcal/mol, respectively, as compared to CCSD(T)/CBS benchmarks where it is the first time to have CCSD(T)/CBS benchmarks for the interactions between various sizes of water cages and guest molecules. On the other hand, XSAPT(KS)+D2+DC with ω_{IP} tuning and B97-D work surprisingly well in clathrate hydrates with MAEs of 0.3 and 0.1 kcal/mol, respectively, therefore, it is recommended to use both methods in clathrate hydrates. One should note that the DC definition used here does not include any non-additive effects in noble gas clusters (one noble gas atom as a monomer) since each noble gas atom has zero charge.

The energy components and total binding energies for XSAPT(KS) based on a series of PBE-based functionals have also been tested here with respect to SAPT2+(3)/aTZ and CCSD(T)/CBS benchmarks, respectively. For S22 and S66, all tested functionals perform similar in electrostatics and induction, and the tuned LRC-ωPBE performs much better in exchange than PBE, asymptotically-corrected LFAs-PBE, and untuned LRC-ωPBE. As compared to CCSD(T)/CBS total binding energies, XSAPT(KS)+D3 with tuned LRC-ωPBE also performs better than the rest of functionals, and is comparable with SAPT2+(3)/aTZ. XSAPT(KS)+D3 with tuned LRC-ωPBE provides both accurate energy components and total binding energies. SAPT0/jun-cc-pVDZ provides large errors for all energy components, but it gives reasonable total binding due to fortuitous error cancellation.

CHAPTER 11

Energy decomposition analysis with a well-defined charge-transfer term for interpreting intermolecular interactions

11.1 Introduction

As compared to closed-shell complexes, non-covalent interactions are less well understood in molecule–radical and radical–radical complexes, and in open-shell cases one might expect charge transfer (CT) effects to be significant. Open-shell hydrogen bonds in radical complexes regulate electron transfer processes in many enzymes,[460] wherein the alf-filled orbital acts as a proton acceptor, forming a single-electron hydrogen bond.[461] Stabilization of this bond is thought to depend significantly upon electron transfer from the half-filled orbital into the σ^* orbital of the H-bond donor,[462] and the strength of the single-electron H-bond can be controlled via substituents added to the proton donor or acceptor. Protonation of the donor, *e.g.*, $H_3O^+ \cdots CH_3$ rather than $H_2O \cdots CH_3$, can significantly increase single-electron H-bond strength. This can be explained in terms of the reduced distance between proton donor and acceptor as well as a lowering of the σ^* energy, both of which make H-bonding CT interactions more favorable.[463] Variations in the strength of molecule–radical interactions can significantly influence both the rate and outcome of chemical reactions,[464]

e.g., by reducing the reactivity of the radical, resulting in high product selectivity,[465] for enantioselective organocatalysis.[466,467]

Despite the presumed importance of CT interactions in non-covalent complexes, direct experimental measurements of the magnitude of the CT contribution to the interaction energy are scarce. For closed-shell, binary gas-phase complexes involving H_2O, there are a few high-resolution molecular beam scattering experiments that, in conjunction with charge-displacement analysis and back-corrected intermolecular model potentials, have been used to obtain experimental values for the absolute magnitude of the CT interaction energy.[29,468,469] CT energies extracted from these experiments are approximately proportional to the amount of charge that is transferred, averaged over all orientations of the two monomers,[29,468,469] which is seen also in some theoretical analyses.[28,470] Stereospecific CT energies are very difficult to estimate from experiment,[29,468,469] thus a reliable CT interaction energy from theoretical calculations would be quite useful, but this has proven to be problematic.

Good electronic structure calculations can provide accurate values for the *total* non-covalent interaction energy between small- to medium-sized molecules, or between fragments of larger molecules.[57,206,272,471,472] What is often desired, however, is a partition of the interaction energy into physically meaningful components, such as electrostatic interactions, Pauli (or "exchange") repulsion, induction (also know as polarization), dispersion (the London or van der Waals interaction), and CT. Such a partition may aid in predicting how a certain chemical modification will alter the interaction energy, but unfortunately any such partition comes with some degree of

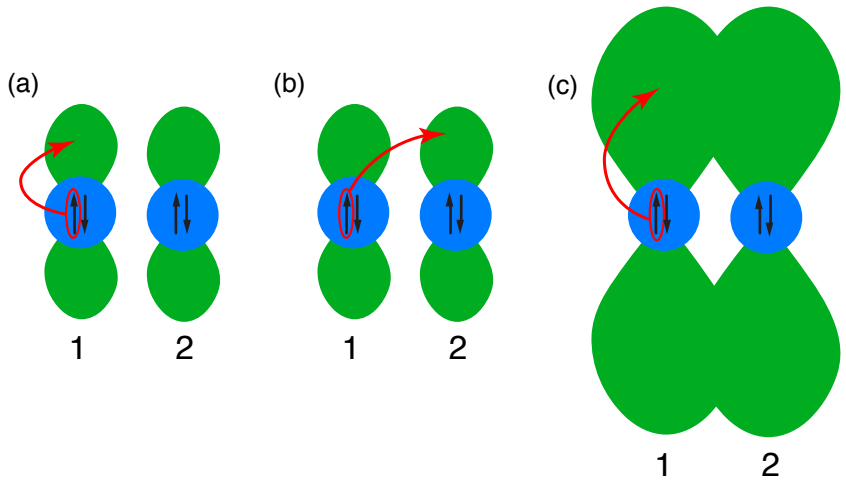

Figure 11.1: Schematic illustration of why the distinction between the CT and the induction energy in SAPT depends strongly on the basis set. In (a), the basis set is small and therefore most MOs are localized on one monomer or the other. Excitation from an occupied MO (in blue) centered on monomer 1 to a virtual MO (in green) creates or enhances a dipole moment whose formation is driven by its interaction with monomer 2; this is the epitome of induction. A CT interaction, in contrast, involves excitation from an occupied MO on monomer 1 into a virtual orbital associated with monomer 2, as in (b). In (c), however, the basis set has been enlarged such that the virtual orbitals on each monomer extend over the other. The distinction between the excitations in (a) and (b) thus becomes muddled, as does the distinction between induction and CT.

arbitrariness. This has given rise to quite a number of energy decomposition analysis (EDA) methods;[292,473–484] see Refs. 485 and 486 for recent reviews.

This work is focused specifically on intermolecular interactions, and for such cases an EDA that we find particularly attractive is symmetry-adapted perturbation theory (SAPT).[115,117,149,150] This approach computes the interaction energy ΔE_{AB} of the

A$\cdots$B dimer system directly, via perturbation theory, rather than by energy difference ($\Delta E_{\rm AB} = E_{\rm AB} - E_{\rm A} - E_{\rm B}$), and is therefore free of the artifactual basis-set superposition error (BSSE) that ordinarily leads to dramatic overestimation of $\Delta E_{\rm AB}$ except when very large basis sets are employed.[46,80] The structure of the perturbation series provides a nature decomposition of the interaction energy into electrostatic, exchange (or Pauli) repulsion, induction, and dispersion components, each with a well-defined basis-set limit.[117] Unfortunately, the CT contribution is not separated in this decomposition but rather is subsumed within the induction term, for reasons that are illustrated schematically in Fig. 11.1. When the basis set is small, there is a clear distinction between excitations (perturbations) that represent polarization originating in the presence of the other monomer, versus those that should be categorized as inter-monomer CT, but this distinction disappears as the unoccupied orbitals centered on one monomer begin to overlap significantly with those centered on the other monomer. For this reason, attempts to isolate the CT energy in SAPT tend to be strongly basis-set dependent,[117,487] and it can be argued that within this framework the CT energy should vanish in the limit of a complete basis set.[487] Attempts to define the CT energy with EDAs based on self-consistent field (SCF) calculations face similar difficulties,[478,488] since the MOs on one monomer will inevitably spread to the other as the quality of the basis set is improved.

For these reasons, there is need for a simple and reliable EDA for intermolecular interactions that includes a well-defined CT contribution, as an interpretative tool for non-covalent chemistry. Ideally, such an EDA should afford CT energies that

correlate with chemical intuition in obvious cases. For example, the CT energy should decrease as a function of intermolecular separation. As suggested and demonstrated in previous computational studies of alkyl and aryl radicals,[30] CT energies should correlate linearly with $(\mathrm{IE} - \mathrm{EA})^{-1}$, where IE is the ionization energy of the donor species and EA is the electron affinity of the acceptor. Finally, the CT energy should not depend strongly on the choice of basis set, at least for medium- and high-quality basis sets.

In this study, we compare the results from several popular EDA approaches that define an explicit CT interaction energy. These CT energies are compared to those obtained from constrained DFT (cDFT),[489] wherein one assigns a particular number of electrons to each monomer within a supermolecular (dimer) DFT calculation, then uses this reference state to quantify net electron flow upon assemblage of the complex from its constituent monomers. Very recently, this has been suggested as a means to define a CT contribution to the interaction energy that is only weakly dependent on the choice of basis set.[490] We will demonstrate that the combination of this cDFT approach with SAPT provides meaningful and robust definitions for all of the various components of the intermolecular interaction energy, and thus may be very useful for providing a theoretical basis for the "chemical intuition" of non-covalent interactions.

11.2 Theory

11.2.1 ALMO-EDA

In this work, we compare SAPT and cDFT to a popular SCF-based EDA, namely, the absolutely-localized MO (ALMO-EDA) procedure of Head-Gordon and co-workers,[478] which is equivalent to the block-localized wave function EDA of Gao and co-workers.[483] In this approach, the MOs on a given monomer are constrained to be formed from linear combinations of Gaussian basis functions centered on the same monomer, hence "absolutely localized". This offers certain advantages relative to some other EDAs, as it allows for a fully self-consistent and variational treatment of polarization and at the same time rigorously separates CT from polarization. (Despite this separation, basis-set dependence persists in the definition of the CT interaction energy.[478,488])

The total energy in ALMO-EDA is partitioned according to[478]

$$\Delta E^{\mathrm{ALMO}} = \Delta E^{\mathrm{ALMO}}_{\mathrm{FRZ}} + \Delta E^{\mathrm{ALMO}}_{\mathrm{POL}} + \Delta E^{\mathrm{ALMO}}_{\mathrm{CT}} \,, \tag{11.1}$$

where the components on the right are the "frozen density" energy, the polarization energy, and the CT energy, respectively. (For simplicity, the energy associated with geometric distortion of the monomers is omitted in this work.) The $\Delta E^{\mathrm{ALMO}}_{\mathrm{FRZ}}$ term includes electrostatics, dispersion, and Pauli repulsion. It is based on an antisymmetric product of monomer wave functions (the Heitler-London wave function) that is *not* variationally optimized, as that optimization would amount to polarization. The polarization term is computed as the energy difference between the frozen-density state and the variationally-optimized intermediate state (Ψ_{ALMO}), using the "SCF

for molecular interactions" (SCF-MI) procedure.[491] Finally. the CT term is defined as

$$\Delta E_{\mathrm{CT}}^{\mathrm{ALMO}} = E[\Psi] - E[\Psi_{\mathrm{ALMO}}] + \Delta E^{\mathrm{BSSE}} \tag{11.2}$$

where Ψ is the fully-optimized SCF wave function for the dimer and ΔE^{BSSE} is the counterpoise correction.[80]

11.2.2 SAPT

When Kohn-Sham (KS) orbitals are used in second-order SAPT,[51,378] the total intermolecular interaction energy can be written as

$$\begin{aligned} E_{\mathrm{int}}^{\mathrm{SAPT(KS)}} = {} & E_{\mathrm{elst}}^{(1)}(\mathrm{KS}) + \mathrm{E}_{\mathrm{exch}}^{(1)}(\mathrm{KS}) + \mathrm{E}_{\mathrm{ind,resp}}^{(2)}(\mathrm{KS}) \\ & + E_{\mathrm{exch\text{-}ind,resp}}^{(2)}(\mathrm{KS}) + \mathrm{E}_{\mathrm{disp}}^{(2)}(\mathrm{KS}) \\ & + E_{\mathrm{exch\text{-}disp}}^{(2)}(\mathrm{KS}) + \delta \mathrm{E}_{\mathrm{int}}^{\mathrm{HF}} \ . \end{aligned} \tag{11.3}$$

The superscripts indicate the order in perturbation theory, and the "KS" indicates that intramolecular correlation is incorporated in a low-cost way via DFT calculations for the monomers. The $\delta E_{\mathrm{int}}^{\mathrm{HF}}$ correction captures polarization effects beyond second order:

$$\begin{aligned} \delta E_{\mathrm{int}}^{\mathrm{HF}} = E_{\mathrm{int}}^{\mathrm{HF}} - \Big[& E_{\mathrm{elst}}^{(10)}(\mathrm{HF}) + \mathrm{E}_{\mathrm{exch}}^{(10)}(\mathrm{HF}) \\ & + E_{\mathrm{ind,resp}}^{(20)}(\mathrm{HF}) + \mathrm{E}_{\mathrm{exch\text{-}ind,resp}}^{(20)}(\mathrm{HF}) \Big] \ . \end{aligned} \tag{11.4}$$

Here, $E_{\mathrm{int}}^{\mathrm{HF}}$ is the counterpoise-corrected Hartree-Fock (HF) binding energy, and the remaining energy components in Eq. (11.4) represent second-order SAPT applied to monomers described at the HF level. If desired, response ("resp") corrections can be included in the induction and exchange-induction terms in Eqs. (11.3) and (11.4), in order to capture orbital relaxation effects,[179,180] at some additional cost. In this

study, the coupled induction and exchange-induction energies with orbital relaxation in SAPT(KS) are approximated as[51]

$$E^{(2)}_{\text{ind,resp}}(\text{KS}) = E^{(2)}_{\text{ind}}(\text{KS}) + \left[E^{(20)}_{\text{ind,resp}}(\text{HF}) - \text{E}^{(20)}_{\text{ind}}(\text{HF})\right] \tag{11.5}$$

and

$$E^{(2)}_{\text{exch-ind,resp}}(\text{KS}) = E^{(2)}_{\text{exch-ind}}(\text{KS}) + \left[E^{(20)}_{\text{exch-ind,resp}}(\text{HF}) - \text{E}^{(20)}_{\text{exch-ind}}(\text{HF})\right] . \tag{11.6}$$

Second-order dispersion is the most expensive term in second-order SAPT(KS), and is also the least accurate.[54,56] (Other terms scale no worse than cubically with respect to system size.[137,375,376]) As in previous work,[55–57] we will therefore replace the SAPT(KS) dispersion energy

$$E_{\text{disp}} = E^{(2)}_{\text{disp}}(\text{KS}) + E^{(2)}_{\text{exch-disp}}(\text{KS}) \tag{11.7}$$

with our third-generation, SAPT-based dispersion potential ("+D3").[57] This is an empirical atom–atom dispersion potential that is fit to reproduce high-level SAPT2+3 dispersion energies for a training set of dimers. (Note that the +D3 dispersion potential for SAPT is unrelated to Grimme's +D3 correction for DFT.[274]) Because the dispersion term is easily separable in SAPT, there is no double-counting problem in SAPT(KS)+D as there is in DFT+D, and the empirical dispersion potential in SAPT represents genuine dispersion interactions, in contrast to the empirical Grimme corrections.[253] This is an especially important point in the context of EDA, because it means that we can take the SAPT-based +D3 correction seriously as the dispersion contribution to the interaction energy.

It has been suggested that a CT energy can be extracted from the SAPT induction energy by taking the difference between SAPT calculations in which the monomer wave functions are computed using the full dimer-centered basis set (DCBS), versus only the monomer-centered basis set (MCBS):[117,487]

$$\begin{aligned}\Delta E_{\rm CT}^{\rm SAPT} = &\left[E_{\rm ind,resp}^{(2)}({\rm DCBS}) + E_{\rm exch\text{-}ind,resp}^{(2)}({\rm DCBS})\right] \\ &- \left[E_{\rm ind,resp}^{(2)}({\rm MCBS}) + E_{\rm exch\text{-}ind,resp}^{(2)}({\rm MCBS})\right] .\end{aligned} \tag{11.8}$$

In principle, however, $\Delta E_{\rm CT}^{\rm SAPT} \rightarrow 0$ in the limit of a complete basis set.[492]

Recently, Řezáč *et al.*[490] have shown that the higher-order induction terms that are included via the $\delta E_{\rm int}^{\rm HF}$ correction in Eq. (11.4) are dominated by CT, and these authors suggest an alternative "SAPT+δSCF" procedure to define CT:

$$\Delta E_{\rm CT}^{\rm SAPT+\delta SCF} = \Delta E_{\rm CT}^{\rm SAPT} + \delta E_{\rm int}^{\rm HF} . \tag{11.9}$$

This should at least provide an upper bound for the CT interaction in second-order SAPT, since the two energy components in Eq. (11.9) are the only places where CT can appear.

11.2.3 Constrained DFT

Yet another way to define the CT interaction energy is to use the difference between the DFT energy of the fully-relaxed dimer system and the energy of a CT-free reference state that is constructed as a superposition of atomic densities, with the resulting molecular densities constrained to integer numbers of electrons using cDFT:[489,493,494]

$$\Delta E_{\rm CT}^{\rm cDFT} = E^{\rm DFT} - E^{\rm cDFT} . \tag{11.10}$$

We expect this definition to exhibit a far weaker dependence on the basis set, as compared to the ALMO- or SAPT-based schemes, because here the basis set serves only to describe the density, and individual MOs need not be assigned to one monomer or the other. The value of $\Delta E_{\mathrm{CT}}^{\mathrm{cDFT}}$ will still depend on how the electron densities of the two monomers are partitioned in real space,[490] *e.g.*, using Mulliken, Löwdin, Hirshfeld, or Becke atomic charges. The Mulliken and Löwdin partitions cannot be recommended for CT analysis, because they are (atomic) orbital-based.[490]

11.3 Computational details

BSSE must be removed in variational approaches such as ALMO-EDA,[478] which we accomplish via counterpoise correction.[80] The value of $\Delta E_{\mathrm{CT}}^{\mathrm{cDFT}}$ is only weakly dependent on basis set, as shown here and in Ref. 490, which indicates that the BSSE approximately cancels in Eq. (11.10), hence no further correction is applied to cDFT calculations.

In ALMO-EDA, the frozen-density energy contains electrostatics, Pauli repulsion, and dispersion. As a result, the this term can be rather sensitive to the choice of density functional, due to the inconsistent treatment of dispersion from one functional to the next. While it is tempting to want to use pairwise empirical dispersion potentials ("DFT+D") in this capacity, it should be recognized that the +D correction is, in Grimme's words, a "model-dependent quantity with no real physical meaning",[253] due to subtleties in the damping function and double-counting of correlation effects in the middle-distance regime.[143] In systems such as $\mathrm{M}^{+}\cdots\mathrm{C_6H_6}$, the empirical DFT+D

potentials differ quite substantially from the pure dispersion potentials that we use in SAPT(KS), predicting a dispersion energy for $K^+ \cdots H_2O$ that is 7 times larger than the SAPT dispersion potential predict, for example.[486]

Recognizing the double-counting problem inherent in DFT+D, Szalewicz and co-workers[6] developed a "dispersionless" density functional (dlDF) by re-parameterizing M05-2X using a data set in which dispersion interactions were specifically removed by subtracting SAPT(DFT)[149,150] dispersion energies. This affords a relatively clean way to separate the dispersion energy in a DFT calculation, and one that we shall examine here. In particular, we can combine dlDF with the dispersion potentials used in SAPT to obtain a DFT+D-type approach (dlDF+D3) with no double-counting problem.

ALMO and cDFT calculations reported here use either the B3LYP functional[495,496] or the ωB97X-D3 functional.[13] The cDFT calculations use the Becke partition of the charges, as in Ref. 497. SCF-MI calculations (needed for ALMO-EDA) are based on the locally-projected equations of Stoll *et al.*,[498] as implemented by Khaliullin *et al.*[491] For SAPT(KS), we use the LRC-ωPBE functional,[227,240] which has no short-range HF exchange, in conjunction with "ω_{GDD}" tuning of the range-separation parameter.[26] The SCF convergence criterion is set to 10^{-7} a.u. and the integral screening threshold is set to 10^{-12} a.u., using a fairly dense Euler-Maclaurin-Lebedev quadrature grid ($N_{\mathrm{radial}} = 75, N_{\mathrm{angular}} = 302$). In the SAPT calculations, the h functions have been removed from aug-cc-pV5Z and both the h and i functions are removed from aug-cc-pV6Z. In the ALMO and cDFT calculations, the i functions in aug-cc-pV6Z are removed. All calculations were performed using a locally-modified version of

Q-Chem.[97]

11.4 Results and discussion

11.4.1 $(H_2O)_2$

The water dimer, which exhibits very little CT,[499] is a prototypical system for application of EDA to understand hydrogen bonding. Using a series of aug-cc-pVXZ basis sets, X = 2–6, we find that the electrostatic, exchange, and induction energies in SAPT are basically converged at the triple-ζ level, changing by < 0.05 kJ/mol for X > 3. The ALMO-EDA polarization energy, however, changes by 1.8 kJ/mol between X = 3 and X = 6. This lack of convergence is a troubling feature, given the simplicity of the system.

Both experiments[29,468,469] and calculations[28,470] suggest that the CT interaction energy is approximately proportional to the amount of charge that is transferred, when averaged over the relative orientations of the interacting partners. The proportionality constant, k (energy per unit of transferred charge) can be estimated from state-of-the-art molecular-beam scattering experiments; the best estimate for $Xe\cdots H_2O$ is $k = 2.6$ meV/me^-.[29] The authors of Ref. 29 suggest that the same value of k is also appropriate for the water dimer, in which case a CT energy for the water dimer of 3.6 kJ/mol is obtained based on the 14.6 me^- of charge that is transferred, according to charge-displacement analysis.[499] ALMO-EDA based on a coupled-cluster wave function [ALMO(CCSD)/aug-cc-pVTZ] affords a CT energy of 3.5 kJ/mol.[27] For reasons of cost, ALMO(CCSD) has only been applied to small, closed-shell dimers

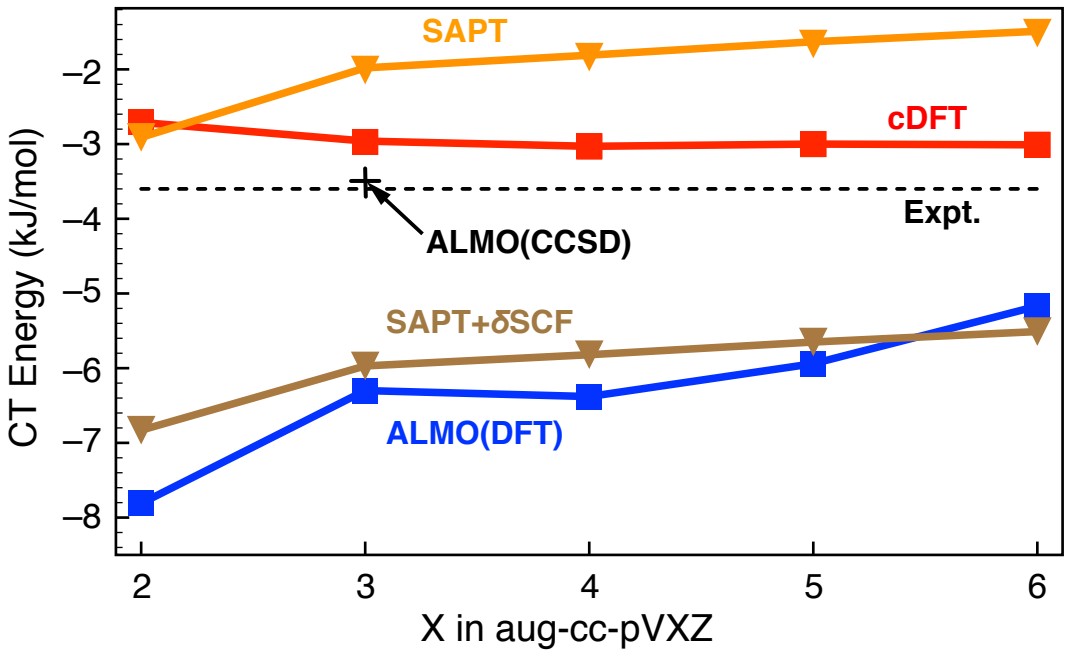

Figure 11.2: CT energies for $(H_2O)_2$ using a sequence of aug-cc-pVXZ basis sets, at the RI-MP2/aug-cc-pVDZ geometry of the dimer. The ωB97X-D3 functional is used for the ALMO(DFT) and cDFT calculations, and the LRC-ωPBE functional (with ω_{GDD} tuning[26]) is used for the SAPT and SAPT+$\delta E_{\mathrm{int}}^{\mathrm{HF}}$ calculations. (Here "δSCF" denotes the $\delta E_{\mathrm{int}}^{\mathrm{HF}}$ correction of Eq. (11.4).)

in triple-ζ basis sets or smaller.[27]

Unsurprisingly, CT energies for the orbital-based EDAs (SAPT and ALMO) are far more dependent on basis set as compared to the other energy components, as shown in Fig. 11.2. Whereas the cDFT value of the CT energy is converged in a triple-ζ basis set, the SAPT- and ALMO-based values change by 1–3 kJ/mol between triple- and hextuple-ζ. Moreover, the cDFT value of $\approx$ 3 kJ/mol is much closer to the experimental estimate as compared to ALMO(DFT) and SAPT. Řezáč *et al.*[490]

report a similar cDFT value (3.3 kJ/mol) using a different functional, basis set, and charge partition scheme.

As discussed above, $\Delta E_{\mathrm{CT}}^{\mathrm{SAPT}}$ should vanish in the complete-basis limit but for $(H_2O)_2$ it remains > 1 kJ/mol in magnitude even in the aug-cc-pV6Z basis set, which illustrates just how slowly atom-centered basis sets converge to the basis-set limit. In contrast, $\Delta E_{\mathrm{CT}}^{\mathrm{SAPT+\delta SCF}}$ (Eq. (11.9)) need *not* vanish in the basis-set limit, as the $\delta E_{\mathrm{int}}^{\mathrm{HF}}$ need not vanish and typically converges in triple-ζ basis sets.

Figure 11.3(a) compares the CT interaction energies computed using cDFT versus ALMO-EDA. Use of B3LYP versus ωB97X-D3 makes only a minor difference of ≈ 1 kJ/mol in the ALMO-EDA case, and essentially no difference in the cDFT case. Figure 11.3(b) examines the amount of charge that is actually transferred in the ALMO calculations, using ALMO charge-transfer analysis.[28] Although the two functionals in question differ by about 0.5 me^- in their predictions (with B3LYP predicting slightly more CT), the dependence on basis set is virtually identical and also quite dramatic, decreasing monotonically as the basis set is enlarged. Only 2.3 me^- is transferred at the B3LYP/aug-cc-pV5Z level (consistent with results in Ref. 470), which seems unrealistic in comparison to the 14.6 me^- estimated from charge-displacement analysis. (The latter reliably reproduces the exponential decay of CT as a function of intermolecular separation.[29,499]) As compared to the experimental results discussed above, it is interesting to note that ALMO-EDA underestimates the fraction of an electron that is transferred, yet overestimates the CT interaction energy, indicating that the proportionality constant k must be significantly overestimated as compared

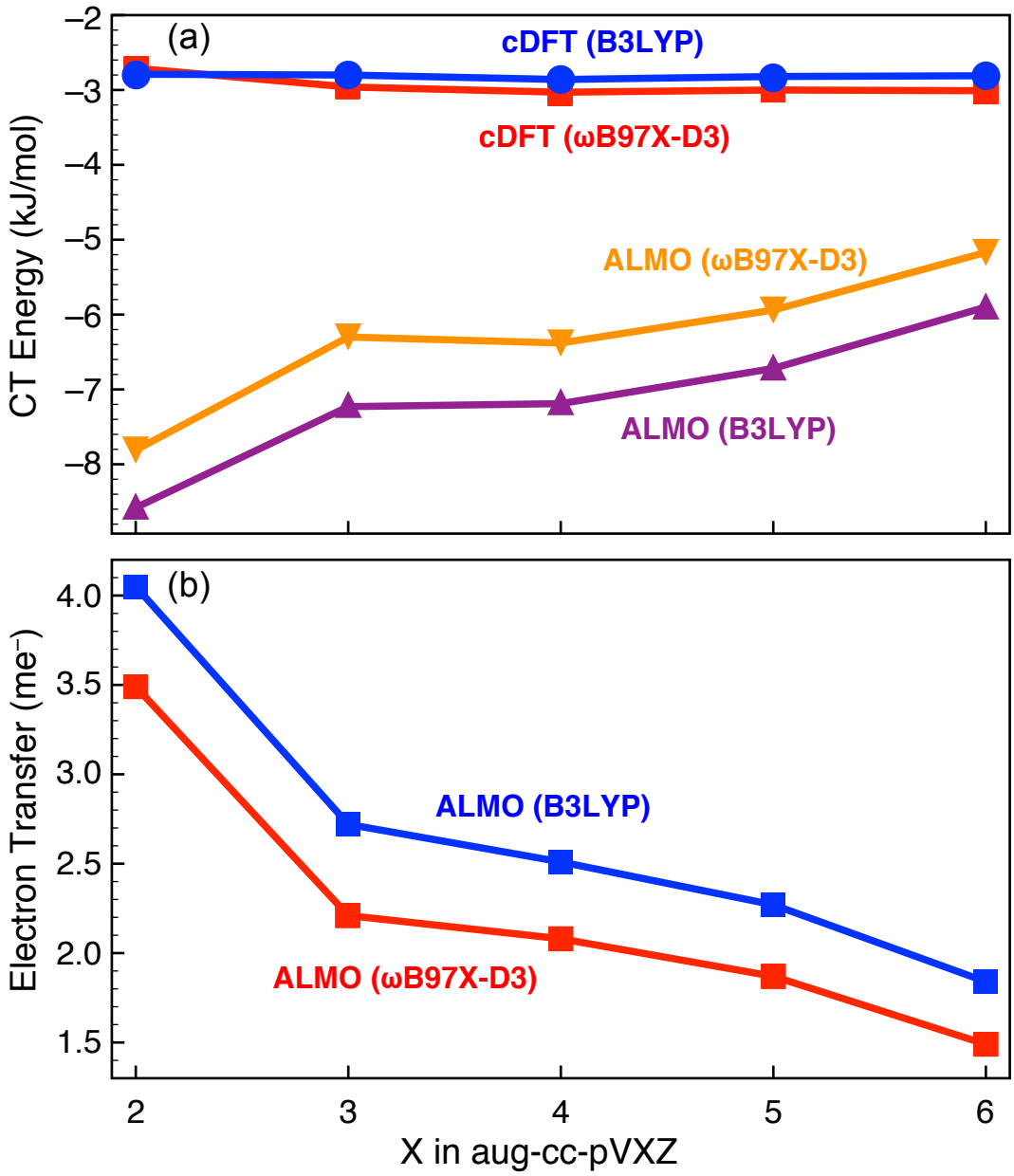

Figure 11.3: (a) CT energies for water dimer in a sequence of aug-cc-pVXZ basis sets. (b) Amount of charge that is transferred, as predicted by ALMO(DFT).

to the value estimated from experiment.

11.4.2 $H_3N\cdots BH_3$

The ammonia–borane complex is prototypical of dative bonding in a Lewis acid/base complex. It is a strongly-interacting system, with SAPT electrostatic, exchange, and induction energies ranging from −200 to −500 kJ/mol, at least nine times larger than the corresponding energy components in water dimer. As in $(H_2O)_2$, these energy components are essentially converged in a triple-ζ basis set, changing by at most 0.8 kJ/mol beyond aug-cc-pVTZ. For ALMO-EDA, the frozen-density and polarization energies range from −100 to −300 kJ/mol, and are at least 25 times larger then the corresponding components in water dimer. However, the ALMO-EDA polarization energy is *not* converged at the triple-ζ level, and changes by 36.3 kJ/mol upon enlarging the basis set from aug-cc-pVTZ to aug-cc-pV6Z. (The frozen-density energy changes only by 0.9 kJ/mol). This is consistent with the idea that the polarization energy is increasingly contaminated by charge transfer in the SCF-MI procedure, as the size of the basis set increases.[488]

CT energies for $H_3N\cdots BH_3$ are shown in Fig. 11.4 as a function of basis set. For reference, the ALMO(CCSD)/aug-cc-pVDZ level of theory affords an estimate of −59.4 kJ/mol for the CT interaction.[27] The cDFT result is close to this value, at about −61 kJ/mol for all basis sets. CT energies from SAPT and ALMO(DFT) exhibit significant basis set dependence, however, with the SAPT result ranging from −54.7 kJ/mol (aug-cc-pVDZ) to −1.2 kJ/mol (aug-cc-pV6Z), again converging towards zero in the basis-set limit. The $\delta E_{\rm int}^{\rm HF}$ term converges to about −44 kJ/mol in

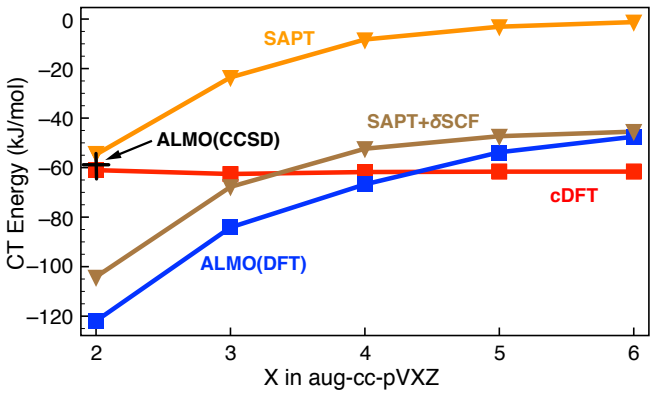

Figure 11.4: CT energies for $H_3N\cdots BH_3$ (RI-MP2/aug-cc-pVDZ geometry), in a sequence of aug-cc-pVXZ basis sets. The ωB97X-D3 functional is used for the ALMO(DFT) and cDFT calculations and the LRC-ωPBE functional is used for the SAPT and SAPT+δSCF. The solid black cross represents the ALMO(CCSD)/aug-cc-pVDZ result.[27]

aug-cc-pVTZ, meaning that the SAPT+$\delta E_{\text{int}}^{\text{HF}}$ method affords a non-zero CT energy even in the basis-set limit. The CT energy predicted by ALMO(DFT) ranges from −121.9 kJ/mol (aug-cc-pVDZ) to −47.5 kJ/mol (aug-cc-pV6Z), decreasing monotonically with basis size but not in a manner that suggests it will converge to zero.

