FEATURE ARTICLE

A Road Map for the Calculation of Molecular Binding Energies

Thom H. Dunning, Jr.

Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory,[†]
Richland, Washington 99352

Received: April 18, 2000; In Final Form: July 20, 2000

During the past decade dramatic progress has been made in calculating the binding energies of molecules. This is the result of two advances reported in 1989: an accurate method for solving the electronic Schrödinger equation that is applicable to a broad range of molecules—the CCSD(T) method—and families of basis sets that systematically converge to the complete basis set limit—the correlation consistent basis sets. The former provides unprecedented accuracy for the prediction of a broad range of molecular properties, including molecular binding energies. The latter provides a means to systematically approach the complete basis set limit, i.e., the exact solutions of approximations to the Schrödinger equation. These two advances combined with a thorough analysis of the errors involved in electronic structure calculations lead to clear guidelines for ab initio calculations of binding energies, ranging from the strong bonds derived from chemical interactions to the extremely weak binding due to dispersion interactions. This analysis has also led to surprises, e.g., it has shown that the Møller−Plesset perturbation theory is unsuitable for calculation of bond energies to chemical accuracy, i.e., with errors of less that 1 kcal/mol. This applies whether one is interested in *absolute* bond energies or *relative* bond energies. Although the analysis presented here is focused on the calculation of molecular binding energies, this same approach can be readily extended to other molecular properties.

1. Introduction

The concept of chemical bonds and the determination of bond energies are central to chemistry. The making and breaking of chemical bonds in molecules governs the behavior of many processes important to our modern world, from the production of energy and pollutants in an automobile engine to the catalytic processes that convert raw materials into materials of value to society. Weaker molecular interactions are also important. Hydrogen bonds play a critical role in a wide range of chemical processes, especially biochemical processes. Both inter- and intramolecular forces determine the properties of polymers, and a wide range of materials has been developed by varying these interactions in a systematic manner. Obtaining a detailed understanding of molecular interactions and molecular binding energies is one of chemistry's "Grand Quests."

With the discovery of the mathematical equation governing the behavior of atoms and molecules in the mid-1920s—the Schrödinger equation—the pathway was opened for calculating molecular binding energies from first principles. In fact, physicists immediately set about computing the binding energy of H_2 with great success. This work not only provided evidence supporting the radical new quantum mechanics, but was the first successful prediction of a chemical bond energy. Unfortunately, what was possible for H_2 was not possible for other molecules, and, as far as the rest of chemistry was concerned, the comment by P. A. M. Dirac in 1929[1] held:

> "The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble."

Despite this difficulty, scientists such as Pauling, Mulliken, and others used the framework provided by quantum mechanics to discover the general laws which govern the structure and energetics of molecules. This work had an enormous impact on chemistry and later led to the award of Nobel Prizes to these two individuals.

In his classic book, *The Nature of the Chemical Bond*, Pauling[2] stated that "there is a chemical bond between two atoms or groups of atoms in 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 system." This definition is still valid today, although what is considered an "independent system" is much different than in Pauling's time. Through the development of sophisticated synthetic techniques and sensitive measurement technologies, experimental physical chemists have prepared and characterized a wide range of weakly bound molecular complexes. This is nowhere better illustrated than in the recent report by Giese, Gentry, and co-workers[3] (see also ref 4) of the synthesis and characterization of the helium dimer—a molecule that is bound by only 1 milliKelvin (0.7 cm^{-1}, 0.002 kcal/mol). This work by experimental physical chemists has greatly increased our understanding of the full range of molecular interactions.

[†] The Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute under Contract No. DE-AC06-76RLO 1830.

10.1021/jp001507z CCC: $19.00 © 2000 American Chemical Society
Published on Web 09/15/2000

In the present article, we will consider all of the types of interactions that can give rise to a stable molecule or molecular complex. Specifically, four types of interactions will be considered.

Chemical interactions, which result from the intimate sharing of electrons between the two atoms involved in the bond. Chemical bonds represent the strongest molecular interactions, and their strengths vary from tens to hundreds of kcal/mol, although most fall in the range of 75 to 150 kcal/mol.

Hydrogen-bond interactions, which result from the sharing of a hydrogen atom between two other atoms. The strengths of hydrogen bonds vary from a few kcal/mol to tens of kcal/mol and are typically weaker than chemical bonds by a factor of 10 or more.

Electrostatic interactions, which arise from the classical electrostatic interactions between the multipole moments of molecules. In these systems, which we shall collectively refer to as "weakly bound molecules," binding energies range from a fraction of a kcal/mol to a few kcal/mol, another order of magnitude weaker than hydrogen bonds.

