CHAPTER 2

Reliable Electronic Structure Computations for Weak Noncovalent Interactions in Clusters

Gregory S. Tschumper

Department of Chemistry and Biochemistry, University of Mississippi, University, Mississippi

INTRODUCTION AND SCOPE

If you are reading this chapter, you are most likely already aware of the importance of weak attractive interactions between molecules (and atoms), such has hydrogen bonding and London dispersion forces, in chemistry and related fields. The relatively weak interactions between uncharged molecules (and/or atoms) are also called nonbonded interactions and sometimes collectively referred to as van der Waals forces. These intermolecular forces are not only prevalent throughout chemistry, but they often provide the governing influence in a wide variety of chemical, physical, and biological processes.[1–7] Some general examples include, but are certainly not limited to, solvation, condensation, crystallization, asymmetric catalysis, bulk-phase properties, directed self-assembly of nanomaterials, chromatographic separation, micelle formation, molecular recognition, drug transport, as well as the structure and function of biomolecules. The initial step in HIV infection, for instance, involves the formation of a weakly bound (noncovalent) complex of the viral envelope and cellular receptor glycoproteins, HIV-gp120 and CD4, respectively.[8,9] The delivery and transport of pharmaceuticals in mammals frequently occurs through subcovalent complexation with blood-soluble proteins such as human serum

Reviews in Computational Chemistry, Volume 26
edited by Kenny B. Lipkowitz and Thomas R. Cundari

albumins.[10] The formation of weakly bound heterogeneous clusters plays a key role in the chemistry of the atmosphere on Earth and elsewhere.[11,12] Noncovalent interactions dictate not only the structure and function of biomolecules, from simple dipeptides to enzymes and DNA,[13–16] but also molecular recognition events.[17] In the closely related and rapidly growing field of nanotechnology, highly selective, directional supramolecular self-assembly can be achieved with the aid of intermolecular hydrogen bonding and π-type interactions.[18,19] Hydrogen bonding also affects the chemical shielding, and therefore the electronic properties, of metal atoms in metalloproteins.[20,21] These weak inter- and intramolecular forces are even used to control diastereoselectivity and mediate catalysis in important classes of organic reactions.[22,23] The very existence of the condensed phase (i.e., solids and liquids) is dependent on the noncovalent interactions between molecules (or atoms), as are phase transitions, liquid structure, diffusion, crystal structure, and solvation/solutions.[24,25] These ubiquitous interactions have even led to the development and refinement of many cardinal chemical concepts such as hydrophilicity and hydrophobicity as well as the very definition of the chemical bond.

Over the past decade, there have been numerous books[26–31] and articles[32–44] reviewing ab initio and density functional theory (DFT) computations of hydrogen bonding and other weak noncovalent interactions. In fact, the very first chapter of this entire review series examines basis sets for noncovalent interactions between atoms and/or molecules,[45] while a chapter in the second volume reviews ab initio methods for hydrogen bonding.[46] Three thematic issues of *Chemical Reviews* have been dedicated to van der Waals interactions (Vol. 88, No. 6, 1988; Vol. 94, No. 7, 1994; and Vol. 100, No. 11, 2000). Two articles in the centennial issue of the *Journal of Physical Chemistry* discuss weakly bound clusters and solvation.[47,48] It is also worth noting that π-type stacking interactions are very topical at the moment and are the subject not only of a separate chapter in this volume of *Reviews in Computational Chemistry*[49] but also of a special issue of *Physical Chemistry Chemical Physics* (Vol. 10, No. 19, 2008).

This chapter is intended to serve two very distinct purposes. Readers new to the subject matter will find a fairly thorough introduction to reliable electronic structure computations for weakly bound clusters (including a step-by-step tutorial). For more experienced readers, this chapter also reviews many of the significant advances made in the field since the turn of the twenty-first century, particularly current state-of-the-art benchmark studies. This work also offers some valuable perspective and will attempt to illustrate the importance of balancing what is possible with what is practical.

Clusters and Weak Noncovalent Interactions

Defining the scope of a chapter for *Reviews in Computational Chemistry* on clusters of molecules (and/or atoms) held together by hydrogen bonding,

London dispersion forces, and/or similar interactions is not a simple task. Chemical bonding, whether noncovalent, covalent, ionic, or metallic, covers a broad, continuous spectrum of electronic interactions and energies. Consequently, the classification of a bond or interaction (e.g., double versus triple[50] or covalent versus noncovalent[51]) is sometimes open to interpretation. As a result, there is no unique criterion or set of criteria that can be used to define *weak* interactions or *noncovalent* interactions. In the second volume of this review series, Scheiner already notes this issue and highlighted the difficulties associated with defining the hydrogen bond.[46] Here, matters are even more complicated because other weak interactions are also considered.

To limit the breadth of the present chapter, it focuses on the most common types of weakly bound clusters, namely those composed of neutral fragments. (The following discussion also assumes the weakly bound clusters are composed of closed-shell fragments that are in their ground electronic states and dominated by a single Hartree–Fock (HF) reference function. It is certainly feasible to perform reliable computations on systems composed of open-shell, excited state, or multireference fragments; however, by assuming the monomers have a "well-behaved" electronic structure, we can focus on computational methods that will accurately describe the weak noncovalent interactions within a cluster.) Clusters containing one or two charged species are mentioned (e.g., solvated ions or ion pairs). However, clusters with numerous charged species (e.g., room temperature ionic liquids[52]) fall outside the scope of this review. This emphasis still leaves a wide spectrum of weak chemical interactions that bind the clusters together (as depicted in Figure 1).

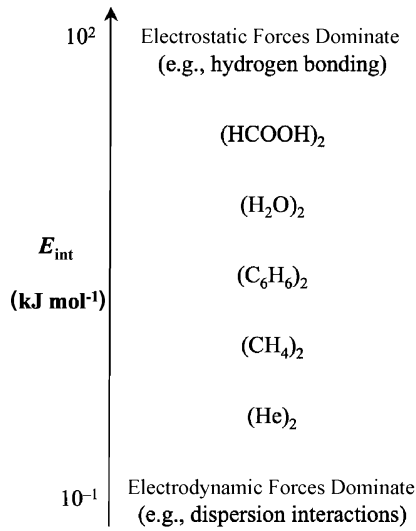

Figure 1 Weak noncovalent interactions between neutral fragments cover a wide spectrum of interactions and energies.

At one extreme (bottom of the figure), one finds complexes held together primarily by dispersion forces (rare gas dimers, He nanodroplets, etc.). On the opposite extreme (top of the figure) are clusters dominated by electrostatic interactions such as hydrogen bonding (formic acid dimer, water clusters, etc.). Of course, most interfragment interactions fall somewhere between these two extremes. [In this work, the term *interfragment* (or *intermonomer*) interaction is used because it is more general than and implicitly includes both *interatomic* and *intermolecular* interactions. Note that some researchers object to the latter adjective when describing weakly bound clusters because it is technically incorrect. For example, if $(HF)_3$ is considered an independent molecular species then, by definition, there can be only intramolecular interactions.] A more detailed analysis of this continuum of weak noncovalent interactions is presented below.

Given the current flurry of activity in the area of π-type interactions and halogen bonding (a specific case of sigma-hole interactions), special attention will be paid to these two types of weak interactions. In fact, an entire chapter in this volume of *Reviews in Computational Chemistry* is dedicated to noncovalent π interactions.[49] It should be noted that, although most examples are for relatively small (dimers, trimers, tetramers, and pentamers) homogeneous clusters, the principles discussed here can readily be extended to larger, heterogeneous systems.

Computational Methods

Although a wide variety of theoretical methods is available to study weak noncovalent interactions such as hydrogen bonding or dispersion forces between molecules (and/or atoms), this chapter focuses on size consistent electronic structure techniques likely to be employed by researchers new to the field of computational chemistry. Not surprisingly, the list of popular electronic structure techniques includes the self-consistent field (SCF) Hartree–Fock method as well as popular implementations of density functional theory (DFT). However, correlated wave function theory (WFT) methods are often required to obtain accurate structures and energetics for weakly bound clusters, and the most useful of these WFT techniques tend to be based on many-body perturbation theory (MBPT) (specifically, Møller–Plesset perturbation theory), quadratic configuration interaction (QCI) theory, and coupled-cluster (CC) theory.

This review concentrates on the fundamentals of supermolecule model chemistries for clusters of atoms/molecules held together by weak chemical forces. The principles behind the appropriate selection of theoretical method and basis set for a particular class of weak noncovalent interactions provide the foundation for understanding more complex computational schemes that might require the user to specify more than just a method and/or basis set, such as highly efficient fragmentation schemes [e.g., the effective fragment potential (EFP) method,[53,54] the fragment molecular orbital (FMO) method,[55,56] the

n-body decomposition (NBD) scheme,[57] and the multicentered integrated method (MC QM:QM or MC ONIOM) for clusters.[58–60]]

Some readers may have noticed that methods based on intermolecular perturbation theory such as symmetry-adapted perturbation theory (SAPT),[61–64] have not been mentioned. These methods are not discussed in this chapter because most versions of SAPT are actually 2-body methods. (See below for a description of 2-body, 3-body, and many-body interactions in clusters.) SAPT is not inherently limited to 2-body interactions; a 3-body implementation exists.[65–67] However, the author is not aware of higher order SAPT programs, let alone a general *n*-body program for clusters composed of *n* fragments. While SAPT can certainly be used to study trimers, tetramers, and larger clusters, such applications require a great deal of a priori knowledge about the nonadditivity (or cooperativity) in the system and are certainly not for novices.

Molecular mechanics methods are also omitted from the present discussion for similar reasons. Although very sophisticated force fields are available for water (including polarizable models), most force fields for weakly bound clusters are essentially 2-body (dimer) potentials that have been adjusted empirically to reproduce bulk-phase properties.[68–71] This procedure leads to very reliable descriptions of liquid water, but diminishes the quality of results for small clusters. Although force fields that include 3-body interactions are beginning to appear,[69,70] the effects of higher order interactions (4-body, 5-body, etc.) are still untested. Furthermore, the composition of a weakly bound cluster, not just its size, is a major concern with molecular mechanics force fields. The highly refined potentials that have been developed for water[68,71] are not necessarily transferable to other weak noncovalent systems (methanol, acetone, etc.).

WEAK NONCOVALENT INTERACTIONS

This section presents an overview of the nature of weak noncovalent interactions between molecules (and atoms). Readers interested in more detail are directed to classic references such as the 1954 text by Hirschfelder, Curtiss and Bird,[72] the 1971 book by Margenau and Kestener,[73] the 1996 monograph by Stone,[4] as well as some more recent sources.[7,27,31,74]

Historical Perspective

Theoretical treatments of attractive forces between molecules (and/or atoms) in the gas phase can be traced as far back as 1873 to the efforts by van der Waals to describe the deviation of real gases from ideal behavior at relatively high densities.[75] By the early 1930s, theoretical explanations of the origins of van der Waals' attractive forces began to emerge from the likes of Keesom,[76,77] Debye,[78,79] Falckenhagen,[80] and London.[81,82] This body of

work established that there are four rigorously defined fundamental components that contribute to interactions between a pair of uncharged molecules or atoms: electrostatic, induction (sometimes referred to as polarization), dispersion, and exchange-repulsion[83] (or simply exchange). The first two contributions to the interaction energy were readily explained in terms of classical electromagnetic theory. Interactions involving two permanent electrostatic multipole moments (dipole, quadrupole, etc.) are relatively easy to understand for anyone who has ever played with a pair of magnets; opposite poles (+/−) attract each other, and like poles (+/+ and −/−) repel each other. Similarly, adhering a balloon to a wall with static electricity provides a macroscopic analog for induction. However, quantum mechanics was required to rationalize the dispersion and exchange energies. The latter is a simple consequence of the Pauli exclusion principle, but an explanation of the dispersion energy is more involved.

London was the first to describe the dispersion interaction.[81,82] Through a quantum mechanical perturbation theory treatment of the interaction energy, he demonstrated that, at second, order attractive terms can arise due to the simultaneous electron correlation between two fragments even if they possess no permanent electrostatic moment (e.g., a pair of rare gas atoms). London dubbed the attraction *dispersion forces* because similar oscillator strengths appear in equations describing the dispersion of electromagnetic radiation (light). The attractive forces of these interactions are typically attributed to fluctuations (thermal or quantum mechanical) in the electron density that give rise to an instantaneous dipole in one fragment that induces a dipole in a neighbor. This semiclassical model was introduced after London's initial work, and its physical significance is not manifest since there are no expressions in the quantum mechanical derivation that can be interpreted as interactions between instantaneous dipoles. At the very least, this fluctuating charge or electrodynamic model provides a useful mnemonic.

As discussed in Paresegian's recent book,[7] the modern view of dispersion interactions has its roots in the the Casimir effect.[84] Rather than charge fluctuations, the phenomenon can be viewed in terms of zero-point electromagnetic-field fluctuations in the vacuum as allowed by the Heisenberg uncertainty principle ($\Delta E\,\Delta t \geq \hbar/2\pi$). Atoms and molecules can absorb some of these frequencies, namely those frequencies that are resonant with transitions between the quantum mechanical energy levels of the system as determined by its electronic structure. This absorption of the electromagnetic fluctuations gives rise to attractive forces between two bodies.

> We now recognize that "empty space" is a turmoil of electromagnetic waves of all frequencies and wavelengths. They wash through and past us in ways familiar from watching the two-dimensional version, a buoy or boat bobbing in rough water. We can turn the dancing charges idea around. From the vacuum point of view, imagine two bodies, such as two boats in rough water or a single boat near a dock, pushed

> by waves from all directions except their wave quelling neighbor. The net result is that the bodies are pushed together. You get close to a dock, you can stop rowing. The waves push you in. We can think of electromagnetic modes between the two bodies as the fluctuations that remain as tiny deviations from the outer turmoil. The extent of quelling is, obviously, in proportion to the material-absorption spectra. So we can think of absorption spectra in two ways: those at which the charges naturally dance; those at which charge polarization quells the vacuum fluctuations and stills the space between the [fragments].[7]

It will become evident in later sections that the nature of the weak noncovalent interactions in a cluster dictate which computational methods will produce accurate results. In particular, it is far more difficult to compute reliable properties for weakly bound clusters in which dispersion is the dominant attractive component of the interaction. For example, Hartree–Fock supermolecule computations are able to provide qualitatively correct data for hydrogen-bonded systems like $(H_2O)_2$[85] even with very small basis sets, but this approach does not even bind Ne_2.[86] What is the origin of this inconsistency? Dispersion is the dominant attractive force in rare gas clusters while the electrostatic component tends to be the most important attractive contribution near the equilibrium structure $(H_2O)_2$. As London's work demonstrated,[81,82,87,88] dispersion interactions are inherently an electron correlation problem and, consequently, cannot be described by Hartree–Fock computations.[89] To this day, dispersion interactions continue to pose a significant challenge in the field of computational chemistry, particularly those involving systems of delocalized π electrons.[49]

Some Notes about Terminology

Because the van der Waals equation of state preceeded "The General Theory of Molecular Forces,"[82] the interactions between molecules and/or atoms became known collectively as *van der Waals* forces. From a historical perspective, *van der Waals interactions* encompass the entire spectrum of weak interactions depicted in Figure 1, from the dispersion forces holding a He nanodroplet together to the hydrogen bonds in a cluster of water molecules. Although many researchers today associate van der Waals forces only with weak dispersion interactions, this review adopts the historical definition of van der Waals interactions and uses the term to collectively refer to all weak chemical interactions between uncharged molecules (and/or atoms).

Additional linguistic dilemmas are encountered in this area of research. For example, these weak chemical forces are sometimes referred to as *nonbonding interactions* despite meeting Pauling's utilitarian definition of a chemical bond introduced in 1939 (which is still one of the most useful and most widely used):[90]

> There is a chemical bond between two atoms or groups of atoms in the case that the forces acting between them are such as to lead to the formation of an

aggregate with sufficient stability to make it convenient for the chemist to consider it as an independent molecular species.

The term *noncovalent interaction* does not completely resolve the matter since ionic interactions (e.g., salt bridges) are frequently included in this category, particularly in the biochemistry community.[15] In this work, the moniker *weak noncovalent interaction* is used to denote the continuum of weak chemical forces between electrically uncharged molecules (and/or atoms).

FUNDAMENTAL CONCEPTS: A TUTORIAL

Model Systems and Theoretical Methods

Because of their relatively small size and high symmetry, the cyclic hydrogen fluoride clusters, $(HF)_n$ where $n = 3 - 5$, are very useful prototypes for studying hydrogen bonding. In this section, these model systems will be used to illustrate several aspects of computations on weakly bound clusters. These planar hydrogen-bonded complexes have C_{nh} symmetry and are shown in Figure 2. Their structures can be specified completely by three internal coordinates: R(HF), which is the length of the HF covalent bond; R(FF), which is the distance between neighboring F atoms; and θ(HFF), which is the small angle the H atoms make out of the ring formed by the F atoms.

The RHF method and aug-cc-pVDZ basis set have been adopted in this tutorial for two practical reasons. All calculations can be run in a few minutes on a reasonably modern desktop or laptop with a few hundred megabytes of memory, and all results should be reproduced readily regardless of the electronic structure software package you happen to be using. In contrast, electronic energies from DFT calculations will differ because the various electronic structure programs often employ different numerical integration grids. It is important to note that this particular model chemistry (RHF method and aug-cc-pVDZ basis set) is not expected to give reliable results. Correlated WFT methods such as second-order Møller–Plesset perturbation theory

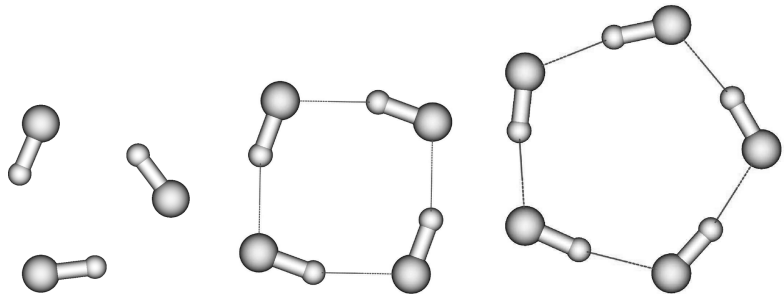

Figure 2 Cyclic hydrogen fluoride trimer, $(HF)_3$, tetramer, $(HF)_4$, and pentamer, $(HF)_5$, are planar structures with C_{3h}, C_{4h}, and C_{5h} symmetry, respectively.

(MP2) or coupled-cluster methods would certainly provide more reliable results, but they are not appropriate for a tutorial because the computations become rather time consuming. Furthermore, matching results from correlated WFT methods can be difficult for users not familiar with the frozen core or deleted virtual approximations because some software packages correlate all electrons by default while others exclude core electrons (i.e., adopt the frozen core approximation) by default. Computations in this work employed spherical harmonic (5*d*, 7*f*, etc.) rather than Cartesian (6*d*, 10*f*, etc.) functions, which gives 32 basis functions per HF monomer.