Figure 11.5(a) compares cDFT and ALMO CT energies in the same sequence of basis sets, and Fig. 11.5(b) plots the fraction of an electron that is actually transferred from ammonia to borane. Comparing ωB97X-D3 to B3LYP, we find that these quantities have only a weak dependence on the functional, but in the ALMO case both quantities are strongly dependent on the basis set. The fraction of the electron that is transferred ranges from 43.2 me^- (ωB97X-D3/aug-cc-pVDZ) to 9.5 me^- (aug-cc-pV6Z), with a large decrease in the CT interaction energy across the same range

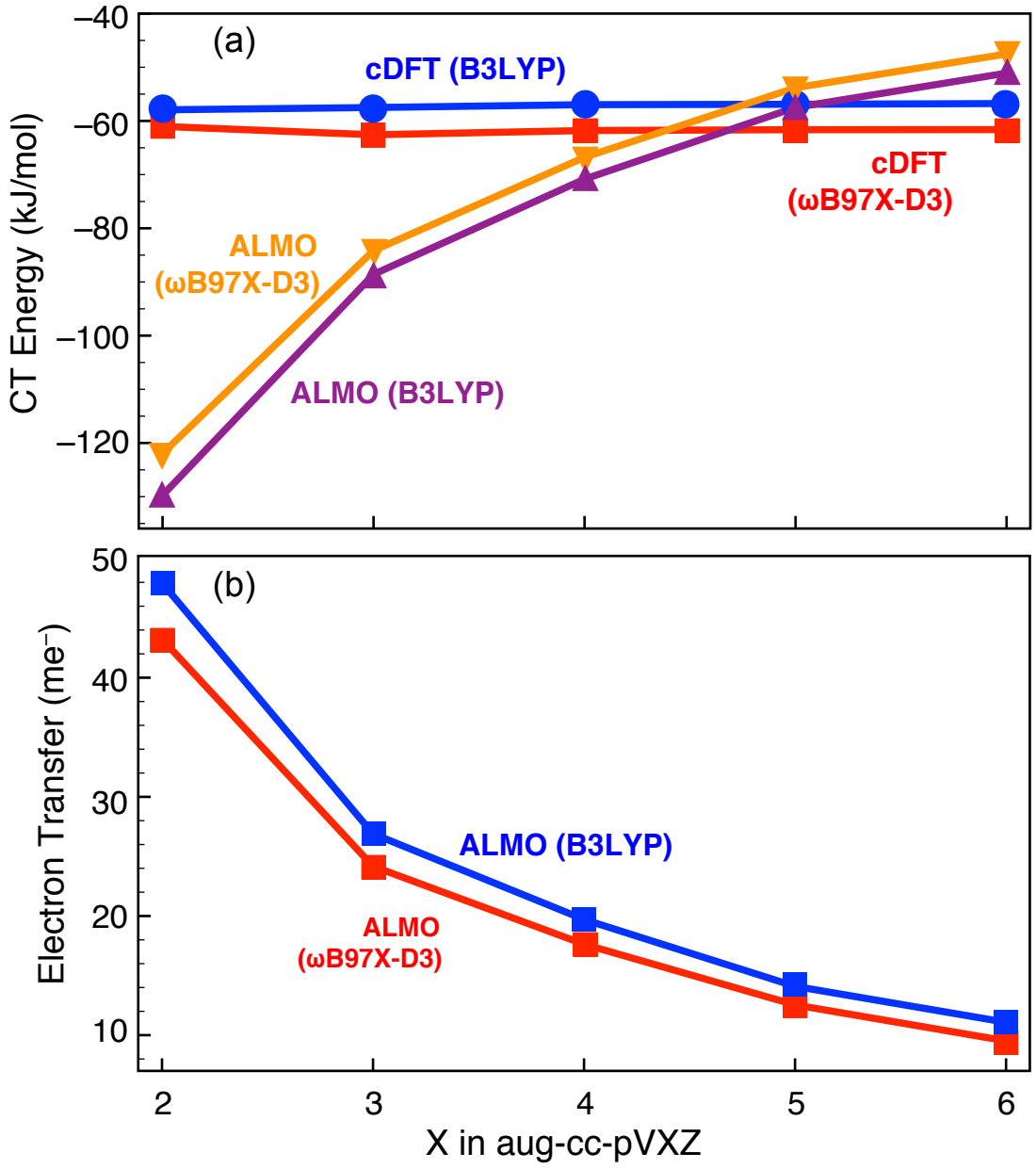

Figure 11.5: (a) CT energies for $H_3N\cdots BH_3$ using ALMO(DFT) and cDFT in a variety of basis sets. (b) Amount of charge that is transferred, as predicted by ALMO CT analysis.[28]

of basis sets. In contrast, cDFT results are basically independent on the basis set used.

11.4.3 $Xe\cdots H_2O$

Experimental data on CT interaction energies for H_2O, NH_3, and H_2S in complexes with small molecules (N_2, O_2, H_2) and rare-gas atoms are available from high-resolution, gas-phase scattering experiments from which angle-averaged intermolecular potentials can be extracted.[29,468,469] Amongst these complexes, the CT interaction energy is largest in $Xe\cdots H_2O$, at -0.98 kJ/mol,[29] which amounts to about 40% of the total interaction energy. The experiments afford a best-fit conversion factor $k = 2.6$ meV/me^- and theoretical charge-displacement analysis suggests 3.9 me^- of charge is transferred from Xe to H_2O. Figure 11.6 shows CT energies in $Xe\cdots H_2O$ computed using ALMO, cDFT, and SAPT-based methods in two different basis sets. The ALMO(DFT) method dramatically overestimates the experimental CT energy in a triple-ζ basis set and affords a much different result in quadruple-ζ basis set.

According to ALMO-based CT analysis, the ALMO(DFT) calculation transfers 1.8 me^- (triple-ζ) or 0.8 me^- (quadruple-ζ), as compared to 3.9 me^- in charge-displacement analysis,[29] nevertheless the ALMO CT interaction energy is larger than the experimental one. As with water dimer, this suggests that ALMO(DFT) significantly overestimates the value of k. (It has previously been suggested that ALMO may underestimate the CT energy due to mixing in with polarization,[30] but the minuteness of the polarization energy for $Xe\cdots H_2O$ means this is not the case here.)

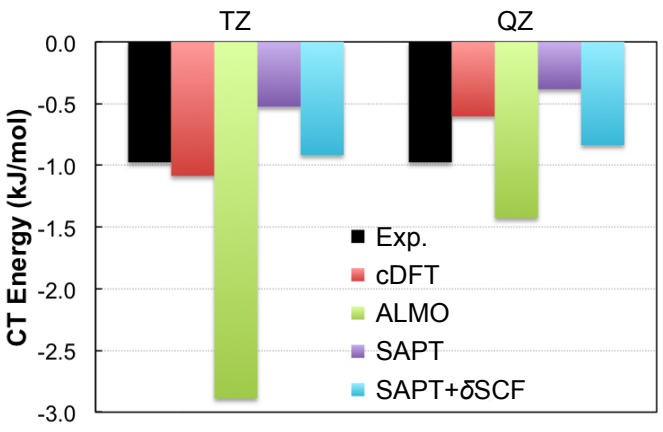

Figure 11.6: CT energies for $Xe\cdots H_2O$, using the ωB97X-D3 functional for the ALMO(DFT) and cDFT calculations and the LRC-ωPBE functional for the SAPT and SAPT+δSCF. The "TZ" basis set is aug-cc-pVTZ for H_2O and def2-TZVPPD for Xe, and "QZ" is aug-cc-pVQZ for H_2O and def2-QZVPPD for Xe. The geometry of the complex is taken from Ref. 29.

Given the dramatic basis-set dependence of the ALMO CT energy, as compared to the cDFT approach, the latter seems more reliable even though both calculations err by the same amount (in opposite directions) when a quadruple-ζ basis set is used.

11.4.4 $M^+\cdots H_2O$

Although it is difficult to obtain reliable reference data to assess the accuracy of various models for the CT interaction energy, an EDA should at least produce results that agree with chemical intuition in obvious cases, exemplified here by $M^+\cdots H_2O$ complexes with M = Li, Na, or K. We call these "obvious" cases because we expect that the smaller alkali metal cations should approach more closely to H_2O and thus more effectively withdraw charge, hence we expect the CT energy to decrease as one

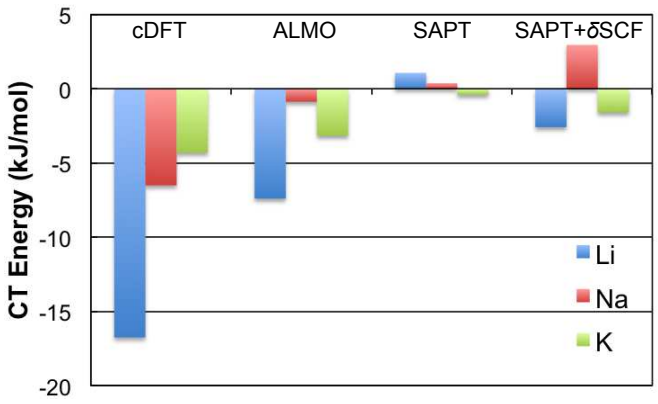

Figure 11.7: CT energies for $M^+ \cdots H_2O$. The ωB97X-D3 functional is used for the ALMO(DFT) and cDFT calculations, and the LRC-ωPBE functional is used for SAPT and SAPT+δSCF. The aug-cc-pVQZ basis set is used for H_2O and the def2-QZVPPD basis set is used for M^+. Geometries are optimized at the RI-MP2 level using the analogous double-ζ basis sets.

moves down the Group I cations, due to increasing ion–molecule distance. Results in Fig. 11.7 reveal that only cDFT reproduces this anticipated trend.

In contrast, ALMO calculations show no trend at all. Different results (in different basis sets) for these same systems have been reported in ALMO calculations by Phipps *et al.*,[486] who found that the CT energy was largest for K^+ even in the smallest basis set employed in that study. In the quadruple-ζ basis set used here, the ALMO CT energy is largest for Li^+. Phipps *et al.* argue that the unphysical behavior they observe might be an artifact of counterpoise correction, but in our calculations the counterpoise corrections for M = Li, Na, and K amount to only 0.06, 0.09, and 0.14 kJ/mol, respectively, all of which are small in comparison to the ALMO CT energies reported in Fig. 11.7.

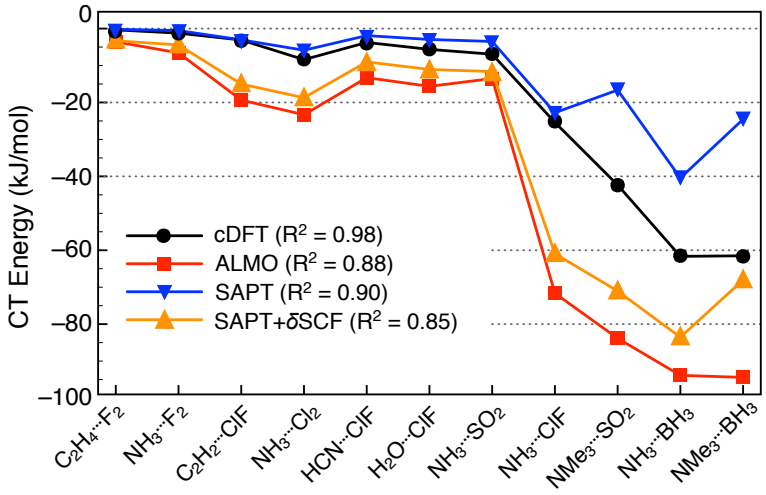

Figure 11.8: CT energies computed using the def2-QZVPPD basis set. The ωB97X-D3 functional is used for the cDFT and ALMO calculations, and LRC-ωPBE is used for the SAPT and SAPT+δSCF calculations. Also listed are the R^2 goodness-of-fit parameters for a linear fit of the CT energy versus the fractional charge δq that is transferred, as determined using NPA atomic charges. The four dimers on the far right are the only ones for which $\delta q > 0.1\ e^-$.

The SAPT CT energies are also rather strange, as they are repulsive for Li^+ and Na^+ and exhibit precisely the opposite trend (in magnitude) as observed using cDFT. The $\delta E_{\text{int}}^{\text{HF}}$ correction is negative for Li^+ and K^+ but positive for $Na^+\cdots H_2O$, with the result that the SAPT+$\delta E_{\text{int}}^{\text{HF}}$ CT energies (like ALMO) show no consistent trend across Group I. In short, none of the orbital-based definitions of the CT energy follow the expected trend across Group I, but this trend *is* reproduced by cDFT.

11.4.5 Charge-transfer complexes

Řezáč *et al.*[490] have assembled a set of 11 dimers that span a wide range of CT energies and used them to test the performance of cDFT using GGA and meta-GGA functionals (but not hybrids); we will use the same data set to test the methods examined here. Following Ref. 490, we estimate the fractional charge transfer δq using atomic charges from natural population analysis[94] (NPA) at the HF/cc-pVTZ level. Figure 11.8 plots the CT energies for this data set. All of the methods examined here are in rough agreement (and exhibit similar trends) only in cases where $\delta q < 0.1\ e^-$, but in the four cases where δq is larger, the CT energies (and even the trends) vary widely. Of the methods in Fig. 11.8, cDFT affords the best linear correlation between δq and the CT interaction energy, with a goodness-of-fit $R^2 = 0.98$.

11.4.6 Cation–alkyl radical complexes

Next we examine CT energies for open-shell systems in order to make contact with the rich chemistry of radicals. Complexes involving alkyl radicals such as $CH_3^\bullet$ and $(CH_3)_3C^\bullet$, stabilized by cations (H_3O^+ or NH_4^+ in this work) are interesting model systems for understanding the stabilization of transition structures in reactions involving proteins.[462,500] These complexes are also interesting from the standpoint of CT analysis, since electron-deficient radicals often have energetically low-lying singly-occupied MOs than can accept some fraction of an electron from a donor molecule. To compare open- versus closed-shell interactions, we will also examine the corresponding closed-shell species, CH_4 and $(CH_3)_3CH$. The 8 complexes considered here

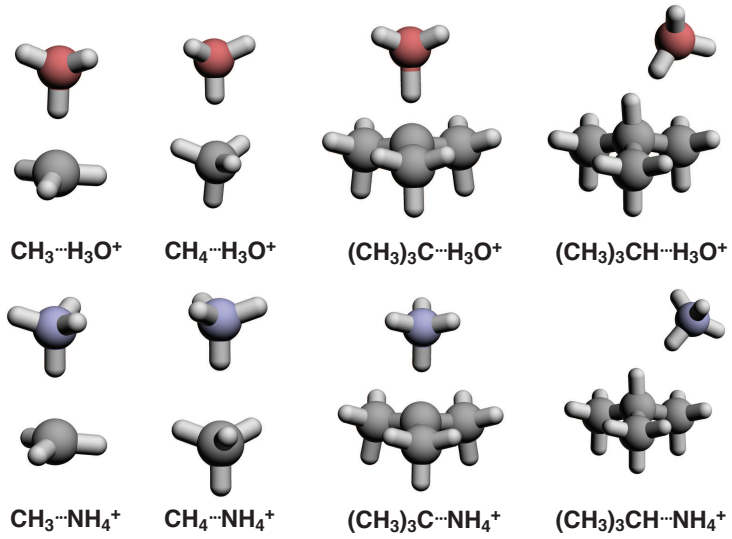

Figure 11.9: Complexes of $CH_3^{\bullet}$, $(CH_3)_3C^{\bullet}$, CH_4, and $(CH_3)_3CH$ with H_3O^+ or NH_4^+. Geometries are taken from Ref. 30.

are shown in Fig. 11.9, and have previously been examined using ALMO-EDA in Ref. 30. That study used B3LYP, whereas we use ωB97X-D3, which systematically reduces the CT interaction energies by an average of 2.5 kJ/mol. This is consistent with the idea that ωB97X-D3 reduces delocalization error,[408,501] which manifests in EDA as an exaggeration of CT interactions.[502,503]

Chemically and qualitatively, one expects the CT interaction energy to be determined by the IE of the donor species and the EA of the acceptor, with a strong dependence on intermolecular distance and mutual orientation.[462] In Fig. 11.10 we plot CT interaction energies versus the inverse donor-acceptor gap, $(\mathrm{IE}-\mathrm{EA})^{-1}$. Very good linear correlations are obtained, with R^2 values close to unity for each method examined. (The R^2 values using SAPT alone are a bit smaller, although $R^2 \approx 1$ upon

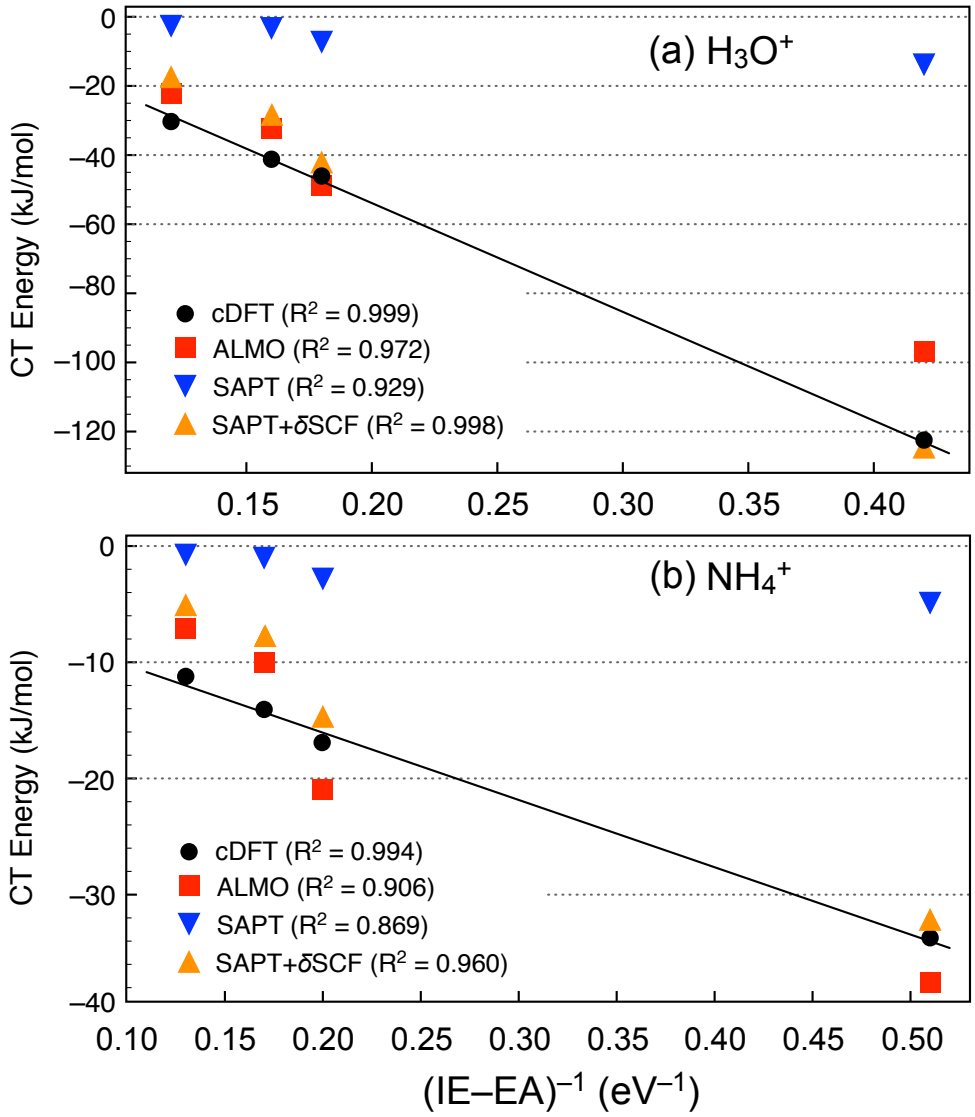

Figure 11.10: CT energies for the cation–molecule and cation–radical complexes in Fig. 11.9, plotted as a function of the inverse donor–acceptor gap. The ωB97X-D3/def2-QZVPPD level of theory is used for ALMO and cDFT calculations, and LRC-ωPBE/def2-QZVPPD is used for SAPT and SAPT+δSCF. Also listed are R^2 values for linear fits of the data, with the linear fit shown for cDFT.

addition of the $\delta E_{\text{int}}^{\text{HF}}$ correction.) Note in particular that cDFT affords the best fits of all, with $R^2 > 0.99$. This suggests an easy way to estimate the magnitude of the CT energy, based on calculations performed on the isolated species.

11.4.7 Nucleophile–$C_6H_5^{\bullet+}$ complexes

Interaction between oxidized aromatic rings (such as damaged DNA bases) and nucleophiles (such as water or alcohol) play key roles in many biochemical reactions.[504,505] The aryl radical, $C_6H_5^{\bullet+}$, has two different binding sites that behave differently during nucleophilic aromatic substitution reactions at the radical center.[506] ALMO-EDA has previously been used to examine these "on-top" versus "side-on" configurations in complexes spanning a wide range of IEs,[30] and we will examine the same complexes here. These complexes are depicted in Fig. 11.11.

CT energies for $C_6H_5^{\bullet+}$ with 8 different nucleophiles oriented in the on-top position are plotted versus $(\text{IE} - \text{EA})^{-1}$ in Fig. 11.12. Linear fits of the data afford $R^2 \approx 1$ but we note in particular that inclusion of the $\delta E_{\text{int}}^{\text{HF}}$ term in the SAPT calculation improves the value of R^2 significantly, as was the case for the cation–alkyl radical complexes. While the *correlations* are significant, the various energy decompositions yield very different *values* for the CT energy, especially when the IE/EA gap is small. Excluding the SAPT results without $\delta E_{\text{int}}^{\text{HF}}$ (since the CT energies in this approach are all relatively small), the remaining three methods examined in Fig. 11.12 span a range of 39 kJ/mol for $C_6H_5^{\bullet+} \cdots PH_3$, the system for which the gap is smallest.

relationship with the inverse gap, regardless of which EDA is used; see the data in Table S2 of the Supporting Information. Values of the CT interactions for these

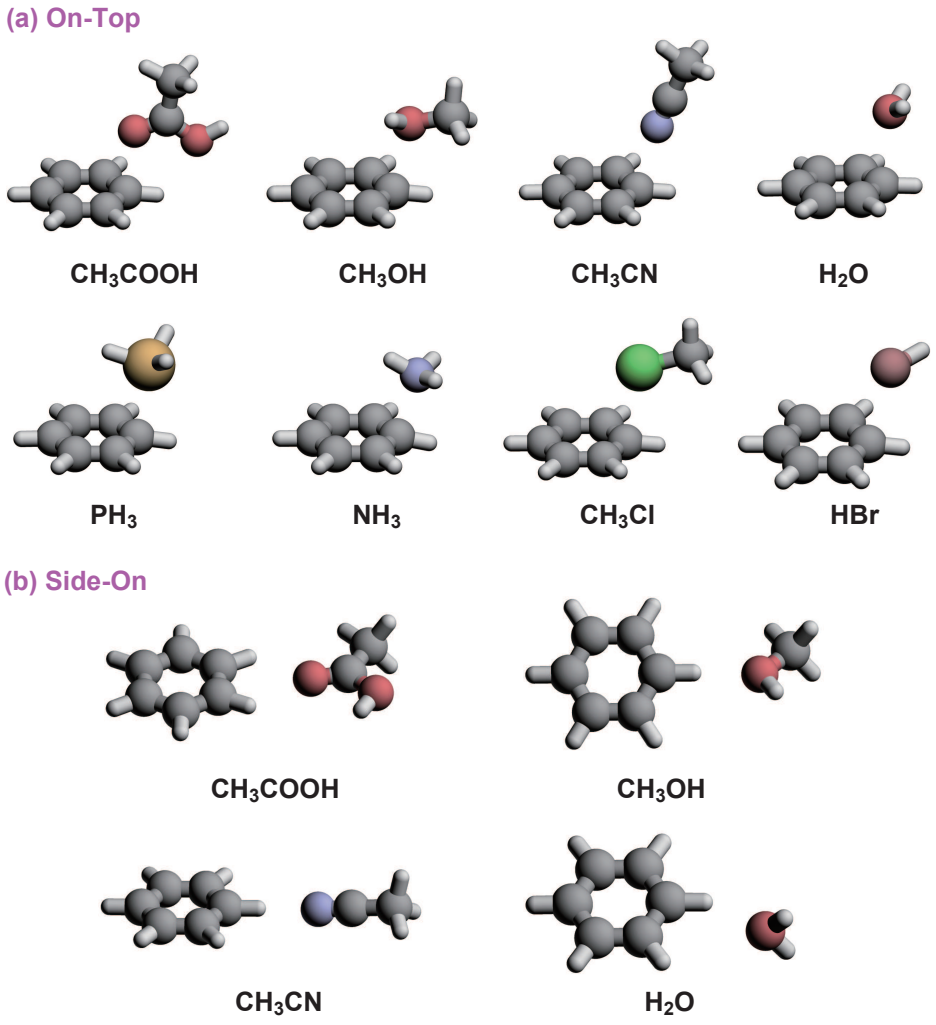

Figure 11.11: $C_6H_5^{\bullet+}$ complexes with various nucleophiles in the "on-top" and "side-on" configurations. Geometries are taken from Ref. 30.

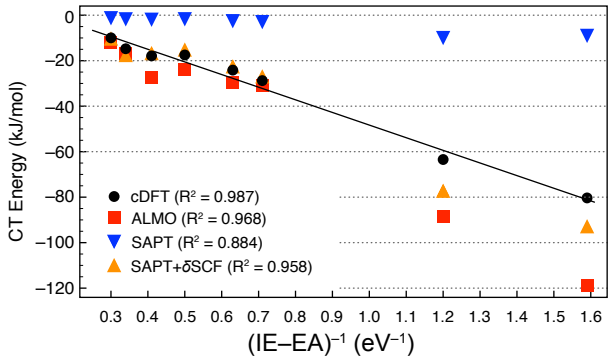

Figure 11.12: CT energies for $C_6H_5^{\bullet+}$ complexes with various nucleophiles in the "on-top" configuration, as a function of the inverse gap $(\mathrm{IE} - \mathrm{EA})^{-1}$, where IE is the ionization energy of the nucleophile and EA denotes the electron affinity of $C_6H_5^{\bullet+}$, that is, the IE of benzene. (IE and EA data are taken from Ref. 31.) The ωB97X-D3/def2-QZVPPD method is used for the ALMO and cDFT calculations and LRC-ωPBE/def2-QZVPPD is used for SAPT and SAPT+δSCF.

configurations range from -4.0 to -5.6 kJ/mol in cDFT calculations, considerably smaller than the values for the on-top orientations in Fig. 11.12, which is probably due to unfavorable overlap. It may be that a strong linear relationship simply does not emerge over such a small energy range, since only four side-on complexes are considered.

The binding energy of $C_6H_5^{\bullet+} \cdots H_2O$ is nearly the same in both the on-top and side-on orientations, and has previously been analyzed using ALMO-EDA with the conclusion that the side-on interaction is dominated by the frozen density term while the on-top interaction has a large CT component.[30] (Note that the information available from ALMO-EDA is limited because the frozen density term contains all energy components except polarization and CT.) We repeated the B3LYP calculations from

Table 11.1: ALMO energy components[a] for $C_6H_5^{\bullet+} \cdots H_2O$ in two orientations.

Orientation	Functional	Energy Component (kJ/mol)				
		FRZ	POL	CT	DISP[b]	Total
On-Top	B3LYP	−11.1	−7.9	−17.0		−36.0
	ωB97X-D3	−21.6	−8.4	−12.1		−42.1
	dlDF	−4.9	−8.3	−8.0	−20.1	−41.3
Side-On	B3LYP	−25.3	−6.9	−4.7		−36.9
	ωB97X-D3	−30.4	−7.2	−4.0		−41.5
	dlDF	−20.8	−6.8	−3.5	−11.1	−42.2

[a] 6-311++G(3df,3pd) basis set. [b] The dispersion energy is normally unavailable for ALMO, but for dlDF is evaluated at the SAPT2+3(CCD)/aug-cc-pVTZ level for the closed-shell $C_6H_6 \cdots H_2O$ complex.

Ref. 30 for these two isomers, and in Table 11.1 we also report ALMO-EDA using ωB97X-D3 and the dlDF functional. All three functionals predict similar values for the total binding energy, and all three predict a frozen-density energy that is larger in the side-on orientation but a CT energy that is larger in the on-top configuration. (The ωB97X-D3 functional does predict less CT as compared to B3LYP, consistent with a larger delocalization error for the latter functional.) However, the dlDF approach predicts a frozen-density energy that is far less attractive as compared to the other two functionals, which is compensated by the dispersion energy. For use with dlDF, we estimate the dispersion energy via a high-level SAPT2+3(CCD)/aug-cc-pVTZ calculation for the closed-shell $C_6H_6 \cdots H_2O$ complex, since SAPT beyond second order has only been implemented for closed-shell systems. Calculations using

Table 11.2: Energy components for $C_6H_5^{\bullet+}\cdots H_2O$ based on open-shell SAPT.[a]

Orientation	Energy Component (kJ/mol)					
	ELST	EXCH	IND		DISP[c]	Total
			Total	CT[b]		
On-Top	−48.5	43.2	−17.2	−10.1	−20.1	−42.6
Side-On	−46.7	25.1	−10.9	−4.0	−11.1	−43.5

[a]Using LRC-ωPBE/def2-QZVPPD for the ELST, EXCH, and IND components. [b]Based on cDFT at the level of ωB97X-D3/def2-QZVPPD. [c]Evaluated at the SAPT2+3(CCD)/aug-cc-pVTZ level for $C_6H_6\cdots H_2O$.

the Becke-Johnson exchange-dipole model (XDM),[355,357] which is available for open-shell systems, suggest that substituting the closed-shell system changes the dispersion energy by < 1 kJ/mol; see the Supporting Information. Unfortunately, it is not possible to analyze these discrepancies amongst functionals in any greater detail, because the frozen-density energy cannot be further decomposed in ALMO-EDA, whereas SAPT provides a more detailed energy decomposition.

SAPT results for the two $C_6H_5^{\bullet+}\cdots H_2O$ complexes are listed in Table 11.2. As in the DFT calculations discussed above, the binding energies for the two orientations are similar in either orientation, but in the SAPT case we obtain a more detailed picture. According to SAPT, the electrostatic interactions are similar, but the on-top configuration is destabilized by Pauli repulsion (EXCH) to a much greater extent. This is consistent with a shorter intermolecular distance in the on-top configuration (3.2 Å versus 4.5 Å center-of-mass distances), and explains why the frozen-density energies in Table 11.1 are less stabilizing in the on-top case, as the Pauli repulsion is

contained in this term.

A final remark is that we expect the CT interaction to correlate roughly with the degree of overlap between the donor and acceptor orbitals, which we may judge in the ALMO-EDA case by examining the correlation between the CT energy and the frozen-density energy (which contains the Pauli repulsion), and in the SAPT case as the correlation between the cDFT CT energy and the SAPT exchange energy. These two comparisons are plotted in Fig. 11.13 for both the on-top and side-on orientations of the $C_6H_5^{\bullet+}\cdots$nucleophile data set. In the SAPT case, CT is strongly correlated with the exchange energy ($R^2 = 0.95$) but less so in ALMO-EDA, perhaps because the exchange energy is mixed up with electrostatics and dispersion in the frozen-density energy.