Dispersion interactions, which result from the instantaneous correlation between the fluctuations in the electronic charge clouds of the two interacting systems. Binding energies arising solely from dispersion interactions range from hundredths of a kcal/mol to a few tenths of a kcal/mol. We shall refer to these systems as "very weakly bound molecules."

As can be seen, molecular binding energies vary from hundredths of a kcal/mol to hundreds of kcal/mol, a variation of more than 4 orders of magnitude! The challenge to the quantum chemist, and it is a daunting one, is to develop theoretical and computational approaches that are capable of accurately describing molecular binding energies over this wide range of values.

Although solving the Schrödinger equation is still a challenging task, the development of sophisticated computational methods for solving the equation, coupled with an exponential growth in computing power over the past two decades, has made it possible to compute interaction energies for a broad range of molecules. These capabilities can be found in a number of quantum chemistry codes, including GAUSSIAN,[5] CADPAC,[6] GAMESS,[7] MOLPRO,[8] ACES II,[9] and DALTON.[10] For small molecules, the accuracy of the calculations now rivals that obtained from experiment for some molecular properties. For large molecules, such calculations often serve as invaluable guides to rationalizing the experimental data. Of particular importance is the fact that the mathematical calculations can be extended to classes of molecules, e.g., radicals and ions, whose existence in the laboratory may be too ephemeral to permit accurate measurements. This is important not only for obtaining a quantitative understanding of many chemical processes, e.g., combustion or plasma processing, but it also provides chemists with a more complete picture of the nature of the chemical bond.

Despite the increasingly important role that quantum chemistry calculations play in modern chemical research, molecular calculations often appear to be a "black art." It is difficult for the novice, if not the expert, to confidently thread his or her way through the bewildering array of choices presented by modern-day computational approaches to solving the Schrödinger equation. What method should be used to solve the equation? There are more than a dozen ab initio methods for solving the Schrödinger equation, not counting density functional theory with its large list of exchange and correlation functionals. What basis set should be used in the calculation? There are hundreds of basis sets in the literature, some of general use, others specialized for calculations of one property or another. As we will show, this situation no longer exists. Because of advances in the past decade, it is now possible to provide a clear road map for calculating accurate molecular binding energies. Although the focus of this article is on the calculation of binding energies, the approach used here applies to other molecular properties as well.[11] The same analysis is also applicable to electronic structure methods not explicitly considered here. Thus, this article presents a prototype case study, the essence of which could be repeated for many other molecular properties and computational methods.

In the following section we will discuss the sources of errors in the solution of the electronic Schrödinger equation. It is critical to be clear about this, because the errors can be of different sign and, when they are, their cancellation can lead to confusing results (as we shall see). In this article, we will focus on two very popular techniques for solving the electronic Schrödinger equation: Møller−Plesset perturbation theory and coupled cluster theory. In the third section, we review the essential features of these two methods. Basis set expansion techniques are used to solve the Schrödinger equation and, so, in the fourth section, we discuss the selection of basis sets for molecular calculations. The use of basis sets to convert the Schrödinger equation into a more readily soluble algebraic equation has been a real boon in quantum chemistry. However, it has also been the source of many problems. Finally, with all of this completed, we discuss the calculation of molecular binding energies in the fifth and sixth sections.

In this article, we will focus on the solution of the electronic Schrödinger equation and, therefore, will be concerned with D_e, the negative of the interaction energy at the equilibrium geometry $\{\mathbf{R}_e\}$ of the molecule. By considering D_e, we eliminate errors associated with the solution of the nuclear Schrödinger equation, although the need to extract D_e from the experimental data can lead to larger experimental uncertainties in D_e than in D_0 (because of uncertainties in the vibrational frequencies).

2. Errors in Electronic Structure Calculations

There are two sources of errors in the solution of the electronic Schrödinger equation: the *basis set convergence error* and the *electronic structure method error*. In practice these errors are often intermingled to the extent that it may not be clear, even to the experienced researcher, what the relative importance of the two types of errors is (this knowledge is usually gained from experience over the course of years of work). It is nonetheless essential to distinguish between these two types of errors if we are to understand the limitations of atomic and molecular calculations.