The geometrical parameters given in the top half of Table 1 for the HF clusters are from RHF/aug-cc-pVDZ optimizations and have been rounded off to three significant figures for bond lengths and two significant figures for bond angles. Although the values differ appreciably from the "best estimates" of Ref. 91, the bond lengths and angles are appropriate for the computational methods adopted for this tutorial. The electronic energies of these fixed structures (i.e., single-point energies) listed in Table 1 are from RHF computations with the aug-cc-pVDZ basis set. Step 1 in this tutorial is to reproduce the RHF/aug-cc-pVDZ electronic energies in Table 1. Sample input files for several popular software packages are available online.[92]

Rigid Monomer Approximation

Frequently, computational studies of weakly bound clusters employ the rigid monomer approximation (RMA). The RMA assumes that geometries of the monomers do not change as they coalesce to form the cluster. Because the interactions between the fragments of such clusters are, by their very definition, weak, the electronic structure, and hence the geometry, of the monomers does not change appreciably. This approach can simplify dramatically theoretical descriptions of the cluster because the intramolecular geometrical

Table 1 Geometrical Parameters and Electronic Energies of $(HF)_n$, $n = 1, 3-5$[a]

	Symmetry	R(FF)	R(HF)	θ(HFF)	*E*
HF	$C_{\infty v}$	n/a	0.900	n/a	−100.033816
		Fully Optimized Clusters			
$(HF)_3$	C_{3h}	2.71	0.910	25	−300.120482
$(HF)_4$	C_{4h}	2.64	0.915	13	−400.169746
$(HF)_5$	C_{5h}	2.61	0.916	6.6	−500.216512
		Rigid Monomer Approximation			
$(HF)_3$	C_{3h}	2.72	0.900	26	−300.120162
$(HF)_4$	C_{4h}	2.65	0.900	13	−400.168826
$(HF)_5$	C_{5h}	2.63	0.900	7.1	−500.215113

[a]Bond lengths (R) in Å, bond angles (θ) in degrees, and electronic energies (*E*) in E_h are from RHF/aug-cc-pVDZ calculations.

parameters are fixed. For example, by employing the RMA, geometry optimizations of weakly bound clusters need only consider the interfragment degrees of freedom. For a system as simple as $(H_2O)_2$, the RMA already reduces the full 12-dimensional problem to a more tractable 6-dimensional intermolecular potential energy hypersurface.

This approximation is typically valid for clusters held together by hydrogen bonds or van der Waals forces because the geometrical distortions tend to be modest and do not qualitatively change the structure of the monomers. As can be seen in bottom half of Table 1, fixing the intramolecular R(HF) distance at 0.900 Å for the HF trimer, tetramer, and pentamer has relatively little effect on the optimized interfragment parameters [R(FF) changes by no more than 0.02 Å and θ(HFF) by less than a degree]. This constraint also has relatively little effect on the electronic energies of $(HF)_3$, $(HF)_4$, and $(HF)_5$, which increase by only ≈ 1 mE_h on average.

These limited results demonstrate that the RMA can be accurate even for relatively strong hydrogen bonds, which can induce some of the largest geometrical distortions in weakly bound molecular clusters. The effect of the RMA on interaction energies will be discussed next. However, the RMA can break down if large qualitative geometrical changes occurs as the complex forms (e.g., conformational changes or isomerization).

Supermolecular Dissociation and Interaction Energies

Within the supermolecule approach, the dissociation energy (D_e) or interaction energy (E_{int}) of a cluster is obtained by calculating the energy difference between the cluster and the noninteracting fragments. This energy difference is depicted in Figure 3. Note that D_e and E_{int} are essentially the same quantity. The only significant difference is the sign ($D_e = -E_{\text{int}}$). A more subtle, technical distinction is that the term *dissociation energy* should be applied only to minima on the potential energy surface (PES) while *interaction energies* are more general and can describe any point on the surface.

When using a size-consistent method, the dissociation of homogeneous system such as $(HF)_n$ into n identical HF monomers [$(HF)_n \rightarrow n$HF] can be determined by computing the energy of the cluster and the energy of the monomer:

$$E_{\text{int}} = E[(HF)_n] - nE[HF] \qquad [1]$$

In the more general case of a heterogeneous cluster composed of N fragments ($f_1f_2f_3 \ldots f_N \rightarrow f_1 + f_2 + f_3 + \cdots + f_N$), up to $N + 1$ computations need to be performed to determine E_{int} or D_e:

$$\begin{aligned} E_{\text{int}} &= E[f_1f_2f_3 \ldots f_N] - E[f_1] - E[f_2] - E[f_3] - \cdots - E[f_N] \\ &= E[f_1f_2f_3 \ldots f_N] - \sum_{i=1}^{N} E[f_i] \end{aligned} \qquad [2]$$

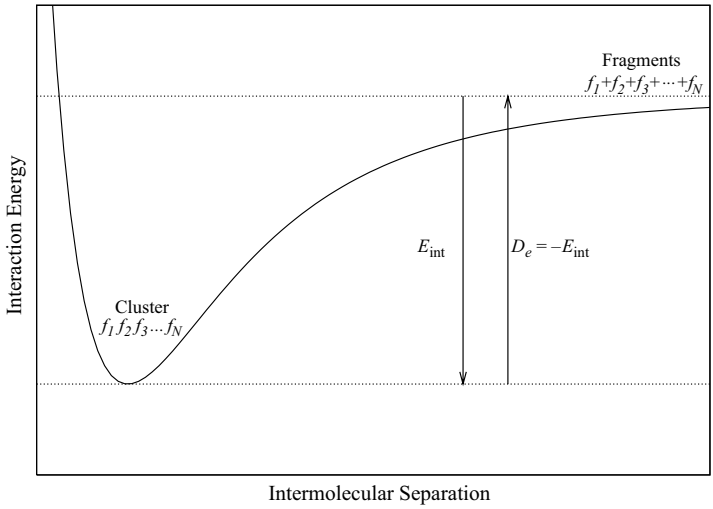

Figure 3 Interaction energies (E_{int}) and dissociation energies (D_e) are simply the energy difference between the cluster and the isolated fragments.

For example, to obtain E_{int} for the HF···H_2O···CH_3OH trimer requires four separate computations: E[HF], E[H_2O], E[CH_3OH], as well as the electronic energy of the cluster. The number of computations can be reduced if some of the fragments are identical (e.g., HF···H_2O···HF).

Interaction Energies: Rigid Monomers vs. Fully Optimized Clusters

The interaction energies reported toward the left side of Table 2 for the HF clusters were calculated with the electronic energies from Table 1 and Eq. [2] (which reduces to Eq. [1] for these homogeneous HF clusters). A simple conversion factor was used to convert E_{int} from E_{h} to kilojoule per mole ($1E_{\text{h}} \approx 2625.5\,\text{kJ mol}^{-1}$). Step 2 of this tutorial is to calculate the interaction energies on the left side of Table 2.

Because the RMA had relatively little effect on the electronic energies of the HF clusters (Table 1), it is not surprising that the approximation has only a modest effect on the interaction energies (Table 2). For all three clusters, the magnitude of E_{int} and $E_{\text{int}}^{\text{CP}}$ decreases slightly when the RMA is employed. Comparison of the values for E_{int} from the top of Table 2 to those at the bottom reveals that the change does not exceed 3% [−49.97 vs. −49.13 or 1.7% for $(HF)_3$, −90.53 vs. −88.12 or 2.7% for $(HF)_4$ and −124.53 vs. −120.86 or 2.9% for $(HF)_5$], which is tolerable for many applications. For systems with even smaller cluster-induced geometrical perturbations, such as π-type van der Waals interactions, the RMA has almost no discernable effect on the interaction energies.[93]

Note that the RMA must, by definition, decrease the magnitude of the interaction energy (for variational electronic structure methods). This result

Table 2 Interaction Energies of $(HF)_n$, $n = 3 - 5$ Computed at the RHF/aug-cc-pVDZ Level with (E_{int}^{CP}) and without (E_{int}) a Counterpoise Correction[a]

	E_{int}[b]	$E[HF]^{\text{cluster basis}}_{\text{cluster geom}}$	$E[HF]^{\text{monomer basis}}_{\text{cluster geom}}$	E_{int}^{CP}[c]
		Fully Optimized Clusters		
$(HF)_3$	−49.97	−100.034072	−100.033695	−47.00
$(HF)_4$	−90.53	−100.034030	−100.033543	−85.41
$(HF)_5$	−124.53	−100.034011	−100.033506	−117.90
		Rigid Monomer Approximation		
$(HF)_3$	−49.13	−100.034175	*d*	−46.30
$(HF)_4$	−88.12	−100.034287	*d*	−83.17
$(HF)_5$	−120.86	−100.034292	*d*	−114.61

[a]The monomer electronic energies for the counterpoise correction are also listed ($E[HF]$). Electronic energies are in E_h, and the interaction energies are in kJ mol^{-1}.
[b]Obtained via application of Eq. [1] or [2] to energies in Table 1.
[c]Obtained via application of Eq. [7] or [8] to energies in Tables 1 and 2.
[d]Equal to monomer energy from Table 1 which implies $E_{RLX} = 0$ in Eq. [7].

is readily illustrated with Figure 3. Although the RMA does not affect the asymptote associated with the noninteracting fragments, it does shift the bottom of the well up because the cluster is not allowed to reach its optimal geometry, which necessarily decreases the magnitude of E_{int} and D_e.

Counterpoise Corrections for Basis Set Superposition Error

The procedure outlined in Eq. [2] introduces an inconsistency when small, finite basis sets are used. Effectively, the monomers are using a larger basis set when the computation is performed on the cluster than when the computation is performed on the isolated monomer fragment. In the cluster calculation, monomer A can utilize the basis functions on monomers B, C, etc. When the computation is performed on the isolated monomer A, those basis functions are no longer available. This inconsistency was noted as early as 1968[94] and later termed basis set superposition error (BSSE).[95] A tutorial covering the theory and practice of basis set superposition errors has appeared earlier in this book series.[96] The most common procedure to correct for BSSE is the counterpoise (CP) procedure developed independently by Jansen and Ros in 1969[97] and Boys and Bernardi in 1970.[98] BSSE and CP corrections are discussed in greater detail below, and this portion of the tutorial merely demonstrates how to perform the necessary computations. Before proceeding, however, it is worth noting that BSSE is not limited to weakly interacting systems. It is a concern in any type of dissociation process (such as breaking a covalent bond) where the energies of fragments are compared to those of the whole system.

In this part of the tutorial, the standard Boys–Bernardi CP correction is applied to the interaction energies of the HF clusters in Figure 2 using the geometrical parameters in Table 1. Unfortunately, these corrections call for some rather hideous notation that denotes both the geometry and the basis set employed for computations on the monomers. The basic goal of the CP correction is to compute the energy of the monomer in the basis set of the cluster ($E[\mathrm{HF}]^{\text{cluster basis}}_{\text{monomer geom}}$). This is readily accomplished within the rigid monomer approximation because the geometry of the monomer is the same in the complex as in the isolated fragment ($E[\mathrm{HF}]^{\text{cluster basis}}_{\text{monomer geom}} = E[\mathrm{HF}]^{\text{cluster basis}}_{\text{cluster geom}}$), and the CP-corrected interaction energy within the RMA is simply

$$E_{\text{int}}^{\text{CP,RMA}} = E[(\mathrm{HF})_n] - nE[\mathrm{HF}]^{\text{cluster basis}}_{\text{cluster geom}} \qquad [3]$$

Again, this expression for the HF clusters can readily be generalized for the case of a heterogeneous cluster composed of N fragments:

$$E_{\text{int}}^{\text{CP,RMA}} = E[f_1 f_2 f_3 \ldots f_N] - \sum_{i=1}^{N} E[f_i]^{\text{cluster basis}}_{\text{cluster geom}} \qquad [4]$$

Let us use $(\mathrm{HF})_3$ in Figure 2 to illustrate the procedure. To perform a CP correction on the bottom HF unit in the trimer, the computations must place H and F basis functions, but not nuclei or electrons, at the appropriate coordinates of the other HF monomers at the top of the figure. In most computational chemistry programs this is accomplished with the use of ghost atoms or ghost orbitals. (Note, dummy atoms are also used to designate coordinates where nuclei are not present, but dummy atoms do not place basis functions at those locations.) Frequently, the input file for the CP-corrected monomer computation is created by modifying the input file from a cluster calculation such that the charges of all atoms are set to zero (i.e., the ghost atoms) except those in the monomer of interest. Because each computational chemistry software program has its own set of keywords for the specification of ghost atoms and nuclear charge, some sample input files for the CP corrections to the $(\mathrm{HF})_n$ interaction energies are available online.[92]

When the monomers are allowed to relax as the complex forms, the procedure becomes a bit more complicated because there is no straightforward, consistent manner by which a computation on the optimized monomer can be performed in the basis set of the cluster. Consequently, $E[\mathrm{HF}]^{\text{cluster basis}}_{\text{monomer geom}} \neq E[\mathrm{HF}]^{\text{cluster basis}}_{\text{cluster geom}}$ when the RMA is not employed. In other words, the energy of the monomer in the cluster basis set is too high (too positive) because the monomer is not at its optimal geometry. This overestimation of the monomer energy can be corrected easily by

calculating the energy liberated as the distorted monomers (complex geom) relax to their optimal structures in the (monomer geom) in the monomer basis set,

$$E_{\mathrm{RLX}}[\mathrm{HF}] = E[\mathrm{HF}]^{\text{monomer basis}}_{\text{monomer geom}} - E[\mathrm{HF}]^{\text{monomer basis}}_{\text{cluster geom}} \quad [5]$$

or more generally

$$E_{\mathrm{RLX}}[f_i] = E[f_i]^{\text{monomer basis}}_{\text{monomer geom}} - E[f_i]^{\text{monomer basis}}_{\text{cluster geom}} \quad [6]$$

Comparing the nearly optimal energy of the HF monomer in Table 1 to $E[\mathrm{HF}]^{\text{monomer basis}}_{\text{cluster geom}}$ in Table 2 reveals that the monomer energies at the distorted cluster geometries are too high by a range of roughly 0.1 $\mathrm{mE_h}$ for the trimer to 0.3 $\mathrm{mE_h}$ for the pentamer. This relaxation energy (E_{RLX}) can then be used to correct the monomer contributions to the CP-corrected interaction energy:

$$\begin{aligned} E^{\mathrm{CP}}_{\mathrm{int}} &= E[(\mathrm{HF})_n] - n(E[\mathrm{HF}]^{\text{cluster basis}}_{\text{cluster geom}} + E_{\mathrm{RLX}}[\mathrm{HF}]) \\ &= E[(\mathrm{HF})_n] - n(E[\mathrm{HF}]^{\text{cluster basis}}_{\text{cluster geom}} + E[\mathrm{HF}]^{\text{monomer basis}}_{\text{monomer geom}} - E[\mathrm{HF}]^{\text{monomer basis}}_{\text{cluster geom}}) \end{aligned} \quad [7]$$

Returning to the general case of a heterogeneous cluster composed of N fragments, an equivalent expression is obtained by summing over the fragments of the system:

$$\begin{aligned} E^{\mathrm{CP}}_{\mathrm{int}} &= E[f_1 f_2 f_3 \ldots f_N] - \sum_{i=1}^{N} (E[f_i]^{\text{cluster basis}}_{\text{cluster geom}} + E_{\mathrm{RLX}}[f_i]) \\ &= E[f_1 f_2 f_3 \ldots f_N] - \sum_{i=1}^{N} (E[f_i]^{\text{cluster basis}}_{\text{cluster geom}} + E[f_i]^{\text{monomer basis}}_{\text{monomer geom}} - E[f_i]^{\text{monomer basis}}_{\text{cluster geom}}) \end{aligned} \quad [8]$$

Note that when the RMA is applied, E_{RLX} vanishes since $E[f_i]^{\text{monomer basis}}_{\text{monomer geom}} = E[f_i]^{\text{monomer basis}}_{\text{cluster geom}}$ so that Eq. [8] reduces to Eq. [4] (and Eq. [7] reduces to Eq. [3]).

Step 3 in this tutorial is to reproduce the RHF/aug-cc-pVDZ electronic energies in Table 2. Sample input files for several popular software packages are available online.[92] To get the most out of this tutorial, it is recommended that you do not utilize features in some software packages that automatically perform the CP corrections for you. Step 4 of this tutorial is to calculate the interaction energies on the right side of Table 2 by applying Eq. [7] (or more generally Eq. [8]) to the electronic energies in Tables 1 and 2.

Interaction Energies: CP Corrected vs. CP Uncorrected

The rightmost column of Table 2 contains the CP-corrected interaction energies. The values of $E_{\text{int}}^{\text{CP}}$ were obtained by applying Eq. [8] to the electronic energies in Tables 1 and 2. With the relatively small aug-cc-pVDZ basis set, the BSSE at the RHF level is nearly $3\,\text{kJ mol}^{-1}$ (or 6%) for $(\text{HF})_3$, and it creeps up to more than $6\,\text{kJ mol}^{-1}$ (or 5%) for $(\text{HF})_5$. These values can change dramatically when different theoretical methods and/or basis sets are employed.

The magnitude of $E_{\text{int}}^{\text{CP}}$ is smaller than that of E_{int} for all three clusters. As with the RMA (discussed above), CP corrections for the BSSE also tend to decrease the magnitude of the interaction energy but for a different reason. With CP corrections, it is the monomer calculation that changes rather than the cluster calculation. Consequently, the bottom of the well in Figure 3 is unaffected by the procedure. The asymptote, however, is generally shifted downward because the larger basis set lowers the energies of the monomers. While this trend ($|E_{\text{int}}^{\text{CP}}| < |E_{\text{int}}|$) is generally true of most WFT conventional methods, CP corrections can actually increase the magnitude of the interaction energy when using resolution of the identity or techniques that employ auxiliary basis sets.[99]

A cautionary note concerning the deleted virtual approximation and CP corrections is offered. For appropriately constructed basis sets, high-lying unoccupied orbitals can be excluded from post-Hartree–Fock correlated computation in the same manner that low-lying core orbitals are omitted in the frozen core approximation. When performing monomer computations in the basis set of the cluster (particularly the heterogeneous variety), the ghost orbitals can sometimes have higher energies than the virtual orbitals centered on the monomer of interest. In such cases, most default procedures for the deleted virtual approximation will exclude the wrong virtual orbitals and special care must be taken to ensure the correct unoccupied orbitals are deleted (such as by reordering the orbitals).

Two-Body Approximation and Cooperative/Nonadditive Effects

Interactions in weakly bound clusters can frequently be dominated by 2-body interactions between pairs of fragments within the system. In other words, the sum of the interactions between each pair of fragments within the cluster can be used to approximate E_{int}. Conceptually, this 2-body approximation is straightforward when presented within the rigid monomer approximation because the geometries of the isolated monomers ($[f_1]$, $[f_2] \ldots [f_N]$) are identical to the monomer geometries in the cluster ($[f_1^*]$, $[f_2^*] \ldots [f_N^*]$). (In this section, an asterisk denotes that the fragment or group of fragments is at the cluster geometry.) For a trimer $f_1 f_2 f_3$, the pairwise or 2-body interaction

energy ($E_{\text{int}}^{\text{2-body}}$) is simply the sum of the interaction energies of each pair of fragments ($f_1f_2^*$, $f_1f_3^*$, and $f_2f_3^*$) within the cluster. This RMA prescription for a trimer can be extended readily to a cluster of arbitrary size, N, and composition in which there are $\binom{N}{2} = N!/(2!(N-2)!) = N(N-1)/2$ unique pairwise interactions:

$$E_{\text{int}}^{\text{2-body}}[f_1f_2\ldots f_N] = \sum_{i=1}^{N-1}\sum_{j>i}^{N} E[f_if_j^*] - E[f_i^*] - E[f_j^*] \qquad [9]$$

The pairwise approximation tends to be accurate in weakly coupled systems. For example, Tauer and Sherrill demonstrated that more than 98% of the interaction energy of various benzene tetramer structures can be recovered by simply adding together the pairwise interactions (or "dimers") in the system.[100] Despite ignoring higher order cooperative effects (3-body and 4-body in this case), $E_{\text{int}}^{\text{2-body}}$ differs from E_{int} by no more than 2% for the benzene tetramer configurations examined in the study. Because the higher order contributions account for deviations between the pairwise additive 2-body approximation and E_{int}, they are also frequently called nonadditive or cooperative effects (or just the nonadditivity or cooperativity). These many-body terms will be defined more precisely in the next section.