11.4.8 $C_{10}H_8^{\bullet+}\cdots C_6H_6$

A final example is the complex between benzene and the naphthalene radical cation, for which $\Delta H^\circ - 33 \pm 4$ kJ/mol has recently been measured experimentally.[32] This and related cation radical dimers of aromatic hydrocarbons are good model systems to understand the basis of photoconductivity and ferromagnetism in organic materials.[507–509] Enthalpies of binding for the homo-dimers $C_6H_5^{\bullet+}\cdots C_6H_6$ ($\Delta H^\circ =$ 71 kJ/mol) and $C_{10}H_8^{\bullet+}\cdots C_{10}H_8$ ($\Delta H^\circ = 74.5$ kJ/mol) are much larger,[510] and the smaller binding energy for $C_{10}H_8^{\bullet+}\cdots C_6H_6$ has been explained in terms of a large degree of charge delocalization in naphthalene as compared to benzene, which reduces the CT interaction in the former.[32] The large difference in IE between naphthalene (8.1 eV) and benzene (9.2 eV) is taken as evidence of the lack of charge-resonance

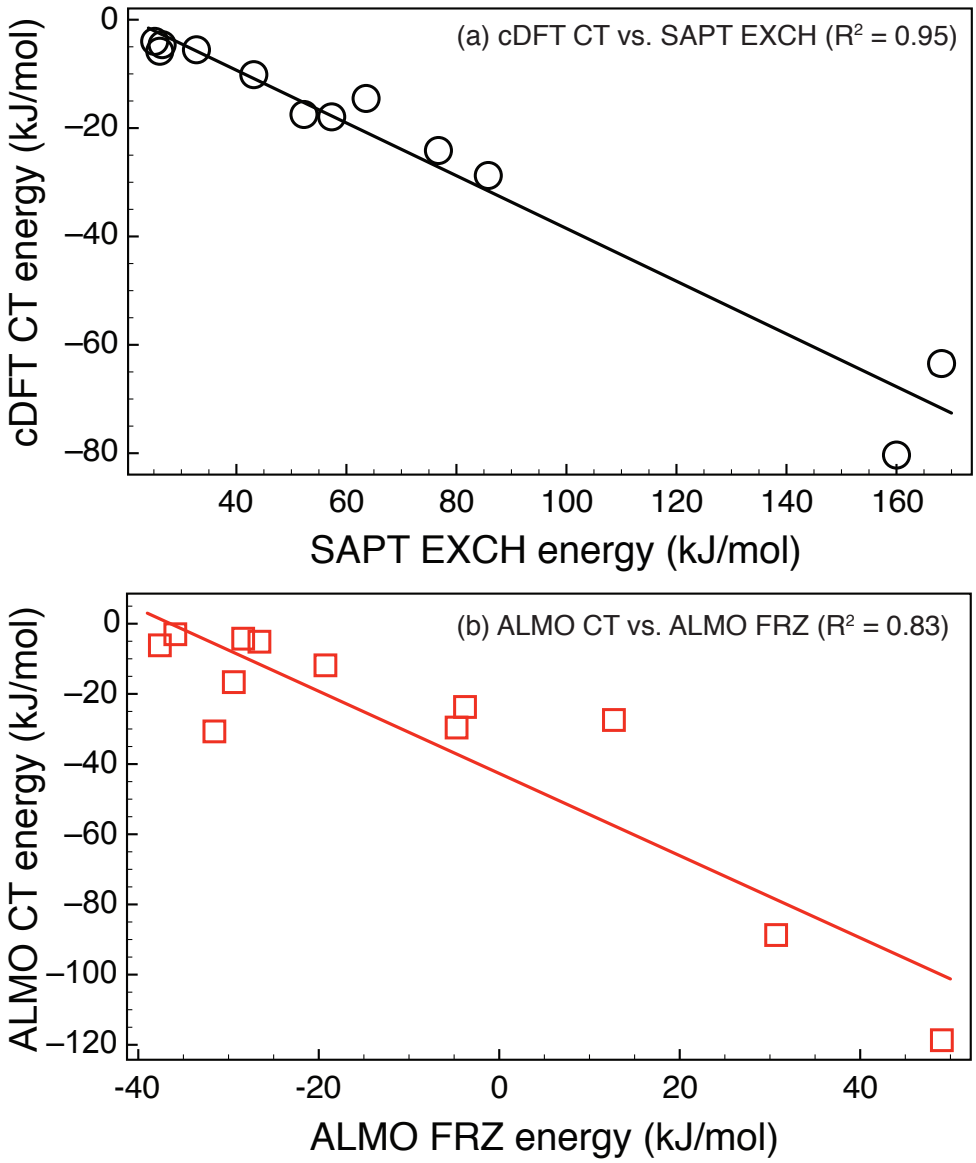

Figure 11.13: (a) CT energies from cDFT versus SAPT exchange energies, and (b) CT energies from ALMO-EDA versus ALMO frozen-density energies, for all 12 $C_6H_5^{\bullet+} \cdots$ nucleophile complexes in Fig. 11.11. The ωB97X-D3/def2-QZVPPD level of theory is used for the ALMO and cDFT calculations and LRC-ωPBE/def2-QZVPPD is used for SAPT EXCH calculations. Linear fits and corresponding R^2 values are indicated.

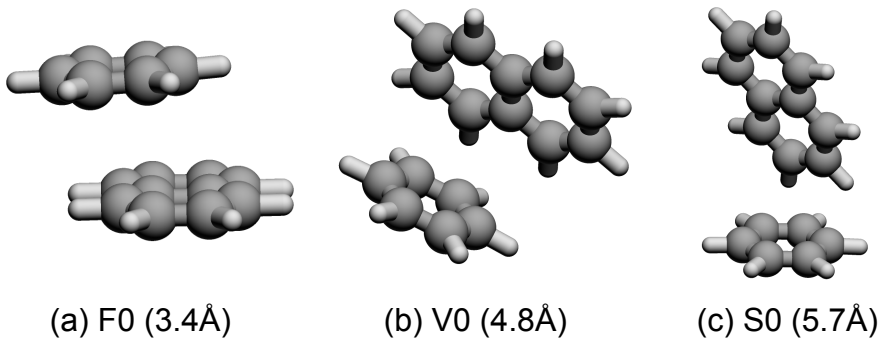

Figure 11.14: Lowest-energy structures for each of three classes of isomers of $C_{10}H_8^{\bullet+} \cdots C_6H_6$: (a) face-to-face, (b) V-shaped, and (c) face-to-side. Geometries are taken from Ref. 32 and the value in parenthesis represents the distance between monomer centers of mass.

interaction in (naphthalene)$^{\bullet+} \cdots$ benzene,[32] whereas this effect is obviously in play for the homo-dimers. Given such a large difference in IEs, however, one would not expect much CT either, but ALMO-EDA with the M11/cc-pVTZ functional[421] predicts a surprisingly large CT interaction for this complex in a face-to-face arrangement.[32] Three different isomers of $C_{10}H_8^{\bullet+} \cdots C_6H_6$ (as shown in Fig. 11.14) will be considered here, in order to assess whether this unusually large CT interaction is a general trend or an artifact of ALMO-EDA.

ALMO-EDA results for these three isomers, using both ωB97X-D3 and dlDF, are listed in Table 11.3. The polarization interaction is similar for all three isomers but the frozen-density term varies, and results for the two functionals are quite different. Using dlDF we can safely disentangle the dispersion energy from the frozen-density energy, and we accomplish this using an atom–atom dispersion potential that is fit

Table 11.3: ALMO-EDA results[a] for three isomers of (naphthalene)$^{\bullet+}\cdots$(benzene).

Isomer	Functional	Energy Component (kJ/mol)				
		FRZ[b]	POL	CT	DISP[c]	Total
F0	ωB97X-D3	15.2	−14.0	−17.7	−24.6	−41.1
	dlDF-D3	57.3	−9.4	−16.0	−66.9	−35.0
V0	ωB97X-D3	−1.6	−15.0	−6.9	−14.4	−38.0
	dlDF-D3	22.2	−10.9	−7.1	−36.2	−32.0
S0	ωB97X-D3	−3.1	−12.4	−5.8	−11.1	−32.4
	dlDF-D3	16.5	−8.9	−5.4	−29.2	−27.1

[a]def2-TZVPPD basis set. [b]DISP energy subtracted from FRZ for ωB97X-D3. [c]Computed using the Grimme-style D3 correction for ωB97X-D3, and using the SAPT-based +D3 potential of Ref. 57 for dlDF-D3.

to high-level SAPT calculations for a training set of small dimers, namely, the +D3 potential of Ref. 57. Results suggest that the face-to-face F0 arrangement exhibits a dispersion interaction about twice as large as the other two configurations, which is consistent with results for benzene dimer,[511] where π-stacked configurations exhibit larger dispersion interactions as compared to T-shaped configurations. This, in turn, is due to the rapid decay of dispersion with distance and the fact that stacked configurations put more of the π electron clouds of the two monomers into close proximity.

Despite the empirical nature of the Grimme-style D3 correction in ωB97X-D3, as discussed in Section 11.3, we have listed the value of this correction as the DISP energy for ωB97X-D3 in Table 11.3, and subtracted this value from the FRZ energy. Whereas the frozen-density term in dlDF is repulsive for all three isomers, once the D3

Table 11.4: CT interaction energies (in kJ/mol) for (naphthalene)$^{\bullet+}\cdots$(benzene).

Method	F0	V0	S0
cDFT[a]	−9.0	−6.4	−5.0
ALMO[a]	−17.7	−6.9	−5.8
SAPT[b]	−0.5	−0.4	−0.3
SAPT[b]+$\delta E_{\text{int}}^{\text{HF}}$	−17.4	−8.8	−6.4

[a] ωB97X-D3/def2-TZVPPD level.

[b] LRC-ωPBE/def2-TZVPPD level with ω_{GDD} tuning.

correction is subtracted from the FRZ energy in the case of ωB97X-D3, this frozen-density contribution is far less repulsive (for isomer F0) or even slightly attractive (for V0 and S0). That the total binding energies are about the same for both functionals is a consequence of the fact that the dispersion correction in ωB97X-D3 is much less attractive than the one in dlDF-D3. Given the documented accuracy of the SAPT-based +D3 correction,[57] this example provides a vivid illustration of why empirical DFT+D dispersion corrections should not be conflated with true dispersion energies. Considering the dlDF+D3 results, we see that the frozen-density term is twice as repulsive for the F0 isomer than for the other two isomers, consistent with the shorter intermolecular separation in this case (see Fig. 11.14). However, the ALMO(dlDF) calculations still do not separate electrostatics from Pauli repulsion.

Consistent with results in Ref. 32, our ALMO-EDA calculations with two different functionals suggest a CT interaction that is twice as large in F0 as in V0, and three times as large as that in S0. This is what the authors of Ref. 32 call "the unexpectedly high CT" in the face-to-face arrangement. Other estimates of CT interaction energies

Table 11.5: Energy components for three isomers of $C_{10}H_8^{\bullet+} \cdots C_6H_6$ based on open-shell SAPT(KS)+D3/cDFT calculations.

Isomer	Energy Component (kJ/mol)					
	ELST[a]	EXCH[a]	IND		DISP[c]	Total
			Total[a]	CT[b]		
F0	−39.3	79.4	−25.2	−9.0	−66.9	−52.0
V0	−29.2	41.7	−19.1	−6.4	−36.2	−42.8
S0	−25.3	33.4	−15.1	−5.0	−29.2	−36.1

[a]From SAPT(KS) using LRC-ωPBE/def2-TZVPPD

[b]Using cDFT at the ωB97X-D3/def2-TZVPPD level.

[c]Using the SAPT-based +D3 potential of Ref. 57.

for these three dimers can be found in Table 11.4. By themselves, SAPT values for the CT energies are negligible in all three isomers, but are larger (and similar to ALMO-EDA results) when the $\delta E_{\text{int}}^{\text{HF}}$ correction is included. However, the more reliable cDFT approach affords CT energies that are considerably smaller and amount to 22% of the total interaction energy for F0 and 15% for S0. This should be compared to ALMO(ωB97X-D3) results where 43% of the F0 interaction energy is assigned to CT and 18% for S0. As such, cDFT calculations do not support the idea of unusually large CT in the face-to-face arrangement, suggesting that such a result may be an artifact of an orbital-based definition of CT.

In Table 11.5, we put forward a hybrid energy decomposition scheme (as applied to $C_{10}H_8^{\bullet+} \cdots C_6H_6$) in which a SAPT(KS)+D3 calculation is used to obtain the electrostatic, Pauli repulsion, induction, and dispersion components, with the latter corresponding precisely to the SAPT-based +D3 correction.[57] At this level, the CT

energy is contained within the induction term, but can be separated by defining the CT interaction based on a cDFT calculation. We propose that the SAPT(KS) induction energy, less the cDFT estimate of the CT energy, be taken as the actual (CT-free) induction energy.

In a previous ALMO-EDA study of $C_{10}H_8^{\bullet+} \cdots C_6H_6$,[32] it was observed that F0 has the smallest (least attractive) frozen-density interaction, consistent with our ALMO-EDA results in Table 11.3. This fact was previously attributed to electrostatics,[32] with the authors speculating that perhaps the charge–quadrupole interaction is smallest in the F0 isomer. However, with SAPT we can extract the actual electrostatic component, which in ALMO is buried in the frozen density term, and we find that the SAPT electrostatic energy is actually *most* stabilizing for F0. This is consistent with SAPT2+/aug-cc-pVTZ results for protonated benzene dimer,[512] which exhibits similar geometric binding motifs: in that system, the electrostatic interaction is strongest in the parallel-displaced arrangement that is qualitatively similar to the F0 isomer of $C_{10}H_8^{\bullet+} \cdots C_6H_6$.

The energy decomposition in Table 11.5 paints an interesting picture of the balance of forces that determine the relative energies of F0, V0, and S0. The F0 isomer exhibits the largest dispersion interaction, as a result of the relatively close proximity of the π-electron clouds as compared to the V0 and S0 isomers. This is probably also the reason why the induction interaction is most favorable in F0, due to polarization of benzene's π electrons from the positive charge on the naphthalene cation. At the same time, F0 exhibits the shortest intermolecular separation (measured as the

distance between monomer centers of mass), and hence the largest value of the Pauli repulsion, and this partially offsets the more favorable electrostatic, induction, and dispersion interactions, to the point where total interaction energies amongst the three isomers are not so disparate. We also note that the CT contributions to the binding energies (computed using our hybrid SAPT/cDFT scheme) are rather small, consistent with chemical intuition given the large difference in IEs between the two monomers.

11.5 Conclusions

We have examined several orbital-based EDAs that are designed to decompose intermolecular interaction energies into physically-meaningful components. These have been compared to constrained DFT calculations, specifically in the context of how these methods interpret CT interactions. Based on applications to a variety of very different systems, the cDFT approach is the only one that we can recommend for evaluating the CT interaction energy, for a variety of reasons:

- It exhibits very little basis-set dependence. This stands in contrast to orbital-based (ALMO-EDA or SAPT) definitions, for which the definition of the CT interaction energy depends strongly on the basis set.

- It provides quantitative CT results for $(H_2O)_2$, $H_3N\cdots BH_3$, and $Xe\cdots H_2O$, as compared to the best values in the literature, including experimental estimates of the CT interaction energy.

- It provides the correct trend in CT across Group I in $M^+ \cdots H_2O$ complexes (M = Li, Na, or K), whereas ALMO and SAPT do not.

- Predicted CT energies exhibit excellent correlation with the inverse donor-acceptor gap, $(IE-EA)^{-1}$, for a variety of complexes whose binding is dominated by CT.

- It can rationalize the relative binding energies for isomers of the radical cation complex $(naphthalene)^{\bullet+} \cdots (C_6H_6)$, and resolves an anomaly wherein ALMO-EDA predicts curiously large CT interaction energies that do not correlate with IE/EA differences.

Moreover, CT interaction energies computed with the cDFT procedure endorsed here, which is based on hybrid functionals, correlate better with the amount of charge that is transferred, as compared to the alternative cDFT procedure used by Řezáč *et al.*[490] that employs GGA functionals.

In summary, we recommend a composite SAPT(KS)+D3/cDFT procedure as a robust procedure for EDA. Within this approach, the SAPT(KS) calculation provides the electrostatic, Pauli repulsion ("exchange"), and induction (polarization) energies, although the latter contains the CT energy. Polarization and CT are separated by using cDFT to define the latter, in a manner that is only weakly dependent on basis set and has a well-defined, non-zero value in the limit of a complete basis . Finally, the dispersion energy is provided by the accurate SAPT-based +D3 atom–atom dispersion potential,[57] which has been fit to high-level third-order SAPT calculations. If the

+D3 potential is suspect, the dispersion energy could be computed directly using third-order SAPT, albeit at increased cost.

CHAPTER 12

Conclusion and Future Directions

Since binding interactions in biological and materials systems are too large to model in full with high accuracy quantum chemistry, we must make intelligent approximations in order to make a computationally feasible model. This has led us to develop a fragmentation approach called XSAPT where the full system is divided into small subsystems and the computational bottleneck in the full system can be eliminated.

We demonstrated that the XSAPT fragmentation method based on long-range-corrected Kohn-Sham (KS) orbitals combined with dispersion potentials (+D), XSAPT(KS)+D, not only reduces the computational cost (only linear scaling) with respect to the numbers of fragments but also obtains the accurate binding energy (with chemical accuracy) between molecules. The accuracy of XSAPT(KS)+D method is typically superior to alternative *ab initio* methods with similar scaling.

For the biologically-relevant data set, XSAPT(KS)+D method exhibits reasonably small errors and compete with other high-level wavefunction methods. It also bears mention that all of the methods that outperform the XSAPT(KS)+D method exhibit at least $\mathcal{O}(N_s^5)$ with respect to the size of the *whole system*, whereas the XSAPT(KS) family of methods scale as $\mathcal{O}(N_f^3)$ with respect to size of a *fragment*. For the real

application, we consider intercalation of the anti-cancer agent ellipticine into DNA, which involves insertion between two Watson-Crick CG base pairs, linked by their respective phosphate sugar puckers which consists of 157 atoms. The binding energy predicted by our XSAPT method is within the statistical error bars of the quantum Monte Carlo benchmark. Thus, the accurate binding energy between DNA and drug molecule can be reached. Furthermore, the XSAPT method has also been applied in molecular and ionic clusters, clathrates, and supramolecular complexes.

Moreover, the monomer-based nature of XSAPT calculations makes them trivially parallelizable, such that wall times scale linearly with respect to the number of monomer units. These characteristics make XSAPT a promising method for use in different fields, such as in fragment-based drug design to prescreen of large numbers of potential drug molecule.

A number of future extensions of this work are currently underway and will be reported in the near future. We summarize them here. XSAPT can potentially be used in drug design. We will use low-cost classical molecular dynamics method to generate a series of reasonable binding positions between drug molecules and the binding pocket of biomolecules, and build a molecular binding library. Then, XSAPT can be used to predict the placement and binding energy of the drug-biomolecule in that library and compare the performance of those traditional empirical scoring functions. Hopefully, our efficient and accurate XSAPT method can replace the low-level empirical scoring functions or better empirical scoring functions can be designed based on our XSAPT method in the future. The XSAPT methodology can also be extended to be used in

molecular crystals by implementing Ewald summation for the long-range Coulomb interactions. Finally, dispersion terms in XSAPT can be calculated in full *ab initio* way based on frequency-dependent density susceptibilities (FDDS). In principle, XSAPT combined with these FDDS-based dispersion interactions should improve the accurate description of intermolecular interactions. The research described in this dissertation along with these continuing efforts opens the door to both qualitative and quantitative studies of noncovalent interactions by XSAPT-based methods in clusters, biomolecules, and condensed-phase systems.

APPENDIX A

Supporting Information for: "Understanding the many-body expansion for large systems"

Tables A.1 and A.2 present what we believe is a nearly exhaustive list of MBE(2) and MBE(3) results from the literature, subject to the following caveats:

- We consider only clusters with $N > 10$ monomers.
- We only consider cases where the MBE calculation was compared against a supersystem benchmark computed at the same level of theory.
- Multi-layer methods, and those based on overlapping fragments, have been excluded from this comparison for simplicity, although many such benchmarks for water clusters can be found in the literature.[72,78,513–518]

That said, direct comparison of the various literature calculations presented in these tables must proceed with caution. In particular, the results in the tables correspond to different structures (*e.g.*, the 16-mer examined in Ref. 519 is different from the one considered here). Differences in electrostatic embeddings and, in a few cases, the use of two H_2O molecules per fragment are noted in footnotes to Tables A.1 and A.2. Unless care is taken to ensure that all of these points are addressed similarly between

implementations, finite-precision issues can easily cause a user to obtain results that are different from those in the literature.[48]

Table A.1: Mean absolute errors (in kcal/mol) in EE-2B interaction energies for $(H_2O)_N$ clusters, as predicted by various fragment-based methods. Only published calculations with $N > 10$ are considered.

N	Ref.	Method	Basis Set					
			STO-3G	3-21G	6-31G*	6-31++G**	6-311++G(3df,2p)	other
16	519	HF[a]	0.15	5.70	4.97	1.00		
	519	HF[a,b]	0.10	4.12	3.54	0.42		
	520	B3LYP[a]			9.83	6.78		
	520	B3LYP[a,b]			7.06	4.42		
	517	MP2[a]			4.64	2.88	10.88	
	87	CCSD(T)						3.59[c]
20	517	MP2[a]			6.20	4.00	14.00	
	521	MP2[c]						2.01[c]
21	84	MP2						2.97[d]
32	519	HF[a]	0.48	15.66	13.2	4.51		
	519	HF[a,b]	0.43	13.1	10.69	3.22		
	520	B3LYP[a]			26.27	23.58		
	520	B3LYP[a,b]			21.5	18.28		
	517	MP2[a]			10.56	12.16	26.88	
57	72	B3LYP						20.52[e]
64	519	HF[a]	1.29	38.6	32.73	12.06		
	519	HF[a,b]	0.94	27.37	23.15	8.81		
	520	B3LYP[a]			65.63	64.24		
	520	B3LYP[a,b]			46.85	48.2		
	517	MP2[a]			22.4	27.52		

[a] EE is defined to be density embedding at close range, density embedding with the Mulliken approximation at mid-range, and Mulliken point-charge embedding at long range.[522]

[b] Two water molecules per fragment.

[c] Basis set is aug-cc-pVTZ.

[d] Basis set is aug-cc-pVTZ for oxygen and cc-pVTZ for hydrogen.

[e] Basis set is 6-31+G(d,2p).

Table A.2: Mean absolute errors (kcal/mol) in EE-3B interaction energies for water clusters, as predicted by various fragment-based methods. Only published calculations with $N > 10$ are considered.

N	Ref.	Method	Basis Set					
			STO-3G	3-21G	6-31G*	6-31++G**	6-311++G(3df,2p)	other
	519	HF[a]	0.01	0.28	0.39	1.17		
	519	HF[a,b]	0.01	0.23	0.26	0.40		
16	520	B3LYP[a]			0.51	3.12		
	520	B3LYP[a,b]			0.36	0.78		
	517	MP2[a]			1.44	0.80	0.64	
	87	CCSD(T)						1.52[c]
20	517	MP2[a]			1.8	0.40	0.40	
	521	MP2[d]						1.56[c]
21	84	MP2						0.38[e]
	519	HF[a]	0.04	0.80	1.33	3.71		
	519	HF[a,b]	0.02	0.57	0.93	1.67		
32	520	B3LYP[a]			1.85	14.71		
	520	B3LYP[a,b]			1.36	0.39		
	517	MP2[a]			3.52	2.24	1.28	
64	517	MP2[a]		6.4	3.2			

[a] EE is defined to be density embedding at close range, density embedding with the Mulliken approximation at mid-range, and Mulliken point-charge embedding at long range.[522]

[b] Two water molecules per fragment.

[c] Basis set is aug-cc-pVTZ.

[d] EE uses point dipoles instead of point charges.

[e] Basis set is aug-cc-pVTZ for oxygen and cc-pVTZ for hydrogen.

APPENDIX B

Supporting Information for: "Breakdown of the single-exchange approximation in third-order symmetry-adapted perturbation theory"[B.1]

B.1 CCSD(T)/CBS benchmarks

B.1.1 Helium dimer

The CCSD(T) single-point energy calculations were carried out at each point on potential energy curves using the aug-cc-pVTZ (aTZ) and aug-cc-pVQZ (aQZ) basis sets. Then the CCSD(T) interaction energies calculated by both basis sets were extrapolated to the complete basis set (CBS) limit by using the two-point scheme suggested by Halkier *et al.*:[190,523]

$$E_{\mathrm{CBS}}(X,Y) = \frac{X^3 E_X - Y^3 E_Y}{X^3 - Y^3} \tag{B.1}$$

where $X = 4$ (aQZ) and $Y = 3$ (aTZ) in this case.

B.1.2 Anion–water and water–water dimers

In this case, MP2/aTZ and MP2/aQZ single-point energy calculations were carried out at each point on the potential energy curves and extrapolated to the CBS limit

[B.1] This chapter appeared as the Supporting Information to the full article in the *Journal of Physical Chemistry A*, in 2012, volume 116, pages 3042–3047.

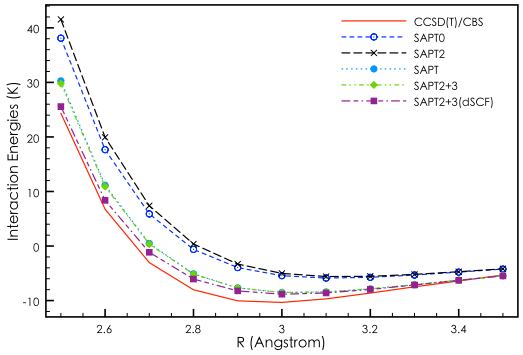

Figure B.1: Comparison of performance of different levels of SAPT on the He$\cdots$He system. SAPT2+3_dSCF means the interaction energies are calculated at the SAPT2+3 level with $E^{(30)}_{\rm ind}$ and $E^{(30)}_{\rm exch-ind}$ substituted by $\delta E^{(2)}_{\rm int}$. R is the distance between two helium atoms.

using Eq. (B.1). From these MP2/CBS energies, a triples correction was estimated by comparing MP2 and CCSD(T) results in the aug-cc-pVTZ basis set. The CCSD(T)/CBS result is thus estimated according to

$$E_{\rm CBS}[{\rm CCSD(T)}] = E_{\rm CBS}[{\rm MP2}] + E_{\rm aTZ}[{\rm CCSD(T)}] - E_{\rm aTZ}[{\rm MP2}] \ . \tag{B.2}$$

This procedure is based on the assumption that the difference between interaction energies by CCSD(T) and MP2 converges rapidly with respect to basis set.

B.2 Additional figures

A variety of potential energy curves for $F^-(H_2O)$ were presented in the paper. Below, we plot all of the same quantities for $Cl^-(H_2O)$, $HO^-(H_2O)$, $(H_2O)_2$, and He_2.

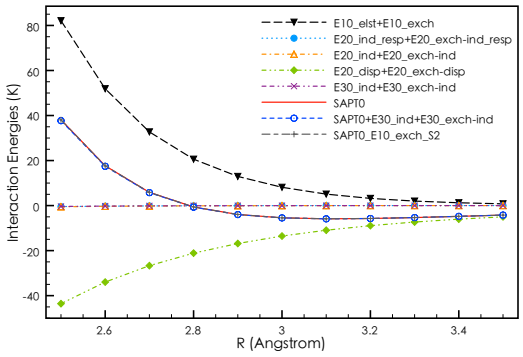

Figure B.2: SAPT decomposition of the interaction energy of the He $\cdots$ He system into different orders of electrostatic (elst), induction (ind), and dispersion (disp) with their corresponding exchange (exch) contributions. SAPT0+$E_{\rm ind}^{(30)}$+$E_{\rm exch-ind}^{(30)}$ means the interaction energies are calculated at the SAPT0 level plus $E_{\rm ind}^{(30)}$ and $E_{\rm exch-ind}^{(30)}$. R is the distance between two helium atoms.

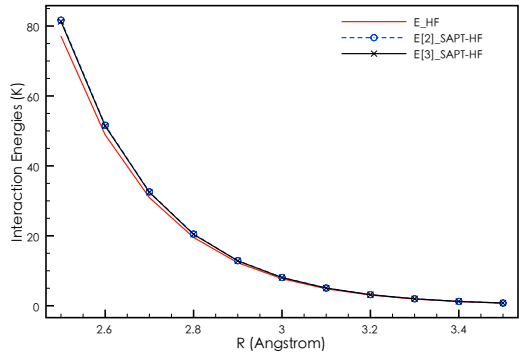

Figure B.3: The supermolecular Hatree-Fock interaction energies $E_{\rm int}^{\rm HF}$ and approximate Hartree-Fock interaction energies composed by SAPT terms in second ($E_{\rm SAPT-HF}^{[2]}$) and third ($E_{\rm SAPT-HF}^{[2]}$) order for He $\cdots$ He system. R is the distance between two helium atoms.

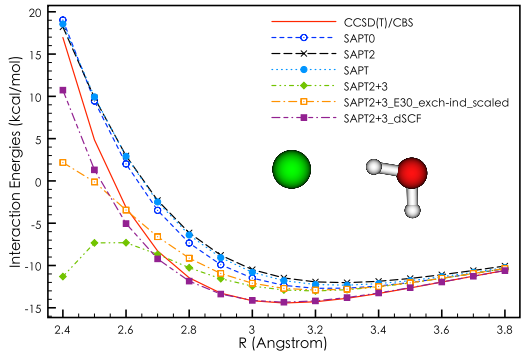

Figure B.4: Comparison of performance of different levels of SAPT on the $Cl^- \cdots H_2O$ system. SAPT2+3_E30_exch-ind_scaled means the interaction energies are calculated at the SAPT2+3 level with $E^{(30)}_{\text{exch}-\text{ind}}$ scaled by Eq. (7). SAPT2+3_dSCF means the interaction energies are calculated at the SAPT2+3 level with $E^{(30)}_{\text{ind}}$ and $E^{(30)}_{\text{exch}-\text{ind}}$ substituted by $\delta E^{(2)}_{\text{int}}$. R is the distance between chloride and oxygen atom.

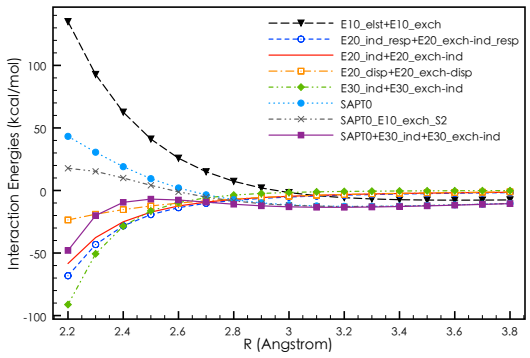

Figure B.5: SAPT decomposition of the interaction energy of the $Cl^- \cdots H_2O$ system into different orders of electrostatic (elst), induction (ind), and dispersion (disp) with their corresponding exchange (exch) contributions. SAPT0+E30_ind+E30_exch-ind means the interaction energies are calculated at the SAPT0 level plus $E^{(30)}_{\text{ind}}$ and $E^{(30)}_{\text{exch}-\text{ind}}$. R is the distance between chloride and oxygen atom.

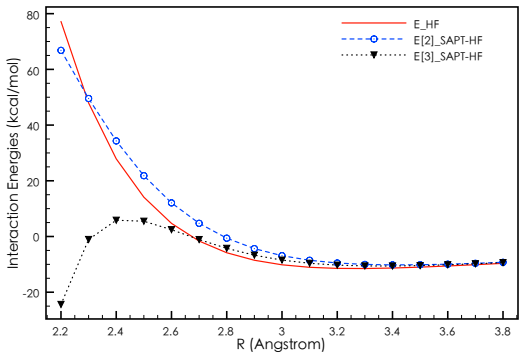

Figure B.6: The supermolecular Hatree-Fock interaction energies $E_{\text{int}}^{\text{HF}}$ and approximate Hartree-Fock interaction energies composed by SAPT terms in second ($E_{\text{SAPT}-\text{HF}}^{[2]}$) and third ($E_{\text{SAPT}-\text{HF}}^{[2]}$) order for $Cl^- \cdots H_2O$ system. R is the distance between chloride and oxygen atom.

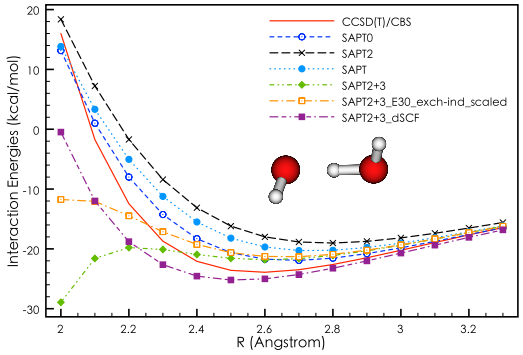

Figure B.7: Comparison of performance of different levels of SAPT on the $OH^- \cdots H_2O$ system. SAPT2+3_E30_exch-ind_scaled means the interaction energies are calculated at the SAPT2+3 level with $E_{\text{exch}-\text{ind}}^{(30)}$ scaled by Eq. (7). SAPT2+3_dSCF means the interaction energies are calculated at the SAPT2+3 level with $E_{\text{ind}}^{(30)}$ and $E_{\text{exch}-\text{ind}}^{(30)}$ substituted by $\delta E_{\text{int}}^{(2)}$. R is the distance between two oxygen atoms.

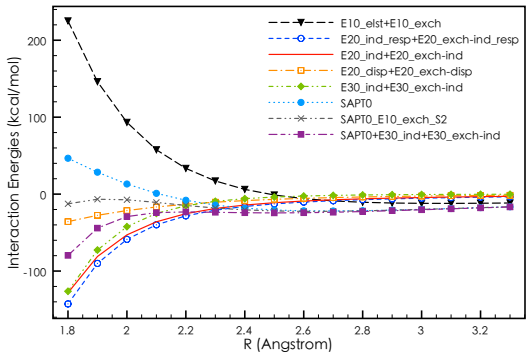

Figure B.8: SAPT decomposition of the interaction energy of the $OH^- \cdots H_2O$ system into different orders of electrostatic (elst), induction (ind), and dispersion (disp) with their corresponding exchange (exch) contributions. SAPT0+E30_ind+E30_exch-ind means the interaction energies are calculated at the SAPT0 level plus $E_{\rm ind}^{(30)}$ and $E_{\rm exch-ind}^{(30)}$. R is the distance between two oxygen atoms.