Assume that a family of basis sets is used to solve the electronic Schrödinger equation, where the members of the family are specified by the label "n" and that, as "n" increases, the basis set becomes more and more complete. For any given basis set "n," the basis set convergence error in D_e is

$$\Delta D_e^{\mathrm{bs}}(\mathrm{M};n) = D_e(\mathrm{M};n) - D_e(\mathrm{M};\infty) \tag{1}$$

where $D_e(\mathrm{M};n)$ is the value of the binding energy obtained with basis set "n" and method "M", and $D_e(\mathrm{M};\infty)$ is the value obtained with a complete basis set, CBS, i.e., as $n \rightarrow \infty$. In other words, $D_e(\mathrm{M};\infty)$ is the value of the binding energy obtained by exactly solving the Schrödinger equation using electronic structure method "M." Because of the ordering of the basis sets, $\Delta D_e^{\mathrm{bs}}(\mathrm{M};n)$ will decrease to zero as "n" increases. The detailed form of $\Delta D_e^{bs}(M;n)$ depends, of course, on the electronic structure method ("M") being used.

The basis set convergence error is different than the error arising from the use of a given electronic structure method "M" to solve the Schrödinger equation (Hartree−Fock, singles and doubles configuration interaction, second-order perturbation theory, etc.). The electronic structure method error for D_e is

$$\Delta D_e^{M} = D_e(M;\infty) - D_e(\text{expt'l}) \qquad (2)$$

where $D_e(M;\infty)$ is defined above and D_e(expt'l) is the experimental value of the binding energy. ΔD_e^M is also referred to as the *intrinsic error* in the binding energy for method "M." It is the error in the binding energy that would result if the Schrödinger equation was solved exactly using method "M"; it does not depend on the basis set (n).

To complete the list of error types involved in the solution of the electronic Schrödinger equation, we also need to define the error associated with a given calculation, i.e., a given choice of electronic structure method ("M") and basis set ("n"). The calculational error is given by

$$\Delta D_e^{calc}(M;n) = D_e(M;n) - D_e(\text{expt'l}) \qquad (3)$$

We sometimes refer to $\Delta D_e^{calc}(M;n)$ as the *apparent error* to clearly distinguish it from the *intrinsic error* for the method "M". Note that from eqs 1 and 2 the calculational error is simply the sum of the basis set and method error:

$$\Delta D_e^{calc}(M;n) = \Delta D_e^{bs}(M;n) + \Delta D_e^{M} \qquad (4)$$

If the basis set convergence error and the electronic structure method error have different signs, which, as we shall see below, they sometimes do, the calculational error can be less than the intrinsic error. This point is illustrated in Figure 1, which shows two common error types.[12] For Type I, the calculated D_e approaches the experimental value from below, never reaching the experimental value even for a complete basis set. In this case $\Delta D_e^{calc}(M;n)$ is always greater than ΔD_e^M. For Type II, on the other hand, the calculated D_e overshoots the experimental D_e for sufficiently large n. In this case, $\Delta D_e^{calc}(M;n)$ can be less than ΔD_e^M for some n. In fact, in the figure $\Delta D_e^{calc}(M;n) \approx 0$ for $n = 3$, leading the unsuspecting researcher to incorrectly conclude that method "M" and basis set $n = 3$ satisfactorily describe the binding in this molecule. Unfortunately, this situation is not rare, as will be illustrated by examples discussed in Section 5.

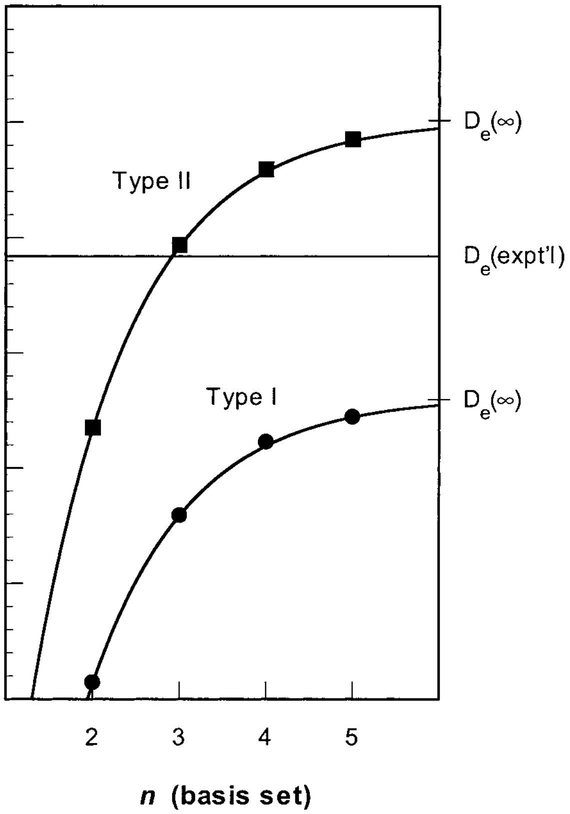

Figure 1. Type I and II errors for the calculation of molecular binding energies. $D_e(\infty)$ refers to the binding energy at the complete basis set limit.