The nonadditivity tends to increase for more strongly coupled systems (sometimes dramatically), and, consequently, the quality of the 2-body approximation deteriorates.[101,102] In clusters of HF and/or H_2O, the nonadditivity can account for more than half of E_{int}, which necessarily implies that the error associated with the 2-body approximation can exceed 50%.[59] This section of the tutorial will use $(HF)_3$, $(HF)_4$, and $(HF)_5$ to demonstrate the procedure for calculating these 2-body interactions as well as higher order (3-body, ... N-body) contributions via a many-body decomposition of E_{int}.

Many-Body Decomposition

The most common rigorous many-body decomposition scheme for weakly bound clusters is based upon the approach introduced by Hankins, Moskowitz, and Stillinger in 1970.[103] Two lucid descriptions of the procedure can be found in Ref. 104 and 105. Technically, a many-body decomposition of E_{int} decomposes the energy of the cluster $E[f_1f_2\ldots f_N]$ into 1-body (E_1), 2-body (E_2), ..., N-body (E_N) contributions:

$$\begin{aligned} E_{\text{int}} &= E[f_1f_2f_3\ldots f_N] - \sum_{i=1}^{N} E[f_i] \\ &= \{E_1 + E_2 + E_3 + \cdots + E_N\} - \sum_{i=1}^{N} E[f_i] \end{aligned} \qquad [10]$$

Each n-body term is obtained by adding together the energies of each unique subset of n fragments within the cluster and subtracting from that the lower order (1-body, 2-body, ..., $(n-1)$-body) contributions. The first-order correction is merely a sum of "monomer" energies in the cluster:

$$E_1 = \sum_{i=1}^{N} E[f_i^*] \qquad [11]$$

Note that the monomer energies (at the cluster geometry) in this summation can be combined with monomer energies (at monomer geometry) from the summation in Eq. [10] to obtain the energy associated with the distortion of the monomers from their optimal structure to their geometry in the cluster (much like the relaxation energy for CP corrections in Eq. [6]). By definition, the contribution of E_{DIST} to the interaction energy is positive (net repulsive effect) when the clusters are fully optimized while $E_{\text{DIST}} = 0$ in the rigid monomer approximation:

$$E_{\text{DIST}} = \sum_{i=1}^{N} (E[f_i^*] - E[f_i]) \qquad [12]$$

$$E_{\text{int}} = E_{\text{DIST}} + E_2 + E_3 + \cdots + E_N \qquad [13]$$

The second-order term is obtained from the energies of each of the $\binom{N}{2} = N(N-1)/2$ unique pairs of fragments (or "dimers") from each of which $\binom{2}{1}$ 1-body contributions must be subtracted:

$$E_2 = \sum_{i=1}^{N-1} \sum_{j>i}^{N} \Delta_2 E[f_i f_j^*] \qquad [14]$$

$$\Delta_2 E[f_i f_j^*] = E[f_i f_j^*] - (E[f_i^*] + E[f_j^*]) \qquad [15]$$

One should recognize this expression for E_2 since it is identical to the 2-body interaction energy from Eq. [9] (i.e., $E_2 = E_{\text{int}}^{\text{2-body}}$). This relationship provides a rigorous definition of the nonadditivity or cooperativity:

$$\begin{aligned} E_{\text{int}} &= E_{\text{DIST}} + E_{\text{int}}^{\text{2-body}} + E_3 + \cdots + E_N \\ &= E_{\text{DIST}} + E_{\text{int}}^{\text{2-body}} + \delta E^{\text{nonadd}} \qquad [16] \\ &= E_{\text{DIST}} + E_{\text{int}}^{\text{many-body}} \qquad [17] \end{aligned}$$

Continuing to the third-order expression gives $\binom{N}{3} = N(N-1)(N-2)/6$ unique "trimers." There are $\binom{3}{2}$ 2-body and $\binom{3}{1}$ 1-body contributions that must be removed from each trimer energy:

$$E_3 = \sum_{i=1}^{N-2} \sum_{j>i}^{N-1} \sum_{k>j}^{N} \Delta_3 E[f_i f_j f_k^*] \quad [18]$$

$$\Delta_3 E[f_i f_j f_k^*] = E[f_i f_j f_k^*] - (\Delta_2 E[f_i f_j^*] + \Delta_2 E[f_i f_k^*] + \Delta_2 E[f_j f_k^*]) - (E[f_i^*] + E[f_j^*] + E[f_k^*]) \quad [19]$$

For the 4-body contribution, there are $\binom{N}{4}$ "tetramer" energies from each of which $\binom{4}{3}$ 3-body, $\binom{4}{2}$ 2-body, and $\binom{4}{1}$ 1-body terms are subtracted:

$$E_4 = \sum_{i=1}^{N-3} \sum_{j>i}^{N-2} \sum_{k>j}^{N-1} \sum_{l>k}^{N} \Delta_4 E[f_i f_j f_k f_l^*] \quad [20]$$

$$\Delta_4 E[f_i f_j f_k f_l^*] = E[f_i f_j f_k f_l^*] - (\Delta_3 E[f_i f_j f_k^*] + \Delta_3 E[f_i f_j f_l^*] + \Delta_3 E[f_i f_k f_l^*] + \Delta_3 E[f_j f_k f_l^*]) - (\Delta_2 E[f_i f_j^*] + \Delta_2 E[f_i f_k^*] + \Delta_2 E[f_i f_l^*] \Delta_2 E[f_j f_k^*] + \Delta_2 E[f_j f_l^*] + \Delta_2 E[f_k f_l^*]) - (E[f_i^*] + E[f_j^*] + E[f_k^*] + + E[f_l^*]) \quad [21]$$

A new indexing notation is introduced to help provide a generalized expression for the n-body contribution to the (interaction) energy of a cluster with N components. The indices $i, j, k, \ldots$ are replaced with $i_1, i_2, i_3, \ldots, i_{n-1}, i_n$ to emphasize that this n-body component, E_n, contains n nested summations giving rise to $\binom{N}{n} = N!/(n!(N-n)!$ terms:

$$E_n = \sum_{i_1=1}^{N-n+1} \sum_{i_2>i_1}^{N-n+2} \cdots \sum_{i_{n-1}>i_{n-2}}^{N-1} \sum_{i_n>i_{n-1}}^{N} \Delta_n E[f_{i_1} f_{i_2} f_{i_3} \ldots f_{i_{n-1}} f_{i_n}^*] \quad [22]$$

Again, each $\Delta_n E$ term is obtained by removing the lower order contributions from the electronic energy of the n-mers composed of fragments $f_{i_1} f_{i_2} f_{i_3} \ldots f_{i_{n-1}} f_{i_n}$. That is, one must subtract $\binom{n}{1}$ 1-body terms, $\binom{n}{2}$ 2-body terms, ..., $\binom{n}{n-1}(n-1)$-body terms from $E[f_{i_1} f_{i_2} f_{i_3} \ldots f_{i_{n-1}} f_{i_n}^*]$. To denote these terms; the indices $a_1, a_2, \ldots, a_{n-1}$ are used to run over the values of a particular set of fragment indices, $\mathbb{S} = \{i_1, i_2, i_3, \ldots, i_{n-1}, i_n\}$. Note that the index a_j corresponds to the jth element in $\mathbb{S}$ and, therefore, fragment f_{i_j}. It does not necessarily have anything to do with fragment f_j. This is a subtle but important distinction. For example, f_1 does not appear in $\Delta_3 E[f_3 f_5 f_7]$ even though a loops

over 1, 2, 3 to give $i_{a_1} = 3$, $i_{a_2} = 5$, and $i_{a_3} = 7$.

$$\begin{aligned}
&\Delta_n E[f_{i_1} f_{i_2} f_{i_3} \ldots f_{i_{n-1}} f_{i_n}^*] \\
&= E[f_{i_1} f_{i_2} f_{i_3} \ldots f_{i_{n-1}} f_{i_n}^*] \\
&\quad - \sum_{a=1}^{n} E[f_{i_a}^*] \\
&\quad - \sum_{a_1=1}^{n-1} \sum_{a_2>a_1}^{n} \Delta_2 E[f_{i_{a_1}} f_{i_{a_2}}^*] \\
&\quad - \sum_{a_1=1}^{n-2} \sum_{a_2>a_1}^{n-1} \sum_{a_3>a_2}^{n} \Delta_3 E[f_{i_{a_1}} f_{i_{a_2}} f_{i_{a_3}}^*] \\
&\quad - \cdots \\
&\quad - \sum_{1\le a_1<a_2\cdots<a_{n-2}} \sum^{n} \cdots \sum \Delta_{n-2} E[f_{i_{a_1}} f_{i_{a_2}} f_{i_{a_3}} \ldots f_{i_{a_{n-2}}}^*] \\
&\quad - \sum_{1\le a_1<a_2\cdots<a_{n-2}<a_{n-1}} \sum^{n} \cdots \sum \sum \Delta_{n-1} E[f_{i_{a_1}} f_{i_{a_2}} f_{i_{a_3}} \ldots f_{i_{a_{n-2}}} f_{i_{n-1}}^*]
\end{aligned} \quad [23]$$

By expanding $\Delta_2 E, \Delta_3 E, \ldots, \Delta_n E$, the expressions for the components of the cluster energy (and therefore interaction energy via Eq. [10] and [13]) can be simplified. In the following form, it is easier to see a connection between this many-body decomposition and the inclusion–exclusion principle (also known as the sieve principle) from combinatorial mathematics:

$$\begin{aligned}
E_2 &= \sum_{i=1}^{N-1} \sum_{j>i}^{N} E[f_i f_j^*] \\
&\quad - (E[i^*] + E[j^*])
\end{aligned} \quad [24]$$

$$\begin{aligned}
E_3 &= \sum_{i=1}^{N-2} \sum_{j>i}^{N-1} \sum_{k>j}^{N} E[f_i f_j f_k^*] \\
&\quad - (E[f_i f_j^*] + E[f_i f_k^*] + E[f_j f_k^*]) \\
&\quad + (E[f_i^*] + E[f_j^*] + E[f_k^*])
\end{aligned} \quad [25]$$

$$\begin{aligned}
E_4 &= \sum_{i=1}^{N-3} \sum_{j>i}^{N-2} \sum_{k>j}^{N-1} \sum_{l>k}^{N} E[f_i f_j f_k f_l^*] \\
&\quad - (E[f_i f_j f_k^*] + E[f_i f_j f_l^*] + E[f_i f_k f_l^*] + E[f_j f_k f_l^*]) \\
&\quad + (E[f_i f_j^*] + E[f_i f_k^*] + E[f_i f_l^*] + E[f_j f_k^*] + E[f_j f_l^*] + E[f_k f_l^*]) \\
&\quad - (E[f_i^*] + E[f_j^*] + E[f_k^*] + E[f_l^*])
\end{aligned} \quad [26]$$

Again, a general expression for E_n can be obtained in the same manner from Eq. [22]:

$$\begin{aligned}
E_n = & \sum_{1\le i_1<i_2\cdots<i_{n-2}<i_{n-1}<i_n}^{N}\cdots\sum\sum\sum E[f_{i_1}f_{i_2}f_{i_3}\ldots f_{i_{n-2}}f_{i_{n-1}}f_{i_n}^*] \\
& + (-1)^1 \sum_{1\le i_1<i_2\cdots<i_{n-2}<i_{n-1}}^{N}\cdots\sum\sum E[f_{i_1}f_{i_2}f_{i_3}\ldots f_{i_{n-2}}f_{i_{n-1}}^*] \\
& + (-1)^2 \sum_{1\le i_1<i_2\cdots<i_{n-2}}^{N}\cdots\sum E[f_{i_1}f_{i_2}f_{i_3}\ldots f_{i_{n-2}}^*] \\
& + \cdots \\
& + (-1)^{n-3}\sum_{i_1=1}^{N-2}\sum_{i_2>i_1}^{N-1}\sum_{i_3>i_2}^{N} E[f_{i_1}f_{i_2}f_{i_3}^*] \\
& + (-1)^{n-2}\sum_{i_1=1}^{N-1}\sum_{i_2>i_1}^{N} E[f_{i_1}f_{i_2}^*] \\
& + (-1)^{n-1}\sum_{i_1=1}^{N} E[f_{i_1}^*]
\end{aligned} \qquad [27]$$

Application to HF Trimer, Tetramer, and Pentamer

Because the cyclic $(HF)_n$ clusters ($n = 3 - 5$) used in this tutorial are symmetric, the number of computations required to perform a many-body decomposition of the interaction energy is reduced dramatically. In general, application of the decomposition procedure to a pentamer could require as many as 25 additional calculations: $\binom{5}{4} = 5$ for the tetramer subsets, $\binom{5}{3} = 10$ for the trimer subsets, and $\binom{5}{2} = 10$ for the dimer subsets. For $(HF)_5$, however, symmetry reduces this to 5 calculations (1 unique tetramer computation, 2 unique trimer computations, and 2 unique dimer computations).

For $(HF)_3$, there is only a single unique 2-body energy since $E[f_1f_2^*] = E[f_1f_3^*] = E[f_2f_3^*]$ while there are two such quantities for $(HF)_4$ and $(HF)_5$ ($E[f_1f_2^*] = E[f_1f_4^*] = E[f_2f_3^*] = E[f_3f_4^*] \equiv E[f_if_j^*]$ and $E[f_1f_3^*] = E[f_2f_4^*] \equiv E[f_if_k^*]$). These values are reported Table 3 for both relaxed and rigid monomers. One finds a single unique 3-body energy ($E[f_1f_2f_3^*] = E[f_1f_2f_4^*] = E[f_1f_3f_4^*] = E[f_2f_3f_4^*]$) for $(HF)_4$ but two for $(HF)_5$ ($E[f_1f_2f_3^*] = E[f_1f_2f_5^*] = E[f_1f_4f_5^*] = E[f_2f_3f_4^*] = E[f_3f_4f_5^*] \equiv E[f_if_jf_k^*]$ and $E[f_1f_2f_4^*] = E[f_1f_3f_4^*] = E[f_1f_3f_5^*] = E[f_2f_3f_5^*] = E[f_2f_4f_5^*] \equiv E[f_if_jf_l^*]$). Of course, $(HF)_5$ has only one unique 4-body energy, which is given in Table 3 along with all of the 2- and 3-body energies.

Step 5 in this tutorial is to reproduce the RHF/aug-cc-pVDZ electronic energies in Table 3. Sample input files for several popular software packages are available online.[92] It is worth noting that these computations could just as easily be performed in the entire basis set of the complex, thereby yielding a

Table 3 Unique Many-Body Electronic Energies (in E_h) for $(HF)_n$, $n = 3 - 5$ Computed at the RHF/aug-cc-pVDZ Level

	$E[f_i f_j^*]$	$E[f_i f_k^*]$	$E[f_i f_j f_k^*]$	$E[f_i f_j f_l^*]$	$E[f_i f_j f_k f_l^*]$
			Fully Optimized Clusters[a]		
$(HF)_3$[b]	−200.072923	—	—	—	—
$(HF)_4$[b]	−200.072748	−200.069328	−300.116161	—	—
$(HF)_5$[b]	−200.072425	−200.068570	−300.115003	−300.109494	−400.160289
			Rigid Monomer Approximation[c]		
$(HF)_3$[b]	−200.073038	—	—	—	—
$(HF)_4$[b]	−200.073100	−200.069755	−300.116246	—	—
$(HF)_5$[b]	−200.072990	−200.069093	−300.115406	−300.110113	−400.160256

[a]See Tables 1 and 2 for 1-body energies.
[b]See Table 1 for the full n-body energies for $(HF)_n$.
[c]See Table 1 for 1-body energy.

CP-corrected many-body decomposition. However, that would increase the time of the computations for this tutorial substantially.

The monomer energies from Tables 1 and 2 have been used to determine the E_{DIST} values (Eq. [12]) shown in Table 4. (Again, a conversion factor of $1E_h \approx 2625.5\,kJ\,mol^{-1}$ has been adopted.) For these symmetric cyclic $(HF)_n$ $(n = 3 - 5)$ clusters, E_{DIST} is simply $n \times (E[HF^*] - E[HF]) = n \times (-E_{RLX})$. The many-body interaction energy can then be calculated from E_{int} and E_{DIST} via Eq. [17]. Recall that within the RMA, $E_{DIST} = 0$ so that in the bottom half of Table 4 $E_{int}^{many\text{-}body}$ is the same as E_{int}.

The 2-body through 5-body contributions to the many-body interaction energy shown in Table 4 are relatively simple to compute because there are only a few symmetry-unique terms. As mentioned earlier, there exist at most two unique 2-body energies [in $(HF)_4$ and $(HF)_5$] and two unique 3-body energies [in $(HF)_5$]. Furthermore, all monomers in a given cluster are identical, and the corresponding energies can be obtained from Table 2 ($E[HF^*] = E[HF]_{\text{cluster geom}}^{\text{monomer basis}}$).

Table 4 Many-Body Decomposition of E_{int} for $(HF)_n$, $n = 3 - 5$[a]

	E_{dist}	$E_{int}^{many\text{-}body}$	E_2	E_3	E_4	E_5	δE^{nonadd}
				Fully Optimized Clusters			
$(HF)_3$	+0.96	−50.93	−43.58	−7.34	—	—	−7.34
$(HF)_4$	+2.87	−93.40	−71.23	−20.65	−1.51	—	−22.17
$(HF)_5$	+4.08	−128.61	−91.52	−33.43	−3.38	−0.27	−37.08
				Rigid Monomer Approximation			
$(HF)_3$	—	−49.13	−42.58	−6.55	—	—	−6.55
$(HF)_4$	—	−88.12	−68.56	−18.27	−1.28	—	−19.55
$(HF)_5$	—	−120.86	−89.51	−28.45	−2.66	−0.24	−31.35

[a]All values were computed at the RHF/aug-cc-pVDZ level and are reported in $kJ\,mol^{-1}$.