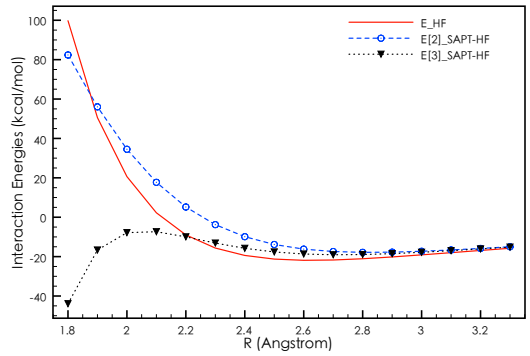

Figure B.9: The supermolecular Hatree-Fock interaction energies $E_{\rm int}^{\rm HF}$ and approximate Hartree-Fock interaction energies composed by SAPT terms in second ($E_{\rm SAPT-HF}^{[2]}$) and third ($E_{\rm SAPT-HF}^{[2]}$) order for $OH^- \cdots H_2O$ system. R is the distance between two oxygen atoms.

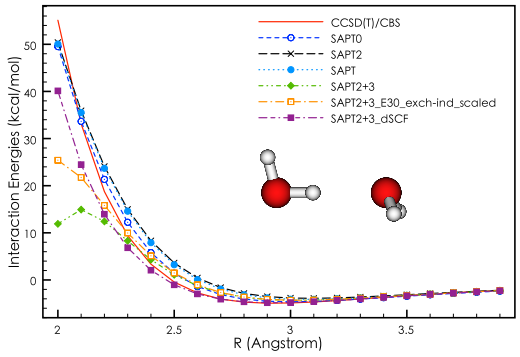

Figure B.10: Comparison of performance of different levels of SAPT on the $H_2O\cdots H_2O$ system. SAPT2+3_E30_exch-ind_scaled means the interaction energies are calculated at the SAPT2+3 level with $E^{(30)}_{\mathrm{exch-ind}}$ scaled by Eq. (7). SAPT2+3_dSCF means the interaction energies are calculated at the SAPT2+3 level with $E^{(30)}_{\mathrm{ind}}$ and $E^{(30)}_{\mathrm{exch-ind}}$ substituted by $\delta E^{(2)}_{\mathrm{int}}$. R is the distance between two oxygen atoms.

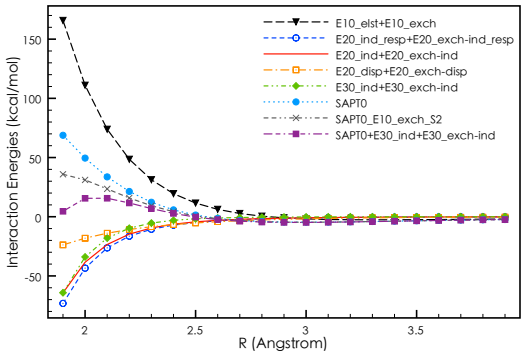

Figure B.11: SAPT decomposition of the interaction energy of the $H_2O\cdots H_2O$ system into different orders of electrostatic (elst), induction (ind), and dispersion (disp) with their corresponding exchange (exch) contributions. SAPT0+E30_ind+E30_exch-ind means the interaction energies are calculated at the SAPT0 level plus $E^{(30)}_{\mathrm{ind}}$ and $E^{(30)}_{\mathrm{exch-ind}}$. R is the distance between two oxygen atoms.

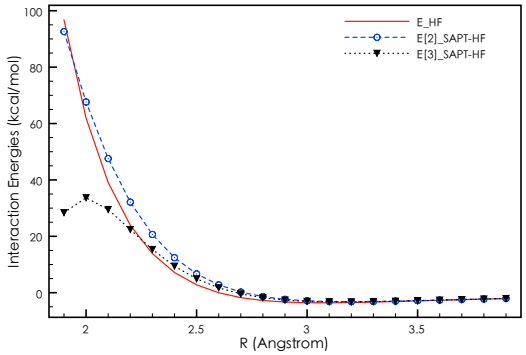

Figure B.12: The supermolecular Hatree-Fock interaction energies $E_{\rm int}^{\rm HF}$ and approximate Hartree-Fock interaction energies composed by SAPT terms in second ($E_{\rm SAPT-HF}^{[2]}$) and third ($E_{\rm SAPT-HF}^{[2]}$) order for $H_2O\cdots H_2O$ system. R is the distance between two oxygen atoms.

APPENDIX C

Supporting Information for: "Accurate description of intermolecular interactions involving ions using symmetry-adapted perturbation theory"[C.1]

Summary

- Errors engendered by the single-exchange approximation (SEA) for various exchange energies terms are listed in Table C.1 (aDZ basis set) and Table C.2 (aQZ basis set). These should be compared to the aTZ results in Table III of the paper.

- For the AHB21 data set, the δMP2 correction as well as the scaled-exchange corrections $[p_{\mathrm{ex}}(\alpha = 2) - 1]E^{(20)}_{\text{exch-ind,resp}}$ and $[p_{\mathrm{ex}}(\alpha = 2) - 1]{}^{t}E^{(22)}_{\text{exch-ind}}$ are shown in Figure C.1 (aDZ basis set) and Figure C.2 (aQZ basis set). These should be compared to the aTZ results in Fig. 3 of the paper.

- For the CHB6 data set, the δMP2 correction as well as the scaled-exchange corrections $[p_{\mathrm{ex}}(\alpha = 2) - 1]E^{(20)}_{\text{exch-ind,resp}}$, and $[p_{\mathrm{ex}}(\alpha = 2) - 1]{}^{t}E^{(22)}_{\text{exch-ind}}$ are shown in Figure C.3 (aDZ basis set) and Figure C.4 (aQZ basis set). These should be compared to the aTZ results in Fig. 3 of the paper.

[C.1] This chapter appeared as the Supporting Information to the full article in the *Journal of Chemical Theory and Computation*, in 2015, volume 11, pages 2473–2486.

- Table C.3 shows the performance of the SAPT2+3-δMP2/aug-cc-pVXZ method (for X = D, T, Q), as applied to AHB21, for different versions of the scaling of the second-order exchange energies. These varieties include no scaling at all, or else $\alpha = 1$ or $\alpha = 3$ in Eq. (4) of the paper. For comparison, with the aTZ basis the value $\alpha = 2$ affords a MAE of 0.49 kcal/mol (see Table IV).

Table C.1: Mean absolute errors (MAEs) and maximum errors in exchange energies based on HF-SAPT/aDZ and DFT-SAPT/aDZ with the S^2 approximation for the AHB21 and CHB6 data sets, with respect to exchange energies without the S^2 approximation.

Exchange Term	error / kcal mol^{-1}		
	MAE	Maximum	
		value	system
——AHB21——			
HF-SAPT			
$E_{\text{exch}}^{(10)}(S^2)$	1.04	6.31	Cl^-(HCl)
$E_{\text{exch-ind,resp}}^{(20)}(S^2)$	1.01	8.22	Cl^-(HCl)
$p_{\text{ex}}(\alpha = 2.07)E_{\text{exch-ind,resp}}^{(20)}(S^2)$	0.08	0.36	F^-(HF)
$E_{\text{exch-disp}}^{(20)}(S^2)$	0.07	0.35	F^-(HF)
${}^tE_{\text{exch-ind}}^{(22)}(S^2)$	0.13	0.63	Cl^-(HCl)
$p_{\text{ex}}(\alpha = 2){}^tE_{\text{exch-ind}}^{(22)}(S^2)$	0.01	0.09	F^-(HF)
DFT-SAPT			
$E_{\text{exch}}^{(1)}(S^2)$	1.14	6.30	Cl^-(HCl)
$E_{\text{exch-ind}}^{(2)}(S^2)$	1.27	9.02	Cl^-(HCl)
$p_{\text{ex}}(\alpha = 2.06)E_{\text{exch-ind}}^{(2)}(S^2)$	0.09	0.61	Cl^-(HCl)
$E_{\text{exch-disp}}^{(2)}(S^2)$	0.12	0.53	F^-(HF)
——CHB6——			
HF-SAPT			
$E_{\text{exch}}^{(10)}(S^2)$	0.03	0.05	$Li^+(C_6H_6)$
$E_{\text{exch-ind,resp}}^{(20)}(S^2)$	0.06	0.08	$Li^+(C_6H_6)$
$p_{\text{ex}}(\alpha = 2.20)E_{\text{exch-ind,resp}}^{(20)}(S^2)$	0.00	0.01	$K^+(C_6H_6)$
$E_{\text{exch-disp}}^{(20)}(S^2)$	0.00	0.01	$K^+(C_6H_6)$
${}^tE_{\text{exch-ind}}^{(22)}(S^2)$	0.00	0.01	$K^+(H_2O)$
$p_{\text{ex}}(\alpha = 2){}^tE_{\text{exch-ind}}^{(22)}(S^2)$	0.00	0.01	$Li^+(H_2O)$
DFT-SAPT			
$E_{\text{exch}}^{(1)}(S^2)$	0.03	0.05	$Li^+(H_2O)$
$E_{\text{exch-ind}}^{(2)}(S^2)$	0.06	0.08	$Li^+(C_6H_6)$
$p_{\text{ex}}(\alpha = 2.19)E_{\text{exch-ind}}^{(2)}(S^2)$	0.00	0.01	$K^+(C_6H_6)$
$E_{\text{exch-disp}}^{(2)}(S^2)$	0.00	0.01	$K^+(C_6H_6)$

Table C.2: Mean absolute errors (MAEs) and maximum errors in exchange energies based on HF-SAPT/aQZ and DFT-SAPT/aQZ with the S^2 approximation for the AHB21 and CHB6 data sets, with respect to exchange energies without the S^2 approximation.

Exchange Term	error / kcal mol^{-1}		
	MAE	Maximum	
		value	system
——AHB21——			
HF-SAPT			
$E_{\text{exch}}^{(10)}(S^2)$	1.05	6.36	Cl^-(HCl)
$E_{\text{exch-ind,resp}}^{(20)}(S^2)$	1.06	8.51	Cl^-(HCl)
$p_{\text{ex}}(\alpha = 2.04)E_{\text{exch-ind,resp}}^{(20)}(S^2)$	0.07	0.31	F^-(HF)
$E_{\text{exch-disp}}^{(20)}(S^2)$	0.07	0.35	F^-(HF)
${}^tE_{\text{exch-ind}}^{(22)}(S^2)$	0.09	0.47	F^-(HF)
$p_{\text{ex}}(\alpha = 2){}^tE_{\text{exch-ind}}^{(22)}(S^2)$	0.01	0.06	F^-(HF)
DFT-SAPT			
$E_{\text{exch}}^{(1)}(S^2)$	1.15	6.32	Cl^-(HCl)
$E_{\text{exch-ind}}^{(2)}(S^2)$	1.33	9.34	Cl^-(HCl)
$p_{\text{ex}}(\alpha = 2.07)E_{\text{exch-ind}}^{(2)}(S^2)$	0.09	0.84	Cl^-(HCl)
$E_{\text{exch-disp}}^{(2)}(S^2)$	0.11	0.51	F^-(HF)
——CHB6——			
HF-SAPT			
$E_{\text{exch}}^{(10)}(S^2)$	0.03	0.05	$Li^+(H_2O)$
$E_{\text{exch-ind,resp}}^{(20)}(S^2)$	0.06	0.10	$Li^+(C_6H_6)$
$p_{\text{ex}}(\alpha = 2.19)E_{\text{exch-ind,resp}}^{(20)}(S^2)$	0.00	0.01	$K^+(C_6H_6)$
$E_{\text{exch-disp}}^{(20)}(S^2)$	0.00	0.01	$K^+(C_6H_6)$
${}^tE_{\text{exch-ind}}^{(22)}(S^2)$	0.00	0.01	$K^+(H_2O)$
$p_{\text{ex}}(\alpha = 2){}^tE_{\text{exch-ind}}^{(22)}(S^2)$	0.00	0.00	$Li^+(H_2O)$
DFT-SAPT			
$E_{\text{exch}}^{(1)}(S^2)$	0.03	0.06	$Li^+(H_2O)$
$E_{\text{exch-ind}}^{(2)}(S^2)$	0.07	0.10	$Li^+(H_2O)$
$p_{\text{ex}}(\alpha = 2.19)E_{\text{exch-ind}}^{(2)}(S^2)$	0.00	0.01	$K^+(C_6H_6)$
$E_{\text{exch-disp}}^{(2)}(S^2)$	0.00	0.01	$K^+(C_6H_6)$

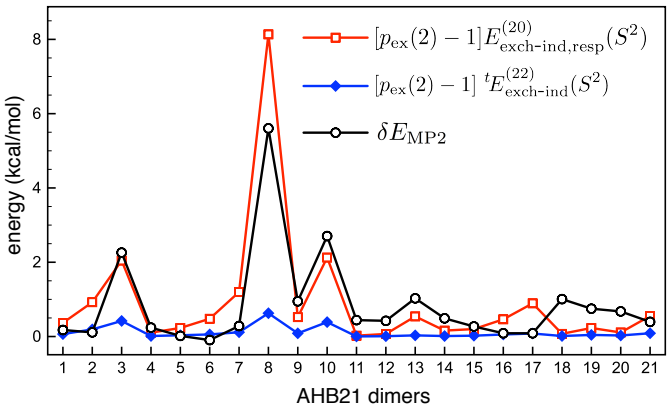

Figure C.1: Magnitude of the δMP2 correction and the scaled-exchange corrections $[p_{ex}(2.0)-1]E^{(20)}_{exch\text{-}ind,resp}$ and $[p_{ex}(2.0)-1]{}^tE^{(22)}_{exch\text{-}ind}$, for the AHB21 data set using the aDZ basis set.

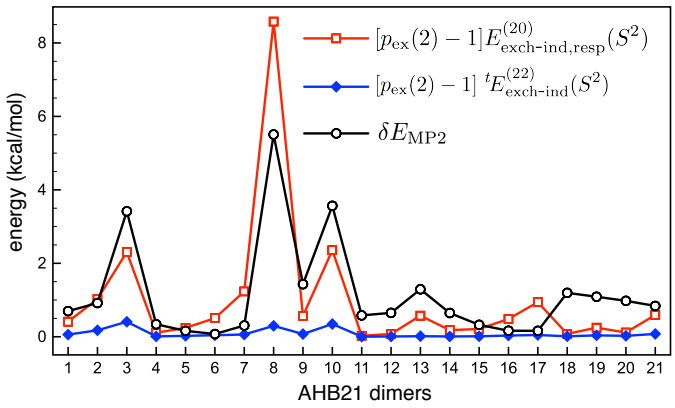

Figure C.2: Magnitude of the δMP2 correction and the scaled-exchange corrections $[p_{ex}(2.0)-1]E^{(20)}_{exch\text{-}ind,resp}$ and $[p_{ex}(2.0)-1]{}^tE^{(22)}_{exch\text{-}ind}$, for the AHB21 data set using the aQZ basis set.

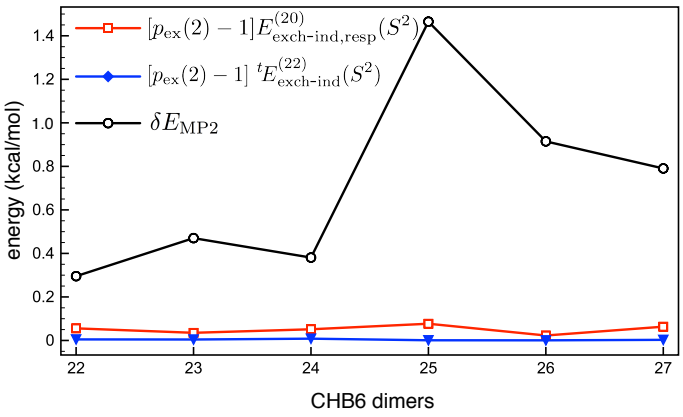

Figure C.3: Magnitude of the δMP2 correction, the exchange correction $[p_{ex}(\alpha)-1]E^{(20)}_{exch\text{-}ind,resp}$, and $[p_{ex}(\alpha)-1]{}^tE^{(22)}_{exch\text{-}ind}$, where $p_{ex}(\alpha)$ is the scaling factor and $\alpha = 2$ is employed, for the CHB6 data set, using the aDZ basis set.

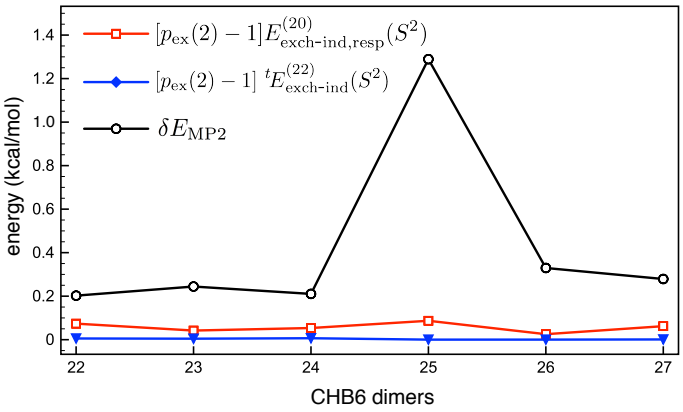

Figure C.4: Magnitude of the δMP2 correction, the exchange correction $[p_{ex}(\alpha)-1]E^{(20)}_{exch\text{-}ind,resp}$, and $[p_{ex}(\alpha)-1]{}^tE^{(22)}_{exch\text{-}ind}$, where $p_{ex}(\alpha)$ is the scaling factor and $\alpha = 2$ is employed, for the CHB6 data set, using the aQZ basis set.

Table C.3: MAEs (in kcal/mol) for the AHB21 data set computed at the SAPT2+3-δMP2 level with different scaled versions of the exchange corrections $E^{(20)}_{\text{exch-ind,resp}}$ and ${}^{t}E^{(22)}_{\text{exch-ind}}$.

Basis Set	all AHB21			$X^- \cdots H^+ \cdots X^-$ only		
	no scaling	$\alpha = 1$	$\alpha = 3$	no scaling	$\alpha = 1$	$\alpha = 3$
aDZ	0.96	1.03	1.17	1.00	1.24	1.76
aTZ	0.49	0.45	0.36	1.80	1.54	0.96
aQZ	0.73	0.66	0.53	2.63	2.35	1.73

APPENDIX D

Supporting Information for: "Symmetry-adapted perturbation theory with Kohn-Sham orbitals using non-empirically tuned, long-range-corrected density functionals"[D.1]

D.1 Range separation parameters

Tables D.1 and D.2 list the "tuned" value of the range separation parameter, ω, for each of the monomers considered in this work. These values were tuned using the aug-cc-pV6Z basis set for He, the aug-cc-pV5Z basis set for Ne, the aug-cc-pVQZ basis set for the SS41 data set, and the aug-cc-pVTZ basis set for both the S22 and S66 data sets. Optimized values are computed separately for the LRC-ωPBE functional[240] and the LRC-ωPBEh functional.[227]

[D.1] This chapter appeared as the Supporting Information to the full article in the *Journal of Chemical Physics*, in 2014, volume 140, pages 044108:1–8.

D.2 Benchmark energy components for the SS41 data set

Energy components for SS41 computed at the SAPT2+(3)/aug-cc-pVQZ level are listed in Table D.3. The SAPT2+(3) is defined as follows:[117]

$$\begin{aligned} E_{\text{SAPT2+(3)}} = & E_{\text{elst}}^{(10)} + E_{\text{elst,resp}}^{(12)} + E_{\text{elst,resp}}^{(13)} \\ & + E_{\text{exch}}^{(10)} + E_{\text{exch}}^{(11)} + E_{\text{exch}}^{(12)} \\ & + E_{\text{ind,resp}}^{(20)} + E_{\text{exch-ind,resp}}^{(20)} + {}^{t}E_{\text{ind}}^{(22)} + {}^{t}E_{\text{exch-ind}}^{(22)} + \delta E_{\text{int,resp}}^{\text{HF}} \\ & + E_{\text{disp}}^{(20)} + E_{\text{exch-disp}}^{(20)} + E_{\text{disp}}^{(21)} + E_{\text{disp}}^{(22)} + E_{\text{disp}}^{(30)} \ . \end{aligned} \tag{D.1}$$

The energy components are grouped as follows:[37]

$$E_{\text{electrostatic}} = E_{\text{elst}}^{(10)} + E_{\text{elst,resp}}^{(12)} + E_{\text{elst,resp}}^{(13)} \ , \tag{D.2}$$

$$E_{\text{exchange}} = E_{\text{exch}}^{(10)} + E_{\text{exch}}^{(11)} + E_{\text{exch}}^{(12)} \ , \tag{D.3}$$

$$E_{\text{induction}} = E_{\text{ind,resp}}^{(20)} + E_{\text{exch-ind,resp}}^{(20)} + {}^{t}E_{\text{ind}}^{(22)} + {}^{t}E_{\text{exch-ind}}^{(22)} + \delta E_{\text{int,resp}}^{\text{HF}} \ , \tag{D.4}$$

$$E_{\text{dispersion}} = E_{\text{disp}}^{(20)} + E_{\text{exch-disp}}^{(20)} + E_{\text{disp}}^{(21)} + E_{\text{disp}}^{(22)} + E_{\text{disp}}^{(30)} \ . \tag{D.5}$$

	ω/bohr^{-1}	
Monomer	LRC-ωPBE	LRC-ωPBEh
—**He_2 and Ne_2**—		
He	1.025	0.900
Ne	0.800	0.650
—**SS41 molecules**—		
Ar	0.550	0.475
ethene	0.375	0.300
methane	0.450	0.400
borane	0.475	0.425
ethyne	0.400	0.350
water	0.500	0.400
methanol	0.425	0.350
CH_3F	0.500	0.425
ethane	0.425	0.375
F_2	0.675	0.550
formaldehyde	0.450	0.350
formamide	0.375	0.300
formic acid	0.400	0.325
CH_3NH_2	0.400	0.325
ammonia	0.450	0.350
HCN	0.450	0.400
HF	0.600	0.500

Table D.1: Tuned range separation parameters for various monomers in He_2, Ne_2, and SS41 data set. The basis sets used are aug-cc-pV6Z basis set for He, aug-cc-pV5Z for Ne, and aug-cc-pVQZ basis set for the monomers in the SS41 data set.

	ω/bohr^{-1}	
Monomer	LRC-ωPBE	LRC-ωPBEh
—S22 molecules—		
adenine	0.275	0.225
2-aminopyridine	0.300	0.225
benzene	0.275	0.225
ethyne	0.400	0.350
ethene	0.350	0.300
methane	0.450	0.400
formamide	0.375	0.300
formic acid	0.425	0.325
water	0.500	0.400
HCN	0.550	0.500
indole	0.275	0.225
ammonia	0.450	0.350
phenol	0.300	0.250
pyrazine	0.375	0.300
2-pyridoxine	0.300	0.250
thymine	0.625	0.525
uracil	0.475	0.375
—S66 molecules—		
methylamine	0.400	0.325
methanol	0.425	0.350
$AcNH_2$	0.350	0.275
AcOH	0.375	0.300
cyclopentane	0.425	0.375
neopentane	0.300	0.250
pentane	0.300	0.250
peptide	0.325	0.250
pyridine	0.300	0.225

Table D.2: Tuned range separation parameters for various monomers. The aug-cc-pVTZ basis set in each case.

system	Electrostatics	Exchange-Repulsion	Induction	Dispersion	Total
		—A24 dimers—			
H_2O-NH_3	-10.99	11.25	-3.76	-3.06	-6.57
H_2O dimer	-7.90	7.59	-2.35	-2.37	-5.03
HCN dimer	-5.91	4.10	-1.54	-1.74	-5.10
HF dimer	-6.41	6.04	-2.36	-1.83	-4.56
NH_3 dimer	-4.75	4.48	-0.86	-2.04	-3.17
HF-CH_4	-1.39	2.51	-1.25	-1.50	-1.62
NH_3-CH_4	-0.95	1.42	-0.38	-0.93	-0.83
H_2O-CH_4	-0.70	1.06	-0.27	-0.80	-0.71
CH_2O dimer	-6.57	8.43	-2.37	-4.28	-4.78
H_2O-C_2H_4	-3.45	4.38	-1.28	-2.08	-2.42
CH_2O-C_2H_4	-1.70	2.55	-0.49	-2.09	-1.73
C_2H_2 dimer (TS)	-1.83	2.06	-0.54	-1.31	-1.62
NH_3-C_2H_4	-1.67	2.35	-0.48	-1.51	-1.31
C_2H_4 dimer (TS)	-0.84	2.11	-0.30	-2.02	-1.06
CH_4-C_2H_4	-0.42	0.92	-0.13	-0.87	-0.50
BH_3-CH_4	-1.50	3.88	-1.08	-2.68	-1.38
CH_4-C_2H_6 (non-linear)	-0.36	1.22	-0.07	-1.66	-0.86
CH_4-C_2H_6 (linear)	-0.23	0.87	-0.05	-1.23	-0.64
CH_4-dimer	-0.20	0.77	-0.04	-1.09	-0.56
Ar-CH_4	-0.17	0.59	-0.03	-0.79	-0.41
Ar-C_2H_4	-0.23	0.72	-0.05	-0.74	-0.31
C_2H_4-C_2H_2	-0.32	3.68	-0.25	-2.11	1.00
C_2H_4 dimer (PS)	-0.52	4.45	-0.28	-2.40	1.25
C_2H_2 dimer (PS)	0.23	3.04	-0.21	-1.89	1.17
		—S22 dimers—			
HCOOH dimer	-32.23	40.78	-18.90	-10.04	-20.38
$HCONH_2$ dimer	-25.39	27.79	-11.25	-8.10	-16.95
		—S66 dimers—			
H_2O-CH_3OH	-8.82	9.15	-2.88	-3.11	-5.65
H_2O-CH_3NH2	-12.07	12.84	-4.31	-3.53	-7.06
CH_3OH dimer	-8.92	9.67	-3.04	-3.53	-5.84
CH_3OH-CH_3NH2	-12.69	14.41	-4.81	-4.66	-7.74
CH_3OH-H_2O	-7.78	7.78	-2.45	-2.66	-5.11
CH_3NH2-CH_3OH	-4.12	4.85	-1.11	-2.77	-3.15
CH_3NH2 dimer	-5.82	7.04	-1.65	-3.85	-4.28
CH_3NH2-H_2O	-12.21	13.25	-4.44	-4.00	-7.40
C_2H_2-H_2O	-3.93	3.37	-1.01	-1.54	-3.11
		—X40 dimers—			
CH_4-F_2	-0.37	0.93	-0.13	-0.91	-0.48
CH_3F-CH_4	-0.37	1.09	-0.11	-1.42	-0.80
CH_3F dimer	-1.45	1.29	-0.23	-1.33	-1.72
HF-CH_3OH	-13.85	14.04	-6.34	-3.58	-9.74
HF-CH_3NH_2	-23.29	25.65	-11.82	-5.23	-14.69
CH_3F-CH_3OH	-5.18	5.36	-1.39	-2.65	-3.86

Table D.3: The energy components and total binding energies calculated by SAPT2+(3)/aug-cc-pVQZ method for SS41 data set.

APPENDIX E

Supporting Information for: " Accurate intermolecular interactions at dramatically reduced cost: XPol+SAPT with empirical dispersion"[E.1]

Optimized values of the range separation parameter, ω, for use with the LRC-ωPBEh functional are provided in Table E.1 for each monomer unit consider in this work. Figure E.1 shows some Ar$\cdots$Ne potential energy curves computed at various levels of theory.

[E.1] This chapter appeared as the Supporting Information to the full article in the *Journal of Physical Chemistry Letters*, in 2012, volume 3, pages 3241–3248.

Table E.1: The optimized range separation parameters (ω) for monomers.

Monomer	ω (bohr^{-1})	Ionization Potential (eV)
adenine[a]	0.05	9.8587
2-aminopyridine	0.10	8.2416
benzene	0.10	9.2163
ethyne	0.15	11.1043
ethene	0.15	10.2657
methane	0.20	14.2371
formamide[a]	0.05	10.1339
formic acid[a]	0.05	11.3846
water	0.10	12.3393
HCN	0.35	13.9377
indole	0.10	7.8064
ammonia	0.10	10.6694
phenol	0.10	8.5962
pyrazine	0.20	10.4075
2-pyridoxine	0.10	8.3881
thymine	0.25	10.0360
uracil	0.20	10.0986
acetate	0.05	3.9296
formaldehyde	0.10	10.7603
guanidinium	0.05	16.1640
imidazolium[a]	0.05	17.1969
methylamine	0.10	9.5708
methylammonium	0.20	20.6299
methanol	0.10	10.9272
$AcNH_2$[a]	0.05	9.6747
AcOH[a]	0.05	10.7562
cyclopentane	0.20	11.7103
neopentane	0.10	11.3508
pentane	0.10	11.3075
peptide[a]	0.05	9.4611
pyridine	0.10	9.6925

[a] This monomer does not meet the condition $\varepsilon_{\mathrm{HOMO}} = -\mathrm{IP}$.
Therefore, we set ω = 0.05 bohr^{-1} to let them as close as possible.

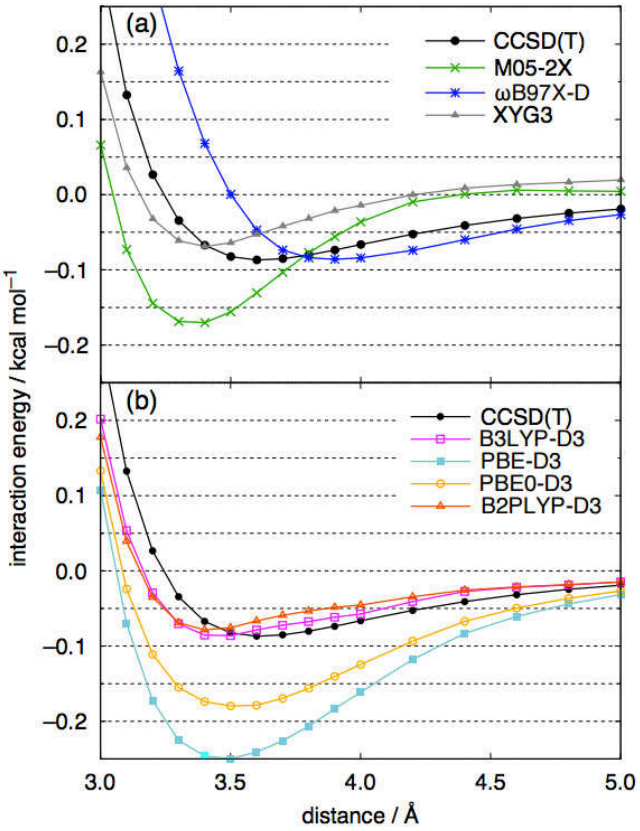

Figure E.1: Potential energy curves for Ar$\cdots$Ne, computed using the aug-cc-pVTZ basis set with counterpoise correction.

APPENDIX F

Supporting Information for: "An improved treatment of empirical dispersion and a many-body energy decomposition scheme for the explicit polarization plus symmetry-adapted perturbation theory (XSAPT) method"[F.1]

F.1 Parameters

Table F.1 lists the "tuned" value of the range separation parameter, ω, for each of the monomers considered in this work. These values were tuned using the haTZVPP basis set (a "heavy augmented" version of def2-TZVPP as described in the paper), and optimized values for both the LRC-ωPBE[240] and LRC-ωPBEh[227] functionals are listed. The "h" in LRC-ωPBEh indicates that the short-range exchange functional is a hybrid, and in our previous paper on XSAPT(KS)+D,[55] we used 60% Hartree-Fock exchange in the short-range ωPBEh component of this functional. This choice afforded good binding energies, but such a large value leads to errors in the exchange components of the interaction energy, as discussed in the present paper.

[F.1] This chapter appeared as the Supporting Information to the full article in the *Journal of Chemical Physics*, in 2013, volume 139, pages 034107:1–16.

In the second-generation (D2) version of the method, we use 20% Hartree-Fock exchange with LRC-ωPBEh (and 0% with LRC-ωPBE), which is consistent with the original parameterization of these functionals.[227,240] Notably, whereas in Ref. 55 we encountered seven cases in the S22 data set where the ω-tuning condition

$$\varepsilon_{\mathrm{HOMO}}(\omega) = -\mathrm{IP}(\omega) \tag{F.1}$$

could not be met, because the $\varepsilon_{\mathrm{HOMO}}(\omega)$ and $-\mathrm{IP}(\omega)$ curves did not cross, we encounter no such cases in the present work.

Table F.2 lists the fitted values of the s_β parameter for use in Hesselmann's dispersion potential,[16] which is the one used in our previous paper.[55] This is the potential called "+D" (as opposed to "+D2") in this work. Values of s_β in Table F.2 were optimized to reproduce S22A binding energies, in a basis-set-specific way.

F.2 Water hexamer benchmarks

We next wish to explain our choice of binding energy benchmarks for isomers of $(H_2O)_6$. As mentioned in the paper, a primary goal in examining $(H_2O)_6$ is to make comparison to energy decomposition analysis (EDA) results in Ref. 8, as a means to validate XSAPT-based EDA. These EDA benchmarks were computed at the MP2/a5Z-h//MP2/aTZ level and *do not* include monomer relaxation energy. On the other hand, prediction of binding energies is a major focus of this work and the most accurate binding energy benchmarks for $(H_2O)_6$ are probably the ones in Ref. 41, which were computed at the CCSD(T)/CBS//MP2/haTZ level. (The customized basis sets used in these benchmarks are defined as follows: a5Z-h means aug-cc-pV5Z

with h functions omitted, whereas haTZ is means aug-cc-pVTZ for oxygen and cc-pVTZ for hydrogen.)