3. Approximate Methods for Solving the Electronic Schrödinger Equation

Since the discovery of the Schrödinger equation in 1925−26, a number of approximate methods have been developed to solve the electronic Schrödinger equation for molecules. In this work we will focus on two very popular approaches: the Møller−Plesset perturbation theory (MP2, MP3, MP4, ...) and the coupled cluster methods (CCSD, CCSDT, ...). Unlike the Hartree−Fock wave function on which they are based, perturbation theory and coupled cluster methods take the detailed effects of electron correlation into account, a "must" for accurate molecular predictions. They are also size-extensive (see refs 13 and 14). This means that the definition of the computed interaction energies is unambiguous. We will not discuss configuration interaction methods in this article. "Less than full" CI methods do not usually provide size-extensive descriptions of the interactions between closed shell molecules, although it should be noted that the carefully designed CI methods of Liu and McLean[15] (or generalizations thereof[16]) do provide a means of using CI techniques to address problems such as those discussed here.

3.1. Perturbation Theory Methods. In Møller−Plesset perturbation theory, it is assumed that electron correlation is a perturbation to the Hartree−Fock Hamiltonian, i.e.,

$$H = H_0 + \lambda H_1 \qquad (5)$$

where H_0 is the Hartree−Fock Hamiltonian and λH_1, the perturbation, is the difference between the Hartree−Fock averaged interelectronic interaction and the exact $\Sigma\Sigma^1/r_{ij}$ interaction in the full Hamiltonian. With the partitioning in eq 5, the wave function and energy can also be written as a power series in λ

$$\Psi = \Psi_0 + \lambda\Psi_1 + \lambda^2\Psi_2 + ... \qquad (6)$$

$$E = E_0 + \lambda E_1 + \lambda^2 E_2 + \lambda^3 E_3 + \lambda^4 E_4 + ... \qquad (7)$$

and the wave function and energy are computed order by order. For sufficiently small perturbations, only Ψ_0 (the Hartree−Fock wave function) and Ψ_1 (the first-order wave function) are important. For larger perturbations, Ψ_2, Ψ_3, ... must also be taken into account.

A wave function through nth order in perturbation theory is sufficient to calculate the energy to $(2n+1)$th order. Thus, $E_0 + E_1$ is computed by taking the expectation value of the Hamiltonian, eq 5, over the Hartree−Fock wave function and is just the Hartree−Fock energy. Similarly, Ψ_0 and Ψ_1 determine both E_2(MP2) and E_3(MP3); Ψ_0, Ψ_1, and Ψ_2

determine E_4(MP4) and E_5(MP5); etc. Because H_1 contains at most two-electron operators, Ψ_1 just contains double excitations (D) relative to the Hartree–Fock wave function and E_2 and E_3 only take double excitations into account. In E_2 the contributions of the double excitations to the energy are independent of one another; in E_3, they are coupled. The second-order correction to the wave function, Ψ_2, contains single, double, triple, and quadruple (SDTQ) excitations, and so E_4 and E_5 take single, triple, and quadruple excitations as well as double excitations into account.

The computational cost of perturbation theory calculations increases rapidly with the order of perturbation theory and the size of the basis set. If the number of functions in the basis set is N, formal analysis of the perturbation theory equations shows that second-order perturbation (MP2) theory scales as N^5, third-order perturbation (MP3) theory as N^6, and fourth-order perturbation (MP4) theory as N^7. Most correlated molecular calculations reported to date have used MP2 theory, which corrects for many of the major deficiencies in the Hartree–Fock method. With the increasing capabilities offered by modern computing technology, MP4 calculations have become more common. There have recently been implementations of both fifth- and sixth-order[17] perturbation theory in the Gaussian program. However, the steep dependence of these techniques on the size of the basis set (N^8 and N^9, respectively) has prevented their use in all but a restricted set of calculations.