Consequently, Eq. [27] and the preceding equations end up with fairly simple forms for $(HF)_n$, $n = 3 - 5$:

$$E_{DIST}[HF_n] = -n \times E_{RLX}[HF] \quad [28]$$

$$E_2[HF_3] = 3E[f_i f_j^*] - 6E[HF^*] \quad [29]$$

$$E_2[HF_4] = 4E[f_i f_j^*] + 2E[f_i f_k^*] - 12E[HF^*] \quad [30]$$

$$E_2[HF_5] = 5E[f_i f_j^*] + 5E[f_i f_k^*] - 20E[HF^*] \quad [31]$$

$$E_3[HF_4] = 4(E[f_i f_j f_k^*] - \{2E[f_i f_j^*] + E[f_i f_k^*]\} + 3E[HF^*]) \quad [32]$$

$$E_3[HF_5] = 5(E[f_i f_j f_k^*] + E[f_i f_j f_l^*] - 3\{E[f_i f_j^*] + E[f_i f_k^*]\} + 6E[HF^*]) \quad [33]$$

$$E_4[HF_5] = 5(E[f_i f_j f_k f_l^*] - 2\{E[f_i f_j f_k^*] + E[f_i f_j f_k^*]\} + 3\{E[f_i f_j^*] + E[f_i f_k^*]\} - 4E[HF^*]) \quad [34]$$

The full n-body contribution to each $(HF)_n$ cluster can be obtained in different ways, the easiest of which is to subtract the lower order contributions $(E_2, E_3, \ldots, E_{n-1})$ from $E_{int}^{many\text{-}body}$. Alternatively, Eq. [27] can be simplified in the same manner as the lower order terms:

$$E_3[HF_3] = E[HF_3] - 3E[f_i f_j^*] + 3E[HF^*] \quad [35]$$

$$E_4[HF_4] = E[HF_4] - 4E[f_i f_j f_k^*] + 4E[f_i f_j^*] + 2[f_i f_k^*] - 4E[HF^*] \quad [36]$$

$$E_5[HF_5] = E[HF_5] - 5E[f_i f_j f_k f_l^*] + 5E[f_i f_j f_k^*] + 5[f_i f_j f_l^*] - 5E[f_i f_j^*] - 5[f_i f_k^*] + 5E[HF^*] \quad [37]$$

Although every contribution to the interaction energies of these HF clusters are attractive, this does not always hold. In certain cases, some of the many-body components may actually be repulsive.[100] In step 6, you should calculate components of E_{int} in Table 4 $(E_{DIST}, E_2, E_3, E_4, E_5)$ by applying Eqs. [12], [16], [17], [25]–[27] to the electronic energies in the Tables 1, 2, and 3.

Size Consistency and Extensivity of the Energy

Energy is an extensive property. This fundamental thermodynamic principle is introduced early in most general chemistry textbooks, and it provides the foundation for the supermolecule description of intermolecular interactions. Unfortunately, not all electronic structure techniques are size consistent[106] (or more generally size extensive[107]). That is, the energy computed by some methods does not scale properly with the number of noninteracting fragments. Readers interested in more detail may be interested in the sections discussing size consistency and extensivity in the review of coupled-cluster theory by Crawford and Schaefer.[108]

To illustrate the point, consider two noninteracting (i.e., well-separated) HF molecules. (This can effectively be achieved in a computation by placing the HF molecules on the z axis and separating them by ≈ 1000 Å so that the F atoms are at 0 and 1000 Å while the H atoms are at 0.900 and 1000.900 Å.) Using the data in Table 1, we know the electronic energy of a single HF molecule is $-100.033816\,E_h$ when computed with the SCF method and the aug-cc-pVDZ basis set. As you would expect, the corresponding energy of two noninteracting HF molecules (i.e., separated by 1000 Å) is exactly $2 \times -100.033816 = -200.067632\,E_h$ but only because the SCF method is size consistent. Truncated configuration interaction (CI) methods such as the one including only single and double configurations (CISD) are not. The CISD/aug-cc-pVDZ energy is $-100.253275\,E_h$ if all electrons are correlated while that of two monomers separated by 1000 Å is $-200.488703\,E_h$, which is significantly different than $2 \times -100.253275 = -200.506551\,E_h$. As mentioned in the Introduction, this review only focuses on size-consistent electronic structure techniques. In the final step (step 7) of this tutorial, you should compute the SCF and CISD (all electrons correlated) energies of two HF monomers separated by 1000 Å. Compare these energies to those of a single monomer.

Summary of Steps in Tutorial

1. Reproduce electronic energies in Table 1.
2. Calculate E_{int} values on the left side of Table 2 (Eq. [2]).
3. Reproduce electronic energies in Table 2.
4. Calculate E_{int}^{CP} values on the right side of Table 2 (Eq. [8]).
5. Reproduce electronic energies in Table 3.
6. Calculate components of E_{int} in Table 4 (Eqs. [12], [16], [17], [25]–[27]).
7. Reproduce the electronic energies discussed in the proceedings section and check the size consistency of the results obtained with the SCF and CISD methods.

Sample input files for various software packages are available online.[92] If the electronic energies that you compute do not agree with those presented here, make sure that the energy is converged to at least eight decimal places in the SCF procedure and that tolerances for integral screening are no larger than 10^{-10}. Additionally, discrepancies on the order of $1 \times 10^{-6}\,E_h$ have been observed in some cases that can be attributed to differences in the conversion factor used to change angstroms to bohrs. Finally, rounding errors may lead to discrepancies on the order of $0.01\,kJ\,mol^{-1}$ for the other data presented in Tables 2 and 4.

HIGH-ACCURACY COMPUTATIONAL STRATEGIES

Although quantum mechanical studies of weak interactions can be traced back to Slater's 1929 work on He,[109] the first supermolecule ab initio

investigations of hydrogen bonding (reminiscent of those outlined above) were conducted approximately four decades ago.[85,110,111] Accuracy has been and continues to be one of the major challenges facing theoreticians (as well as experimentalists) working with weakly bound clusters. Consider, for example, covalent versus noncovalent interactions. An error of a few kilojoules per mole (*chemical accuracy*) per covalent bond may be acceptable because it typically represents a relative error of just a few percent. However, for weak noncovalent bonding an absolute error of a few kilojoules per mole could easily amount to a relative error in excess of 100%. Fortunately, by carefully applying the arsenal of sophisticated electronic structure techniques available today, it is possible to reduce the major sources of error (basis sets and electron correlation) to acceptable levels.

One of the most important lessons learned over the years is that not all weak noncovalent interactions are created equal. A particular quantum model chemistry that provides quantitatively reliable results for hydrogen bonding may yield qualitatively incorrect results for something like π stacking. For example, second-order Møller–Plesset (MP2) perturbation theory and several popular density functional (DFT) techniques can characterize the water dimer and trimer with a reasonable degree of accuracy. However, the former method overestimates π-stacking interactions in benzene by a factor of 2, while the latter fail to yield any sort of attractive interaction between two stacked benzene molecules. Consequently, it is imperative that "the right answer" is obtained for "the right reason" rather than relying on (or hoping for) some sort of error cancellation. Fortunately, well-established procedures exist by which one can converge to "the right answer." The most common of these convergent approaches to high-accuracy computational chemistry systematically improve (i) the correlated electronic structure techniques and (ii) the atomic orbital (AO) basis sets. This dual extrapolation scheme is depicted in Figure 4.

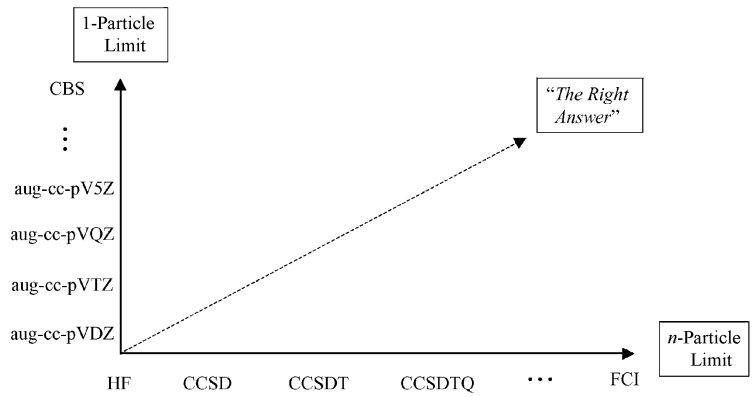

Figure 4 Example of convergent quantum chemistry scheme that employs AO basis sets that systematically approach the 1-particle or complete basis set (CBS) limit along with correlated electronic structure techniques that systematically approach the *n*-particle or full configuration interaction (FCI) limit.

Primer on Electron Correlation

The first component of convergent approaches to quantum chemistry is the computational procedure used to treat the electron correlation problem (depicted along the horizontal axis of Figure 4). A tutorial on treating electron correlation has been published earlier in this book series.[112] References 113 and 114 provide two additional excellent overviews of the subject. In any system with n interacting bodies (classical or quantum), the instantaneous motions of the bodies are correlated. Except for the simplest cases (e.g., certain one-electron systems), exact solutions to this n-particle (or many-body) problem cannot be obtained. Mean-field approximations (such as Hartree–Fock theory) neglect the instantaneous correlated motions of the bodies. The "missing" energy that corresponds to these simultaneous and instantaneous interactions is the correlation energy. In electronic structure theory, the correlation energy is typically (although not unambiguously) defined as the difference between the exact (nonrelativistic) electronic energy and the Hartree–Fock energy.[115]

Configuration interaction (CI) theory, coupled-cluster (CC) theory, and many-body perturbation theory (MBPT), of which Møller–Plesset (MP) perturbation theory is a specific case, are three of the most popular and relevant approaches that have been developed to systematically improve the computational description of electron correlation that is absent in Hartree–Fock theory. (Density functional methods are not mentioned here. Although DFT provides a very cost-effective means of recovering part of the electron correlation energy, the systematic improvement of individual functionals is problematic.) The missing correlation energy is recovered by constructing the wave function out of many different electron configurations (or Slater determinants) that are generated by "exciting" electrons from the occupied orbitals of the Hartree–Fock reference configuration to unoccupied (or virtual) orbitals. These additional (or excited) configurations are typically classified by excitation (or substitution) level: S for single excitations/substitutions, D for double, T for triple, etc.

Approximate many-electron wave functions are then constructed from the Hartree–Fock reference and the excited-state configurations via some sort of expansion (e.g., a linear expansion in CI theory, an exponential expansion in CC theory, or a perturbative power series expansion in MBPT). When all possible excitations have been incorporated (S, D, T, $\ldots, n$ for an n-electron system), one obtains the exact solution to the nonrelativistic electronic Schrödinger equation for a given AO basis set. This n-particle limit is typically referred to as the full CI (FCI) limit (which is equivalent to the full CC limit). As Figure 5 illustrates, several WFT methods can, at least in principle, converge to the FCI limit by systematically increasing the excitation level (or perturbation order) included in the expansion technique.

It is particularly important to note that while the linear CI expansion necessarily converges and all evidence suggests the exponential CC expansion always converges, the MBPT (or MP) series does diverge occasionally.[116,117]

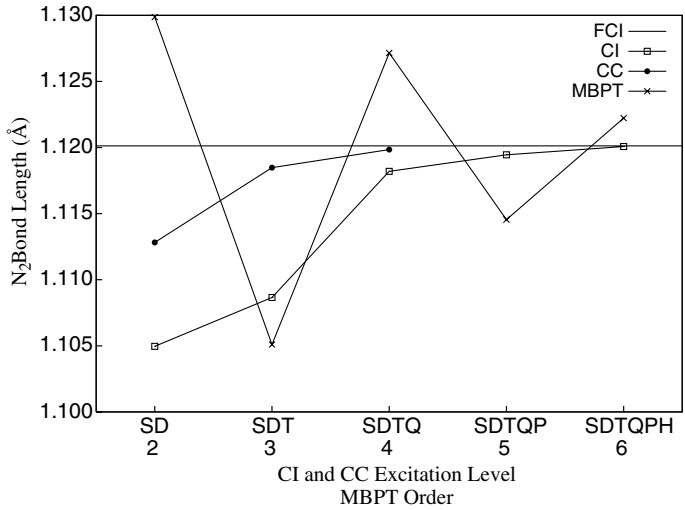

Figure 5 Convergence of correlated methods to the FCI bond length of N_2. All results were obtained with the cc-pVDZ basis set. The CCSDTQ bond length was computed for this work while all other data points were taken from Ref. 116.

Consequently, the most popular progression toward the FCI limit is MBPT2 (or MP2) → CCSD(T) rather than MBPT2 (or MP2) → MP3 → MP4. The CCSD(T) method, which includes a perturbative estimate of triple substitutions, is often referred to as the "gold standard" of quantum chemistry because (i) it generally provides results that are close to the FCI limit, especially for the systems that are the focus of this chapter, and (ii) it can be applied feasibly to moderately sized systems (a few dozen atoms). [Estimates of higher order correlation effects (e.g., quadruple substitutions) suggest that the CCSD(T) method provides converged results for the entire spectrum of noncovalent interactions.[118–120]] The CCSD → CCSD(T) sequence is also useful but is less commonly used because the CCSD method has more significant computational demands than MP2. These computational demands (or computational overhead) are the topic of the discussion below on the scaling problem.

Primer on Atomic Orbital Basis Sets

To introduce the concepts of a basis set and basis functions, we begin with a simple (unknown) function of a single variable, $f(x)$. A variety of procedures can be used to "fit" (or estimate) this function. For example, a simple power series could be used to approximate this simple function of the variable x:

$$f(x) \approx c_0 + c_1 x + c_2 x^2 + c_3 x^3 \qquad [38]$$

Here we are using a *basis set* to approximate the unknown function $f(x)$. The basis functions are $\{x^i : i = 0, 1, 2, 3\}$. The expansion coefficients, c_i, are determined by some sort of procedure that adjusts their values in order to obtain the best fit to the function $f(x)$. The approximation can generally be improved by using a larger basis set,

$$f(x) \approx c_0 + c_1x + c_2x^2 + c_3x^3 + c_4x^4 + c_5x^5 \qquad [39]$$

and it becomes exact in the limit of an infinitely large or complete basis set (CBS):

$$f(x) = c_0 + c_1x + c_2x^2 + c_3x^3 + \cdots = \sum_{i=0}^{\infty} c_ix^i \qquad [40]$$

In quantum chemistry we are concerned with approximating a molecular wave function, ψ, rather than a simple function of a single variable, $f(x)$. In the Hartree–Fock approximation, the many-electron wave function, ψ, is approximated with the antisymmetrized product of one-electron molecular orbitals (MOs). As you might expect, powers of x are not necessarily the best choice for a basis set in which to expand these one-electron functions. It does not require too much chemical intuition to recognize that the analytical wave functions for one-electron atoms (i.e., the s, p, and d orbitals shown in general chemistry textbooks) might provide a good set of basis functions in which to expand the molecular orbitals. After all, molecules are made of atoms. So, why not build molecular orbitals out of atomic orbitals? This is, of course, the familiar linear combination of atomic orbitals to form molecular orbitals (LCAO-MO) approximation. Both Slater and Gaussian atomic orbitals (AOs) provide fairly convenient basis functions for electronic structure computations. Of course, not all basis sets need to have a chemically motivated origin. For example, plane wave basis sets owe their success to computational efficiency.

Unfortunately, a bigger AO basis set does not necessarily give better results. Bigger is better only if the basis sets are properly constructed. In 1989, Dunning introduced the correlation-consistent family of basis sets,[121–123] which was a huge advance in the field of convergent quantum chemistry. They were designed to converge systematically to the complete basis set (CBS) or 1-particle limit. These basis sets are typically denoted cc-pV*X*Z where X denotes the maximum angular momentum of the Gaussian atomic orbitals in the basis set (2 for d functions, 3 for f functions, etc.) and is also referred to as the cardinal number of the basis (D for double-ζ basis set, T for triple-ζ, etc.). Because of the convergence properties of these basis sets, we can expect the larger basis sets to be more reliable. [$X = 4$ (or Q) is better than $X = 3$ (or T), which is better than $X = 2$ (or D).] The same is not true of other families of basis sets [e.g., 6-311G($2df, 2pd$) vs. 6-31G(d, p) or TZ2P vs. DZP].

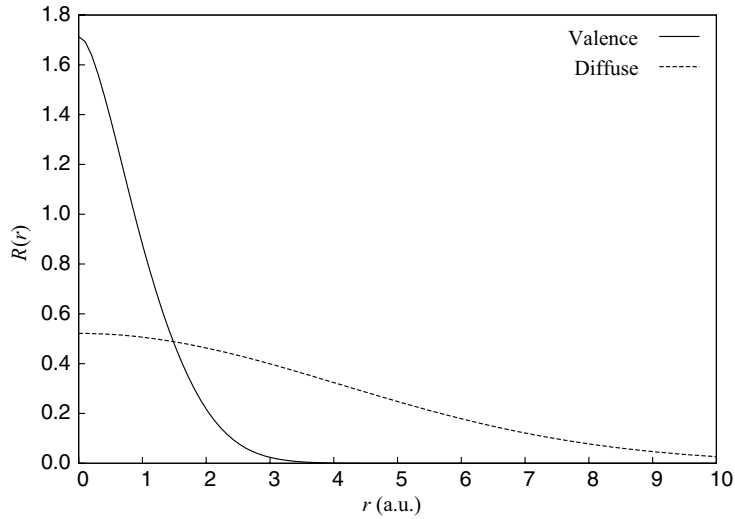

Figure 6 Diffuse basis functions are more spatially extended than their valence counterparts. The dashed curve represents the radial plot of the normalized diffuse 1*s* Gaussian basis function for H from the aug-cc-pVDZ basis set while the solid curve is for the corresponding contracted valence 1*s* basis function.

In the examples shown along the *y* axis in Figure 4, the "aug-" prefix indicates that the correlation-consistent basis sets have been augmented with diffuse basis functions.[122] These functions have small orbital exponents and are therefore spatially extended as can be seen in the radial plots shown in Figure 6. Diffuse functions are useful in computations for weakly bound clusters because they help describe the long-range interactions between fragments. (Basis sets augmented with diffuse functions are also commonly used to improve the description of negatively charged ions.) In fact, diffuse functions are practically essential for such applications. In some situations, diffuse functions need only be added to nonhydrogen atoms. However, results for weak noncovalent interactions obtained without diffuse functions must be analyzed carefully because CP corrections for basis set superposition error can still leave an unacceptably large basis set incompleteness error (see discussion below).

Extrapolation Techniques

Because of the systematic nature of the correlation-consistent family of basis sets, it is possible to use extrapolation techniques to estimate the CBS limit. For example, the three-parameter exponential function introduced by Feller nicely describes the convergence of SCF energy to the CBS limit with respect to the cardinal number of the basis set, X.[124,125] By fitting data from

three or more basis sets to this function, one can estimate the SCF CBS limit:

$$E_{\mathrm{SCF}} = E_{\mathrm{SCF}}^{\mathrm{CBS}} + a\exp(-bX) \qquad [41]$$

Various formulas have been proposed to describe the convergence behavior of the correlation energy (E_{corr}), which is distinctly different from and slower than that for the SCF energy. Here a few common examples are introduced. Note that the following extrapolation procedures are applied only to the correlation energy, not the total energy. In other words, $E_{\mathrm{corr}} = E_{\mathrm{total}} - E_{\mathrm{SCF}}$. The simple two-parameter formula suggested by Helgaker et al.[126] is popular:

$$E_{\mathrm{corr}} = E_{\mathrm{corr}}^{\mathrm{CBS}} + \frac{b}{X^3} \qquad [42]$$

because it can be manipulated into an expression that utilizes only the two most accurate data points. No fitting is required with the resulting equation. One simply inserts the correlation energies from the two largest basis sets with cardinal numbers $X_{\max}$ and $X_{\max} - 1$:

$$E_{\mathrm{corr}}^{\mathrm{CBS}} = \frac{E_{\mathrm{corr}}^{X_{\max}}(X_{\max})^3 - E_{\mathrm{corr}}^{X_{\max}-1}(X_{\max}-1)^3}{(X_{\max})^3 - (X_{\max}-1)^3} \qquad [43]$$

For the fitting of correlation energies obtained with extremely large hextuple- and heptuple-zeta (6Z and 7Z) basis sets, the following revision to Eq. [43] has been suggested based on MP2 pair energies:[127]

$$E_{\mathrm{corr}}^{\mathrm{CBS}} = \frac{E_{\mathrm{corr}}^{X_{\max}}(X_{\max}+0.5)^3 - E_{\mathrm{corr}}^{X_{\max}-1}(X_{\max}-0.5)^3}{(X_{\max}+0.5)^3 - (X_{\max}-0.5)^3} \qquad [44]$$

Also, Martin has proposed a two-parameter fit to a quartic polynomial (Schwartz4) and a three-parameter fit to a sixth-degree polynomial (Schwartz6):[128]

$$E_{\mathrm{corr}} = E_{\mathrm{corr}}^{\mathrm{CBS}} + \frac{a}{(X+\frac{1}{2})^4} + \frac{b}{(X+\frac{1}{2})^6} \quad \begin{cases} b = 0 \text{ Schwartz4} \\ b \neq 0 \text{ Schwartz6} \end{cases} \qquad [45]$$

Although many more incarnations describing the convergence of the correlation energy with respect to X can be found in the literature, they tend to adhere to the same philosophy as that adopted in Eqs. [42]–[45]. It is worth noting that better results are generally obtained when only the most accurate

data points are used in the extrapolation.[129,130] For example, if MP2 correlation energies are available for cc-pVDZ, cc-pVTZ, cc-pVQZ, cc-pV5Z, and cc-pV6Z basis sets, then the most reliable extrapolation schemes can be obtained by fitting the Q/5/6 data points or just the 5/6 data points. Some examples of these extrapolation procedures are presented in the tutorial below.