Benchmark binding energies for eight isomers of $(H_2O)_6$, taken from the aforementioned references, are listed in Table F.3. The same eight isomers of $(H_2O)_6$ are considered in both papers, differing only in whether the aTZ of haTZ basis was used for the MP2 geometry optimizations. The presence or absence of diffuse functions on the hydrogen atoms likely makes very little difference in the actual geometries, as confirmed in Ref. 8. Three sets of binding energies are taken from Ref. 8, with each set computed at the MP2/a5Z-h//MP2/aTZ level of theory but differing based on whether counterpoise correction is included and whether monomer relaxation is included. In Ref. 41, monomer relaxation is always included but counterpoise correction is not (because all results are extrapolated to the CBS limit); MP2/CBS//MP2/haTZ and CCSD(T)/CBS//MP2/haTZ results are shown. Table F.4 lists some energy differences between these benchmarks.

Our strategy for benchmarking XSAPT(KS)+D2 in the case of $(H_2O)_6$ is as follows. We will use the MP2/aTZ geometries from Ref. 8, because those are the geometries at which the EDA has been performed, and for comparing energy components to XSAPT(KS)+D2 results we will use the MP2/a5Z-h EDA results in Ref. 8. On the other hand, we should examine binding energies as well (since we examine binding energies for all of the other systems considered in this paper), and we wish to do this at the same set of geometries, namely, the MP2/aTZ ones. To obtain approximate CCSD(T)/CBS //MP2/aTZ binding energies benchmarks, we add a triples

correction

$$\delta_{\mathrm{CCSD(T)}} = \mathrm{BE}_{\mathrm{CCSD(T)}} - \mathrm{BE}_{\mathrm{MP2}} \tag{F.2}$$

to the counterpoise-corrected MP2/a5Z-h//MP2/aTZ binding energies from Ref. 8. This correction is computed using the CCSD(T)/CBS and MP2/CBS binding energies (BEs) from Ref. 41. Finally, because we have computed $(H_2O)_6$ binding energies at a variety of levels of theory *without* considering monomer relaxation (*i.e.*, the binding energy is computed relative to monomers that retain their cluster geometries), we subtract the monomer relaxation energies δ_{relax} that are computed at the MP2/a5Z-h level in Ref. 8 and listed in Table F.4.

The resulting "corrected" binding energies are listed in Table F.5, under the column titled "+ triples − relaxation". (In the paper, this is the data set labeled "MP2/a5Z-h + $\delta_{\mathrm{CCSD(T)}}$" in Fig. 4.) The primary source of error in these data, relative to a true CCSD(T)/CBS //MP2/aTZ benchmark, is likely the fact that we have not extrapolated the MP2 energy to the basis-set limit, but rather have simply used MP2/a5Z-h values. Table F.4 shows the difference, δ_{MP2}, between the MP2/a5Z-h //MP2/aTZ binding energies from Ref. 8 and the MP2/CBS//MP2/haTZ results from Ref. 41; this difference is consistently 0.7–0.8 kcal/mol in magnitude for the eight isomers that are considered. The differences in cluster geometries likely have a very minor impact on this difference, so most of δ_{MP2} probably arises from basis-set incompleteness at the a5Z-h level. (The MP2/CBS results in Ref. 41 are converged to within 0.1 kcal/mol of MP2-R12 calculations performed in a modified aug-cc-pV5Z basis,[41] and are therefore likely within 0.1 kcal/mol of the actual basis-set limit.) Residual

basis-set superposition error is unlikely a major contribution to δ_{MP2}, since the MP2/CBS binding energies are larger than the counterpoise-corrected MP2/a5Z-h values (see Table F.5), and counterpoise-corrected binding energies converge from below in this system.[46]

Lastly, we may consider simply using the CCSD(T)/CBS//MP2/haTZ binding energy benchmarks from Ref. 41 at the MP2/aTZ geometries from Ref. 8. To compare with other results in the paper, we want binding energies using unrelaxed monomer geometries, so the CCSD(T)/CBS binding energies in Ref. 41 have been corrected using the MP2/a5Z-h values of δ_{relax} in Table F.4. These results constitute the final column of data in Table F.5, which is the data set labeled "CCSD(T)/CBS//MP2/haTZ" in Fig. 4 of the paper.

F.3 Halide–water benchmarks

Figure F.1 shows binding energies for the ten $F^-(H_2O)_{10}$ isomers that were considered in the paper, computed at various levels of theory including the benchmark RI-CCSD(T)/CBS results. (Geometries of these isomers have been optimized at the B3LYP/6-31G* level and Cartesian coordinates, in Angstroms, are provided as part of this Supplementary Material, in a separate text file.) This figure also serves to document that, for the most part, the various XSAPT and DFT methods considered here parallel the benchmark results, meaning that relative isomer energies are predicted more accurately than are absolute binding energies, for which there are some systematic errors.

RI-MP2/CBS and RI-CCSD(T)/CBS binding energies for $F^-(H_2O)_n$ and $Cl^-(H_2O)_n$, up to $n = 6$, are shown in Tables F.6 and F.7. These geometries were optimized at the RI-MP2/aug-cc-pVTZ level and are available elsewhere in this Supplementary Material. Both two- and three-body approximations to the triples correction, $\delta_{\mathrm{RI-CCSD(T)}}$, are shown for comparison; the three-body approximation is used as the benchmark in this work. Table F.8 provides the analogous information for $F^-(H_2O)_{10}$ isomers whose geometries were optimized at the B3LYP/6-31G* level.

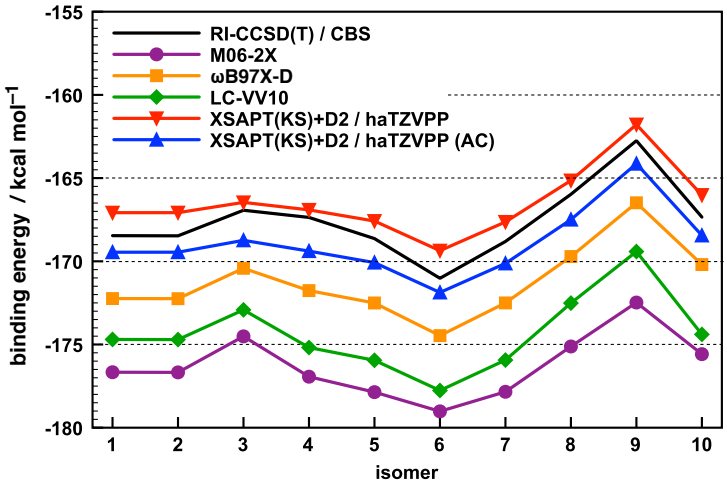

Figure F.1: Binding energies for ten isomers of $F^-(H_2O)_{10}$ isomers optimized at the B3LYP/6-31G* level. The RI-CCSD(T)/CBS benchmarks are compared to DFT/def2-QZVP results (computed using Boys-Bernardi counterpoise correction) and to XSAPT(KS)+D2/results (computed using the $\delta E_{\mathrm{int}}^{\mathrm{HF}}$ correction).

	ω/bohr^{-1}	
Monomer	LRC-ωPBE	LRC-ωPBEh
—S22 molecules—		
adenine	0.275	0.225
2-aminopyridine	0.300	0.225
benzene	0.275	0.225
ethyne	0.400	0.350
ethene	0.350	0.300
methane	0.450	0.400
formamide	0.375	0.300
formic acid	0.425	0.325
water	0.500	0.400
HCN	0.550	0.500
indole	0.275	0.225
ammonia	0.450	0.350
phenol	0.300	0.250
pyrazine	0.375	0.300
2-pyridoxine	0.300	0.250
thymine	0.625	0.525
uracil	0.475	0.375
—S66 molecules—		
methylamine	0.400	0.325
methanol	0.425	0.350
$AcNH_2$	0.350	0.275
AcOH	0.375	0.300
cyclopentane	0.425	0.375
neopentane	0.300	0.250
pentane	0.300	0.250
peptide	0.325	0.250
pyridine	0.300	0.225
—Misc. molecules & ions—		
Ar	0.550	0.475
Ne	0.800	0.650
F^-	0.475	0.375
Cl^-	0.375	0.300

Table F.1: Tuned range separation parameters for various monomers.

Basis set	s_β
aug-cc-pVDZ′	0.7267[a]
TZVPP	0.8439[a]
haTZVPP	0.8587[b]
haTZVPP (AC)	0.8547[b]

[a] From Ref. 55.

[b] Introduced in this work.

Table F.2: Values of the dimensionless parameter s_β used in the "first generation" XSAPT(KS)+D method. These values have been optimized against binding energies in the S22A data set, using a KS functional equal to LRC-ωPBEh with 60% short-range Hartree-Fock exchange and a tuned value of ω. The exact AC scheme uses monomer-specific values of ω; in other cases, ω is tuned according to Eq. (F.1) using the lowest supersystem IP.

	Chen & Li (Ref. 8)			Bates & Tschumper (Ref. 41)	
	MP2/a5Z-h			MP2/CBS	CCSD(T)/CBS
Includes BSSE correction?	Yes	Yes	No	No	No
Includes relaxation energy?	No	Yes	Yes	Yes	Yes
prism	−47.41	−45.05	−46.09	−45.86	−45.92
cage	−47.33	−44.98	−46.05	−45.80	−45.67
book1	−47.15	−44.73	−45.79	−45.53	−45.20
book2	−46.87	−44.41	−45.46	−45.22	−44.90
cyclic chair	−46.27	−43.94	−44.95	−44.65	−44.12
bag	−46.41	−43.83	−44.89	−44.63	−44.30
cyclic boat1	−45.29	−42.95	−43.96	−43.66	−43.13
cyclic boat2	−45.17	−42.87	−43.87	−43.58	−43.07

Table F.3: Comparison of binding energies (in kcal/mol) for isomers of $(H_2O)_6$. All three sets of calculations from Ref. 8 are performed at the MP2/a5Z-h//MP2/aTZ level but differ depending on whether or not the Boys- Bernardi counterpoise correction was applied, and whether or not monomer relaxation is included in the binding energy. The calculations from Ref. 41 use MP2/haTZ geometries, and single-point energies at both the MP2/CBS and CCSD(T)/CBS level are listed here. All data come from Refs. 8 and 41.

isomer	correction / kcal mol^{-1}			
	δ_{MP2}[a]	$\delta_{\mathrm{CCSD(T)}}$[b]	$\delta_{\mathrm{MP2}}+\delta_{\mathrm{CCSD(T)}}$[c]	δ_{relax}[d]
prism	−0.81	−0.06	−0.87	2.36
cage	−0.82	+0.13	−0.69	2.35
book1	−0.80	+0.33	−0.47	2.42
book2	−0.81	+0.32	−0.49	2.46
cyclic chair	−0.71	+0.53	−0.18	2.33
bag	−0.80	+0.33	−0.47	2.58
cyclic boat1	−0.71	+0.53	−0.18	2.34
cyclic boat2	−0.71	+0.51	−0.20	2.30

[a] Difference between MP2/CBS and MP2/a5Z-h binding energies.

[b] Difference between CCSD(T)/CBS and MP2/CBS binding energies.

[c] The sum of these two corrections equals the difference between CCSD(T)/CBS and MP2/a5Z-h binding energies.

[d] Monomer relaxation energy, estimated from counterpoise-corrected MP2/a5Z-h results.

Table F.4: Various differences in binding energies that can be extracted from the data in Table F.3. See the footnotes and the text for an explanation of each quantity. Note that δ_{MP2} includes the effects of the slightly different geometries used in Ref. 8 versus those in Ref. 41.

isomer	binding energy / kcal mol^{-1}			
	MP2/a5Z-h results			CCSD(T)/CBS
	MP2/a5Z-h// MP2/aTZ[a]	+triples correction[b]	+triples −relaxation[c]	−relaxation[d]
prism	−45.05	−45.11	−47.47	−48.28
cage	−44.98	−44.85	−47.20	−48.02
book1	−44.73	−44.40	−46.82	−47.62
book2	−44.41	−44.09	−46.55	−47.36
cyclic chair	−43.94	−43.41	−45.74	−46.45
bag	−43.83	−43.50	−46.08	−46.88
cyclic boat1	−42.95	−42.42	−44.76	−45.47
cyclic boat2	−42.87	−42.36	−44.66	−45.37

[a] Counterpoise-corrected result including monomer relaxation (see Table F.3).
[b] MP2/a5Z-h result plus $\delta_{\mathrm{CCSD(T)}}$ from Table F.4.
[c] MP2/a5Z-h result plus $\delta_{\mathrm{CCSD(T)}}$ minus δ_{relax} from Table F.4.
[d] CCSD(T)/CBS//MP2/haTZ result from Ref. 41, corrected to remove monomer relaxation using MP2/a5Z-h values of δ_{relax}.

Table F.5: MP2/a5Z-h//MP2/aTZ binding energies from Ref. 8 and "corrected" values based on CCSD(T) results in Ref. 41. See the footnotes and the text for an explanation of the corrections.

n	binding energy / kcal mol^{-1}		$\delta_{\mathrm{RI-CCSD(T)}}$ / kcal mol^{-1}	
	RI-MP2/CBS	RI-CCSD(T)/CBS	two-body	three-body
1	−32.25	−32.32	−0.07	—
2	−51.86	−52.22	−0.44	−0.36
3	−69.08	−70.07	−1.27	−0.99
4	−83.82	−85.11	−1.53	−1.29
5	−99.52	−100.97	−1.82	−1.45
6	−114.75	−116.21	−1.83	−1.46

Table F.6: Binding energies for isomers of $F^-(H_2O)_n$, computed at the RI-MP2/CBS and RI-CCSD(T)/CBS levels using RI-MP2/aug-cc-pVTZ geometries. Also shown is the triples correction, $\delta_{\mathrm{RI-CCSD(T)}}$, computed using either a two-body or a three-body approximation. and the corresponding two-body and three-body triples corrections $\delta_{\mathrm{RI-CCSD(T)}}$ are also provided. The RI-CCSD(T)/CBS binding energy is the sum of RI-MP2/CBS binding energy (obtained by extrapolation) and the three-body triples correction.

n	binding energy / kcal mol^{-1}		$\delta_{\mathrm{RI-CCSD(T)}}$ / kcal mol^{-1}	
	RI-MP2/CBS	RI-CCSD(T)/CBS	two-body	three-body
1	−15.75	−15.48	0.27	—
2	−31.28	−30.94	0.30	0.34
3	−47.72	−47.50	0.02	0.22
4	−58.38	−58.10	0.16	0.28
5	−74.92	−74.58	0.10	0.34
6	−86.91	−86.45	0.21	0.46

Table F.7: Binding energies for isomers of $Cl^-(H_2O)_n$, computed at the RI-MP2/CBS and RI-CCSD(T)/CBS levels using RI-MP2/aug-cc-pVTZ geometries. Also shown is the triples correction, $\delta_{\mathrm{RI-CCSD(T)}}$, computed using either a two-body or a three-body approximation. and the corresponding two-body and three-body triples corrections $\delta_{\mathrm{RI-CCSD(T)}}$ are also provided. The RI-CCSD(T)/CBS binding energy is the sum of RI-MP2/CBS binding energy (obtained by extrapolation) and the three-body triples correction.

n	binding energy / kcal mol^{-1}		$\delta_{\mathrm{RI-CCSD(T)}}$ / kcal mol^{-1}	
	RI-MP2/CBS	RI-CCSD(T)/CBS	two-body	three-body
1	−166.67	−168.46	−2.47	−1.79
2	−166.68	−168.47	−2.47	−1.79
3	−165.15	−166.93	−2.29	−1.79
4	−165.54	−167.36	−2.38	−1.82
5	−166.89	−168.64	−2.33	−1.74
6	−169.50	−171.02	−2.17	−1.52
7	−167.06	−168.81	−2.33	−1.75
8	−163.91	−165.97	−2.73	−2.07
9	−160.73	−162.75	−2.81	−2.03
10	−165.72	−167.34	−2.06	−1.63

Table F.8: RI-MP2/CBS and RI-CCSD(T)/CBS binding energies for ten different isomers of $F^-(H_2O)_{10}$. Geometries were optimized at the B3LYP/6-31G* level and the numbering scheme reflects the order of these isomers in the Cartesian coordinate file that is provided as part of this Supplementary Material. The corresponding two-body and three-body triples corrections $\delta_{\mathrm{RI-CCSD(T)}}$ are also provided. RI-CCSD(T)/ CBS binding energy are the sum of the RI-MP2/CBS binding energy (obtained by extrapolation) and the three-body triples correction.

APPENDIX G

Supporting Information for: "Accurate and efficient quantum chemistry calculations for noncovalent interactions in many-body systems: The XSAPT family of methods"[G.1]

G.1 Basis set convergence tests

Table G.1 shows the errors in the first-order energy components $E^{(1)}_{\text{elst}}$ and $E^{(1)}_{\text{exch}}$, with respect to SAPT2+(3)/aTZ benchmarks. The errors are averages over four representative systems: $F^-(H_2O)$, $(H_2O)_2$, and the T-shaped and parallel-displaced isomers of $(C_6H_6)_2$. Results are shown for 21 different AO basis sets, using either the dimer-centered SAPT basis or else the "projected" (pseudocanonicalized monomer-centered) SAPT basis.

G.2 Tuned values of the range separation parameter

Table G.2 lists the tuned value of the range separation parameter, ω, for each of the monomers considered in this work. These values were tuned using the LRC-ωPBE/ hpTZVPP method.

[G.1] This chapter appeared as the Supporting Information to the full article in the *Journal of Physical Chemistry A*, in 2015, volume 119, pages 235–252.

G.3 Empirical dispersion parameters

In addition to high-level SAPT2+(3)/aTZ, a more robust treatment of dispersion based on a coupled-cluster approach is available.[183,184] The error in the dispersion energy for the parallel-displaced benzene dimer is about 0.5 kcal/mol for SAPT2+3/aDZ as compared to SAPT2+3(CCD)/aDZ.[117] Our own tests on the S66 data set suggest that the MAE in dispersion energies between the SAPT2+(3)/aTZ and SAPT2+(3)(CCD)/aTZ methods is about 0.19 kcal/mol. The maximum discrepancy occurs for the π-stacked uracil dimer, for which the SAPT2+(3)/aTZ method overestimates the dispersion energy by 0.8 kcal/mol as compared to SAPT2+(3)(CCD)/aTZ. Furthermore, SAPT2+(3)(CCD)/aTZ works slightly better than SAPT2+(3)/aTZ for the total binding energy. In comparison to the CCSD(T)/CBS benchmarks for S66 binding energies, the SAPT2+(3)(CCD)/aTZ method affords a MAE of 0.13 kcal/mol and the SAPT2+(3)/aTZ method affords a MAE of 0.20 kcal/mol. Due to some favorable error cancellation, however, the XSAPT(KS)+D method that is fit to SAPT2+(3) dispersion energies performs slightly better than the corresponding method whose dispersion potentials are fit to SAPT2+(3)(CCD) dispersion energies. For this reason, and in consideration of computational efficiency, we use SAPT2+(3)/aTZ as our dispersion benchmark for the purpose fo fitting the D3 dispersion potential. The sole exceptions to this protocol are the two anionic systems, $F^-(H_2O)$ and $Cl^-(H_2O)$, for which SAPT2+(3)(CCD)/aTZ is used as the benchmark.

The training set used to determine the D3 parameters consists of 74 dimers: 22

dimers from the S22 data set,[186] 22 dimers from the X40 data set,[36] 13 dimers containing divalent sulfur,[524] 8 dimers proposed by Zhao and Truhlar[525] in the KB49 data set[526] (without the H_2S dimer, which is already included as part of the divalent-sulfur data set), and 9 additional dimers optimized at the MP2/aTZ level, including $HCl\cdots H_2O$, $HCl\cdots NH_3$, $HF\cdots CH_3SH$, $HF\cdots H_2O$, $HF\cdots H_2S$, $HF\cdots HCl$, $HF\cdots NH_3$, $F^-\cdots H_2O$, and $Cl^-\cdots H_2O$. For each dimer, 5 different radial geometries corresponding to the same angular configuration were considered. For the 22 dimers from the X40 dataset and the 13 dimers containing divalent sulfur, each dimer contains five data points with relative displacements of 0.90, 1.00, 1.25, 1.50, and 2.00 with respect to the equilibrium geometry. For the rest of dimers, each dimer contains five data points with relative displacements of 0.9, 1.0, 1.2, 1.5, and 2.0 with respect to the equilibrium geometry. There are 370 training geometries in total.

The parameters of D3 were optimized using the least-squares method with deviations

$$\chi^2 = \frac{1}{370}\left[\sum_{i=1}^{370}\left(E_{\text{disp}}^{(i)}(\text{D3}) - E_{\text{disp}}^{(i)}(\text{SAPT})\right)^2\right], \tag{G.1}$$

where $E_{\text{disp}}^{(i)}$ is the dispersion energy for the ith dimer, computed either using the benchmark SAPT calculation or else the D3 dispersion potential. We used a genetic algorithm followed by simplex optimization to fit the parameters, and the final deviation was $\chi = 0.1377$ kcal/mol. Optimized values of $C_{6,i}$, $C_{8,i}$, and β_i are listed in Table G.3.

G.4 New benchmarks

Benchmark binding energies for $(H_2O)_6$, $(H_2O)_{20}$, $F^-(H_2O)_{n\leq 6}$, $Cl^-(H_2O)_{n\leq 6}$, and $F^-(H_2O)_{10}$ are shown in Tables G.4, G.5, G.6, G.7, and G.8, respectively. The $(H_2O)_6$ geometries are available from Ref. 8, where they were optimized at the MP2/aug-cc-pVTZ level, and the $(H_2O)_{20}$ structures are available from Ref. 339, where they were optimized using the TIP4P force field. Coordinates for the remaining structures are available in this work, as a separate attachment. For $X^-(H_2O)_n$ structures with $n \leq 6$, the geometries were optimized at the RIMP2/aug-cc-pVTZ level; $F^-(H_2O)_{10}$ geometries were optimized at the B3LYP/6-31G* level.

G.5 Software

All XSAPT, att-MP2, and DFT calculations were performed using a locally-modified copy of Q-Chem v. 4.2.[97] [The att-MP2 and XSAPT(KS)+D2 methods are available in the current release of Q-Chem, v. 4.2, and the XSAPT(KS)+D3 and sd-XSAPT methods will be released in v. 4.3 in 2015.] CCSD(T)-F12 calculations for $(H_2O)_6$, $F^-(H_2O)_{n\leq 6}$, and $Cl^-(H_2O)_{n\leq 6}$ were performed using Orca 3.0.2,[188] and the CCSD(T)-F12 and MP2-F12 calculations for $(H_2O)_{20}$ and $F^-(H_2O)_{10}$ were performed using Molpro 2012.1.[189] All terms in $\delta E_{\text{int}}^{\text{HF}}$, along with the CCSD(T) calculations for $X^-(H_2O)$ and the rest of the MP2 and CCSD(T) calculations, were performed using Psi4 v. beta5.[187] All supersystem calculations are counterpoise corrected, with the exception of the DFT and att-MP2 calculations, as well as the CCSD(T)-F12 and MP2-F12 calculations for $(H_2O)_{20}$ and $F^-(H_2O)_{10}$.

AO Basis Set	DCBS		Projected	
	$E^{(1)}_{\text{elst}}$	$E^{(1)}_{\text{exch}}$	$E^{(1)}_{\text{elst}}$	$E^{(1)}_{\text{exch}}$
cc-pVDZ	17.73	8.65	30.46	51.52
cc-pVTZ	10.05	3.98	18.38	29.51
cc-pVQZ	7.51	3.16	12.07	15.94
heavy-aug-cc-pVDZ	3.58	3.97	10.76	10.46
heavy-aug-cc-pVTZ	3.93	4.47	7.09	8.02
heavy-aug-cc-pVQZ	3.95	4.55	4.81	5.15
aug-cc-pVDZ	3.17	4.31	9.71	14.55
aug-cc-pVTZ	3.95	4.52	6.81	8.04
aug-cc-pVQZ	3.94	4.57	4.85	5.15
def2-SVP	16.91	8.77	30.15	50.69
def2-TZVPP	7.85	2.95	11.76	13.93
def2-QZVPP	5.40	4.01	6.80	6.99
def2-SVPD	5.46	4.74	23.53	24.21
def2-TZVPPD	3.97	4.52	9.42	7.01
def2-QZVPPD	4.05	4.56	7.11	6.50
aug-def2-TZVPP	3.90	4.50	4.63	4.69
aug-def2-QZVPP	3.94	4.55	3.37	3.86
heavy-aug-def2-TZVPP	3.91	4.48	4.52	4.53
heavy-aug-def2-QZVPP	4.00	4.55	3.64	3.98
Pople-def2-TZVPP	4.54	4.27	4.98	4.46
heavy-Pople-def2-TZVPP	4.58	4.25	5.09	4.72

Table G.1: Mean absolute percentage errors (in kcal/mol) for the energy components $E^{(1)}_{\text{elst}}$ and $E^{(1)}_{\text{exch}}$ computed using SAPT(KS)/LRC-ωPBE, as compared to SAPT2+(3)/aTZ benchmarks. The test systems are $F^-(H_2O)$, $(H_2O)_2$, and the T-shaped and parallel-displaced isomers of $(C_6H_6)_2$. Results are shown for both the dimer-centered SAPT basis (DCBS) as well as the "projected" (pseudocanonicalized monomer-centered) basis, only the latter of which is appropriate for XSAPT.

Monomer	ω / a_0^{-1}	
	$\Delta\omega = 0.025\ a_0^{-1}$	$\Delta\omega = 0.005\ a_0^{-1}$
—S22 molecules—		
adenine	0.275	—
2-aminopyridine	0.300	—
benzene	0.275	—
ethyne	0.400	—
ethene	0.350	—
methane	0.450	—
formamide	0.475	—
formic acid	0.425	—
water	0.500	0.485
HCN	0.450	—
indole	0.275	—
ammonia	0.450	—
phenol	0.275	—
pyrazine	0.375	—
2-pyridoxine	0.300	—
thymine	0.275	—
uracil	0.300	—
—S66 molecules—		
methylamine	0.400	—
methanol	0.450	—
$AcNH_2$	0.450	—
AcOH	0.375	—
cyclopentane	0.450	—
neopentane	0.300	—
pentane	0.325	—
peptide	0.350	—
pyridine	0.325	—
—ions—		
F^-	0.475	0.480
Cl^-	0.375	0.370

Table G.2: Tuned values of the range separation parameter, ω, for various monomers, where the tuning was performed at the LRC-ωPBE/hpTZVPP level. In a few cases, a finer spacing of $\Delta\omega = 0.005\ a_0^{-1}$ was used to scan the $\varepsilon_{\mathrm{HOMO}}(\omega)$ and $-\mathrm{IP}(\omega)$ curves, and in these cases we used the more finely-tuned value of ω.

Element	C_6/ J nm^6 mol^{-1}	C_8/ J nm^8 mol^{-1}	β/ a_0^{-1}
H–C	0.0087	1.9024	0.3962
H–N	0.0670	2.4646	0.3245
H–O	0.0847	3.3700	0.3021
H–F	0.0272	0.8556	0.3904
H–S	0.0000	18.5511	0.2456
H–Cl	0.0001	4.1514	0.3122
C	2.4618	0.0109	4.4611
N	1.2111	0.0059	8.0263
O	0.5450	0.0010	10.4192
F	0.3468	0.0005	12.8513
S	14.1799	0.0075	8.4090
Cl	19.6013	0.0001	15.8640

Table G.3: Fitting parameters that define the D3 dispersion potential for various elements. Note that the hydrogen parameters depend upon the atom to which it is bonded.

Isomer	Binding Energy / kcal mol^{-1}	
	CCSD(T)/CBS	MP2/CBS
prism	−48.31	−48.22
cage	−48.02	−48.13
book1	−47.61	−47.93
book2	−47.33	−47.65
cyclic chair	−46.52	−47.02
bag	−46.87	−47.19
cyclic boat1	−45.52	−46.04
cyclic boat2	−45.42	−45.91

Table G.4: Binding energies isomers of $(H_2O)_6$, using geometries from from Ref. 8.

Isomer	Binding Energy / kcal mol^{-1}	
	CCSD(T)/CBS	MP2/CBS
1	−200.54	−197.97
2	−199.20	−196.68
3	−198.93	−196.51
4	−197.89	−195.02
5	−198.15	−195.58
6	−198.17	−195.24
7	−197.67	−194.62
8	−197.44	−194.90
9	−197.03	−194.29
10	−196.59	−193.92

Table G.5: Binding energies isomers of $(H_2O)_{20}$, using geometries from from Ref. 18.

n	Binding Energy / kcal mol^{-1}	
	CCSD(T)/CBS	MP2/CBS
1	−32.31	−32.25
2	−52.27	−51.86
3	−70.14	−69.09
4	−85.24	−83.84
5	−101.09	−99.53
6	−116.58	−114.98

Table G.6: Binding energies for optimized (MP2/aTZ) geometries of $F^-(H_2O)_n$.

n	Binding Energy / kcal mol^{-1}	
	CCSD(T)/CBS	MP2/CBS
1	−15.50	−15.75
2	−31.01	−31.30
3	−47.62	−47.74
4	−58.20	−58.40
5	−74.73	−74.94
6	−86.86	−87.16

Table G.7: Binding energies for optimized (MP2/aTZ) geometries of $Cl^-(H_2O)_n$.

Isomer	Binding Energy / kcal mol^{-1}	
	CCSD(T)/CBS	MP2/CBS
1	−169.02	−166.97
2	−169.03	−166.98
3	−168.01	−165.93
4	−169.07	−167.06
5	−169.91	−167.98
6	−171.42	−169.68
7	−169.92	−167.99
8	−166.65	−164.34
9	−163.59	−161.25
10	−168.22	−166.47

Table G.8: Binding energies for optimized (B3LYP/6-31G*) geometries of ten different isomers of $F^-(H_2O)_{10}$.

APPENDIX H

Supporting Information for: "Atomic orbital implementation of extended symmetry-adapted perturbation theory (XSAPT) and benchmark calculations for large supramolecular complexes"

H.1 Tuned values of the range separation parameters

Tuned values of ω for complexes in L7 and S12L, using both the ω_{IP} and ω_{GDD} tuning procedures, are shown in Tables H.1 and H.2. For the ω_{GDD} tuning procedure, there is in addition a cutoff value (μ) to define the region of localized molecular orbitals; see Ref. 26. Following that work, we use $\omega = 0.3\ a_0^{-1}$ as an initial guess, then determine μ to normalize $\langle d_{\mathrm{x}}^2 \rangle$. A series of molecules was used to determine the constant C in

$$\omega_{\mathrm{GDD}} = C \langle d_{\mathrm{x}}^2 \rangle^{-1/2} \ , \tag{H.1}$$

and subsequently ω_{GDD} and μ were determined for each molecule of interest (with C fixed), with results in Tables H.1 and H.2.

H.2 Energies and other properties

Listed in Table H.3 are binding energies for complexes in L7 and S12L, computed using XSAPT(KS)+D3(ω_{IP}), XSAPT(KS)+D3(ω_{GDD}), sd-XSAPT(KS)(ω_{IP}), and sd-XSAPT(KS)(ω_{GDD}). Also shown are binding energies computed at the ωB97X-V/aTZ and B97M-V/(h)aTZ levels of theory.

Components of the interaction energies for L7 and S12L are shown in Table H.4, computed using the XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$/hpTZVPP($\omega_{\mathrm{GDD}}$) method.

Deformation energies for the S12L complexes, computed at various levels of theory, are shown in Table H.5.

Dipole moments for the polarized and unpolarized wave functions for molecules in the L7 and S12L data sets are listed in Tables H.6 (for L7) and H.7 (for S12L), where they are computed using XSAPT(KS)(ω_{GDD}).

Table H.1: Tuned values of the range separation parameter, ω (in units of a_0^{-1}) for the L7 data set, at two different LRC-DFT levels of theory. For ω_{GDD} tuning there is also a cutoff, μ (see Ref. 26) that is listed in parentheses. A spacing of $\Delta\omega = 0.025\ a_0^{-1}$ was used to scan the $\varepsilon_{\mathrm{HOMO}}(\omega)$ and $-\mathrm{IP}(\omega)$ curves for ω_{IP}.

Complex	$\omega_{\mathrm{IP}}{}^a$	ω_{GDD} $(\mu)^a$	$\omega_{\mathrm{IP}}{}^b$	ω_{GDD} $(\mu)^b$
guanine	0.275	0.316 (0.072)	0.250	0.338 (0.075)
guanine dimer	0.250	0.289 (0.047)	0.225	0.313 (0.051)
cytosine	0.300	0.328 (0.082)	0.250	0.359 (0.083)
adenine	0.275	0.316 (0.078)	0.250	0.338 (0.080)
Watson-Crick GC	0.250	0.291 (0.051)	0.225	0.314 (0.055)
circumcoronene	0.175	0.246 (0.033)	0.150	0.266 (0.036)
coronene	0.225	0.271 (0.053)	0.175	0.289 (0.055)
octadecane	0.250	0.260 (0.038)	0.200	0.269 (0.040)
phenylalanine	0.250	0.290 (0.054)	0.200	0.308 (0.057)
phenylalanine dimer	0.225	0.268 (0.037)	0.200	0.285 (0.037)

[a] LRC-ωPBE/hpTZVPP. [b] LRC-ωPBEh/6-31G(d,2p).

Table H.2: Tuned values of the range separation parameter, ω (in units of a_0^{-1}) for the S12L data set, calculated at the level of LRC-ωPBE/hpTZVPP. For ω_{GDD} tuning there is also a cutoff, μ (see Ref. 26) that is listed in parentheses. A spacing of $\Delta\omega = 0.025\ a_0^{-1}$ was used to scan the $\varepsilon_{\mathrm{HOMO}}(\omega)$ and $-\mathrm{IP}(\omega)$ curves for ω_{IP}.

System	Complex	ω_{IP}	ω_{GDD} (μ)
2a/2b	host	0.175	0.248 (0.032)
2a	guest	0.225	0.296 (0.057)
2b	guest	0.275	0.313 (0.080)
3a/3b	host	0.175	0.251 (0.027)
3a	guest	0.200	0.296 (0.048)
3b	guest	0.250	0.315 (0.063)
4a/4b	host	0.225	0.234 (0.029)
4a	guest	0.200	0.243 (0.038)
4b	guest	0.200	0.237 (0.035)
5a/5b	host	0.300	0.236 (0.023)
5a	guest	0.600	0.332 (0.079)
5b	guest	0.400	0.336 (0.082)
6a/6b	host	0.425	0.256 (0.027)
6a	guest	0.375	0.339 (0.085)
6b	guest	0.375	0.360 (0.091)
7a/7b	host	0.375	0.248 (0.024)
7a	guest	0.200	0.288 (0.045)
7b	guest	0.275	0.304 (0.067)

Table H.3: Binding energies (in kcal/mol) for L7 and S12L complexes computed at various XSAPT and DFT levels of theory. The hpTZVPP basis set is used for XSAPT(KS)+D3, and the 6-31G(d,2p) basis set for sd-XSAPT(KS). The aTZ basis set is used for DFT calculations except the B97M-V calculations use the haTZ basis set for **4a**, **4b**, **7a**, and **7b**. Deformation energies for DFT methods are computed using the corresponding functional. Deformation energies for XSAPT-based methods are computed using the NLDFT functional.

System	XSAPT(KS)+D3		sd-XSAPT(KS)		ωB97X-V	B97M-V	B97M-V (CP)
	(ω_{IP})	(ω_{GDD})	(ω_{IP})	(ω_{GDD})			
			—L7—				
GGG	−2.50	−2.60	−1.47	−1.08	−2.52	−2.33	−1.94
C2C2PD	−26.06	−26.39	−24.07	−21.70	−22.40	−22.32	−21.54
GCGC	−15.68	−16.06	−14.96	−13.93	−15.61	−15.46	−14.51
PHE	−23.36	−23.96	−24.48	−24.43	−26.26	−25.87	−25.21
CBH	−9.45	−9.55	−9.12	−8.45	−12.55	−12.40	−11.72
C3A	−20.11	−20.46	−18.42	−17.09	−17.73	−17.47	−16.56
C3GC	−33.03	−33.56	−31.17	−28.52	−30.70	−31.05	−29.27
			—S12L—				
2a	−39.24	−40.92	–	−34.76	−34.62	−32.85	−31.23
2b	−25.80	−26.73	–	−22.15	−22.79	−21.88	−20.58
3a	−27.11	−28.80	–	−29.19	–	−29.85	−27.11
3b	−23.02	−24.28	–	−23.57	–	−22.88	−20.95
4a	−49.96	−50.18	–	−52.14	–	−40.54	−37.60
4b	−51.35	−51.66	–	−55.41	–	−42.58	−39.38
5a	−44.86	−38.21	–	−36.29	–	−36.70	−35.16
5b	−26.60	−24.67	–	−24.82	–	−25.67	−24.30
6a	−94.47	−90.35	–	−82.16	–	−87.07	−85.79
6b	−90.19	−86.49	–	−78.70	–	−82.58	−81.49
7a	−146.54	−142.60	–	−134.32	–	−142.09	−139.85
7b	−38.51	−37.72	–	−25.62	–	−34.74	−33.05

Table H.4: Interaction energy components for L7 and S12L complexes computed using the XSAPT(KS)+D3+$E_{\mathrm{disp,3B}}^{\mathrm{ATM(TS)}}$/hpTZVPP($\omega_{\mathrm{GDD}}$) method, in kcal/mol. The interaction types are dispersion-dominated ("D"), electrostatics-dominated ("E"), or of mixed influence (dispersion and electrostatics, "M").

System	Elst	Exch	Ind	Disp	Total[a]	Type
			—L7—			
GGG	2.70	6.52	−1.29	−9.43	−1.51	D
C2C2PD	−7.85	26.65	−2.46	−35.00	−18.67	D
GCGC	−8.94	24.62	−1.93	−26.09	−12.33	D
PHE	−27.56	31.93	−11.01	−16.15	−22.79	M
CBH	−3.54	14.75	−1.24	−16.99	−7.02	D
C3A	−7.58	19.95	−1.96	−25.79	−15.38	D
C3GC	−13.08	37.49	−4.42	−44.50	− 24.51	D
			—S12L—			
2a	−27.23	53.17	−11.84	−50.07	−35.97	M
2b	−17.18	36.60	−6.01	−36.15	−22.74	D
3a	−22.41	58.02	−7.29	−59.36	−31.04	D
3b	−18.77	38.05	−5.83	−42.38	−28.92	D
4a	−26.95	69.02	−7.23	−72.88	−39.43	D
4b	−28.11	73.25	−8.17	−76.39	−38.03	D
5a	−48.50	69.21	−20.36	−41.33	−40.98	M
5b	−33.53	54.73	−12.43	−35.68	−26.92	M
6a	−68.15	39.00	−24.80	−35.70	−89.65	M
6b	−65.78	33.26	−24.69	−30.26	−87.46	E
7a	−104.05	57.97	−24.43	−67.96	−138.47	M
7b	−10.07	29.28	−3.05	−45.82	−29.67	D

[a]Relaxation energies are not included.

Table H.5: Deformation energies for complexes in S12L datasets computed at the NLDFT/def2-QZVP, MP2/CBS, SCS-MP2/CBS, and B97M-V/aTZ levels. The B97M-V results are new; the rest are from Ref. 42.

System	deformation energy (kcal/mol)			
	NLDFT	MP2	SCS-MP2	B97M-V
2a	3.08	3.25	3.12	3.56
2b	1.66	2.03	1.82	2.08
3a	12.70	23.63	17.67	15.11
3b	11.06	21.34	15.60	13.86
4a	1.24	1.24	1.21	0.99
4b	2.24	2.56	2.55	1.72
5a	6.58	8.05	7.21	7.60
5b	5.72	6.49	5.66	6.21
6a	4.90	4.09	4.37	5.01
6b	5.45	4.53	4.86	5.62
7a	7.30	0.48	2.49	6.08
7b	1.08	0.14	0.33	0.21

Table H.6: Dipole moments (in Debye) for molecules in the L7 data set computed at the XSAPT(KS)/hpTZVPP(ω_{GDD}) level.

Complex	Fragment	Unpolarized	Polarized	Difference
GGG	guanine	6.98	6.29	−0.68
	guanine dimer	13.31	12.51	−0.80
C2C2PD	coronene	0.00	0.72	0.72
	coronene	0.00	0.72	0.72
GCGC	Watson-Crick guanine-cytosine	5.94	6.57	0.63
	Watson-Crick guanine-cytosine	5.94	6.57	0.62
PHE	phenylalanine	2.12	3.25	1.13
	phenylalanine dimer	2.46	3.76	1.29
CBH	octadecane	0.00	0.05	0.05
	octadecane	0.00	0.05	0.05
C3A	adenine	2.62	2.79	0.17
	circumcoronene	0.36	1.11	0.76
C3GC	Watson-Crick guanine-cytosine	6.27	6.55	0.28
	circumcoronene	0.09	2.48	2.39

Table H.7: Dipole moments (in Debye) for molecules in the L7 data set computed at the XSAPT(KS)/hpTZVPP(ω_{GDD}) level.

Complex	Fragment	Unpolarized	Polarized	Difference
2a	host	1.36	1.66	0.30
	guest	0.51	0.60	0.08
2b	host	1.22	1.67	0.45
	guest	0.10	0.48	0.38
3a	host	15.44	16.43	1.00
	guest	1.22	0.95	−0.27
3b	host	15.71	17.35	1.64
	guest	4.92	6.29	1.37
4a	host	1.48	1.52	0.04
	guest	0.02	1.30	1.28
4b	host	1.54	1.49	−0.05
	guest	0.01	1.39	1.38
5a	host	0.71	0.79	0.08
	guest	0.23	0.28	0.05
5b	host	0.71	0.82	0.11
	guest	0.02	0.05	0.03
6a	host	0.35	6.58	6.24
	guest	13.28	14.29	1.01
6b	host	0.26	6.62	6.36
	guest	12.78	13.94	1.16
7a	host	0.11	0.29	0.17
	guest	1.25	1.50	0.25
7b	host	0.08	0.12	0.04
	guest	1.74	1.66	−0.07

APPENDIX I

Supporting Information for: "A Simple Correction for Non-Additive Dispersion within Extended Symmetry-Adapted Perturbation Theory"

I.1 Tuned values of the range separation parameter

Table I.1 lists the tuned value of the range separation parameter, ω, for each of the monomers considered in this work using ω_{IP} and ω_{GDD} tunings. These values were tuned using the LRC-ωPBE/hpTZVPP method based on the monomer geometries optimized at the level of RIMP2/aug-cc-pVDZ.

I.2 Empirical dispersion parameters

We use SAPT2+(3) combined with def2-QZVPP basis set for Li, Na, and K elements and aug-cc-pVTZ basis set for the remaining elements as our dispersion benchmark for the purpose of fitting the D3 dispersion parameters for Li, Na, and K elements. The training set used to determine the their D3 parameters consists of 21 dimers optimized at the RIMP2 method combined with def2-QZVPP basis set for Li, Na, and K elements and aug-cc-pVTZ basis set for the remaining elements, including $Li^+\cdots C_6H_6$, $Na^+\cdots C_6H_6$, $K^+\cdots C_6H_6$, $Li^+\cdots C_2H_2$, $Na^+\cdots C_2H_2$, $K^+\cdots C_2H_2$,

$Li^+\cdots C_2H_4$, $Na^+\cdots C_2H_4$, $K^+\cdots C_2H_4$, $Li^+\cdots CH_3NH_2$, $Na^+\cdots CH_3NH_2$, $K^+\cdots CH_3NH_2$, $Li^+\cdots CH_3OH$, $Na^+\cdots CH_3OH$, $K^+\cdots CH_3OH$, $Li^+\cdots H_2O$, $Na^+\cdots H_2O$, $K^+\cdots H_2O$, $Li^+\cdots NH_3$, $Na^+\cdots NH_3$, and $K^+\cdots NH_3$. For each dimer, 5 different radial geometries corresponding to the same angular configuration were considered. Each dimer contains five data points with relative displacements of 0.9, 1.0, 1.2, 1.5, and 2.0 with respect to the equilibrium geometry. There are 105 training geometries in total.

The parameters of D3 were optimized using the least-squares method with deviations

$$\chi^2 = \frac{1}{105}\left[\sum_{i=1}^{105}\left(E_{\text{disp}}^{(i)}(\text{D3}) - E_{\text{disp}}^{(i)}(\text{SAPT})\right)^2\right] , \tag{I.1}$$

where $E_{\text{disp}}^{(i)}$ is the dispersion energy for the ith dimer, computed either using the benchmark SAPT calculation or else the D3 dispersion potential. The dispersion parameters for H, C, N, and O elements obtained before[57] are set as constants during the fitting. We used a genetic algorithm followed by simplex optimization to fit the parameters, and the final deviation was $\chi = 0.1310$ kcal/mol. Optimized values of $C_{6,i}$, $C_{8,i}$, and β_i are listed in Table I.2.

I.3 Benchmark energy components for the S66 data set

Energy components for S66 computed at the SAPT2+(3)/aug-cc-pVTZ level are listed in Table I.3, Table I.4, and Table I.5 for hydrogen bonded, dispersion bonded,

and mixed complexes, respectively. The SAPT2+(3) is defined as follows:[117]

$$\begin{aligned} E_{\text{SAPT2+(3)}} &= E^{(10)}_{\text{elst}} + E^{(12)}_{\text{elst,resp}} + E^{(13)}_{\text{elst,resp}} \\ &+ E^{(10)}_{\text{exch}} + E^{(11)}_{\text{exch}} + E^{(12)}_{\text{exch}} \\ &+ E^{(20)}_{\text{ind,resp}} + E^{(20)}_{\text{exch-ind,resp}} + {}^{t}E^{(22)}_{\text{ind}} + {}^{t}E^{(22)}_{\text{exch-ind}} + \delta E^{\text{HF}}_{\text{int,resp}} \\ &+ E^{(20)}_{\text{disp}} + E^{(20)}_{\text{exch-disp}} + E^{(21)}_{\text{disp}} + E^{(22)}_{\text{disp}} + E^{(30)}_{\text{disp}} \ . \end{aligned} \tag{I.2}$$

The energy components are grouped as follows:[37]

$$E_{\text{electrostatic}} = E^{(10)}_{\text{elst}} + E^{(12)}_{\text{elst,resp}} + E^{(13)}_{\text{elst,resp}} \ , \tag{I.3}$$

$$E_{\text{exchange}} = E^{(10)}_{\text{exch}} + E^{(11)}_{\text{exch}} + E^{(12)}_{\text{exch}} \ , \tag{I.4}$$

$$E_{\text{induction}} = E^{(20)}_{\text{ind,resp}} + E^{(20)}_{\text{exch-ind,resp}} + {}^{t}E^{(22)}_{\text{ind}} + {}^{t}E^{(22)}_{\text{exch-ind}} + \delta E^{\text{HF}}_{\text{int,resp}} \ , \tag{I.5}$$

$$E_{\text{dispersion}} = E^{(20)}_{\text{disp}} + E^{(20)}_{\text{exch-disp}} + E^{(21)}_{\text{disp}} + E^{(22)}_{\text{disp}} + E^{(30)}_{\text{disp}} \ . \tag{I.6}$$

I.4 Binding benchmarks

CCSD(T)/CBS benchmark binding energies for clusters $(H_2O)_6$, $F^-(H_2O)_{10}$, and $(H_2O)_{20}$ are shown in Tables I.6, I.7, and I.8, respectively. CCSD(T)/CBS benchmark binding energies for clusters $F^-(H_2O)_{n\leq 6}$, $Cl^-(H_2O)_{n\leq 6}$, $SO_4^{2-}(H_2O)_{n\leq 6}$, $Li^+(H_2O)_{n\leq 6}$, $Na^+(H_2O)_{n\leq 6}$, and $K^+(H_2O)_{n\leq 6}$ are shown in Tables I.9.

The $(H_2O)_6$ geometries are available from Ref. 8, where they were optimized at the MP2/aug-cc-pVTZ level, and the $(H_2O)_{20}$ structures are available from Ref. 339, where they were optimized using the TIP4P force field. For $F^-(H_2O)_{n\leq 6}$ and $Cl^-(H_2O)_{n\leq 6}$ structures, the geometries were optimized at the RIMP2/aug-cc-pVTZ level; For $SO_4^{2-}(H_2O)_{n\leq 6}$, the geometries were optimized at the RIMP2/aug-cc-pVDZ

level; For $Li^+(H_2O)_{n\leq 6}$, $Na^+(H_2O)_{n\leq 6}$, and $K^+(H_2O)_{n\leq 6}$ structures, the geometries were optimized at the TPSS/def2-TZVPP level; $F^-(H_2O)_{10}$ geometries were optimized at the B3LYP/6-31G* level. Coordinates for all structures are available in this work, as a separate attachment.

Monomer	ω_{IP} / a_0^{-1}	ω_{GDD} / a_0^{-1}
adenine	0.271	0.316
2-aminopyridine	0.293	0.326
benzene	0.280	0.337
ethyne	0.397	0.428
ethene	0.359	0.420
methane	0.454	0.443
formamide	0.460	0.406
formic acid	0.412	0.429
water	0.502	0.488
HCN	0.452	0.472
indole	0.267	0.306
ammonia	0.440	0.450
phenol	0.292	0.330
pyrazine	0.367	0.350
2-pyridoxine	0.294	0.335
thymine	0.284	0.325
uracil	0.295	0.336
$MeNH_2$	0.397	0.404
MeOH	0.438	0.421
$AcNH_2$	0.453	0.371
AcOH	0.381	0.384
cyclopentane	0.420	0.331
neopentane	0.287	0.324
pentane	0.365	0.320
peptide	0.341	0.347
pyridine	0.316	0.349
F^-	0.480	0.447
Cl^-	0.372	0.324
$SO4^{2-}$	0.344	0.316
Li^+	2.006	1.305
Na^+	1.049	0.839
K^+	0.755	0.589

Table I.1: Tuned values of the range separation parameter, ω, for various monomers optimized at the level of RIMP2/aug-cc-pVDZ, where the tuning was performed at the LRC-ωPBE/hpTZVPP level using ω_{IP} and ω_{GDD} tuning procedures.

Element	C_6/ J nm^6 mol^{-1}	C_8/ J nm^8 mol^{-1}	β/ a_0^{-1}
Li	0.0002	0.0220	0.7674
Na	0.0184	0.6891	0.5271
K	0.4950	25.5348	0.4024

Table I.2: Fitting parameters that define the D3 dispersion potential for for Li, Na, and K elements.

system	Electrostatics	Exchange-Repulsion	Induction	Dispersion	Total
—S66 hydrogen bonded dimers—					
H_2O-H_2O	-7.77	7.57	-2.32	-2.32	-4.84
H_2O-MeOH	-8.79	9.28	-2.88	-3.07	-5.46
H_2O-$MeNH_2$	-12.09	13.03	-4.32	-3.51	-6.90
H_2O-peptide	-12.37	13.16	-4.44	-4.53	-8.19
MeOH-MeOH	-8.90	9.81	-3.05	-3.50	-5.65
MeOH-$MeNH_2$	-12.72	14.61	-4.83	-4.63	-7.57
MeOH-peptide	-12.59	14.24	-4.74	-5.31	-8.39
MeOH-H_2O	-7.75	7.89	-2.45	-2.64	-4.95
$MeNH_2$-MeOH	-4.11	4.92	-1.11	-2.75	-3.05
$MeNH_2$-$MeNH_2$	-5.83	7.15	-1.66	-3.83	-4.17
$MeNH_2$-peptide	-6.88	8.47	-1.98	-5.20	-5.59
$MeNH_2$-H_2O	-12.22	13.43	-4.46	-3.97	-7.23
peptide-MeOH	-7.87	8.27	-2.41	-4.24	-6.26
peptide-$MeNH_2$	-10.84	12.12	-3.75	-5.08	-7.54
peptide-peptide	-11.21	12.27	-3.91	-6.10	-8.95
peptide-H_2O	-6.67	6.07	-1.85	-2.75	-5.21
uracil-uracil-bp	-25.98	29.77	-12.50	-9.38	-18.09
H_2O-pyridine	-11.21	12.18	-4.13	-3.74	-6.91
MeOH-pyridine	-11.85	13.56	-4.60	-4.63	-7.51
AcOH-AcOH	-32.36	40.13	-18.06	-10.06	-20.36
$AcNH_2$-$AcNH_2$	-25.54	27.98	-11.11	-8.23	-16.91
AcOH-uracil	-30.57	35.30	-15.64	-9.70	-20.61
$AcNH_2$-uracil	-28.86	31.11	-13.28	-9.07	-20.10

Table I.3: The energy components and total binding energies calculated by SAPT2+(3)/aug-cc-pVTZ method for the hydrogen bonded subset of S66 data set.

system	Electrostatics	Exchange-Repulsion	Induction	Dispersion	Total
—S66 dispersion bonded dimers—					
bz-bz-pi	-1.47	6.23	-0.68	-6.94	-2.86
pyridine-pyridine-pi	-2.93	7.29	-0.82	-7.53	-3.99
uracil-uracil-pi	-8.93	12.61	-1.85	-12.59	-10.76
bz-pyridine-pi	-2.33	6.92	-0.77	-7.33	-3.51
bz-uracil-pi	-4.75	10.10	-1.18	-10.25	-6.08
pyridine-uracil-pi	-5.79	10.04	-1.26	-10.31	-7.33
bz-ethene	-0.79	4.22	-0.49	-4.13	-1.19
uracil-ethene	-3.12	5.74	-0.59	-5.34	-3.31
uracil-ethyne	-3.91	5.56	-0.63	-4.88	-3.86
pyridine-ethene	-1.47	4.78	-0.54	-4.39	-1.62
pentane-pentane	-1.80	6.07	-0.51	-7.57	-3.80
neopentane-pentane	-1.30	4.30	-0.39	-5.26	-2.65
neopentane-neopentane	-0.75	2.92	-0.29	-3.71	-1.83
cyclopentane-neopentane	-1.29	4.29	-0.40	-5.07	-2.47
cyclopentane-cyclopentane	-1.47	4.82	-0.44	-5.98	-3.06
bz-cyclopentane	-2.24	6.15	-0.66	-6.88	-3.63
bz-neopentane	-1.75	4.72	-0.55	-5.40	-2.97
uracil-pentane	-2.75	7.55	-0.89	-9.08	-5.18
uracil-cyclopentane	-2.21	6.43	-0.66	-7.97	-4.41
uracil-neopentane	-2.44	5.49	-0.55	-6.46	-3.97
ethene-pentane	-1.05	3.41	-0.35	-3.97	-1.97
ethyne-pentane	-1.13	3.03	-0.34	-3.26	-1.71
peptide-pentane	-2.56	6.86	-0.99	-7.79	-4.48

Table I.4: The energy components and total binding energies calculated by SAPT2+(3)/aug-cc-pVTZ method for the dispersion bonded subset of S66 data set.

system	Electrostatics	Exchange-Repulsion	Induction	Dispersion	Total
		—S66 mixed dimers—			
bz-bz-ts	-1.84	4.23	-0.62	-4.79	-3.01
pyridine-pyridine-ts	-2.78	4.91	-0.79	-5.06	-3.72
bz-pyridine-ts	-2.32	4.45	-0.77	-4.87	-3.50
bz-ethyne	-2.19	3.29	-0.89	-3.24	-3.03
ethyne-ethyne	-1.83	2.10	-0.54	-1.31	-1.58
bz-AcOH-ohpi	-4.07	6.27	-2.07	-5.02	-4.89
bz-$AcNH_2$	-4.82	6.51	-1.68	-4.54	-4.53
bz-H_2O	-2.89	3.81	-1.11	-3.12	-3.31
bz-MeOH	-3.32	5.33	-1.32	-4.87	-4.18
bz-$MeNH_2$	-2.42	4.65	-0.75	-4.71	-3.23
bz-peptide	-3.90	6.49	-1.47	-6.51	-5.39
pyridine-pyridine-chn	-4.75	5.81	-1.47	-4.10	-4.51
ethyne-H_2O	-3.91	3.41	-1.01	-1.52	-3.03
ethyne-AcOH	-7.06	8.45	-2.68	-3.77	-5.06
pentane-AcOH	-1.71	4.73	-0.59	-5.46	-3.03
pentane-$AcNH_2$	-2.31	6.01	-1.06	-6.35	-3.71
bz-AcOH	-3.12	5.87	-0.87	-5.80	-3.93
peptide-ethene	-2.69	4.64	-0.83	-4.20	-3.08
pyridine-ethyne	-5.92	6.13	-1.91	-2.62	-4.32
$MeNH_2$-pyridine	-4.31	6.49	-1.13	-4.96	-3.91

Table I.5: The energy components and total binding energies calculated by SAPT2+(3)/aug-cc-pVTZ method for the mixed subset of S66 data set.

Isomer	Binding Energy / kcal mol^{-1}	
	CCSD(T)/CBS	MP2/CBS
prism	−48.31	−48.22
cage	−48.02	−48.13
book1	−47.61	−47.93
book2	−47.33	−47.65
cyclic chair	−46.52	−47.02
bag	−46.87	−47.19
cyclic boat1	−45.52	−46.04
cyclic boat2	−45.42	−45.91

Table I.6: Binding energies isomers of $(H_2O)_6$, using geometries from from Ref. 8.

Isomer	Binding Energy / kcal mol^{-1}	
	CCSD(T)/CBS	MP2/CBS
1	−169.02	−166.97
2	−169.03	−166.98
3	−168.01	−165.93
4	−169.07	−167.06
5	−169.91	−167.98
6	−171.42	−169.68
7	−169.92	−167.99
8	−166.65	−164.34
9	−163.59	−161.25
10	−168.22	−166.47

Table I.7: Binding energies for optimized (B3LYP/6-31G*) geometries of ten different isomers of $F^-(H_2O)_{10}$.

Isomer	Binding Energy / kcal mol^{-1}	
	CCSD(T)/CBS	MP2/CBS
1	−200.54	−197.97
2	−199.20	−196.68
3	−198.93	−196.51
4	−197.89	−195.02
5	−198.15	−195.58
6	−198.17	−195.24
7	−197.67	−194.62
8	−197.44	−194.90
9	−197.03	−194.29
10	−196.59	−193.92

Table I.8: Binding energies isomers of $(H_2O)_{20}$, using geometries from from Ref. 18.

n	CCSD(T)/CBS Binding Energy / kcal mol^{-1}					
	$F^-(H_2O)_n$	$Cl^-(H_2O)_n$	$SO_4^{2-}(H_2O)_n$	$Li^+(H_2O)_n$	$Na^+(H_2O)_n$	$K^+(H_2O)_n$
1	−32.31	−15.50	−32.12	−34.43	−23.92	−17.89
2	−52.27	−31.01	−60.33	−64.15	−44.93	−33.73
3	−70.14	−47.62	−81.37	−87.81	−63.47	−49.33
4	−85.24	−58.20	−105.15	−105.35	−78.48	−62.57
5	−101.09	−74.73	−126.44	−120.59	−93.01	−78.12
6	−116.58	−86.86	−144.62	−134.75	−107.27	−92.35

Table I.9: Binding energies for clusters $F^-(H_2O)_n$, $Cl^-(H_2O)_n$, $SO_4^{2-}(H_2O)_n$, $Li^+(H_2O)_n$, $Na^+(H_2O)_n$, and $K^+(H_2O)_n$. $F^-(H_2O)_n$ and $Cl^-(H_2O)_n$ clusters were optimized at the level of RIMP2/aug-cc-pVTZ. $SO_4^{2-}(H_2O)_n$ clusters were optimized at the level of RIMP2/aug-cc-pVDZ. $Li^+(H_2O)_n$, $Na^+(H_2O)_n$, and $K^+(H_2O)_n$ were optimized at the level of TPSS/def2-TZVPP.

APPENDIX J

Supporting Information for: "Energy decomposition analysis with a well-defined charge-transfer term for interpreting intermolecular interactions"

J.1 Examination of basis-set dependence

J.1.1 Set of 11 charge-transfer complexes

In Fig. J.1 we examine the basis-set dependence of CT energies for this set of complexes. For this comparison, we choose the 6-311++G(3df,3pd) basis set, which is used frequently in ALMO-EDA,[30] along with def2-TZVPPD and def2-QZVPPD, which have been recommended for the cDFT scheme.[490] The cDFT method exhibits essentially no basis set dependence, whereas the basis-set dependence of the ALMO and SAPT methods is significant for the four complexes where the extent of CT is large ($\delta q > 0.1\ e^-$). CT energies for ALMO and SAPT vary by as much as 26.7 kJ/mol (ALMO-EDA) and 16.5 kJ/mol (SAPT) for the $H_3N\cdots BH_3$ complex. For the ALMO-EDA and SAPT calculations, the maximum deviations amongst the three basis sets occurs between def2-TZVPPD and def2-QZVPPD, indicating once again the extreme dependence on basis set size associated with these definitions of CT.

J.1.2 Cation–alkyl radical complexes

Using the same three basis sets for this set of 8 complexes, we once again find that the cDFT method exhibits very mild basis-set dependence, with CT energies that vary by no more than 1.4 kJ/mol amongst these three basis sets. ALMO-EDA results, in contrast, vary by as much as 27.3 kJ/mol across these basis sets and SAPT results vary by as much as 8.2 kJ/mol. The large basis-set dependence of the ALMO-EDA results cannot be explained by BSSE, as the counterpoise corrections are no larger than 2.5 kJ/mol.

J.1.3 Nucleophile–$C_6H_5^{\bullet+}$ complexes

In most cases, CT energies calculated using B3LYP are greatly overestimated as compared to the those computed using ωB97X-D3. For example, the mean unsigned deviation for ALMO-EDA CT energies, comparing B3LYP/ and ωB97X-D3/ 6-311++G(3df,3pd) results is 9.3 kJ/mol, and the corresponding mean unsigned deviation for cDFT calculations is 6.0 kJ/mol. We attribute the larger B3LYP values of the CT energies to greater delocalization error with this functional, which is known to exaggerate the extent of CT.[502,503]

Examining the same three basis sets considered above for this set of 12 complexes, we find that the maximum basis-set deviation in the CT energy is 4.5 kJ/mol for ALMO-EDA and 2.1 kJ/mol for cDFT. For SAPT and SAPT+δSCF, the maximum deviations are 3.7 and 3.9 kJ/mol, respectively, for $C_6H_5^{\bullet+}\cdots HBr$ and $C_6H_5^{\bullet+}\cdots PH_3$. As in all of the results reported in this work, cDFT results are far less dependent on

the choice of basis set as compared to orbital-based EDAs.

J.2 Exchange-dipole model for $C_6H_5^{\bullet+}\cdots H_2O$

The high-level SAPT2+3(CCD)/aug-cc-pVTZ method was used to estimate the dispersion energy (for use with dlDF) for the $C_6H_5^{\bullet+}\cdots H_2O$ complex. These calculations were performed using the Psi4 program,[187] and because this method is only available for closed-shell species, we substitute the closed shell $C_6H_6\cdots H_2O$ complex for these calculations. To assess the magnitude of this approximation, we turn to the Becke-Johnson exchange-dipole model (XDM),[355,357] a DFT-based way to compute dispersion interactions that is available for both closed- and open-shell species. At the B3LYP/aug-cc-pVTZ level of theory, we find that the XDM dispersion energy for the open-shell $C_6H_5^{\bullet+}\cdots H_2O$ complex is smaller than that of the closed-shell $C_6H_6\cdots H_2O$ complex, but only by 0.7 kJ/mol (on-top orientation) or 0.3 kJ/mol (side-on orientation). Assuming that these differences are transferable to SAPT calculations, they amount to only 3% of the SAPT2+3(CCD)/aug-cc-pVTZ dispersion energy for $C_6H_6\cdots H_2O$, in either orientation. This justifies our use of the closed-shell complex to obtain an accurate dispersion correction to augment the dlDF functional.

J.3 Charge-transfer interactions in alkyl and aryl complexes

CT energies in alkyl and aryl complexes are shown in Tables J.3 and J.2, respectively.

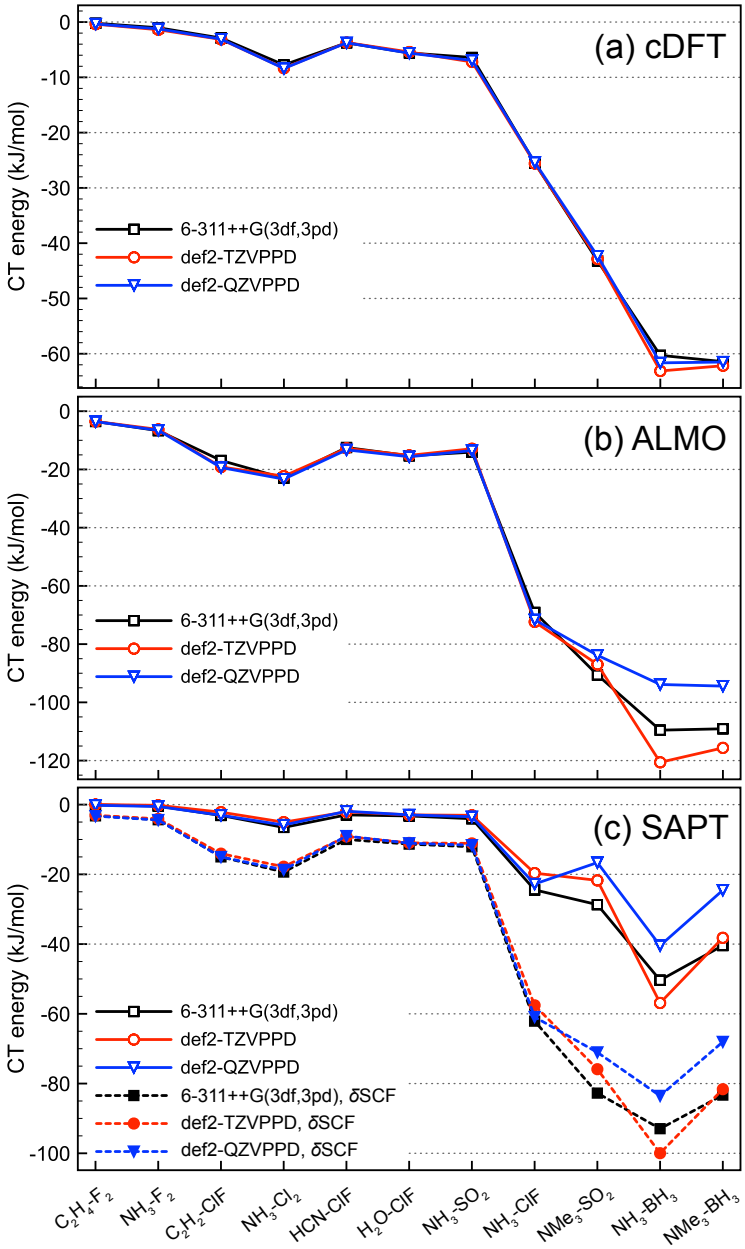

Figure J.1: CT energies computed in a variety of basis sets using (a) cDFT with the ωB97X-D3 functional, (b) ALMO-EDA using the ωB97X-D3 functional, and (c) SAPT with and without the $\delta E_{\text{int}}^{\text{HF}}$ ("δSCF") correction, using the LRC-ωPBE functional.