Explicitly Correlated Methods

A serious implication of the equations presented in the previous section is that the correlation energy converges to the CBS limit slowly with respect to the cardinal number (or angular momentum) of the basis set. In response, dramatic progress has been made in the development of explicitly correlated R12 methods that " bypass the slow convergence of conventional methods, by augmenting the traditional orbital expansions with a small number of terms that depend explicitly on the interelectronic distance $r12$."[131] Through various approximations (e.g., the resolution of the identity) and by changing the linear $r12$ dependence to a different functional form ($f12$), these R12 and F12 methods can provide correlation energies (typically at the MP2 level) that are converged to the CBS limit with only TZ or possibly even DZ quality basis sets. Readers interested in more details are strongly encouraged to consult the outstanding review by Klopper and co-workers.

Scaling Problem

Thanks to the concurrent development of more efficient computer algorithms and affordable high-performance computing hardware, the sophisticated electronic structure techniques described in the primer on electron correlation above can be brought to bear on weakly bound clusters of ever-increasing size (and with larger/better basis sets such as the correlation-consistent basis sets described in the preceding section). The drawback of these correlated electronic structure techniques is that their computational demands (memory, CPU time, disk space) increase sharply with the size of the system. For example, the "gold standard" of single-reference, ground-state quantum chemistry [i.e., the CCSD(T) method] scales as the 7th power of the size of the system, $\mathcal{O}(N^7)$. The practical consequences of this are devastating. Suppose you have the facilities to perform a CCSD(T) computation on the water hexamer with the aug-cc-pVTZ basis set (552 basis functions). By the time this chapter is printed, high-performance personal computers *might* be fast enough to perform a serial (nonparallel) computation of this magnitude in approximately 1 week. If one wishes to perform the same calculation on $(H_2O)_{18}$, the computational requirements will increase by $3^7 = 2187$ since the system has tripled in size. That means the time required to perform the computation would increase from 1 week to more than 42 years on the same computer. (A nice overview of the scaling requirements of some popular methods can be

found in Ref. 113, and a tutorial on linear scaling in quantum chemistry has appeared in this book series.[132]) Parallelization does not solve the issue. If a parallel CCSD(T) code executed on a high-performance cluster can reduce the time for the $(H_2O)_6$ computation to 1 day, the $(H_2O)_{18}$ calculation will still take 6 years to finish. Even if time was not a factor, these ludicrous computations would not be feasible because memory and disk requirements also increase by the same factor of 2187.

Estimating E_{int} at the CCSD(T) CBS Limit: Another Tutorial

High-accuracy model chemistries (for all types of chemical systems, not just weakly bound clusters) typically rely on additive schemes because of the hefty computational demands of highly correlated electronic structure techniques (see preceding section). In this section we demonstrate how to reliably estimate the CCSD(T) CBS limit even though it is generally not feasible to compute CCSD(T) energies for weakly bound clusters with basis sets large enough to yield a meaningful extrapolation to the CBS limit. This feat can be achieved because contributions from higher order (triple, quadruple, etc.) excitations tend to converge quickly with respect to the size of the AO basis set even though the total correlation energy converges slowly to the CBS limit (discussed above). As a result, the general strategy is to combine the CBS limit for a less demanding correlated method that includes only lower order excitations (e.g., MP2, CCSD, CISD) with a correction for higher order correlation effects obtained with small basis sets. The most popular combination is to use the MP2 CBS limit with a CCSD(T) correction.

Table 5 contains the MP2, CCSD, and CCSD(T) correlation energies for the HF monomer and trimer obtained with a series of correlation-consistent basis sets where diffuse functions are added only to the heavy (nonhydrogen) atoms, denoted ha*XZ*. The frozen core approximation was adopted for all of the calculations (i.e., electrons in the 1*s*-like orbitals on F were not included in the correlation procedure). The SCF energies converge very quickly. The ha5Z and ha6Z data points are within 1 mE_h of the SCF CBS limit that was obtained by fitting all five SCF energies (haDZ–ha6Z) to Eq. [41]. In contrast, the MP2 correlation energy converges more slowly. The ha5Z values are still more than 9 mE_h away from the MP2 CBS limit that was obtained by simply applying Eq. [43] to the ha5Z and ha6Z MP2 correlation energies (not the total MP2 electronic energies). The CCSD and CCSD(T) correlation energies are also provided for the haDZ, haTZ, and haQZ basis sets. No CCSD and CCSD(T) CBS limits are given, however, because extrapolations with smaller basis sets tend not to be as reliable as when larger correlation-consistent basis sets (e.g., pentuple- or sextuple-ζ) are used to obtain the correlation energies.[129,130,133] Unfortunately, such computations are often prohibitively demanding.

Table 5 SCF Electronic Energies and MP2, CCSD, and CCSD(T) Correlation Energies (in E_h) of HF and $(HF)_3$ Obtained with a Series of Correlation Consistent Basis Sets

Basis set[a]	E_{SCF}	E_{MP2}	E_{CCSD}	$E_{CCSD(T)}$
		HF Monomer		
haDZ (28)	−100.033348	−0.220691	−0.224524	−0.228425
haTZ (60)	−100.061354	−0.278496	−0.279840	−0.287214
haQZ (110)	−100.068993	−0.300097	−0.299628	−0.307789
ha5Z (182)	−100.071047	−0.309009	—	—
ha6Z (280)	−100.071251	−0.313005	—	—
CBS	−100.071625[b]	−0.318494[c]	—	—
		HF Trimer		
haDZ (84)	−300.119318	−0.665993	−0.676952	−0.689508
haTZ (180)	−300.202077	−0.840519	−0.844137	−0.867192
haQZ (330)	−300.224732	−0.905371	−0.903564	−0.929013
ha5Z (546)	−300.230791	−0.932136	—	—
ha6Z (840)	−300.231389	−0.944120	—	—
CBS	−300.232512[b]	−0.960582[c]	—	—

[a]haXZ denotes cc-pVXZ for H and aug-cc-pVXZ for F. Number of basis functions in parentheses.
[b]Obtained by fitting the haDZ–ha6Z data to Eq. [41].
[c]Obtained by applying Eq. [43] to the ha5Z and ha6Z MP2 correlation energies.

Step 1 in the tutorial associated with this section is to reproduce the SCF and correlation energies in Table 5. However, this step is optional because several of the computations require a good deal of time and resources. As such, it may not be worth the effort for most readers interested in the tutorial.

Step 2 in the tutorial, however, is more important and should not be considered optional. Readers should be able to reproduce the MP2 CBS limit by plugging the ha5Z and ha6Z data from Table 5 into Eq. [43]:

$$E_{MP2}^{CBS} = \frac{216(E_{MP2}^{ha6Z}) - 125(E_{MP2}^{ha5Z})}{91} \quad [46]$$

Similarly, you should be able to reproduce the SCF CBS limit by fitting all five SCF energies (haDZ–ha6Z) to Eq. [41]. This process is a bit more involved since it generally requires software capable of performing a nonlinear fit. Fortunately, a variety of freely available programs (including `gnuplot`[134]) can fit data to nonlinear equations.

The data in Table 5 have been used to compute the interaction energies (E_{int}) of $(HF)_3$ shown in Table 6. Here it is easier to achieve rapid convergence of higher order correlation effects. While the MP2 and CCSD(T) interaction energies continue to change appreciably as X increases, the *difference* between the two ($\delta_{MP2}^{CCSD(T)}$) converges very quickly to $\approx -1.5\,kJ\,mol^{-1}$. Thus the CCSD

Table 6 SCF, MP2, and CCSD(T) Interaction Energies (E_{int}) of $(HF)_3$ Obtained with a Series of Correlation-Consistent Basis Sets[a]

Basis set[b]	SCF	MP2	$\delta_{MP2}^{CCSD(T)}$	CCSD(T)
haDZ	−50.60	−60.89	−0.82	−61.71
haTZ	−47.30	−60.50	−1.37	−61.87
haQZ	−46.61	−59.94	−1.49	−61.44
ha5Z	−46.34	−59.75	—	—
ha6Z	−46.30	−59.71	—	—
CBS	−46.31[c]	−59.70[c]	[−1.49]	[−61.19]

[a]CCSD(T) corrections for higher order correlations effects are also reported relative to the MP2 values ($\delta_{MP2}^{CCSD(T)}$). All values are in kj mol^{-1}. Square brackets denote values obtained with the additive approximation described in the text.

[b]haXZ denotes cc-pVXZ for H and aug-cc-pVXZ for F.

[c]Obtained from data in Table 5.

(T) can be reasonably estimated by adding the $\delta_{MP2}^{CCSD(T)}$ correction obtained from smaller basis sets to the MP2 CBS limit of E_{int}:

$$E_{int}^{CCSD(T)/CBS} \approx E_{int}^{MP2/CBS} + \delta_{MP2}^{CCSD(T)} \quad [47]$$

For this example, the CCSD(T) correction of $-1.49\,kJ\,mol^{-1}$ obtained with the haQZ basis set is combined with the MP2 CBS limit to produce an estimate of $-61.19\,kJ\,mol^{-1}$ for the CCSD(T) CBS interaction energy of $(HF)_3$. Square brackets have been placed around these numbers in Table 6 to denote that they are based upon the additive approximation in Eq. [47] rather than an extrapolation of the correlation energy.

Step 3 in the tutorial associated with the section is to calculate the SCF and MP2 interaction energies (including the CBS values) in Table 6 from the energies given in Table 5.

Step 4 in the tutorial is to calculate the CCSD(T) E_{int} and $\delta_{MP2}^{CCSD(T)}$ values for the haDZ, haTZ, and haQZ basis sets.

Step 5 in the tutorial is to use Eq. [47] to estimate the CCSD(T) CBS limit of E_{int} for $(HF)_3$.

Accurate Potential Energy Surfaces

The systematic computational strategy outlined in this section of the review is necessary albeit demanding. The approach provides an accurate description of the entire spectrum of noncovalent interactions between fragments in a cluster. One can be confident in the calculated results regardless of cluster composition [i.e., whether examing the $(H_2O)_6$, $(C_6H_6)_2$, or a mixture of the two]. Less obviously but more importantly, one can also be confident in the calculated results across the entire (intermolecular) potential energy

surface (PES). The nature of the interactions not only depends on the identities of the fragments, but it is also highly sensitive to their separations and relative orientations in the cluster.[135]

The practical consequences of these dependencies can be severe. Consider, for example, the pairing of nucleic acid bases through both hydrogen bonding and stacking interactions.[136] If a particular quantum model chemistry does not properly describe both the dispersion interactions that play a major role in the latter configuration and the electrostatic interactions that dominate in the former orientation, qualitatively incorrect conclusions about the relative stability of stacked and hydrogen-bonded base pairs will be derived. The benzene dimer also illustrates this point nicely. The convergent approach outlined in this section has identified two isoenergetic, low-energy configurations on the PES, a T-shaped structure and a parallel displaced stacked structure.[137–139] Even at the CBS limit, the MP2 method overestimates the E_{int} for both structures. The more serious problem, however, is that the error is much larger for the stacked structure than for the parallel displaced structure because the nature of the interactions in the two configurations is different. By not considering high-order correlation effects, one arrives at the specious conclusion that the parallel displaced configuration is nearly 1.5 kcal mol^{-1} (6 kJ mol^{-1}) more stable than the T-shaped structure. Clearly, it is imperative that any computational strategy give consistent results across the entire PES, not just at one particular configuration (e.g., the global or a local minimum).

LESS DEMANDING COMPUTATIONAL STRATEGIES

The previous section outlined demanding computational procedures that provide the right answer for the right reason. Those convergent techniques[32,36] provide very accurate interaction energies across the entire PES for weakly bound clusters that can then be used as benchmarks[41,42,140–142] against which less demanding computational procedures may be measured. In this section, we review the performance of less demanding quantum model chemistries for different classes of weak noncovalent interactions, focusing on MP2 and DFT methods.

Second-Order Møller–Plesset Perturbation Theory

In general, second-order Møller–Plesset perturbation theory (a specific case of second-order many-body perturbation theory) is the workhorse of electronic structure techniques for weakly bound systems because the method tends to provide a reliable description of a wide range of weak interactions. For most hydrogen-bonding scenarios, MP2 energetics are extremely accurate and nearly identical to those from CCSD(T) computations with the same basis set. In fact, a recent study revealed that MP2 interaction energies obtained with an appropriate triple-ζ basis set agree favorably with CCSD(T) CBS

benchmark values.[143] Even for high-energy saddle points and structures with bifurcated hydrogen bonds, the deviations still tend to be less than a few tenths of a kilocalorie per mole per hydrogen bond.[118,119] Only for cyclic hydrogen-bonding motifs (like that found in the formic acid dimer), do MP2 interaction energies deviate substantially from CCSD(T) values.[143]

At the other end of the spectrum of weak interactions (Figure 1), the MP2 method can still provide a reasonable description of dispersion-bound clusters despite having a tendency to slightly overestimate the interaction energies between molecules relative to CCSD(T) results. For example, MP2 calculations yield interaction energies that are just a few tenths of a kilocalorie per mole larger than the CCSD(T) values for *n*-alkane dimers[142] and even some simple π-stacked dimers such as $(N_2)_2$ and $(C_2H_2)_2$.[120] If the π systems are delocalized, however, the MP2 errors can become massive (*vida infra*). Even the interactions between rare gas atoms are described reasonably well by the MP2 method.[144] However, MP2 tends to underbind the Ne_2 and He_2 by an amount that is small in an absolute sense but large in a relative sense, particularly in the case of He_2.

The MP2 method is not perfect, however, and the most notable (and fairly dramatic) failure of the MP2 method in the field of weak noncovalent interactions occurs for delocalized π stacking.[137–139] In fact, a separate chapter of this volume is dedicated to these π-type interactions.[49] The MP2 method overestimates dramatically the stability of "face-to-face" or "stacked"-type configurations relative to "edge-to-face" or "T-shaped" orientations that leads to a qualitatively incorrect description of delocalized π stacking as illustrated with the benzene dimer in described above. Consider the parallel-displaced stacked and T-shaped configurations of the diacetylene dimer $(H{-}C{\equiv}C{-}C{\equiv}C{-}H)_2$ (shown in Figure 7) along with the analogous configurations of the acetylene

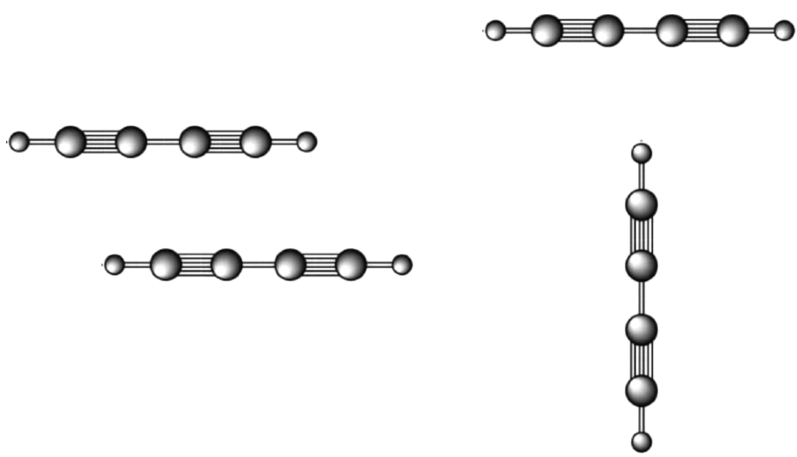

Figure 7 Parallel displaced and T-shaped configurations of the diacetylene dimer.

Table 7 MP2 and CCSD(T) Interaction Energies of the Stacked and T-shaped Dimers of Acetylene and Diacetylene[a]

Orientation	Dimer	CCSD(T)	MP2	$\delta_{MP2}^{CSD(T)}$
Stacked	$(H{-}C{\equiv}C{-}C{\equiv}C{-}H)_2$[b]	−1.31	−2.38	+1.07
T-shaped	$(H{-}C{\equiv}C{-}C{\equiv}C{-}H)_2$[b]	−1.63	−1.94	+0.31
Stacked	$(H{-}C{\equiv}C{-}H)_2$[c]	−1.72	−1.99	+0.17
T-shaped	$(H{-}C{\equiv}C{-}H)_2$[c]	−2.32	−2.45	+0.13

[a] All values are in kcal mol^{-1}.
[b] Reference 93
[c] Reference 120

dimer $(H{-}C{\equiv}C{-}H)_2$. In agreement with the "gold standard" of quantum chemistry [i.e., the CCSD(T) method], MP2 calculations correctly predict an attractive interactions in each case. (See data in Table 7.) However, for the dimers composed of fragments containing a delocalized π electron network, the MP2 method substantially overestimates the interaction energy (by more than 1 kcal mol^{-1}, which represents a relative error in excess of 80%) while the difference between the CCSD(T) and MP2 results ($\delta_{MP2}^{CCSD(T)}$) is modest for $(H{-}C{\equiv}C{-}H)_2$ (less than 0.17 kcal mol^{-1} or 10%). Matters are made even worse by the fact that the MP2 error is not uniform across the entire potential surface, which can lead to qualitatively incorrect conclusions about the nature of delocalized π-type interactions. With MP2 computations, one would conclude that the stacked configuration of $(H{-}C{\equiv}C{-}C{\equiv}C{-}H)_2$ is more stable than the T-shaped structure. However, the opposite (and correct) conclusion is reached when the effects of higher order excitations are included via CCSD (T) calculations.

Spin-Scaled MP2

Several approaches have been introduced that attempt to address this shortcoming of MP2 for delocalized π interactions. In 2003, Grimme introduced the spin-component-scaled second-order Møller–Plesset perturbation theory (SCS-MP2) method where the parallel spin (↑↑) and antiparallel spin (↑↓) pair correlation energies (which are related to the singlet and triplet components of the correlation energy) were assigned different weights.[145] The empirical scaling parameters ($p_{\uparrow\uparrow} = \frac{6}{5}$ and $p_{\uparrow\downarrow} = \frac{1}{3}$) improve MP2 results significantly not only for a variety of reaction energies and atomization energies but also for interaction energies between delocalized π systems (including the benzene dimer). The approach has no additional computational overhead relative to MP2, and it preserves the size consistency of the MP2 method. Other variations of these scaling parameters have since been introduced in an attempt to further improve the description of noncovalent interactions and/or reduce the computational demands.[146–149] Of particular interest is the SCSN-MP2 method of Platts and Hill in which the scaling parameters were optimized using

benchmark interaction energies for nucleic acid bases.[149] More recently, Takatani and Sherrill assessed the performance of several of these spin-scaled MP2 methods and found that the SCSN parameters reproduce nicely CCSD(T)/CBS interaction energies across the entire PES for a variety of dispersion-bound dimers.[150]

RI-MP2 and LMP2

Because the MP2 method has been so successful, there has also been a great deal of work done to reduce its computational demands, thereby further improving its price-to-performance ratio. These include schemes that employ localized orbitals (such as LMP2)[151–153] as well as the resolution of the identity approximation (RI), which is also known as density fitting (DF).[154–158] Werner, Manby, and Knowles have even introduced the DF-LMP2 method that incorporates both local approximations and density fitting.[159] (Those interested in more details about localization and RI techniques should consult their paper and the references contained therein.) Readers should be aware of a potential pitfall of some localization schemes. At certain points, the domain definitions can change, which leads to discontinuous PESs.[160] In large systems with many degrees of freedom, such discontinuities could lead to difficulties with geometry optimization procedures needed to locate minima or other stationary points on the PES. Fortunately, it is possible to construct local correlation models that are free of this problem.[161]

By reducing the overhead of MP2 computations substantially, these methods are helping to extend the domain of reliable electronic structure computations for noncovalent interactions to larger and larger systems (within the limits of the MP2 method). The RI-MP2 technique has been shown to yield structures and interaction energies that are virtually identical to canonical MP2 results when an extended basis set containing *f* functions is used while reducing the computational time by up to an order of magnitude.[44,162] The DF-LMP2 method has already been coupled with the aforementioned SCS approach of Grimme. The spin component scaled for nucleobases (SCSN) parameters were obtained from DF-LMP2 interaction energies.[149]

Density Functional Theory

As the previous millennium drew to a close, it became clear that density functional theory was not sufficiently reliable for the study of noncovalent interactions (including hydrogen bonding).[32,163–166] It was widely accepted at the time that the MP2 method and a high-quality triple-ζ (or larger) basis set were required to obtain chemically significant results for noncovalent interactions. Subsequently, a massive effort has been directed toward the development and calibration of new density functionals for noncovalent interactions.

Before proceeding, readers should be aware that caution must be exercised when drawing conclusions or making generalizations from a particular

systematic analysis (or calibration) of DFT methods, particularly in the case of weak interactions. Results depend heavily not only on the identity of the density functional[167] and the basis set[168,169] but also on the numerical integration grid, which differs from one quantum chemistry software package to the next (sometimes substantially).[170] Additionally, the criteria for a "reliable" DFT method are very different for someone trying to discern the energetic order of several low-lying minima than from someone merely trying to determine that a stable structure exists. While it is clear that some progress has been made, this section demonstrates that DFT methods are often reliable in the latter sense but rarely in the former sense.

For noncovalent interactions dominated by the electrostatic component of the interaction energy, it is likely that one of the well-established density functionals will produce reliable results. For example, it has been known for some time that DFT methods can provide an accurate description of charge–charge and charge–dipole interactions.[166,171,172] DFT methods can sometimes provide a reasonable description of hydrogen bonding when suitable basis sets are used,[39,166,173–177] especially for water.[178,179] Interestingly, despite the tendency of DFT interaction energy to converge rapidly to the 1-particle CBS limit (typically with aug-cc-pVTZ or aug-cc-pVQZ basis sets),[168] results are still sensitive to the type basis set used for the calculations.[169] The "best" functional tested with the correlation-consistent aug-cc-pVDZ and aug-cc-pVTZ basis sets will be far from the best when the split-valence Pople-style basis sets (e.g., 6-31++G*) are employed.

Some of the most disturbing news regarding the applicability of density functional methods to hydrogen bonding came in 2004 when Ireta, Neugebauer, and Scheffler noticed that, for the Perdew, Burke, Ernzerhof (PBE) functional, errors in E_{int} increase as hydrogen bonds deviate from linearity.[180] Shortly thereafter, Cybulski and Severson examined 12 popular functionals and reported a closely related observation that none can describe intermolecular PESs properly because they failed to reproduce the angular and distance dependencies of E_{int}.[181] A study from the author's lab noted that DFT methods can have problems characterizing the nature of transition states and higher order saddle points (i.e., the correct number of imaginary frequencies) for a system as simple as the water dimer.[182] Nevertheless, these studies suggest that the functionals can be used to identify minima in cases where electrostatic interactions dominate in that region of the PES.

Density functional theory methods should not be used when dispersion plays a significant role in the noncovalent interactions. Conventional DFT methods do not include dispersion, and the chemical consequences of this deficiency have been known for quite some time,[183,184] but the renewed interest in π-type interactions has fueled a great deal of work in this area.[141,185–188] A recent overview of these efforts can be found in Ref. 188. Of the three main ways to incorporate dispersion into DFT (empirical, nonempirical, and modification of existing functionals), the DFT method with an empirical

dispersion term (DFT-D) appears to be promising.[187–189] The DFT-D method, however, is still not nearly as reliable as high-accuracy WFT approaches such as those outlined above. While the nonempirical approaches (such as those of Becke[190–192]) are certainly appealing, they currently do not outperform DFT-D. The last approach, which attempts to include dispersion in DFT via modification of existing functionals, is by far the most popular of the three.[174,177,193–195] Unfortunately, it seems that modifications to the exchange functional leading to a successful description of dispersion interactions also happen to destroy any ability to reliably describe hydrogen bonding. When examining the performance of newly developed functionals, it is important to note that rare gas dimers are a poor model for dispersion-bound molecular clusters. Just because a functional gives reasonable results for rare gas dimers does not mean that similar performance can be expected for molecular dimers.

Guidelines

Although being an oversimplification, noncovalent interactions can be divided into three categories to help select less demanding computational procedures when studying a particular weakly bound cluster.

Category 1 (*Easy*) Strong noncovalent interactions that are dominated by the electrostatic component of the interaction energy tend to be fairly easy to compute. *MP2 will provide excellent results while most DFT methods will generally provide reliable results near minima on the PES.*

Category 2 (*Hard*) Interactions in which dispersion plays a non-negligible role tend to be more difficult to compute. *MP2 will provide reasonable results while conventional DFT methods will not even provide a qualitatively correct description of these interactions across the PES.*

Category 3 (*Problematic*) Dispersion interactions involving one or more delocalized π electron systems are exceptionally difficult to describe. *MP2 will overbind in a manner that is inconsistent across the PES, and conventional DFT methods still provide an unphysical description of the interactions.*

OTHER COMPUTATIONAL ISSUES

Basis Set Superposition Error and Counterpoise Corrections

As noted when introducing the Boys–Bernardi CP correction above, BSSE is a concern whenever the supermolecule method is used to compare the energies of fragments to the energy of the entire cluster [i.e., when computing the dissociation (D_e) or interaction (E_{int}) energy].

Although many arguments for and against CP corrections have been made over the years, this part of the chapter will briefly illustrate three important concepts related to this issue.

The CP procedure can lead to unphysical descriptions of PESs that are not easily corrected.

CP corrections often do not improve calculated interaction energies. In fact, the procedure can even make results worse.

BSSE is a poor diagnostic for the quality of a computed D_e or E_{int}. The crucial quantity is the basis set completeness error (BSCE).

While the CP procedure is useful when examining specific structures, it is sometimes desirable to examine a path across the PES or reaction profile that might include, for example, reactants, products, and a transition state (TS). Generating a CP-corrected potential energy curve or reaction profile can introduce a new set of problems in certain circumstances. As demonstrated by Lendvay and Mayer, the Boys–Bernardi CP procedure (and variations thereof) can actually produce discontinuities in the PES near the TS and even give different TS energies for the forward and reverse reactions.[196] Fortunately, this unphysical behavior is normally limited to regions near the TS.

To illustrate the last two points, consider the D_e of $(H_2O)_2$ and $(H_2O)_3$. The MP2 CBS limits have been determined from explicitly correlated MP2-R12 computations with the K2−− basis set to be 20.76 kJ mol^{-1} for the C_s global minimum of $(H_2O)_2$ and 66.12 kJ mol^{-1} for the cyclic C_1 global minimum of $(H_2O)_3$.[118,119,197] (The K2−− basis set is constructed by removing functions with the two highest angular momentum values from Klopper's K2 basis set,[198] which corresponds to *f* and *g* functions for H, and *g* and *h* functions for O.) Figures 8 and 9 show how CP-corrected and uncorrected MP2 calculations deviate from the corresponding CBS limit when using the aug-cc-pV*X*Z and haug-cc-pV*X*Z (diffuse functions only added to O atoms) families of basis sets. The height of a particular bar above the *x* axis indicates the basis set completeness error (BSCE) for that basis set while the combined height of a bar above and its CP-corrected counterpart below the *x* axis represents the magnitude of BSSE for that basis set. For both $(H_2O)_2$ and $(H_2O)_3$, the MP2 dissociation energies converge systematically to the CBS limit from below when the CP procedure is applied and from above when it is not.

Closer examination of the data presented in Figures 8 and 9 reveals that the errors associated with the CP-corrected dissociation energies are almost always larger than those that are not corrected for BSSE (i.e., the bars below the *x* axis are larger than those above it). Only for aug-cc-pVQZ, aug-cc-pV5Z, and aug-cc-pV6Z D_e values for $(H_2O)_2$ does the CP procedure offer any improvement (and merely on the order of 0.1 kJ mol^{-1}). In these particular cases, the CP corrections are clearly not worth the substantial additional effort.

Although these trends have been observed elsewhere,[133,199] they do not necessarily apply to all weakly bound complexes. Sinnokrot and Sherrill have

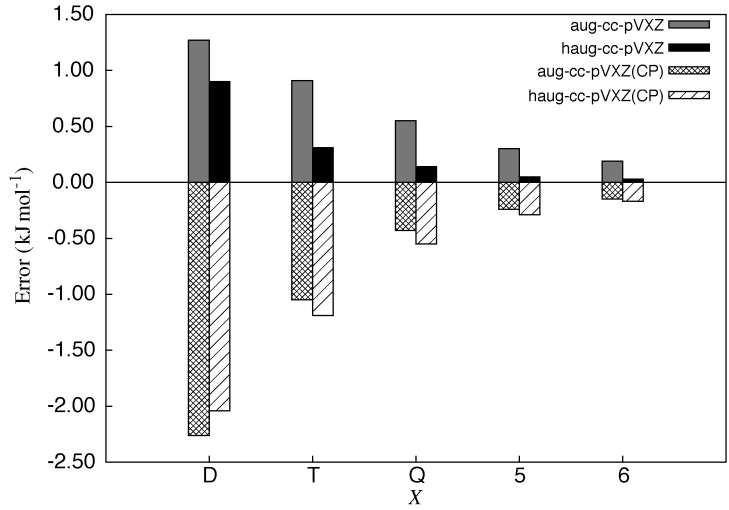

Figure 8 Errors relative to the MP2 CBS limit D_e for $(H_2O)_2$ using data from Ref. 118.

noted the opposite trend for benzene dimer interaction energies, and they report that the CP-corrected values converge to the CBS limit more rapidly than the uncorrected ones.[135]

The data in Figure 9 demonstrate nicely why BSSE should not be used to judge the quality of a particular D_e. The CP-corrected haug-cc-pVTZ D_e

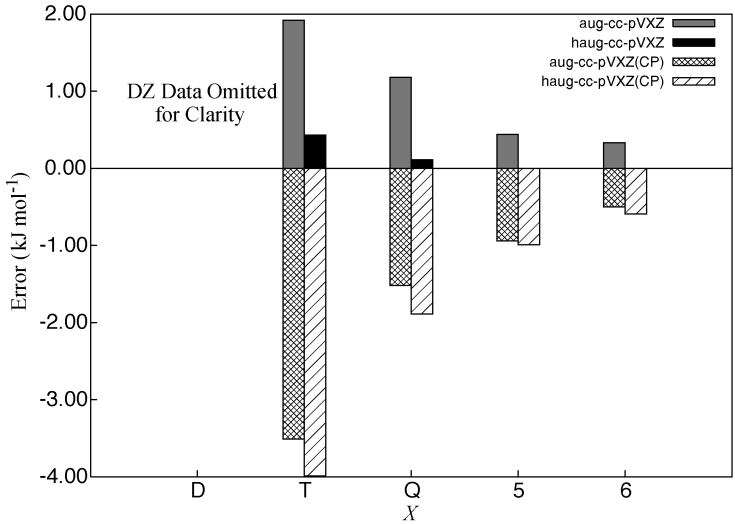

Figure 9 Errors relative to the MP2 CBS limit D_e for $(H_2O)_3$ using data from this work and Ref. 119. The errors with the double-ζ basis sets are so large that they have been omitted.

underestimates the MP2 CBS limit by 3.99 kJ mol^{-1} (striped bar) while the uncorrected value overestimates it by 0.43 kJ mol^{-1} (solid black bar). The difference between the two values, 4.42 kJ mol^{-1}, is the BSSE. If the MP2 CBS limit was not known, one would conclude that the haug-cc-pVTZ basis set is rather poor because of this large BSSE. However, the haug-cc-pVTZ is an excellent basis set for the D_e of the water trimer because the basis completeness error (BSCE) is only 0.43 kJ mol^{-1} (or $\approx$ 0.1 kcal mol^{-1}).

Beyond Interaction Energies: Geometries and Vibrational Frequencies

Care must be taken when interpreting computed equilibrium geometries and vibrational frequencies[44] when dealing with the extremely flat PESs of weakly bound clusters. For example, the vibrational frequencies associated with the large amplitude intermolecular motions are highly anharmonic. Consequently, estimates of the zero-point vibrational energy (ZPE or ZPVE) based on harmonic vibrational frequencies can be misleading. Furthermore, these harmonic frequencies for floppy vibrational modes can be sensitive to both the electronic structure method and basis set used, even for simple hydrogen bonding prototypes.[182,200]

In the case of equilibrium structures, the flat nature of the PESs can actually be an asset because it implies that the energy of a weakly bound cluster is fairly insensitive to changes in the intermolecular geometrical parameters. Consequently, MP2 optimizations with a sufficiently flexible triple-ζ basis set [e.g, TZ2P(f, d)+dif, aug-cc-pVTZ, TZVPP] usually provide sufficiently reliable structures for an accurate assessment of the interaction energy.[44,201] [Although "sufficiently flexible" is not a very precise description, it is generally agreed that the basis set needs to include two sets of polarization functions as well as at least one set of higher angular momentum functions (e.g., f functions for C, N, O, and F) and possibly diffuse functions.] However, the physical significance of these structures is not always clear since the ZPVE of the cluster may be larger than the barrier(s) separating minima on the PES.

Concluding Remarks

As with all areas of computational chemistry, the study of noncovalent interactions in weakly bound van der Waals clusters has benefited from the rapid improvements in hardware and software. As a result, high-accuracy benchmark databases that span the entire spectrum of weak interactions are now available.[42,177] Although relatively new, these collections of interaction energies have already been used to calibrate less demanding quantum model chemistries in the hopes of identifying methods that can be applied confidently to larger systems. However, a need exists for similar high-accuracy data for clusters (trimers, tetramers, etc.), not just dimers. Given the significant role

that cooperative effects can play in hydrogen-bonding networks, it is imperative that methods be able to describe the nonadditivity as well as the 2-body interactions. A consistent benchmark cluster database for the entire range of noncovalent interactions will play an essential role in the development of practical computational strategies for large clusters and explicit solvation models.

ACKNOWLEDGMENTS

I thank the current and former members of my research group at the University of Mississippi for their assistance and input during the preparation of this review, and I would also acknowledge the National Science Foundation for financial support (CHE-0517067, EPS-0132618).

REFERENCES

1. G. C. Pimentel and A. L. McClellan, *The Hydrogen Bond*, Freeman, San Francisco, 1960.
2. S. N. Vinogradov and R. H. Linnell, *Hydrogen Bonding*, Van Nostrand Reinhold, New York, 1971.
3. G. A. Jeffrey and W. Saenger, *Hydrogen Bonding in Biological Structures*, Springer, Berlin, 1991.
4. A. J. Stone, *The Theory of Intermolecular Forces*, Oxford University Press, Oxford, UK, 1996.
5. G. A. Jeffrey, *An Introduction to Hydrogen Bonding*, Oxford University Press, Oxford, UK, 1997.
6. C. Desfrançois, S. Carles, and J. Schermann, *Chem. Rev.*, **100**, 3943 (2000). Weakly Bound Clusters of Biological Interest.
7. V. A. Parsegian, *Van der Waals Forces*, Cambridge University Press, New York, 2006, page 10.
8. J. S. McDougal, M. S. Kennedy, J. M. Sligh, S. P. Cort, A. Mawle, and J. K. A. Nicholson, *Science*, **231**, 382 (1986). Binding of HLV-III/LAV to T4+ Cells by a Complex of the 110K Viral Protein and the T4 Molecule.
9. C. Borchers and K. B. Tomer, *Biochemistry*, **38**, 11734 (2000). Characterization of the Noncovalent Complex of Human Immunodeficiency Virus Glycoprotein 120 with Its Cellular Receptor CD4 by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry.
10. T. Peters, Ed., *All about Albumin: Biochemistry, Genetics and Medical Applications*, Academic, San Diego, 1996.
11. V. Vaida and J. Headrick, *J. Phys. Chem. A*, **104**, 5401 (2000). Physicochemical Properties of Hydrated Complexes in the Earth's Atmosphere.
12. W. Klemperer and V. Vaida, *Proc. Nat. Acad. Sci. U.S.A.*, **103**, 10584 (2006). Cluster Chemistry and Dynamics Special Feature: Molecular Complexes in Close and Far Away.
13. P. Schuster and P. Wolschann, *Monat. Chem.*, **130**, 947 (1999). Hydrogen Bonding: From Small Clusters to Biopolymers.
14. C.-D. Poon and E. T. Samulski, *J. Am. Chem. Soc.*, **122**, 5642 (2000). Do Bridging Water Molecules Dictate the Structure of a Model Dipeptide in Aqueous Solution?
15. A. Karshikoff, *Non-Covalent Interactions in Proteins*, Imperial College Press, London, 2006.
16. J. Černý and P. Hobza, *Phys. Chem. Chem. Phys.*, **9**, 5281 (2007). Non-Covalent Interactions in Biomacromolecules.
17. E. A. Meyer, R. K. Castellano, and F. Diederich, *Angew. Chem. Int. Ed. Engl.*, **42**, 1210 (2003). Interactions with Aromatic Rings in Chemical and Biological Recognition.