Cation	Alkyl Species	$(IE - EA)^{-1}$/ eV^{-1}	CT Energy / kJ mol^{-1}			
			cDFT	ALMO	SAPT	SAPT+δSCF
H_3O^+	CH_4	0.12	−30.30	−22.05	−2.73	−17.50
	$(CH_3)_3CH$	0.16	−41.55	−32.42	−3.37	−28.80
	$CH_3^\bullet$	0.18	−45.46	−48.84	−7.34	−41.75
	$(CH_3)_3C^\bullet$	0.42	−122.52	−96.96	−13.89	−124.98
NH_4^+	CH_4	0.13	−11.22	−7.11	−0.78	−5.21
	$(CH_3)_3CH$	0.17	−14.06	−10.06	−1.05	−7.80
	$CH_3^\bullet$	0.20	−16.91	−20.97	−2.84	−14.81
	$(CH_3)_3C^\bullet$	0.51	−33.71	−37.53	−4.92	−32.16

Table J.1: CT energies for the cation–alkyl complexes.

Orientation	Nucleophile	$(IE - EA)^{-1}$/ eV^{-1}	CT Energy / kJ mol^{-1}			
			cDFT	ALMO	SAPT	SAPT+δSCF
On-Top	PH_3	1.59	−80.34	−118.69	−9.10	−93.07
	NH_3	1.20	−63.45	−88.72	−10.05	−77.17
	CH_3COOH	0.71	−28.76	−30.69	−3.04	−27.33
	CH_3OH	0.63	−24.14	−29.69	−2.64	−22.48
	CH_3Cl	0.50	−17.49	−23.73	−1.79	−15.46
	HBr	0.41	−17.93	−27.44	−2.00	−16.76
	CH_3CN	0.34	−14.53	−16.65	−1.82	−17.89
	H_2O	0.30	−10.14	−11.86	−1.34	−10.78
Side-On	CH_3COOH	0.71	−4.60	−3.00	−0.65	−3.54
	CH_3OH	0.63	−5.28	−5.13	−1.05	−5.85
	CH_3CN	0.34	−5.55	−6.16	−1.31	−7.18
	H_2O	0.30	−3.95	−4.27	−0.91	−4.36

Table J.2: CT energies for $C_6H_5^{\bullet+} \cdots$ nucleophile complexes, considering two orientations.

Bibliography

[1] Z. Chen et al., J. Biol. Chem. **269**, 26344 (1994).

[2] M. N. Ucisik, D. S. Dashti, J. C. Faver, and K. M. Merz, Jr., J. Chem. Phys. **135**, 085101:1 (2011).

[3] D. J. Wales and M. P. Hodges, Chem. Phys. Lett. **286**, 65 (1998).

[4] S. Zahn, D. R. MacFarlane, and E. I. Izgorodina, Phys. Chem. Chem. Phys. **15**, 13664 (2013).

[5] C. D. Sherrill, T. Takatani, and E. G. Hohenstein, J. Phys. Chem. A **113**, 10146 (2009).

[6] K. Pernal, R. Podeszwa, K. Patkowski, and K. Szalewicz, Phys. Rev. Lett. **103**, 263201:1 (2009).

[7] L. Grafova, M. Pitonak, J. Řezác, and P. Hobza, J. Chem. Theory Comput. **6**, 2365 (2010).

[8] Y. Chen and H. Li, J. Phys. Chem. A **114**, 11719 (2010).

[9] J. Han, D. G. Truhlar, and J. Gao, Theor. Chem. Acc. **131**, 1161:1 (2012).

[10] J. Řezác, K. E. Riley, and P. Hobza, J. Chem. Theory Comput. **7**, 2427 (2011).

[11] L. Goerigk, H. Kruse, and S. Grimme, ChemPhysChem **12**, 3421 (2011).

[12] O. A. Vydrov and T. Van Voorhis, J. Chem. Theory Comput. **8**, 1929 (2012).

[13] Y.-S. Lin, G.-D. Li, S.-P. Mao, and J.-D. Chai, J. Chem. Theory Comput. **9**, 263 (2013).

[14] S. Grimme, S. Ehrlich, and L. Goerigk, J. Comput. Chem. **32**, 1456 (2011).

[15] A. Fiethen, G. Jansen, A. Hesselmann, and M. Schütz, J. Am. Chem. Soc. **130**, 1802 (2008).

[16] A. Heßelmann, J. Phys. Chem. A **115**, 11321 (2011).

[17] M. Goldey, A. Dutoi, and M. Head-Gordon, Phys. Chem. Chem. Phys. **15**, 15869 (2013).

[18] K. Wang, W. Li, and S. Li, J. Chem. Theory Comput. **10**, 1546 (2014).

[19] R. Sedlak et al., J. Chem. Theory Comput. **9**, 3364 (2013).

[20] S. Grimme, Chem. Eur. J. **18**, 9955 (2012).

[21] A. Hansen, F. Neese, and S. Grimme, (2015).

[22] R. Sure and S. Grimme, J. Chem. Theory Comput. **11**, 3785 (2015).

[23] T. Risthaus and S. Grimme, J. Chem. Theory Comput. **9**, 1580 (2013).

[24] J. Řezác and P. Hobza, J. Chem. Theory Comput. **7**, 685 (2011).

[25] E. B. Fleischer and K. C. Janda, J. Phys. Chem. A **117**, 4001 (2013).

[26] M. Modrzejewski, L. Rajchel, G. Chalasinski, and M. M. Szczesniak, J. Phys. Chem. A **117**, 11580 (2013).

[27] R. J. Azar and M. Head-Gordon, J. Chem. Phys. **136**, 024103:1 (2012).

[28] R. Z. Khaliullin, A. T. Bell, and M. Head-Gordon, J. Chem. Phys. **128**, 184112:1 (2008).

[29] D. Cappelletti, E. Ronca, L. Belpassi, F. Tarantelli, and F. Pirani, Acc. Chem. Res. **45**, 1571 (2012).

[30] P. R. Horn, E. J. Sundstrom, T. A. Baker, and M. Head-Gordon, J. Chem. Phys. **138**, 134119:1 (2013).

[31] S. G. Lias et al., *"Ion Energetics Data" in NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, National Institute of Standards and Technology, Gaithersburg, MD, 2011.

[32] I. K. Attah et al., J. Phys. Chem. Lett. **6**, 1111 (2015).

[33] J. Řezáč and P. Hobza, J. Chem. Theory Comput. **9**, 2151 (2013).

[34] M. S. Marshall, L. A. Burns, and C. D. Sherrill, J. Chem. Phys. **135**, 194102:1 (2011).

[35] J. Řezác, K. E. Riley, and P. Hobza, J. Chem. Theory Comput. **7**, 3466 (2011).

[36] J. Řezác, K. E. Riley, and P. Hobza, J. Chem. Theory Comput. **8**, 4285 (2012).

[37] J. C. Flick, D. Kosenkov, E. G. Hohenstein, C. D. Sherrill, and L. V. Slipchenko, J. Chem. Theory Comput. **8**, 2835 (2012).

[38] T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall, and C. D. Sherrill, J. Chem. Phys. **132**, 144104:1 (2010).

[39] J. Řezác and P. Hobza, J. Chem. Theory Comput. **8**, 141 (2012).

[40] A. Ambrosetti, D. Alfè, R. A. DiStasio, Jr., and A. Tkatchenko, J. Phys. Chem. Lett. **5**, 849 (2014).

[41] D. M. Bates and G. S. Tschumper, J. Phys. Chem. A **113**, 3555 (2009).

[42] A. Heßelmann and T. Korona, J. Chem. Phys. **141**, 094107:1 (2014).

[43] M. S. Gordon, D. G. Fedorov, S. R. Pruitt, and L. V. Slipchenko, Chem. Rev. **112**, 632 (2012).

[44] L. D. Jacobson, R. M. Richard, K. U. Lao, and J. M. Herbert, Annu. Rep. Comput. Chem. **9**, 25 (2013).

[45] R. M. Richard, K. U. Lao, and J. M. Herbert, Acc. Chem. Res. **47**, 2828 (2014).

[46] R. M. Richard, K. U. Lao, and J. M. Herbert, J. Phys. Chem. Lett. **4**, 2674 (2013).

[47] R. M. Richard, K. U. Lao, and J. M. Herbert, J. Chem. Phys. **139**, 224102:1 (2013).

[48] R. M. Richard, K. U. Lao, and J. M. Herbert, J. Chem. Phys. **141**, 014108:1 (2014).

[49] K. U. Lao, K.-Y. Liu, R. M. Richard, and J. M. Herbert, (2016), submitted.

[50] K. U. Lao and J. M. Herbert, J. Phys. Chem. A **116**, 3042 (2012).

[51] K. U. Lao and J. M. Herbert, J. Chem. Phys. **140**, 044108:1 (2014).

[52] K. U. Lao, R. Schäffer, G. Jansen, and J. M. Herbert, J. Chem. Theory Comput. **11**, 2473 (2015).

[53] K. U. Lao and J. M. Herbert, (2016), (to be submitted).

[54] J. M. Herbert, L. D. Jacobson, K. U. Lao, and M. A. Rohrdanz, Phys. Chem. Chem. Phys. **14**, 7679 (2012).

[55] K. U. Lao and J. M. Herbert, J. Phys. Chem. Lett. **3**, 3241 (2012).

[56] K. U. Lao and J. M. Herbert, J. Chem. Phys. **139**, 034107:1 (2013), Erratum: *ibid.* **139**, 034107 (2013).

[57] K. U. Lao and J. M. Herbert, J. Phys. Chem. A **119**, 235 (2015).

[58] K. U. Lao, A. Pomogaeva, D. M. Chipman, and J. M. Herbert, (2016), (in preparation).

[59] K. U. Lao and J. M. Herbert, (2016), (in preparation).

[60] K. U. Lao and J. M. Herbert, (2016), (in preparation).

[61] K. U. Lao and J. M. Herbert, (2016), (in preparation).

[62] K. U. Lao and J. M. Herbert, (2016), (in preparation).

[63] K. U. Lao and J. M. Herbert, (2016), (in preparation).

[64] E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **3**, 46 (2007).

[65] L. D. Jacobson and J. M. Herbert, J. Chem. Phys. **134**, 094118:1 (2011).

[66] M. A. Collins and R. P. Bettens, Chem. Rev. **115**, 5607 (2015).

[67] W. Kohn, Phys. Rev. Lett. **76**, 3168 (1996).

[68] E. Prodan and W. Kohn, Proc. Natl. Acad. Sci. USA **102**, 11635 (2005).

[69] C. A. Coulson, *Valence*, Oxford University Press, 2nd edition, 1961.

[70] C. J. Cramer, *Essentials of Computational Chemistry: Theories and Methods*, Wiley, New York, 2002.

[71] R. M. Richard and J. M. Herbert, J. Chem. Phys. **137**, 064113:1 (2012).

[72] R. M. Richard and J. M. Herbert, J. Chem. Theory Comput. **9**, 1408 (2013).

[73] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory*, Wiley, New York, 2000.

[74] S. F. Boys and F. Bernardi, Mol. Phys. **19**, 553 (1970).

[75] G. S. Tschumper, Reliable electronic structure computations for weak noncovalent interactions in clusters, in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and T. R. Cundari, volume 26, chapter 2, pages 39–90, Wiley-VCH, 2009.

[76] M. Kamiya, S. Hirata, and M. Valiev, J. Chem. Phys. **128**, 074103:1 (2008).

[77] J. Liu and J. M. Herbert, J. Chem. Theory Comput. **12**, 572 (2016).

[78] W. Li, S. Li, and Y. Jiang, J. Phys. Chem. A **111**, 2193 (2007).

[79] S. Hua, W. Hua, and S. Li, J. Phys. Chem. A **114**, 8126 (2010).

[80] F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, and J. H. van Lenthe, Chem. Rev. **94**, 1873 (1994).

[81] P. Valiron and I. Mayer, Chem. Phys. Lett. **275**, 46 (1997).

[82] J. F. Ouyang and R. P. A. Bettens, J. Chem. Theory Comput. **11**, 5132 (2015).

[83] G. J. O. Beran, J. Chem. Phys. **130**, 164115:1 (2009).

[84] E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **3**, 1342 (2007).

[85] U. Góra, R. Podeszwa, W. Cencek, and K. Szalewicz, J. Chem. Phys. **135**, 224102:1 (2011).

[86] N. J. Mayhall and K. Raghavachari, J. Chem. Theory Comput. **7**, 1336 (2011).

[87] H. W. Qi, H. R. Leverentz, and D. G. Truhlar, J. Phys. Chem. A **117**, 4486 (2013).

[88] J. Friedrich et al., J. Phys. Chem. Lett. **5**, 666 (2014).

[89] S. Kazachenko and A. J. Thakkar, J. Chem. Phys. **138**, 194302:1 (2013).

[90] J. F. Ouyang, M. W. Cvitkovic, and R. P. A. Bettens, J. Chem. Theory Comput. **10**, 3699 (2014).

[91] P. M. W. Gill, B. G. Johnson, and J. A. Pople, Chem. Phys. Lett. **209**, 506 (1993).

[92] H. R. Leverentz and D. G. Truhlar, J. Chem. Theory Comput. **5**, 1573 (2009).

[93] F. L. Hirshfeld, Theor. Chem. Acc. **44**, 129 (1977).

[94] A. E. Reed, R. B. Weinstock, and F. Weinhold, J. Chem. Phys. **83**, 735 (1985).

[95] C. M. Breneman and K. B. Wiberg, J. Comput. Chem. **11**, 361 (1990).

[96] A. V. Marenich, S. V. Jerome, C. J. Cramer, and D. G. Truhlar, J. Chem. Theory Comput. **8**, 527 (2012).

[97] Y. Shao et al., Mol. Phys. **113**, 184 (2015).

[98] M. Feyereisen, G. Fitzgerald, and A. Komornicki, Chem. Phys. Lett. **208**, 359 (1993).

[99] R. A. Kendall and H. A. Früchtl, Theor. Chem. Acc. **97**, 158 (1997).

[100] F. Weigend, M. Häser, J. Patzelt, and R. Ahlrichs, Chem. Phys. Lett. **294**, 143 (1998).

[101] R. A. DiStasio Jr., R. P. Steele, Y. M. Rhee, Y. Shao, and M. Head-Gordon, J. Comput. Chem. **28**, 839 (2007).

[102] F. Weigend, A. Köhn, and C. Hättig, J. Chem. Phys. **116**, 3175 (2002).

[103] D. M. Bates, J. R. Smith, T. Janowski, and G. S. Tschumper, J. Chem. Phys. **135**, 044123:1 (2011).

[104] R. O. Ramabhadran and K. Raghavachari, J. Chem. Theory Comput. **9**, 3986 (2013).

[105] N. Mardirossian and M. Head-Gordon, Phys. Chem. Chem. Phys. **16**, 9904 (2014).

[106] P.-O. Löwdin, Int. J. Quantum Chem. Symp. **19**, 19 (1986).

[107] I. G. Kaplan, R. Santamaria, and O. Novaro, Mol. Phys. **84**, 105 (1995).

[108] B. Paulus, K. Rosciszewski, N. Gaston, P. Schwerdtfeger, and H. Stoll, Phys. Rev. B **70**, 165106:1 (2004).

[109] A. Hermann et al., Phys. Rev. A **76**, 013202:1 (2007).

[110] T. J. Mach and T. D. Crawford, Theor. Chem. Acc. **133**, 1449:1 (2014).

[111] M. P. Hodges, A. J. Stone, and S. S. Xantheas, J. Phys. Chem. A **101**, 9163 (1997).

[112] R. Peverati and D. G. Truhlar, J. Phys. Chem. Lett. **2**, 2810 (2011).

[113] N. Mardirossian and M. Head-Gordon, J. Chem. Theory Comput. **9**, 4453 (2013).

[114] B. Jeziorski et al., SAPT: A program for many-body symmetry-adapted perturbation theory calculations of intermolecular interaction energies, in *Methods and Techniques in Computational Chemistry: METECC-94*, edited by E. Clementi, volume B, chapter 3, pages 79–129, STEF, Cagliari, 1993.

[115] B. Jeziorski, R. Moszynski, and K. Szalewicz, Chem. Rev. **94**, 1887 (1994).

[116] K. Szalewicz, K. Patkowski, and B. Jeziorski, Intermolecular interactions via perturbation theory: From diatoms to biomolecules, in *Intermolecular Forces and Clusters II*, edited by D. J. Wales, volume 116 of *Structure and Bonding*, pages 43–117, Springer-Verlag, Berlin, 2005.

[117] E. G. Hohenstein and C. D. Sherrill, WIREs Comput. Mol. Sci. **2**, 304 (2012).

[118] A. Heßelmann and G. Jansen, Phys. Chem. Chem. Phys. **5**, 5010 (2003).

[119] R. Bukowski, K. Szalewicz, G. Groenenboom, and A. van der Avoird, J. Chem. Phys. **125**, 044301:1 (2006).

[120] R. Podeszwa, R. Bukowski, and K. Szalewicz, J. Phys. Chem. A **110**, 10345 (2006).

[121] R. Podeszwa, R. Bukowski, B. M. Rice, and K. Szalewicz, Phys. Chem. Chem. Phys. **9**, 5561 (2007).

[122] R. Podeszwa, B. M. Rice, and K. Szalewicz, Phys. Rev. Lett. **101**, 115503:1 (2008).

[123] R. Podeszwa, B. M. Rice, and K. Szalewicz, Phys. Chem. Chem. Phys. **11**, 5512 (2009).

[124] D. E. Taylor, F. Rob, B. M. Rice, R. Podeszwa, and K. Szalewicz, Phys. Chem. Chem. Phys. **13**, 16629 (2011).

[125] K. Yu, J. G. McDaniel, and J. R. Schmidt, J. Phys. Chem. B **115**, 10054 (2011).

[126] X. Li, A. Volkov, P. Coppens, and K. Szalewicz, Acta Cryst. D **62**, 639 (2006).

[127] B. Jeziorski, M. Bulski, and L. Piela, Int. J. Quantum Chem. **10**, 281 (1976).

[128] G. Chałasiǹski and B. Jeziorski, Mol. Phys. **32**, 81 (1976).

[129] E. G. Hohenstein, H. M. Jaeger, E. J. Carrell, G. S. Tschumper, and C. D. Sherrill, J. Chem. Theory Comput. **7**, 2842 (2011).

[130] K. Patkowski, K. Szalewicz, and B. Jeziorski, J. Chem. Phys. **125**, 154107:1 (2006).

[131] K. Patkowski, K. Szalewicz, and B. Jeziorski, Theor. Chem. Acc. **127**, 211 (2010).

[132] R. Bukowski et al., *SAPT2008: An Ab Initio Program for Many-Body Symmetry-Adapted Perturbation Theory Calculations of Intermolecular Interaction Energies*, University of Delaware and University of Warsaw, 2008.

[133] V. Saunders and M. Guest, *ATMOL Program Package*, SERC Daresbury Laboratory, Daresbury, Great Britain.

[134] A. J. Sadlej, Mol. Phys. **39**, 1249 (1980).

[135] M. Jeziorska, B. Jeziorski, and J. Čížek, Int. J. Quantum Chem. **32**, 149 (1987).

[136] R. Moszynski, T. Heijmen, and B. Jeziorski, Mol. Phys. **88**, 741 (1996).

[137] A. Heßelmann, G. Jansen, and M. Schütz, J. Chem. Phys. **122**, 014103:1 (2005).

[138] I. G. Kaplan, *Intermolecular Interactions: Physical Picture, Computational Methods and Model Potentials*, John Wiley and Sons, Ltd, Chichester, 2006.

[139] K. Müller-Dethlefs and P. Hobza, Chem. Rev. **100**, 143 (2000).

[140] K. E. Riley, M. Pitoňák, P. Jurečka, and P. Hobza, Chem. Rev. **110**, 5023 (2010).

[141] L. Šimová, J. Řezáč, and P. Hobza, J. Chem. Theory Comput. **9**, 3420 (2013).

[142] Y. Zhao and D. G. Truhlar, Theor. Chem. Acc. **120**, 215 (2008).

[143] S. Grimme, WIREs Comput. Mol. Sci. **1**, 211 (2011).

[144] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist, Phys. Rev. Lett. **92**, 246401:1 (2004).

[145] K. Lee, É. D. Murray, L. Kong, B. I. Lundqvist, and D. C. Langreth, Phys. Rev. B **82**, 081101:1 (2010).

[146] O. A. Vydrov and T. Van Voorhis, Phys. Rev. Lett. **103**, 063004:1 (2009).

[147] O. A. Vydrov and T. Van Voorhis, J. Chem. Phys. **133**, 244103:1 (2010).

[148] A. Heßelmann, J. Chem. Theory Comput. **9**, 273 (2013).

[149] K. Szalewicz, WIREs Comput. Mol. Sci. **2**, 254 (2012).

[150] G. Jansen, WIREs Comput. Mol. Sci. **4**, 127 (2014).

[151] S. Rybak, B. Jeziorski, and K. Szalewicz, J. Chem. Phys. **95**, 6576 (1991).

[152] R. Moszynski, B. Jeziorski, and K. Szalewicz, J. Chem. Phys. **100**, 1312 (1994).

[153] H. L. Williams, E. M. Mas, K. Szalewicz, and B. Jeziorski, J. Chem. Phys. **103**, 7374 (1995).

[154] T. Korona, J. Chem. Phys. **128**, 224104:1 (2008).

[155] T. Korona, Phys. Chem. Chem. Phys. **10**, 6509 (2008).

[156] T. Korona, J. Chem. Theory Comput. **5**, 2663 (2009).

[157] E. G. Hohenstein and C. D. Sherrill, J. Chem. Phys. **132**, 184111:1 (2010).

[158] E. G. Hohenstein and C. D. Sherrill, J. Chem. Phys. **133**, 014101:1 (2010).

[159] E. G. Hohenstein and C. D. Sherrill, J. Chem. Phys. **133**, 104107:1 (2010).

[160] E. G. Hohenstein, R. M. Parrish, C. D. Sherrill, J. M. Turney, and H. F. Schaefer III, J. Chem. Phys. **135**, 174107:1 (2011).

[161] A. J. Misquitta, R. Podeszwa, B. Jeziorski, and K. Szalewicz, J. Chem. Phys. **123**, 214103:1 (2005).

[162] A. J. Misquitta and K. Szalewicz, Chem. Phys. Lett. **357**, 301 (2002).

[163] A. Heßelmann and G. Jansen, Chem. Phys. Lett. **357**, 464 (2002).

[164] A. Heßelmann and G. Jansen, Chem. Phys. Lett. **362**, 319 (2002).

[165] R. Moszynski, B. Jeziorski, and K. Szalewicz, Chem. Phys. **166**, 329 (1992).

[166] D. Kim, P. Tarakeshwar, and K. S. Kim, J. Phys. Chem. A **108**, 1250 (2004).

[167] S. Zahn, F. Uhlig, J. Thar, C. Spickermann, and B. Kirchner, Angew. Chem. Int. Ed. Engl. **47**, 3639 (2008).

[168] N. J. Singh, S. K. Min, D. Y. Kim, and K. S. Kim, J. Phys. Chem. B **5**, 515 (2009).

[169] B. Bankiewicz, P. Matczak, and M. Palusiak, J. Phys. Chem. A **116**, 452 (2012).

[170] P. Matczak, J. Phys. Chem. A **116**, 8731 (2012).

[171] K. Ansorg, M. Tafipolsky, and B. Engels, J. Phys. Chem. B **117**, 10093 (2013).

[172] T. Korona, Mol. Phys. **111**, 3705 (2013).

[173] T. M. Parker, L. A. Burns, R. M. Parrish, A. G. Ryno, and C. D. Sherrill, J. Chem. Phys. **140**, 094106:1 (2014).

[174] R. Schäffer and G. Jansen, Theor. Chem. Acc. **131**, 1235:1 (2012).

[175] R. Schäffer and G. Jansen, Mol. Phys. **111**, 2570 (2013).

[176] M. Goldey and M. Head-Gordon, J. Phys. Chem. Lett. **3**, 3592 (2012).

[177] M. Goldey and M. Head-Gordon, J. Phys. Chem. B **118**, 6519 (2014).

[178] M. Goldey and M. Head-Gordon, Chem. Phys. Lett. **608**, 249 (2014).

[179] R. Moszynski, B. Jeziorski, A. Ratkiewicz, and S. Rybak, J. Chem. Phys. **99**, 8856 (1993).

[180] R. Moszyński, S. M. Cybulski, and G. Chałasiński, J. Chem. Phys. **100**, 4998 (1994).

[181] K. Patkowski, R. Podeszwa, and K. Szalewicz, J. Phys. Chem. A **111**, 12822 (2007).

[182] E. Papajak and D. G. Truhlar, J. Chem. Theory Comput. **7**, 10 (2011).

[183] H. L. Williams, K. Szalewicz, R. Moszynski, and B. Jeziorski, J. Chem. Phys. **103**, 4586 (1995).

[184] R. M. Parrish, E. G. Hohenstein, and C. D. Sherrill, J. Chem. Phys. **139**, 174102:1 (2013).

[185] P. Schah-Mohammedi et al., J. Am. Chem. Soc. **122**, 12878 (2000).

[186] P. Jurečka, J. Šponer, J. Černý, and P. Hobza, Phys. Chem. Chem. Phys. **8**, 1985 (2006).

[187] J. M. Turney et al., WIREs Comput. Mol. Sci. **2**, 556 (2012).

[188] F. Neese, WIREs Comput. Mol. Sci. **2**, 73 (2012).

[189] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, WIREs Comput. Mol. Sci. **2**, 242 (2012).

[190] A. Halkier et al., Chem. Phys. Lett. **286**, 243 (1998).

[191] E. Valeev, Chem. Phys. Lett. **395**, 190 (2004).

[192] F. A. Bischoff, S. Wolfsegger, D. P. Tew, and W. Klopper, Mol. Phys. **107**, 963 (2009).

[193] T. B. Adler, G. Knizia, and H.-J. Werner, J. Chem. Phys. **127**, 221106:1 (2007).

[194] J. Noga and J. Šimunek, Chem. Phys. **356**, 1 (2009).

[195] T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).

[196] R. A. Kendall, T. H. Dunning, Jr., and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).

[197] F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).

[198] E. K. U. Gross, J. F. Dobson, and M. Petersilka, Topics Curr. Chem. **181**, 81 (1996).

[199] M. Grüning, O. V. Gritsenko, S. J. A. van Gisbergen, and E. J. Baerends, J. Chem. Phys. **114**, 652 (2001).

[200] D. Lee and K. Burke, Mol. Phys. **108**, 2687 (2010).

[201] J.-D. Chai and M. Head-Gordon, Phys. Chem. Chem. Phys. **10**, 6615 (2008).

[202] N. Mardirossian and M. Head-Gordon, J. Chem. Phys. **142**, 074111:1 (2015).

[203] N. Mardirossian, D. S. Lambrecht, L. McCaslin, S. S. Xantheas, and M. Head-Gordon, J. Chem. Theory Comput. **9**, 1368 (2013).

[204] A. D. Dutoi and M. Head-Gordon, J. Phys. Chem. A **112**, 2110 (2008).

[205] R. Podeszwa, R. Bukowski, and K. Szalewicz, J. Chem. Theory Comput. **2**, 400 (2006).

[206] L. A. Burns, A. Vázquez-Mayagoitia, B. G. Sumpter, and C. D. Sherrill, J. Chem. Phys. **134**, 084107:1 (2011).

[207] D. Cremer, WIREs Comput. Mol. Sci. **1**, 509 (2011).

[208] H. L. Williams and C. F. Chabalowski, J. Phys. Chem. A **105**, 646 (2001).

[209] G. Jansen and A. Heßelmann, J. Phys. Chem. A **105**, 11156 (2001).

[210] D. J. Tozer and N. C. Handy, J. Chem. Phys. **109**, 10180 (1998).

[211] J. P. Perdew and K. Burke, Int. J. Quantum Chem. **57**, 309 (1996).

[212] G. Chałasiński and M. M. Szczęśniak, Chem. Rev. **94**, 1723 (1994).

[213] A. Heßelmann and G. Jansen, Chem. Phys. Lett. **367**, 778 (2003).

[214] A. J. Misquitta, B. Jeziorski, and K. Szalewicz, Phys. Rev. Lett. **91**, 033201:1 (2003).

[215] R. van Leeuwen and E. J. Baerends, Phys. Rev. A **49**, 2421 (1994).

[216] A. P. Gaiduk and V. N. Staroverov, J. Chem. Phys. **131**, 044107:1 (2009).

[217] C.-R. Pan, P.-T. Fang, and J.-D. Chai, Phys. Rev. A **87**, 052510:1 (2013).

[218] W. Cencek and K. Szalewicz, J. Chem. Phys. **139**, 024104:1 (2013), Erratum: J. Chem. Phys., **140**, 149902 (2014).

[219] H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao, J. Chem. Phys. **115**, 3540 (2001).

[220] T. Yanai, D. P. Tew, and N. C. Handy, Chem. Phys. Lett. **393**, 51 (2004).

[221] O. A. Vydrov, J. Heyd, A. V. Krukau, and G. E. Scuseria, J. Chem. Phys. **125**, 074106:1 (2006).

[222] T. M. Henderson, B. G. Janesko, and G. E. Scuseria, J. Chem. Phys. **128**, 194105:1 (2008).

[223] J.-W. Song, T. Hirosawa, T. Tsuneda, and K. Hirao, J. Chem. Phys. **126**, 154105:1 (2007).

[224] E. Livshits and R. Baer, Phys. Chem. Chem. Phys. **9**, 2932 (2007).

[225] J.-D. Chai and M. Head-Gordon, J. Chem. Phys. **128**, 084106:1 (2008).

[226] M. A. Rohrdanz and J. M. Herbert, J. Chem. Phys. **129**, 034107:1 (2008).

[227] M. A. Rohrdanz, K. M. Martins, and J. M. Herbert, J. Chem. Phys. **130**, 054112:1 (2009).

[228] A. W. Lange, M. A. Rohrdanz, and J. M. Herbert, J. Phys. Chem. B **112**, 6304 (2008).

[229] T. Stein, L. Kronik, and R. Baer, J. Chem. Phys. **131**, 244119:1 (2009).

[230] R. Baer, E. Livshits, and U. Salzner, Annu. Rev. Phys. Chem. **61**, 85 (2010).

[231] S. Refaely-Abramson, R. Baer, and L. Kronik, Phys. Rev. B **84**, 075144:1 (2011).

[232] L. Kronik, T. Stein, S. Refaely-Abramson, and R. Baer, J. Chem. Theory Comput. **8**, 1515 (2012).

[233] M. E. Foster and B. M. Wong, J. Chem. Theory Comput. **8**, 2682 (2012).

[234] H. L. Williams and C. F. Chabalowski, J. Phys. Chem. A **105**, 11158 (2001).

[235] A. J. Misquitta and K. Szalewicz, J. Chem. Phys. **122**, 214109:1 (2005).

[236] T. Korona, H. L. Williams, R. Bukowski, B. Jeziorski, and K. Szalewicz, J. Chem. Phys. **106**, 5109 (1997).

[237] M. Jeziorska, W. Cencek, K. Patkowski, B. Jeziorski, and K. Szalewicz, J. Chem. Phys. **127**, 124303:1 (2007).

[238] M. Jeziorska, W. Cencek, K. Patkowski, B. Jeziorski, and K. Szalewicz, Int. J. Quantum Chem. **108**, 2053 (2008).

[239] T. Korona and B. Jeziorski, J. Chem. Phys. **128**, 144107:1 (2008).

[240] A. W. Lange and J. M. Herbert, J. Am. Chem. Soc. **131**, 3913 (2009).

[241] O. A. Vydrov and G. E. Scuseria, J. Chem. Phys. **125**, 234109:1 (2006).

[242] Y. Shao et al., Phys. Chem. Chem. Phys. **8**, 3172 (2006).

[243] A. I. Krylov and P. M. W. Gill, WIREs Comput. Mol. Sci. **3**, 317 (2013).

[244] R. Podeszwa, K. Pernal, K. Patkowski, and K. Szalewicz, J. Phys. Chem. Lett. **1**, 550 (2010).

[245] W. Cencek and K. Szalewicz, Int. J. Quantum Chem. **108**, 2191 (2008).

[246] R. Podeszwa, W. Cencek, and K. Szalewicz, J. Chem. Theory Comput. **8**, 1963 (2012).

[247] A. Hesselmann, G. Jansen, and M. Schütz, J. Am. Chem. Soc. **128**, 11730 (2006).

[248] P. Yakovchuk, E. Protozanova, and M. D. Frank-Kamenetskii, Nucl. Acids Res. **34**, 564 (2006).

[249] V. F. Lotrich and K. Szalewicz, J. Chem. Phys. **106**, 9668 (1997).

[250] V. F. Lotrich and K. Szalewicz, J. Chem. Phys. **112**, 112 (2000).

[251] N. Turki et al., J. Chem. Phys. **109**, 7157 (1998).

[252] W. Xie, L. Song, D. G. Truhlar, and J. Gao, J. Chem. Phys. **128**, 234108:1 (2008).

[253] S. Grimme, J. Comput. Chem. **25**, 1463 (2004).

[254] S. Grimme, J. Comput. Chem. **27**, 1787 (2006).

[255] I. Adamovic and M. S. Gordon, Mol. Phys. **103**, 379 (2005).

[256] S. S. Xantheas, J. Chem. Phys. **104**, 8821 (1996).