18. S. C. Zimmerman and P. S. Corbin, *Struct. Bonding*, **96**, 63 (2000). Heteroaromatic Modules for Self-Assembly Using Multiple Hydrogen Bonds.
19. S. H. M. Söontjens, R. P. Sijbesma, M. H. P. van Generen, and E. W. Meijer, *J. Am. Chem. Soc.*, **122**, 7487 (2000). Stability and Lifetime of Quadruply Hydrogen Bonded 2-Ureido-4 [1*H*]-pyrimidinone Dimers.
20. K. I. Hagen, C. M. Schwab, J. O. Edwards, J. G. Jones, R. G. Lawler, and D. A. Sweigart, *J. Am. Chem. Soc.*, **110**, 7024 (1988). Cobalt-59 NMR of Cobalt(III) Porphyrin Complexes. 2. Electric Field Gradients, d-Orbital Populations, and Hydrogen Bonding.
21. X. Xu and S. C. F. Au-Yeng, *J. Am. Chem. Soc.*, **122**, 6468 (2000). A DFT and ^{59}Co Solid-State NMR Study of the Chemical Shielding Property and Electronic Interaction in the Metalloporphyrin System.
22. W. Adam, K. Peters, E. Peters, and S. B. Schambony, *J. Am. Chem. Soc.*, **122**, 7610 (2000). Control through Hydrogen Bonding Mediated by the Urea Functionality of Chiral Auxilliaries.
23. K. N. Rankin, J. W. Gauld, and R. J. Boyd, *J. Am. Chem. Soc.*, **122**, 5384 (2000). Catalysis Mediated by Hydrogen Bonding: A Computational Study of the Aminolysis of 6-Chloropyrimidine.
24. D. Chandler, J. D. Weeks, and H. C. Andersen, *Science*, **220**, 787 (1983). Van der Waals Picture of Liquids, Solids, and Phase Transformations.
25. R. Ludwig, *Angew. Chem. Int. Ed. Engl.*, **40**, 1808 (2001). Water: From Clusters to the Bulk.
26. D. Hadži, Ed., *Theoretical Treatments of Hydrogen Bonding*, Wiley, Chichester, UK, 1997.
27. S. Scheiner, Ed., *Molecular Interactions: From van der Waals to Strongly Bound Complexes*, Wiley, Chichester, UK, 1997.
28. W. Gans and J. C. A. Boeyens, Eds., *Intermolecular Interactions*, Plenum, New York, 1998.
29. D. J. Wales, Ed., *Intermolecular Forces and Clusters I*, Vol. 115 of *Structure and Bonding*, Springer, Germany, 2005.
30. D. J. Wales, Ed., *Intermolecular Forces and Clusters II*, Vol. 116 of *Structure and Bonding*, Springer, Germany, 2005.
31. I. G. Kaplan, *Intermolecular Interactions: Physical Picture, Computational Methods and Model Potentials*, Wiley, Chichester, UK, 2006.
32. A. K. Rappe and E. R. Bernstein, *J. Phys. Chem. A*, **104**, 6117 (2000). Ab Initio Calculation of Nonbonded Interactions: Are We There Yet?
33. K. Müller-Dethlefs and P. Hobza, *Chem. Rev.*, **100**, 143 (2000). Noncovalent Interactions: A Challenge for Experiment and Theory.
34. O. Engkvist, P.-O. Åstrand, and G. Karlström, *Chem. Rev.*, **100**, 4087 (2000). Accurate Intermolecular Potentials Obtained from Molecular Wave Functions: Bridging the Gap between Quantum Chemistry and Molecular Simulations.
35. K. Kim, P. Tarakeshwar, and J. Lee, *Chem. Rev.*, **100**, 4145 (2000). Molecular Clusters of π-Systems: Theoretical Studies of Structures, Spectra, and Origin of Interaction Energies.
36. T. H. Dunning, *J. Phys. Chem. A*, **104**, 9062 (2000). A Road Map for the Calculation of Molecular Binding Energies.
37. C. E. Dykstra and J. M. Lisy, *J. Mol. Struct. (THEOCHEM)*, **500**, 375 (2000). Experimental and Theoretical Challenges in the Chemistry of Noncovalent Intermolecular Interaction and Clustering.
38. J. E. Del Bene and M. J. T. Jordan, *J. Mol. Struct. (THEOCHEM)*, **573**, 11 (2001). What a Difference a Decade Makes: Progress in Ab Initio Studies of the Hydrogen Bond.
39. P. Hobza, *Annu. Rep. Prog. Chem., Sect. C*, **100**, 3 (2004). Theoretical Studies of Hydrogen Bonding.
40. S. Tsuzuki, in *Intermolecular Forces and Clusters I*, Vol. 115 of *Structure and Bonding*, D. J. Wales, Ed., Springer, Berlin/Heidelberg Germany, 2005, pp. 149–193. Interactions with Aromatic Rings.

41. M. Sinnokrot and C. Sherrill, *J. Phys. Chem. A*, **110**, 10656 (2006). High-Accuracy Quantum Mechanical Studies of π–π Interactions in Benzene Dimers.

42. P. Jurečka, J. Šponer, J. Černý, and P. Hobza, *Phys. Chem. Chem. Phys.*, **8**, 1985 (2006). Benchmark Database of Accurate (MP2 and CCSD(T) Complete Basis Set Limit) Interaction Energies of Small Model Complexes, DNA Base Pairs, and Amino Acid Pairs.

43. S. Tsuzuki and T. Uchimaru, *Curr. Org. Chem.*, **10**, 745 (2006). Magnitude and Physical Origin of Intermolecular Interactions of Aromatic Molecules: Recent Progress of Computational Studies.

44. P. Hobza, R. Zahradnik, and K. Müller-Dethlefs, *Coll. Czech. Chem. Commun.*, **71**, 443 (2006). The World of Non-Covalent Interactions: 2006.

45. D. Feller and E. R. Davidson, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., VCH, New York, 1990, Vol. 1, pp. 1–44. Basis Sets for *Ab Initio* Molecular Orbital Calculations and Intermolecular Interactions.

46. S. Scheiner, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., VCH, New York, 1991, Vol. 2, pp. 165–218. Calculating the Properties of Hydrogen Bonds by *Ab Initio* Methods.

47. A. Castleman and K. Bowen, *J. Phys. Chem.*, **100**, 12911 (1996). Clusters: Structure, Energetics, and Dynamics of Intermediate States of Matter.

48. Z. Bačić and R. E. Miller, *J. Phys. Chem.*, **100**, 12945 (1996). Molecular Clusters: Structure and Dynamics of Weakly Bound Systems.

49. C. D. Sherrill, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and T. R. Cundari, Eds., Wiley-VCH, Hoboken, NJ, 2008, Vol. 26, pp. 1–38. Computations of Noncovalent π Interactions.

50. Y. Xie, R. Grev, J. Gu, H. Schaefer, P. Schleyer, J. Su, X.-W. Li, and G. Robinson, *J. Am. Chem. Soc.*, **120**, 3773 (1998). The Nature of the Gallium-Gallium Triple Bond.

51. T. Ghanty, V. Staroverov, P. Koren, and E. Davidson, *J. Am. Chem. Soc.*, **122**, 1210 (2000). Is the Hydrogen Bond in Water Dimer and Ice Covalent?

52. M. G. Del Popolo, C. Pinilla, and P. Ballone, *J. Chem. Phys.*, **126**, 144705 (2007). Local and Semilocal Density Functional Computations for Crystals of 1-Alkyl-3-methyl-imidazolium Salts.

53. P. N. Day, J. H. Jensen, M. S. Gordon, S. P. Webb, W. J. Stevens, M. Krauss, D. Garmer, and D. Cohen, *J. Chem. Phys.*, **105**, 1968 (1996). An Effective Fragment Method for Modeling Solvent Effects in Quantum Mechanical Calculations.

54. I. Adamovic and M. S. Gordon, *J. Phys. Chem. A*, **110**, 10267 (2006). Methanol-Water Mixtures: A Microsolvation Study Using the Effective Fragment Potential Method.

55. K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, *Chem. Phys. Lett.*, **313**, 701 (1999). Fragment Molecular Orbital Method: An Approximate Computational Method for Large Molecules.

56. D. G. Fedorov and K. Kitaura, *J. Phys. Chem. A*, **111**, 6904 (2007). Extending the Power of Quantum Chemistry to Large Systems with the Fragment Molecular Orbital Method.

57. R. A. Christie and K. D. Jordan, in *Intermolecular Forces and Clusters II*, Vol. 116 of *Structure and Bonding*, D. J. Wales, Ed., Springer, Berlin/Heidelberg Germany, 2005, pp. 27–41. *n*-Body Decomposition Approach to the Calculation of Interaction Energies of Water Clusters.

58. B. W. Hopkins and G. S. Tschumper, *Chem. Phys. Lett.*, **407**, 362 (2005). Integrated Computational Methods for Extended π Systems: Multicentered QM/QM Studies of the Cyanogen and Diacetylene Trimers.

59. G. S. Tschumper, *Chem. Phys. Lett.*, **427**, 185 (2006). Multicentered Integrated QM:QM Methods for Weakly Bound Clusters: An Efficient and Accurate 2-Body:Many-Body Treatment of Hydrogen Bonding and van der Waals Interactions.

60. A. M. ElSohly, C. L. Shaw, M. E. Guice, B. D. Smith, and G. S. Tschumper, *Mol. Phys.*, **105**, 2777 (2007). Analytic Gradients for the Multicentered Integrated QM:QM Method for Weakly Bound Clusters: Efficient and Accurate 2-Body:Many-Body Geometry Optimizations.

61. B. Jeziorski, R. Moszynski, and K. Szalewicz, *Chem. Rev.*, **94**, 1887 (1994). Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van der Waals Complexes.

62. B. Jeziorski and K. Szalewicz, in *Encyclopedia of Computational Chemistry*, P. v. R. Schleyer, Ed., Wiley, Chichester, UK, 1998, pp. 1376–1398. Intermolecular Interactions by Perturbation Theory.

63. B. Jeziorski and K. Szalewicz, in *Handbook of Molecular Physics and Quantum Chemistry*, S. Wilson, Ed., Wiley, Chichester, UK, 2003, Vol. 3, pp. 232–279. Symmetry-Adapted Perturbation Theory.

64. K. Szalewicz, K. Patkowski, and B. Jeziorski, in *Intermolecular Forces and Clusters II*, Vol. 116 of *Structure and Bonding*, D. J. Wales, Ed., Springer, Berlin/Heidelberg Germany, 2005, pp. 44–117. Intermolecular Interactions via Perturbation Theory: From Diatoms to Biomolecules.

65. R. Moszynski, P. E. S. Wormer, B. Jeziorski, and A. van der Avoird, *J. Chem. Phys.*, **103**, 8058 (1995). Symmetry-Adapted Perturbation Theory of Nonadditive Three-Body Interactions in van der Waals Molecules. I. General Theory.

66. R. Moszynski, P. E. S. Wormer, T. G. A. Heijmen, and A. van der Avoird, *J. Chem. Phys.*, **108**, 579 (1998). Symmetry-Adapted Perturbation Theory of Nonadditive Three-Body Interactions in van der Waals Molecules. II. Application to the Ar_2–HF Interaction.

67. R. Moszynski, P. E. S. Wormer, B. Jeziorski, and A. van der Avoird, *J. Chem. Phys.*, **107**, 672 (1997). Erratum: Symmetry-Adapted Perturbation Theory of Nonadditive Three-Body Interactions in van der Waals Molecules. I. General Theory [J. Chem. Phys. **103**, 8058 (1995)].

68. T. J. Dick and J. D. Madura, in *Annual Reports in Computational Chemistry*, D. C. Spellmeyer, Ed., Elsevier, Amsterdam, 2005, Vol. 1, pp. 59–74. A Review of the TIP4P, TIP4P-Ew, TIP5P, and TIP5P-E Water Models.

69. A. J. Stone, *Science*, **315**, 1228 (2007). Water from First Principles.

70. R. Bukowski, K. Szalewicz, G. C. Groenenboom, and A. van der Avoird, *Science*, **315**, 1249 (2007). Predictions of the Properties of Water from First Principles.

71. A. DeFusco, D. Schofield, P. Daniel, and K. Jordan, *Mol. Phys.*, **105**, 2681 (2007). Comparison of Models with Distributed Polarizable Sites for Describing Water Clusters.

72. J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids*, Wiley, New York, 1954.

73. H. Margenau and N. R. Kestner, *Theory of Intermolecular Forces*, Pergamon, Oxford, UK, 1971.

74. G. Chatasiński and M. M. Szczęśniak, *Chem. Rev.*, **94**, 1723 (1994). Origins of Structure and Energetics of van der Waals Clusters from Ab Initio Calculations.

75. J. D. van der Waals, 1873. Ph.D. Thesis, Leiden University, Leiden, Netherlands, 1873, *Over de Continuïteit van den Gas- en Vloeistoftoestand (On the Continuity of the Gas and Liquid State)*.

76. W. H. Keesom, *Physik. Z.*, **22**, 129 (1921). van der Waals Attractive Force.

77. W. H. Keesom, *Physik. Z.*, **23**, 225 (1922). Die Berechnung der Moleckularen Quadrupolmomente aus der Zustandgleichung.

78. P. Debye, *Physik. Z.*, **21**, 178 (1920). van der Waals' Cohesion Forces.

79. P. Debye, *Physik. Z.*, **22**, 302 (1921). Molecular Forces and Their Electrical Interpretation.

80. H. Falckenhagen, *Physik. Z.*, **23**, 87 (1922). Kohäsion and Zustandsgleichung bei Dipolgasen.

81. F. London, *Z. Phys. Chem. (B)*, **11**, 222 (1930). Über einige Eigenschaften und Anwendungen der Molekularkräfte.

82. F. London, *Trans. Faraday Soc.*, **33**, 8 (1937). The General Theory of Molecular Forces.

83. W. Heitler and F. London, *Z. Phys.*, **44**, 455 (1927). Interaction of Neutral Atoms and Homopolar Binding According to the Quantum Mechanics.

84. H. B. G. Casimir, *Proc. Nederl. Akad. Wetensch.*, **B51**, 793 (1948). On the Attraction between Two Perfectly Conducting Plates.

85. K. Morokuma and L. Pedersen, *J. Chem. Phys.*, **48**, 3275 (1968). Molecular-Orbital Studies of Hydrogen Bonds. An *Ab Initio* Calculation for Dimeric H_2O.

86. T. L. Gilbert and A. C. Wahl, *J. Chem. Phys.*, **47**, 3425 (1967). Single-Configuration Wavefunctions and Potential Curves for the Ground States of He_2, Ne_2, and Ar_2.

87. R. Eisenschitz and F. London, *Z. Physik*, **60**, 491 (1930). The Relation between the van der Waals Forces and the Homeopolar Valence Forces.

88. F. London, *Z. Physik*, **63**, 245 (1930). Theory and Systematics of Molecular Forces.

89. J. H. van Lenthe, J. G. C. M. van Duijneveldt-van de Rijdt, and F. B. van Duijneveldt, *Adv. Chem. Phys.*, **69**, 521 (1987). Weakly Bonded Systems.

90. L. Pauling, *The Nature of the Chemical Bond and The Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry*, 3 ed., Cornell University Press, Ithaca, NY, 1960 (first edition published in 1939), page 12.

91. C. Maerker, P. v. R. Schleyer, K. Liedl, T. -K. Ha, M. Quack, and M. A. Suhm, *J. Comput. Chem.*, **18**, 1695 (1997). A Critical Analysis of Electronic Density Functionals for Structural, Energetic, Dynamic and Magnetic Properties of Hydrogen Fluoride Clusters.

92. Example input files for NWChem, MPQC, PSI3, ACES2 and Gaussian03 are available at http://quantum.chem.olemiss/Tutorials.

93. B. W. Hopkins, A. M. ElSohly, and G. S. Tschumper, *Phys. Chem. Chem. Phys.*, **9**, 1550 (2007). Reliable Structures and Energetics for Two New Delocalized $\pi\cdots\pi$ Prototypes: Cyanogen Dimer and Diacetylene Dimer.

94. N. R. Kestner, *J. Chem. Phys.*, **48**, 252 (1968). He-He Interaction in the SCF-MO Approximation.

95. B. Liu and A. D. McLean, *J. Chem. Phys.*, **59**, 4557 (1973). Accurate Calculation of the Attractive Interaction of Two Ground State Helium Atoms.

96. N. R. Kestner and J. E. Combariza, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 1999, Vol. 13, pp. 99–132. Basis Set Superposition Errors: Theory and Practice.

97. H. B. Jansen and P. Ros, *Chem. Phys. Lett.*, **3**, 140 (1969). Non-Empirical Molecular Orbital Calculations on the Protonation of Carbon Monoxide.

98. S. F. Boys and F. Bernardi, *Mol. Phys.*, **19**, 553 (1970). The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors.

99. G. S. Tschumper, unpublished work.

100. T. P. Tauer and C. D. Sherrill, *J. Phys. Chem. A*, **109**, 10475 (2005). Beyond the Benzene Dimer: An Investigation of the Additivity of π–π Interactions.

101. A. Karpfen, in *Molecular Interactions: From van der Waals to Strongly Bound Complexes*, S. Scheiner, Ed., Wiley, New York, 1997, pp. 265–296. Case Studies in Cooperativity in Hydrogen-Bonded Clusters and Polymers.

102. A. Karpfen, *Adv. Chem. Phys.*, **123**, 469 (2002). Cooperative Effects in Hydrogen Bonding.

103. D. Hankins, J. W. Moskowitz, and F. H. Stillinger, *J. Chem. Phys.*, **53**, 4544 (1970). Water Molecule Interactions.

104. S. S. Xantheas, *J. Chem. Phys.*, **100**, 7523 (1994). Ab Initio Studies of Cyclic Water Clusters $(H_2O)_n$, $n = 1 - 6$. II. Analysis of Many-Body Interactions.

105. M. Quack and M. A. Suhm, in *Conceptual Perspectives in Quantum Chemistry*, J. -L. Calais and E. S. Kryachko, Eds., Kluwer, Dordrecht, 1997, Vol. III, pp. 415–464. Potential Energy Hypersurfaces for Hydrogen Bonded Clusters $(HF)_N$.

106. J. A. Pople, in *Energy, Structure, and Reactivity*, D. W. Smith and W. B. McRae, Eds., Wiley, New York, 1973, pp. 51–61. Theoretical Models for Chemistry.

107. R. J. Bartlett, *Ann. Rev. Phys. Chem.*, **32**, 359 (1981). Many-Body Perturbation Theory and Coupled Cluster Theory for Electron Correlation in Molecules.

108. T. D. Crawford and H. F. Schaefer, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 2000, Vol. 14, pp. 33–136. An Introduction to Coupled Cluster Theory for Computational Chemists.

109. J. C. Slater, *Phys. Rev.*, **32**, 349 (1928). The Normal State of Helium.

110. K. Morokuma and J. R. Winick, *J. Chem. Phys.*, **52**, 1301 (1970). Molecular-Orbital Studies of Hydrogen Bonds. Dimeric H_2O with the Slater Minimal Basis.

111. P. A. Kollman and L. C. Allen, *J. Chem. Phys.*, **52**, 5085 (1970). Theory of the Hydrogen Bond: *Ab Initio* Calculations on Hydrogen Fluoride Dimer and the Mixed Water-Hydrogen Fluoride Dimer.

112. R. J. Bartlett and J. F. Stanton, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 1994, Vol. 5, pp. 65–169. Applications of Post-Hartree-Fock Methods: A Tutorial.

113. K. Raghavachari and J. B. Anderson, *J. Phys. Chem.*, **100**, 12960 (1996). Electron Correlation Effects in Molecules.

114. D. P. Tew, W. Klopper, and T. Helgaker, *J. Comput. Chem.*, **28**, 1307 (2007). Electron Correlation: The Many-Body Problem at the Heart of Chemistry.

115. P. O. Löwdin, *Adv. Chem. Phys.*, **2**, 207 (1959). Correlation Problem in Many-Electron Quantum Mechanics. I. Review of Different Approaches and Discussion of Some Current Ideas.

116. M. L. Leininger, W. D. Allen, H. F. Schaefer, and C. D. Sherrill, *J. Chem. Phys.*, **112**, 9213 (2000). Is Møller–Plesset Perturbation Theory a Convergent *Ab Initio* Method?

117. J. Olsen, O. Christiansen, H. Koch, and P. Jørgensen, *J. Chem. Phys.*, **105**, 5082 (1996). Surprising Cases of Divergent Behavior in Møller-Plesset Perturbation Theory.

118. G. S. Tschumper, M. L. Leininger, B. C. Hoffman, E. F. Valeev, H. F. Schaefer, and M. Quack, *J. Chem. Phys.*, **116**, 690 (2002). Anchoring the Water Dimer Potential Energy Surface with Explicitly Correlated Computations and Focal Point Analyses.

119. J. A. Anderson, K. Crager, L. Fedoroff, and G. S. Tschumper, *J. Chem. Phys.*, **121**, 11023 (2004). Anchoring the Potential Energy Surface of the Cyclic Water Trimer.

120. B. W. Hopkins and G. S. Tschumper, *J. Phys. Chem. A*, **108**, 2941 (2004). *Ab Initio* Studies of $\pi \cdots \pi$ Interactions: The Effects of Quadruple Excitations.

121. T. H. Dunning, *J. Chem. Phys.*, **90**, 1007 (1989). Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen.

122. R. A. Kendall, T. H. Dunning, and R. J. Harrison, *J. Chem. Phys.*, **96**, 6796 (1992). Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions.

123. D. E. Woon and T. H. Dunning, *J. Chem. Phys.*, **98**, 1358 (1993). Gaussian Basis Sets for Use in Correlated Molecular Calculations. III. The Atoms Aluminum through Argon.

124. D. Feller, *J. Chem. Phys.*, **96**, 6104 (1992). Application of Systematic Sequences of Wave Functions to the Water Dimer.

125. D. Feller, *J. Chem. Phys.*, **98**, 7059 (1993). The Use of Systematic Sequences of Wave Functions for Estimating the Complete Basis Set, Full Configuration Interaction Limit in Water.

126. T. Helgaker, W. Klopper, H. Koch, and J. Noga, *J. Chem. Phys.*, **106**, 9639 (1996). Basis-Set Convergence of Correlated Calculations on Water.

127. E. F. Valeev, W. D. Allen, R. Hernandez, C. D. Sherrill, and H. F. Schaefer, *J. Chem. Phys.*, **118**, 8594 (2003). On the Accuracy Limits of Orbital Expansion Methods: Explicit Effects of *k*-Functions on Atomic and Molecular Energies.

128. J. M. L. Martin, *Chem. Phys. Lett.*, **259**, 669 (1996). Ab Initio Total Atomization Energies of Small Molecules—Towards the Basis Set Limit.

129. A. Halkier, T. Helgaker, W. Klopper, P. Jørgensen, and A. G. Császár, *Chem. Phys. Lett.*, **310**, 385 (1999). Comment on "Geometry Optimization with an Infinite Basis Set" [*J. Phys. Chem. A*, **103**, 651 (1999)] and "Basis-Set Extrapolation" [*Chem. Phys. Lett.*, **294**, 45 (1998)].

130. D. Bakowies, *J. Chem. Phys.*, **127**, 084105 (2007). Extrapolation of Electron Correlation Energies to Finite and Complete Basis Set Targets.

131. W. Klopper, F. R. Manby, S. Ten-no, and E. F. Valeev, *Int. Rev. Phys. Chem.*, **25**, 427 (2006). R12 Methods in Explicitly Correlated Molecular Electronic Structure Theory, page 427.

132. C. Ochsenfeld, J. Kussmann, and D. S. Lambrecht, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and T. R. Cundari, Eds., Wiley-VCH, Hoboken, NJ, 2007, Vol. 23, pp. 1–82. Linear-Scaling Methods in Quantum Chemistry.

133. A. Boese, J. Martin, and W. Klopper, *J. Phys. Chem. A*, **111**, 11122 (2007). Basis Set Limit Coupled Cluster Study of H-Bonded Systems and Assessment of More Approximate Methods.

134. http://www.134.info.

135. M. Sinnokrot and C. Sherrill, *J. Phys. Chem. A*, **108**, 10200 (2004). Highly Accurate Coupled Cluster Potential Energy Curves for the Benzene Dimer: Sandwich, T-Shaped, and Parallel-Displaced Configurations.

136. P. Jurecka and P. Hobza, *J. Am. Chem. Soc.*, **125**, 15608 (2003). True Stabilization Energies for the Optimal Planar Hydrogen-Bonded and Stacked Structures of Guanine···Cytosine, Adenine···Thymine, and Their 9- and 1-Methyl Derivatives: Complete Basis Set Calculations at the MP2 and CCSD(T) Levels and Comparison with Experiment.

137. W. Klopper, H. P. Luthi, T. Brupbacher, and A. Bauder, *J. Chem. Phys.*, **101**, 9747 (1994). Ab Initio Computations Close to the One-Particle Basis Set Limit on the Weakly Bound van der Waals Complexes Benzene–Neon and Benzene–Argon.

138. P. Hobza, H. Selzle, and E. Schlag, *J. Phys. Chem.*, **100**, 18790 (1996). Potential Energy Surface for the Benzene Dimer. Results of Ab Initio CCSD(T) Calculations Show Two Nearly Isoenergetic Structures: T-Shaped and Parallel-Displaced.

139. M. Sinnokrot, E. Valeev, and C. Sherrill, *J. Am. Chem. Soc.*, **124**, 10887 (2002). Estimates of the Ab Initio Limit for π–π Interactions: The Benzene Dimer.

140. W. Klopper, M. Quack, and M. A. Suhm, *Mol. Phys.*, **94**, 105 (1998). Explicitly Correlated Coupled Cluster Calculations of the Electronic Dissociation Energies and Barriers to Concerted Hydrogen Exchange of $(HF)_n$ Oligomers ($n = 2, \ldots, 5$).

141. Y. Zhao and D. Truhlar, *J. Chem. Theo. Comput.*, **1**, 415 (2005). Benchmark Databases for Nonbonded Interactions and Their Use to Test Density Functional Theory.

142. S. Tsuzuki, K. Honda, T. Uchimaru, and M. Mikami, *J. Chem. Phys.*, **124**, 114304 (2006). Estimated MP2 and CCSD(T) Interaction Energies of n-Alkane Dimers at the Basis Set Limit: Comparison of the Methods of Helgaker et al. and Feller.

143. K. Riley and P. Hobza, *J. Phys. Chem. A*, **111**, 8257 (2007). Assessment of the MP2 Method, along with Several Basis Sets, for the Computation of Interaction Energies of Biologically Relevant Hydrogen Bonded and Dispersion Bound Complexes.

144. T. Van Mourik, A. K. Wilson, and T. H. Dunning, *Mol. Phys.*, **96**, 529 (1999). Benchmark Calculations with Correlated Molecular Wavefunctions. XIII. Potential Energy Curves for He_2, Ne_2 and Ar_2 Using Correlation Consistent Basis Sets through Augmented Sextuple Zeta.

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

146. Y. Jung, R. C. Lochan, A. D. Dutoi, and M. Head-Gordon, *J. Chem. Phys.*, **121**, 9793 (2004). Scaled Opposite-Spin Second Order Møller-Plesset Correlation Energy: An Economical Electronic Structure Method.

147. R. Lochan, Y. Jung, and M. Head-Gordon, *J. Phys. Chem. A*, **109**, 7598 (2005). Scaled Opposite Spin Second Order Møller-Plesset Theory with Improved Physical Description of Long-Range Dispersion Interactions.

148. J. G. Hill, J. A. Platts, and H.-J. Werner, *Phys. Chem. Chem. Phys.*, **8**, 4072 (2006). Calculation of Intermolecular Interactions in the Benzene Dimer Using Coupled-Cluster and Local Electron Correlation Methods.

149. J. Hill and J. Platts, *J. Chem. Theo. Comput.*, **3**, 80 (2007). Spin-Component Scaling Methods for Weak and Stacking Interactions.

150. T. Takatani and C. D. Sherrill, *Phys. Chem. Chem. Phys.*, **9**, 6106 (2007). Performance of Spin-Component-Scaled Møller–Plesset Theory (SCS-MP2) for Potential Energy Curves of Noncovalent Interactions.

151. S. Saebø and P. Pulay, *Ann. Rev. Phys. Chem.*, **44**, 213 (1993). Local Treatment of Electron Correlation.

152. M. Schütz, G. Hetzer, and H.-J. Werner, *J. Chem. Phys.*, **111**, 5691 (1999). Low-Order Scaling Local Electron Correlation Methods. I. Linear Scaling Local MP2.

153. S. Saebø, in *Computational Chemistry. Review of Current Trends*, J. Lecszsynski, Ed., World Scientific, NJ, 2002, Vol. 7, pp. 63–87. Low-Scaling Methods for Electron Correlation.

154. M. Feyereisen, G. Fitzgerald, and A. Komornicki, *Chem. Phys. Lett.*, **208**, 359 (1993). Use of Approximate Integrals in Ab Initio Theory. An Application in MP2 Energy Calculations.

155. O. Vahtras, J. Almlöf, and M. Feyereisen, *Chem. Phys. Lett.*, **213**, 514 (1993). Integral Approximations for LCAO-SCF Calculations.

156. D. E. Bernholdt and R. J. Harrison, *Chem. Phys. Lett.*, **294**, 477 (1996). Large-Scale Correlated Electronic Structure Calculations: The RI-MP2 Method on Parallel Computers.

157. F. Weigend and M. Häser, *Theor. Chim. Acta*, **97**, 331 (1996). RI-MP2: First Derivatives and Global Consistency.

158. F. Weigend, M. Häser, H. Patzelt, and R. Ahlrichs, *Chem. Phys. Lett.*, **294**, 143 (1998). RI-MP2: Optimized Auxiliary Basis Sets and Demonstration of Efficiency.

159. H.-J. Werner, F. R. Manby, and P. J. Knowles, *J. Chem. Phys.*, **118**, 8149 (2003). Fast Linear Scaling Second-Order Møller-Plesset Perturbation Theory (MP2) Using Local and Density Fitting Approximations.

160. N. J. Russ and T. D. Crawford, *J. Chem. Phys.*, **121**, 691 (2004). Potential Energy Surface Discontinuities in Local Correlation Methods.

161. J. E. Subotnik and M. Head-Gordon, *J. Chem. Phys.*, **123**, 064108 (2005). A Local Correlation Model That Yields Intrinsically Smooth Potential-Energy Surfaces.

162. P. Jurečka, P. Nachtigall, and P. Hobza, *Phys. Chem. Chem. Phys.*, **3**, 4578 (2001). RI-MP2 Calculations with Extended Basis Set: A Promising Tool for Study of H-Bonded and Stacked DNA Base Pairs.

163. J. Del Bene and I. Shavitt, in *Molecular Interactions: From van der Waals to Strongly Bound Complexes*, S. Scheiner, Ed., Wiley, Chichester, UK, 1997, pp. 157–180. The Quest for Reliability in Calculated Properties of Hydrogen-Bonded Systems.

164. H. Guo, S. Sirois, E. I. Proynov, and D. R. Salahub, in *Theoretical Treatments of Hydrogen Bonding*, D. Hadži, Ed., Wiley, Chichester, UK, 1997, pp. 49–74. Density Functional Theory and Its Applications to Hydrogen-Bonded Systems.

165. J. Del Bene, in *Encyclopedia of Computational Chemistry*, P. v. R. Schleyer, Ed., Wiley, Chichester, UK, 1998, pp. 1263–1271. Hydrogen Bonding: 1.

166. A. D. Rabuck and G. E. Scuseria, *Theor. Chim. Acta*, **104**, 439 (2000). Performance of Reccently Developed Kinetic Energy Density Functionals for the Calculation of Hydrogen Binding Strengths and Hydrogen Bonded Structures.

167. J. Van de Vondele, F. Mohamed, M. Krack, J. Hutter, M. Sprik, and M. Parrinello, *J. Chem. Phys.*, **122**, 014515 (2005). The Influence of Temperature and Density Functional Models in Ab Initio Molecular Dynamics Simulation of Liquid Water.

168. G. Csonka, A. Ruzsinszky, and J. Perdew, *J. Phys. Chem. B*, **109**, 21471 (2005). Proper Gaussian Basis Sets for Density Functional Studies of Water Dimers and Trimers.

169. K. Riley, B. Op'tHolt, and K. Merz, *J. Chem. Theo. Comput.*, **3**, 407 (2007). Critical Assessment of the Performance of Density Functional Methods for Several Atomic and Molecular Properties.

170. B. N. Papas and H. F. Schaefer, *J. Mol. Struct. (THEOCHEM)*, **768**, 175 (2006). Concerning the Precision of Standard Density Functional Programs: Gaussian, Molpro, NWChem, Q-Chem, and Gamess.

171. A. Pudzianowski, *J. Phys. Chem.*, **100**, 4781 (1996). A Systematic Appraisal of Density Functional Methodologies for Hydrogen Bonding in Binary Ionic Complexes.

172. M. Meot-Ner(Mautner), *Chem. Rev.*, **105**, 213 (2005). The Ionic Hydrogen Bond.

173. S. Sirois, E. I. Proynov, D. T. Nguyen, and D. R. Salahub, *J. Chem. Phys.*, **107**, 6770 (1997). Hydrogen-Bonding in Glycine and Malonaldehyde: Performance of the Lap1 Correlation Functional.

174. Y. Zhao and D. Truhlar, *J. Phys. Chem. A*, **108**, 6908 (2004). Hybrid Meta Density Functional Theory Methods for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions: The MPW1B95 and MPWB1K Models and Comparative Assessments for Hydrogen Bonding and van der Waals Interactions.

175. J. Sponer, P. Jurecka, and P. Hobza, *J. Am. Chem. Soc.*, **126**, 10142 (2004). Accurate Interaction Energies of Hydrogen-Bonded Nucleic Acid Base Pairs.

176. J. Černý and P. Hobza, *Phys. Chem. Chem. Phys.*, **7**, 1624 (2005). The X3LYP Extended Density Functional Accurately Describes H-Bonding But Fails Completely for Stacking.

177. Y. Zhao, N. Schultz, and D. Truhlar, *J. Chem. Theo. Comput.*, **2**, 364 (2006). Design of Density Functionals by Combining the Method of Constraint Satisfaction with Parametrization for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions.

178. J. Su, X. Xu, and W. Goddard, *J. Phys. Chem. A*, **108**, 10518 (2004). Accurate Energies and Structures for Large Water Clusters Using the X3LYP Hybrid Density Functional.

179. E. Dahlke and D. Truhlar, *J. Phys. Chem. B*, **109**, 15677 (2005). Improved Density Functionals for Water.

180. J. Ireta, J. Neugebauer, and M. Scheffler, *J. Phys. Chem. A*, **108**, 5692 (2004). On the Accuracy of DFT for Describing Hydrogen Bonds: Dependence on the Bond Directionality.

181. S. M. Cybulski and C. E. Seversen, *J. Chem. Phys.*, **122**, 014117 (2005). Critical Examination of the Supermolecule Density Functional Theory Calculations of Intermolecular Interactions.

182. J. A. Anderson and G. S. Tschumper, *J. Phys. Chem. A*, **110**, 7268 (2006). Characterizing the Potential Energy Surface of the Water Dimer with DFT: Failures of Some Popular Functionals for Hydrogen Bonding.

183. S. Kristyán and P. Pulay, *Chem. Phys. Lett.*, **229**, 175 (1994). Can (Semi)local Density Functional Theory Account for the London Dispersion Forces?

184. P. Hobza, J. Šponer, and T. Reschel, *J. Comput. Chem.*, **16**, 1315 (1995). Density-Functional Theory and Molecular Clusters.

185. E. R. Johnson, R. A. Wolkow, and G. A. DiLabio, *Chem. Phys. Lett.*, **394**, 334 (2004). Application of 25 Density Functionals to Dispersion-Bound Homomolecular Dimers.

186. Y. Zhao and D. Truhlar, *J. Phys. Chem. A*, **110**, 5121 (2006). Comparative DFT Study of van der Waals Complexes: Rare-Gas Dimers, Alkaline-Earth Dimers, Zinc Dimer, and Zinc-Rare-Gas Dimers.

187. C. Morgado, M. A. Vincent, I. H. Hillier, and X. Shan, *Phys. Chem. Chem. Phys.*, **9**, 448 (2007). Can the DFT-D Method Describe the Full Range of Noncovalent Interactions Found in Large Biomolecules?

188. P. Jurečka, J. Černý, P. Hobza, and D. R. Salahub, *J. Comput. Chem.*, **28**, 555 (2007). Density Functional Theory Augmented with an Empirical Dispersion Term. Interaction Energies and Geometries of 80 Noncovalent Complexes Compared with Ab Initio Quantum Mechanics Calculations.

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

190. E. R. Johnson and A. D. Becke, *J. Chem. Phys.*, **123**, 024101 (2005). A Post-Hartree-Fock Model of Intermolecular Interactions.

191. E. R. Johnson and A. D. Becke, *J. Chem. Phys.*, **124**, 174104 (2006). A Post-Hartree-Fock Model of Intermolecular Interactions: Inclusion of Higher-Order Corrections.

192. A. D. Becke and E. R. Johnson, *J. Chem. Phys.*, **127**, 124108 (2007). A Unified Density-Functional Treatment of Dynamical, Nondynamical, and Dispersion Correlations.

193. A. D. Boese and N. C. Handy, *J. Chem. Phys.*, **114**, 5497 (2001). A New Parametrization of Exchange-Correlation Generalized Gradient Approximation Functionals.

194. A. D. Boese, A. Chandra, J. M. L. Martin, and D. Marx, *J. Chem. Phys.*, **119**, 5965 (2003). From Ab Initio Quantum Chemistry to Molecular Dynamics: The Delicate Case of Hydrogen Bonding in Ammonia.

195. X. Xu and W. A. Goddard, *Proc. Natl. Acad. Sci. USA*, **101**, 2673 (2004). The X3LYP Extended Density Functional for Accurate Descriptions of Nonbond Interactions, Spin States, and Thermochemical Properties.

196. F. Weigend, M. Häser, H. Patzelt, and R. Ahlrichs, *Chem. Phys. Lett.*, **297**, 365 (1998). RI-MP2: Optimized Auxiliary Basis Sets and Demonstration of Efficiency.

197. Additional computations performed for this work.

198. W. Klopper, *J. Chem. Phys.*, **102**, 6168 (1995). Limiting Values for Møller-Plesset Second-Order Correlation Energies of Polyatomic Systems: A Benchmark Study on Ne, HF, H_2O, N_2, and He$\cdots$He.

199. M. Masamura, *Theor. Chem. Acc.*, **106**, 301 (2001). The Effect of Basis Set Superposition Error on the Convergence of Interaction Energies.

200. G. S. Tschumper, M. D. Kelty, and H. F. Schaefer, *Mol. Phys.*, **96**, 493 (1999). Subtle Basis Set Effects on Hydrogen Bonded Systems.

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