[257] S. S. Xantheas, C. J. Burnham, and R. J. Harrison, J. Chem. Phys. **116**, 1493 (2002).

[258] S. S. Xantheas and E. Aprà, J. Chem. Phys. **120**, 823 (2004).

[259] G. S. Fanourgakis, E. Aprà, and S. S. Xantheas, J. Chem. Phys. **121**, 2655 (2004).

[260] S. Bulusu, S. Yoo, E. Aprà, S. S. Xantheas, and X. C. Zeng, J. Phys. Chem. A **110**, 11781 (2006).

[261] J. M. Herbert and M. Head-Gordon, J. Am. Chem. Soc. **128**, 13932 (2006).

[262] W. K. Olson et al., J. Mol. Biol. **313**, 229 (2001).

[263] S. Grimme, J. Chem. Phys. **118**, 9095 (2003).

[264] J. Antony and S. Grimme, J. Phys. Chem. A **111**, 4862 (2007).

[265] M. Pitonak, K. E. Riley, P. Neogrady, and P. Hobza, ChemPhysChem **9**, 1636 (2008).

[266] R. A. DiStasio, Jr. and M. Head-Gordon, Mol. Phys. **105**, 1073 (2007).

[267] R. Sedlak, P. Jurečka, and P. Hobza, J. Chem. Phys. **127**, 075104:1 (2007).

[268] E. Aprà et al., Liquid water: Obtaining the right answer for the right reasons, in *SC '09: Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, pages 66:1–7, ACM, New York, 2009.

[269] S. Grimme, L. Goerigk, and R. F. Fink, WIREs Comput. Mol. Sci. **2**, 886 (2012).

[270] K. E. Riley, J. A. Platts, J. Řezác, P. Hobza, and J. G. Hill, J. Phys. Chem. A **116**, 4159 (2012).

[271] K. E. Riley, J. Řezác, and P. Hobza, Phys. Chem. Chem. Phys. **13**, 21121 (2011).

[272] P. Hobza, Acc. Chem. Res. **45**, 663 (2012).

[273] J. Klimeš and A. Michaelides, J. Chem. Phys. **137**, 120901:1 (2012).

[274] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104:1 (2010).

[275] M. S. Gordon, L. V. Slipchenko, H. Li, and J. H. Jensen, Annu. Rep. Comput. Chem. **3**, 177 (2007).

[276] D. G. Fedorov, T. Nagata, and K. Kitaura, Phys. Chem. Chem. Phys. **14**, 7562 (2012).

[277] S. Wen, K. Nanda, Y. Huang, and G. J. O. Beran, Phys. Chem. Chem. Phys. **14**, 7578 (2012).

[278] S. M. Cybulski and M. L. Lytle, J. Chem. Phys. **127**, 141102:1 (2007).

[279] G. Chałasiński, M. M. Szczęśniak, and R. A. Kendall, J. Chem. Phys. **101**, 8860 (1994).

[280] J. M. Pedulla, K. Kim, and K. D. Jordan, Chem. Phys. Lett. **291**, 78 (1998).

[281] T. P. Tauer and C. D. Sherrill, J. Phys. Chem. A **109**, 10475 (2005).

[282] A. L. Ringer and C. D. Sherrill, Chem. Eur. J. **14**, 2542 (2008).

[283] K. T. Tang and J. P. Toennies, J. Chem. Phys. **80**, 3726 (1984).

[284] B. Wang and D. G. Truhlar, J. Chem. Theory Comput. **6**, 3330 (2010).

[285] J. J. P. Stewart, J. Mol. Model. **19**, 1 (2013).

[286] R. P. Steele, R. A. DiStasio, Jr., Y. Shao, J. Kong, and M. Head-Gordon, J. Chem. Phys. **125**, 074108:1 (2006).

[287] R. P. Steele, R. A. DiStasio, Jr., and M. Head-Gordon, J. Chem. Theory Comput. **5**, 1560 (2009).

[288] J. J. P. Stewart, *Stewart Computational Chemistry, Colorado Springs, CO, USA, HTTP://OpenMOPAC.net*, 2012.

[289] W. Xie and J. Gao, J. Chem. Theory Comput. **3**, 1890 (2007).

[290] R. Moszynski, B. Jeziorski, S. Rybak, K. Szalewicz, and H. L. Williams, J. Chem. Phys. **100**, 5080 (1994).

[291] T. Korona, R. Moszynski, and B. Jeziorski, Mol. Phys. **100**, 1723 (2002).

[292] K. Kitaura and K. Morokuma, Int. J. Quantum Chem. **10**, 325 (1976).

[293] W. Gao, H. Fen, X. Xuan, and L. Chen, J. Mol. Model. **18**, 4577 (2012).

[294] H. R. Leverentz, H. W. Qi, and D. G. Truhlar, J. Chem. Theory Comput. **9**, 995 (2013).

[295] O. Guvench and A. D. MacKerell, Curr. Opin. Struc. Biol. **19**, 56 (2009).

[296] S. Durdagi, C. Zhao, J. E. Cuervo, and S. Y. Noskov, Curr. Med. Chem. **18**, 2601 (2011).

[297] S. M. Woodley and R. Catlow, Nat. Mater. **7**, 937 (2008).

[298] J. P. Perdew, MRS Bull. **38**, 743 (2013).

[299] S. Grimme, J. Chem. Phys. **124**, 034108:1 (2006).

[300] J. C. Howard and G. S. Tschumper, WIREs Comput. Mol. Sci. **4**, 199 (2014).

[301] M. Del Ben, M. Schönherr, J. Hutter, and J. VandeVondele, J. Phys. Chem. Lett. **4**, 3753 (2013), Erratum: **5**, 3066–3067 (2014).

[302] A. Szabo and N. S. Ostlund, J. Chem. Phys. **67**, 4351 (1977).

[303] A. Tkatchenko, R. A. DiStasio Jr., M. Head-Gordon, and M. Scheffler, J. Chem. Phys. **131**, 094106:1 (2009).

[304] B. W. Hopkins and G. S. Tschumper, J. Phys. Chem. A **108**, 2941 (2004).

[305] M. Dubecký et al., J. Chem. Theory Comput. **9**, 4287 (2013).

[306] D. G. A. Smith, P. Jankowski, M. Slawik, H. A. Witek, and K. Patkowski, J. Chem. Theory Comput. **10**, 3140 (2014).

[307] S. Yoo, E. Aprà, X. C. Zeng, and S. S. Xantheas, J. Phys. Chem. Lett. **1**, 3122 (2010).

[308] M. Pitoňák, J. Řezác, and P. Hobza, Phys. Chem. Chem. Phys. **12**, 9611 (2010).

[309] A. Heßelmann, J. Chem. Phys. **128**, 144112:1 (2008).

[310] M. Pitoňák and A. Heßelmann, J. Chem. Theory Comput. **6**, 168 (2010).

[311] Y. Huang, Y. Shao, and G. J. O. Beran, J. Chem. Phys. **138**, 224112:1 (2013).

[312] Y. Huang, M. Goldey, M. Head-Gordon, and G. J. O. Beran, J. Chem. Theory Comput. **10**, 2054 (2014).

[313] R. Bukowski, R. Podeszwa, and K. Szalewicz, Chem. Phys. Lett. **414**, 111 (2005).

[314] J. G. McDaniel and J. R. Schmidt, J. Phys. Chem. B **118**, 8042 (2014).

[315] V. F. Lotrich and K. Szalewicz, J. Chem. Phys. **106**, 9688 (1997).

[316] J. Gao, J. Phys. Chem. B **101**, 657 (1997).

[317] J. Gao, J. Chem. Phys. **109**, 2346 (1998).

[318] J. Gao et al., Acc. Chem. Res. **47**, 2837 (2014).

[319] Z. C. Holden, R. M. Richard, and J. M. Herbert, J. Chem. Phys. **139**, 244108:1 (2013), Erratum: *J. Chem. Phys.* **142**, 059901:1–2 (2015).

[320] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999).

[321] D. G. Fedorov and K. Kitaura, J. Phys. Chem. A **111**, 6904 (2007).

[322] H. P. Hratchian, P. V. Parandekar, K. Raghavachari, M. J. Frisch, and T. Vreven, J. Chem. Phys. **128**, 034107 (2008).

[323] T. Nagata, K. Brorsen, D. G. Fedorov, K. Kitaura, and M. S. Gordon, J. Chem. Phys. **134**, 124115:1 (2011).

[324] D. G. Fedorov, T. Ishida, M. Uebayasi, and K. Kitaura, J. Phys. Chem. A **111**, 2722 (2007).

[325] W. Li, W. Hua, T. Fang, and S. Li, The energy-based fragmentation approach for ab initio calculations of large systems, in *Computational Methods for Large Systems: Electronic Structure Approaches for Biotechnology and Nanotechnology*, edited by J. R. Reimers, pages 227–258, Wiley, Hoboken, NJ, 2011.

[326] C. J. F. Bottcher, O. V. Belle, P. Bordewijk, and A. Rip, *Theory of Electric Polarization*, volume 1 of *Dielectrics in Static Fields*, Elsevier, 1973.

[327] L. D. Jacobson, C. F. Williams, and J. M. Herbert, J. Chem. Phys. **130**, 124115:1 (2009).

[328] J. O. Hirschfelder, Chem. Phys. Lett. **1**, 325 (1967).

[329] J. O. Hirschfelder and W. J. Meath, Adv. Chem. Phys. **12**, 3 (1967).

[330] K. Wolinski and P. Pulay, J. Chem. Phys. **118**, 9497 (2003).

[331] D. J. Tozer and N. C. Handy, J. Chem. Phys. **109**, 10180 (1998).

[332] S. A. Maurer, M. Beer, D. S. Lambrecht, and C. Ochsenfeld, J. Chem. Phys. **139**, 184104:1 (2013).

[333] G. A. DiLabio, E. R. Johnson, and A. Otero-de-la-Roza, Phys. Chem. Chem. Phys. **15**, 12821 (2013).

[334] Y. Huang, M. Goldey, M. Head-Gordon, and G. J. O. Beran, J. Chem. Theory Comput. **10**, 2054 (2014).

[335] M. S. Gordon, Q. A. Smith, P. Xu, and L. V. Slipchenko, Annu. Rev. Phys. Chem. **64**, 553 (2013).

[336] B. Hartke, Phys. Chem. Chem. Phys. **5**, 275 (2003).

[337] A. Lagutschenkov, G. S. Fanourgakis, G. Niedner-Schatteburg, and S. S. Xantheas, J. Chem. Phys. **122**, 194310:1 (2005).

[338] S. Yoo and S. S. Xantheas, Structures, energetics, and spectroscopic fingerprints of water clusters n = 2-24, in *Handbook of Computational Chemistry*, edited by J. Leszczynski, chapter 21, pages 761–792, Springer Science + Business Media, 2012.

[339] J. K. Kazimirski and V. Buch, J. Phys. Chem. A **107**, 9762 (2003).

[340] D. Das et al., J. Chem. Phys. **117**, 10534 (2002).

[341] M. J. Deible, O. Tuguldur, and K. D. Jordan, J. Phys. Chem. B **118**, 8257 (2014).

[342] P. Kumar and M. Sathyamurthy, J. Phys. Chem. A **115**, 14276 (2011).

[343] K. R. Ramya and A. Venkatnathan, J. Phys. Chem. A **116**, 7742 (2012).

[344] Y. Liu, J. Zhao, F. Li, and Z. Chen, J. Comput. Chem. **34**, 121 (2013).

[345] N. Weill, E. Therrien, V. Campagna-Slater, and N. Moitessier, Curr. Pharm. Design **20**, 3338 (2014).

[346] V. Campagna-Slater, E. Therrien, N. Weill, and N. Moitessier, Curr. Pharm. Design **20**, 3360 (2014).

[347] O. A. von Lilienfeld and M. E. Tuckerman, J. Chem. Theory Comput. **3**, 1083 (2007).

[348] R. A. Bryce and I. H. Hillier, Curr. Pharm. Design **20**, 3293 (2014).

[349] M. Stiborova et al., Cancer Res. **64**, 8374 (2004).

[350] B. A, L. Shulenburger, N. A. Romero, J. Kim, and O. A. von Lilienfeld, J. Chem. Theory Comput. **10**, 3417 (2014).

[351] O. A. von Lilienfeld and A. Tkatchenko, J. Chem. Phys. **132**, 234109:1 (2010).

[352] A. J. Stone, *The Theory of Intermolecular Forces*, Oxford University Press, Oxford, UK, 2nd edition, 2013.

[353] J. G. Brandenburg, M. Hochheim, T. Bredow, and S. Grimme, J. Phys. Chem. Lett. **5**, 4275 (2014).

[354] J. G. Brandenburg, J. Gerit, and S. Grimme, Dispersion corrected Hartree-Fock and density functional theory for organic crystal structure prediction, in *Prediction and Calculation of Crystal Structures*, volume 345 of *Topic in Current Chemistry*, pages 1–23, Springer, 2014.

[355] A. D. Becke and E. R. Johnson, J. Chem. Phys. **122**, 154104:1 (2005).

[356] E. R. Johnson and A. D. Becke, J. Chem. Phys. **123**, 024101:1 (2005).

[357] A. D. Becke and E. R. Johnson, J. Chem. Phys. **123**, 154101:1 (2005).

[358] A. Tkatchenko and M. Scheffler, Phys. Rev. Lett. **102**, 073005:1 (2009).

[359] A. Tkatchenko, R. A. DiStasio Jr., R. Car, and M. Scheffler, Phys. Rev. Lett. **108**, 236402:1 (2012).

[360] R. A. DiStasio Jr., O. A. von Lilienfeld, and A. Tkatchenko, Proc. Natl. Acad. Sci. USA **109**, 14791 (2012).

[361] A. Tkatchenko, A. Ambrosetti, and R. A. DiStasio Jr., J. Chem. Phys. **138**, 074106:1 (2013).

[362] A. Ambrosetti, A. M. Reilly, R. A. DiStasio Jr., and A. Tkatchenko, J. Chem. Phys. **140**, 18A508:1 (2014).

[363] M. O. Sinnokrot and C. D. Sherrill, J. Phys. Chem. A **108**, 10200 (2004).

[364] T. Korona and B. Jeziorski, J. Chem. Phys. **125**, 184109:1 (2006).

[365] W. Wang, Y. Zhang, and Y.-B. Wang, J. Chem. Phys. **140**, 094302:1 (2014).

[366] W. W. T. Sun, Y. Zhang, and Y.-B. Wang, Comp. Theor. Chem. **1046**, 64 (2014).

[367] R. Podeszwa and K. Szalewicz, J. Chem. Phys. **126**, 194101:1 (2007).

[368] Q. Wu and W. Yang, J. Chem. Phys. **116**, 515 (2002).

[369] R. Podeszwa and K. Szalewicz, J. Chem. Phys. **136**, 161102:1 (2012).

[370] J. F. Dobson, Int. J. Quantum Chem. **114**, 1157 (2014).

[371] A. Otero-de-la-Roza and E. R. Johnson, J. Chem. Phys. **137**, 054103:1 (2012).

[372] A. M. Reilly and A. Tkatchenko, J. Phys. Chem. Lett. **4**, 1028 (2013).

[373] A. Tkatchenko, D. Alfe, and K. S. Kim, J. Chem. Theory Comput. **8**, 4317 (2012).

[374] A. Heßelmann, PhD thesis, Universität Duisburg-Essen, 2003.

[375] M. Beer, *Effiziente 'ab-initio' Berechnung molekularer Eigenschaften großer Systeme*, PhD thesis, University of Munich (LMU), 2011.

[376] M. Hapka, P. S. Zuchowski, M. M. Szczęśniak, and G. Chałasiǹski, J. Chem. Phys. **137**, 164104:1 (2012).

[377] C.-W. Tsai, Y.-C. Su, G.-D. Li, and J.-D. Chai, Phys. Chem. Chem. Phys. **15**, 8352 (2013).

[378] M. Hapka, L. Rajchel, M. Modrzejewski, G. Chałasiǹski, and M. M. Szczęśniak, J. Chem. Phys. **141**, 134120:1 (2014).

[379] F. Uhlig, J. M. Herbert, M. P. Coons, and P. Jungwirth, J. Phys. Chem. A **118**, 7507 (2014).

[380] K. Garrett et al., J. Chem. Theory Comput. **10**, 3821 (2014).

[381] C. Riplinger and F. Neese, J. Chem. Phys. **138**, 034106:1 (2013).

[382] C. Riplinger, B. Sandhoefer, A. Hansen, and F. Neese, J. Chem. Phys. **139**, 134101:1 (2013).

[383] M. Sparta and F. Neese, Chem. Soc. Rev. **43**, 5032 (2014).

[384] D. G. Liakos, M. Sparta, M. K. Kesharwani, J. M. L. Martin, and F. Neese, J. Chem. Theory Comput. **11**, 1525 (2015).

[385] J. Calbo, E. Ortí, J. C. Sancho-García, and J. Aragó, J. Chem. Theory Comput. **11**, 932 (2015).

[386] A. Klamt, WIREs Comput. Mol. Sci. **1**, 699 (2011).

[387] R. Sure and S. Grimme, J. Comput. Chem. **34**, 1672 (2013).

[388] T. Bucko, S. Lebegue, J. G. Angyan, and J. Hafner, J. Chem. Phys. **141**, 034114:1 (2014).

[389] A. Li, H. S. Muddana, and M. K. Gilson, J. Chem. Theory Comput. **10**, 1563 (2014).

[390] A. Otero-de-la-Roza and E. R. Johnson, J. Chem. Theory Comput. **11**, 4033 (2015).

[391] P. Bultinck, C. Van Alsenoy, P. W. Ayers, and R. Carbó-Dorca, J. Chem. Phys. **126**, 144111:1 (2007).

[392] Y. Zhao and D. G. Truhlar, J. Phys. Chem. A **109**, 5656 (2005).

[393] T. Anacker and J. Friedrich, J. Comput. Chem. **35**, 634 (2014).

[394] A. Mladek et al., Phys. Chem. Chem. Phys. **15**, 7295 (2013).

[395] M. Steinmetz and S. Grimme, ChemistryOpen **2**, 115 (2013).

[396] J. Antony, R. Sure, and S. Grimme, Chem. Commun. **51**, 1764 (2015).

[397] S. Grimme and M. Steinmetz, Phys. Chem. Chem. Phys. **15**, 16031 (2013).

[398] W. Hujo and S. Grimme, J. Chem. Theory Comput. **7**, 3866 (2011).

[399] W. Hujo and S. Grimme, J. Chem. Theory Comput. **9**, 308 (2013).

[400] M. Steinmetz, A. Hansen, S. Ehrlich, T. Risthaus, and S. Grimme, Top. Curr. Chem. **365**, 1 (2015).

[401] Y. Zhao and D. G. Truhlar, Phys. Chem. Chem. Phys. **10**, 2813 (2008).

[402] C. Muck-Lichtenfeld, S. Grimme, L. Kobryn, and A. Sygula, Phys. Chem. Chem. Phys. **12**, 7091 (2010).

[403] G.-X. Zhang, A. Tkatchenko, J. Paier, H. Appel, and M. Scheffler, Phys. Rev. Lett. **107**, 245501:1 (2011).

[404] W. Iali, P. Petrovic, M. Pfeffer, S. Grimme, and J.-P. Djukic, Dalton Trans. **41**, 12233 (2012).

[405] S. Ehrlich, J. Moellmann, and S. Grimme, Acc. Chem. Res. **46**, 916 (2013).

[406] R. Sure, J. Antony, and S. Grimme, J. Phys. Chem. B **118**, 3431 (2014).

[407] J. O. B. Tempkin, H. R. Leverentz, B. Wang, and D. G. Truhlar, J. Phys. Chem. Lett. **2**, 2141 (2011).

[408] A. J. Cohen, P. Mori-Sanchez, and W. Yang, Chem. Rev. **112**, 289 (2012).

[409] A. Pribram-Jones, D. A. Gross, and K. Burke, Annu. Rev. Phys. Chem. **66**, 283 (2015).

[410] A. M. Reilly and A. Tkatchenko, Chem. Sci. **6**, 3289 (2015).

[411] X. Chu and A. Dalgarno, J. Chem. Phys. **121**, 4083 (2004).

[412] A. D. Becke and E. R. Johnson, J. Chem. Phys. **127**, 154108:1 (2007).

[413] T. Sato and H. Nakai, J. Chem. Phys. **131**, 224104:1 (2009).

[414] M. A. Neumann and M.-A. Perrin, J. Phys. Chem. B **109**, 15531 (2005).

[415] P. Jurečka, J. Černý, P. Hobza, and D. R. Salahub, J. Comput. Chem. **28**, 555 (2007).

[416] L. C. Smeeton, J. D. Farrell, M. T. Oakley, D. J. Wales, and R. L. Johnston, J. Chem. Theory Comput. **11**, 2377 (2015).

[417] A. T. Blades and P. Kebarle, J. Am. Chem. Soc. **116**, 10761 (1994).

[418] X.-B. Wang, J. B. Nicholas, and L.-S. Wang, J. Chem. Phys. **113**, 10837 (2000).

[419] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, Phys. Rev. Lett. **91**, 146401:1 (2003).

[420] D. G. Liakos and F. Neese, J. Chem. Theory Comput. **11**, 4054 (2015).

[421] R. Peverati and D. G. Truhlar, J. Phys. Chem. Lett. **2**, 2810 (2011).

[422] L. Goerigk and S. Grimme, J. Chem. Theory Comput. **7**, 291 (2011).

[423] D. Rappoport and F. Furche, J. Chem. Phys. **1333**, 134105:1 (2010).

[424] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs, Chem. Phys. Lett. **240**, 283 (1995).

[425] F. Weigend, Phys. Chem. Chem. Phys. **8**, 1057 (2006).

[426] F. Weigend, M. Kattannek, and R. Ahlrichs, J. Chem. Phys. **130**, 164106:1 (2009).

[427] C. Sutton, M. S. Marshall, C. D. Sherrill, C. Risko, and J.-L. Brédas, J. Am. Chem. Soc. **137**, 8775 (2015).

[428] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Imprey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).

[429] C. A. Koh, Chem. Soc. Rev. **31**, 157 (2002).

[430] V. V. Struzhkin, B. Militzer, W. L. Mao, H. k. Mao, and R. J. Hemley, Chem. Rev. **107**, 4133 (2007).

[431] D.-Y. Kim, Y. Park, and H. Lee, Catalysis Today **120**, 257 (2007).

[432] V. Buch et al., Phys. Chem. Chem. Phys. **11**, 10245 (2009).

[433] Y. Zhang and W. Yang, Phys. Rev. Lett. **80**, 890 (1998).

[434] Y. Zhao and D. G. Truhlar, Chem. Phys. Lett. **502**, 1 (2011).

[435] A. Khan, J. Mol. Struct. (Theochem) **664-665**, 237 (2003).

[436] S. Patchkovskii and J. S. Tse, Proc. Natl. Acad. Sci. USA **100**, 14645 (2003).

[437] S. Grimme, J. Antony, T. Schwabe, and C. Muck-Lichtenfeld, Org. Biomol. Chem. **5**, 741 (2007).

[438] S. Alavi and J. A. Ripmeester, Angew. Chem. Int. Ed. Engl. **119**, 6214 (2007).

[439] C.-Y. Geng, H. Wen, and H. Zhou, J. Phys. Chem. A **113**, 5463 (2009).

[440] O. Loboda and V. Goncharuk, Chem. Phys. Lett. **484**, 144 (2010).

[441] G. Roman-Perez, M. Moaied, J. M. Soler, and F. Yndurain, Phys. Rev. Lett. **105**, 145901:1 (2010).

[442] Q. Li et al., Phys. Rev. B **84**, 153103:1 (2011).

[443] P. K. Chattaraj, S. Bandaru, and S. Mondal, J. Phys. Chem. A **115**, 187 (2011).

[444] H. K. Srivastava and G. N. Sastry, J. Phys. Chem. A **115**, 7633 (2011).

[445] G. Bravo-Perez and H. Saint-Martin, Int. J. Quantum Chem. **112**, 3655 (2012).

[446] S. S. Park, S. Lee, Y. S. Won, and Y. J. Ahn, Chem. Phys. **441**, 128 (2014).

[447] A. Khan, J. Chem. Phys. **110**, 11884 (1999).

[448] A. Khan, J. Phys. Chem. A **105**, 7429 (2001).

[449] A. Khan, J. Chem. Phys. **116**, 6628 (2002).

[450] F. Lebsira, A. Bouyacouba, D. Bormannb, and A. Krallafa, J. Mol. Struct. (Theochem) **864**, 42 (2008).

[451] D. Bakowies, J. Chem. Phys. **127**, 164109:1 (2007).

[452] X. Cao, Y. Su, and J. Zhao, J. Phys. Chem. A **119**, 7063 (2015).

[453] X. Cao, Y. Su, Y. Liu, J. Zhao, and C. Liu, J. Phys. Chem. A **118**, 215 (2014).

[454] Y. Liu and L. Ojamae, J. Phys. Chem. A **118**, 11641 (2014).

[455] K. R. Ramya, R. Kumar, and A. Venkatnathan, Comput. Theor. Chem. **1039**, 28 (2014).

[456] P. Kumar, B. K. Mishra, and N. Sathyamurthy, Comput. Theor. Chem. **1029**, 26 (2014).

[457] L. Florusse et al., Science **306**, 469 (2004).

[458] H.-J. Werner, T. B. Adler, and F. R. Manby, J. Chem. Phys. **126**, 164102:1 (2007).

[459] K. A. Peterson, T. B. Adler, and H.-J. Werner, J. Chem. Phys. **128**, 084102 (2008).

[460] S. P. de Visser and S. Shaik, J. Am. Chem. Soc. **125**, 7413 (2003).

[461] B.-Q. Wang et al., Chem. Phys. Lett. **375**, 91 (2003).

[462] S. Hammerum, J. Am. Chem. Soc. **131**, 8627 (2009).

[463] Q.-Z. Li and H.-B. Li, Hydrogen bonds involving radical species, in *Noncovalent Forces*, edited by S. Scheiner, volume 19 of *Challenges and Advances in Computational Chemistry and Physics*, pages 107–127, Springer, 2015.

[464] C. E. Boozer and G. S. Hammond, J. Am. Chem. Soc. **76**, 3861 (1954).

[465] G. A. Russell, J. Am. Chem. Soc. **80**, 4987 (1958).

[466] R. R. Knowles and E. N. Jacobsen, Proc. Natl. Acad. Sci. USA **107**, 20678 (2010).

[467] T. Lu and S. E. Wheeler, Science **347**, 719 (2015).

[468] L. Belpassi et al., J. Am. Chem. Soc. **132**, 13046 (2010).

[469] A. Bartocci, D. Cappelletti, F. Pirani, F. Tarantelli, and L. Belpassi, J. Phys. Chem. A **118**, 6440 (2014).

[470] R. Z. Khaliullin, A. T. Bell, and M. Head-Gordon, Chem. Eur. J. **15**, 851 (2009).

[471] K. E. Riley and P. Hobza, WIREs Comput. Mol. Sci. **1**, 3 (2011).

[472] L. A. Burns, M. S. Marshall, and C. D. Sherrill, J. Chem. Phys. **141**, 234111:1 (2014).

[473] P. S. Bagus, K. Hermann, and C. W. Bauschlicher Jr., J. Chem. Phys. **80**, 4378 (1984).

[474] W. J. Stevens and W. H. Fink, Chem. Phys. Lett. **139**, 15 (1987).

[475] W. Chen and M. S. Gordon, J. Phys. Chem. **100**, 14316 (1996).

[476] F. M. Bickelhaupt and E. J. Baerends, Kohn-Sham density functional theory: Predicting and understanding chemistry, in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and D. B. Boyd, volume 15, pages 1–86, Wiley-VCH, 2000.

[477] E. D. Glendening, J. Phys. Chem. A **109**, 11936 (2005).

[478] R. Z. Khaliullin, E. A. Cobar, R. C. Lochan, A. T. Bell, and M. Head-Gordon, J. Phys. Chem. A **111**, 8753 (2007).

[479] P. Reinhardt, J.-P. Piquemal, and A. Savin, J. Chem. Theory Comput. **4**, 2020 (2008).

[480] P. Su and H. Li, J. Chem. Phys. **131**, 014102:1 (2009).

[481] M. P. Mitoraj, A. Michalak, and T. Ziegler, J. Chem. Theory Comput. **5**, 962 (2009).

[482] M. Mandado and J. M. Hermida-Ramón, J. Chem. Theory Comput. **7**, 633 (2011).

[483] Y. Mo, P. Bao, and J. Gao, Phys. Chem. Chem. Phys. **13**, 6760 (2011).

[484] K. Yamada and N. Koga, Theor. Chem. Acc. **131**, 1178:1 (2012).

[485] M. von Hopffgarden and G. Frenking, WIREs Comput. Mol. Sci. **2**, 43 (2012).

[486] M. J. S. Phipps, T. Fox, C. S. Tautermann, and C.-K. Skylaris, Chem. Soc. Rev. **44**, 3177 (2015).

[487] A. J. Stone and A. J. Misquitta, Chem. Phys. Lett. **473**, 201 (2009).

[488] R. J. Azar, P. R. Horn, E. J. Sundstrom, and M. Head-Gordon, J. Chem. Phys. **138**, 084102:1 (2013).

[489] B. Kaduk, T. Kowalczyk, and T. Van Voorhis, Chem. Rev. **112**, 321 (2012).

[490] J. Řezáč and A. de la Lande, J. Chem. Theory Comput. **11**, 528 (2015).

[491] R. Z. Khaliullin, M. Head-Gordon, and A. T. Bell, J. Chem. Phys. **124**, 204105:1 (2006).

[492] A. J. Stone, Chem. Phys. Lett. **211**, 101 (1993).

[493] Q. Wu and T. Van Voorhis, Phys. Rev. A **72**, 024502:1 (2005).

[494] Q. Wu and T. Van Voorhis, J. Chem. Theory Comput. **2**, 765 (2006).

[495] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).

[496] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).

[497] Q. Wu, C.-L. Cheng, and T. Van Voorhis, J. Chem. Phys. **127**, 164119:1 (2007).

[498] H. Stoll, G. Wagenblast, and H. Preuss, Theor. Chem. Acc. **57**, 169 (1980).

[499] E. Ronca, L. Belpassi, and F. Tarantelli, ChemPhysChem **15**, 2682 (2014).

[500] S. Hammerum and C. B. Nielsen, J. Phys. Chem. A **109**, 12046 (2005).

[501] A. J. Cohen, P. Mori-Sanchez, and W. Yang, Science **321**, 792 (2008).

[502] S. N. Steinmann, C. Piemontesi, A. Delachat, and C. Corminboeuf, J. Chem. Theory Comput. **8**, 1629 (2012).

[503] E. R. Johnson, M. Salamone, M. Bietti, and G. A. DiLabio, J. Phys. Chem. A **117**, 947 (2013).

[504] R. N. Barnett et al., J. Am. Chem. Soc. **128**, 10795 (2006).

[505] J. Cadet, T. Douki, and J.-L. Ravanat, Acc. Chem. Res. **41**, 1075 (2008).

[506] K. Mizuse, H. Hasegawa, N. Mikami, and A. Fujii, J. Phys. Chem. A **114**, 11060 (2010).

[507] B. Badger and B. Brocklehurst, Nature **219**, 263 (1968).

[508] H. Kobayashi et al., J. Am. Chem. Soc. **118**, 368 (1996).

[509] J. K. Kochi, R. Rathore, and P. L. Magueres, J. Org. Chem **65**, 6826 (2000).

[510] M. Meot-Ner, J. Phys. Chem. **84**, 2724 (1980).

[511] E. G. Hohenstein, *Implementation and applications of density-fitted symmetry-adapted perturbation theory*, PhD thesis, Georgia Institute of Technology, 2011.

[512] H. M. Jaeger, H. F. Schaefer, E. G. Hohenstein, and C. D. Sherrill, Comput. Theor. Chem. **973**, 47 (2011).

[513] W. Hua, T. Fang, W. Li, J.-G. Yu, and S. Li, J. Phys. Chem. A **112**, 10864 (2008).

[514] A. P. Rahalkar, V. Ganesh, and S. R. Gadre, J. Chem. Phys. **129**, 234101:1 (2008).

[515] Z. Yang, S. Hua, W. Hua, and S. Li, J. Phys. Chem. A **114**, 9253 (2010).

[516] S. R. Gadre, K. V. J. Jose, and A. P. Rahalkar, J. Chem. Sci. **122**, 47 (2010).

[517] S. R. Pruitt, M. A. Addicoat, M. A. Collins, and M. S. Gordon, Phys. Chem. Chem. Phys. **14**, 7752 (2012).

[518] A. Saha and K. Raghavachari, J. Chem. Theory Comput. **10**, 58 (2014).

[519] D. G. Fedorov and K. Kitaura, J. Chem. Phys. **120**, 6832 (2004).

[520] D. G. Fedorov and K. Kitaura, J. Chem. Phys. **121**, 2483 (2004).

[521] S. Hirata et al., Mol. Phys. **130**, 2255 (2005).

[522] T. Nakano et al., Chem. Phys. Lett. **351**, 475 (2002).

[523] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, and J. Olsen, Chem. Phys. Lett. **302**, 437 (1999).

[524] B. J. Mintz and J. M. Parks, J. Phys. Chem. A **116**, 1086 (2012).

[525] Y. Zhao and D. G. Truhlar, J. Chem. Theory Comput. **1**, 415 (2005).

[526] F. O. Kannemann and A. D. Becke, J. Chem. Theory Comput. **6**, 1081 (2010).