There is one last issue to be addressed in this section—the question of the convergence of the perturbation theory expansion. Perturbation theory works well when the perturbation is small, and the theory has been successfully used in numerous areas in molecular quantum mechanics. It has long been recognized that the reliance of conventional perturbation theory on the Hartree–Fock Hamiltonian and wave function limits the radius of convergence of MP*n* expansions to those molecules for which a single configuration is a good approximation.[18,19] More recently, it has become apparent that convergence difficulties in the MP*n* expansion are not limited to molecules whose zero-order wave function is poorly described by a single configuration.[20−22] These latter authors found that such classical single configuration systems as hydrogen fluoride, with a weight of approximately 95% for the Hartree–Fock wave function in the MP*n* expansion, can give rise to expansions that are either slowly converging or divergent. The physical reason for this behavior has yet to be firmly established, but Cremer and He[17] note that it appears to be associated with the presence of multiple sets of lone pairs in the same region of space (for a mathematical analysis of the problem, see refs 20(b) and 23). We will see evidence of this problem in the MP*n* calculations discussed here, although it is less prominent in the calculation of D_e than in calculation of other molecular properties.[21]

3.2. Coupled Cluster Methods. The coupled cluster method is a latecomer to electronic structure theory. It was first developed to treat the nuclear problem in the late 1950s by Coester and Kümmel,[24] adapted by Cizek[25] for quantum chemical applications, and popularized by Bartlett and coworkers.[26−30] In coupled cluster theory, the wave function is given by

$$\Psi = e^{T}\,\Psi_0 \tag{8}$$

where T is a sum of excitation operators

$$T = t_1 + t_2 + t_3 + \ldots \tag{9}$$

In eq 9, t_1 generates all possible single excitations when it operates on Ψ_0, t_2 generates all possible double excitations, etc. Truncation of eq 9 at $t_1 + t_2$ gives the CCSD method, truncation at $t_1 + t_2 + t_3$ gives the CCSDT method, and so on. Because of the exponential form of the excitation operator, the coupled cluster wave function contains not only all of the excitations included in T, but all of the products of excitations included in T. For example, for the CCSD method the wave function is

$$\Psi = \left(1 + t_1 + \frac{1}{2!}t_1t_1 + t_2 + \frac{1}{3!}t_1t_1t_1 + t_1t_2 + \frac{1}{4!}t_1t_1t_1t_1 + \frac{1}{2!}t_1t_1t_2 + \frac{1}{2!}t_2t_2 + \ldots\right)\Psi_0 \tag{10}$$

where the sum extends to the maximum number of excitations allowed. Thus, the CCSD wave function contains the effects of triple excitations that are products of three single excitations or a single excitation times a double excitation (called disconnected triple excitations); disconnected quadruple excitations that are products of four single excitations, two single excitations times a double excitation, or two double excitations; and so on. The inclusion of higher-order excitations in Ψ is very important because, as Sinanoglu[31] first observed, the largest component of higher-order excitations are often products of lower-order excitations. Thus, for most molecules, t_2t_2 is the largest contributor to quadruple excitations. The connected quadruple excitations represented by t_4 are small because they require all four electrons to be in the same region of space.

There is a simple, appealing physical picture provided by the coupled cluster wave function. For a set of noninteracting electron pairs, say M molecules of H_2 at very large intermolecular separations, the CCSD wave function provides an exact description of the system. That is, the *connected* t_1 and t_2 excitations plus the *disconnected* (t_1t_1, t_1t_2, t_2t_2, etc.) excitations included in the CCSD wave function are all that are required to exactly describe the set of M noninteracting electron pairs. As the separation between the electron pairs decreases, the charge clouds begin to overlap and the contribution of the higher order *connected* excitations increases: first t_3 becomes important (three electrons in the same region of space), and then t_4 (four electrons in the same region of space), and so on. Thus, to the extent that molecules are composed of well-localized electron pairs and are well described by the Hartree–Fock wave function, the CCSD wave function should provide an accurate description of their electronic structure. In He_2, van Mourik and Dunning[32] have shown that the contribution of both t_3 and t_4 to D_e increases exponentially with decreasing R and that $\Delta D_e(t_1{+}t_2) \gg \Delta D_e(t_3) \gg \Delta D_e(t_4)$.

Unfortunately, it has been found that, for typical molecular geometries, the effects of connected triple excitations (t_3) must be explicitly included in the coupled cluster wave function to obtain high accuracy. The CCSDT method[29] scales as N^8 and thus is computationally very expensive.[33] This has led to a number of attempts to approximate the effect of connected triple excitations in the coupled cluster CCSDT wave function.[28,34] The CCSD(T) method of Raghavachari et al.,[34] which includes the effects of t_3 perturbationally, has been found to provide an excellent compromise between accuracy and computational cost[35,36] (for additional insight into the underpinnings of this approach, see ref 37). In fact, when used with large basis sets, the CCSD(T) method yields molecular binding energies and many other molecular properties that are comparable to those obtained from all but the most sophisticated experiments (see refs 38, 39, 40, and 41). The CCSD(T) method scales as N^6 for

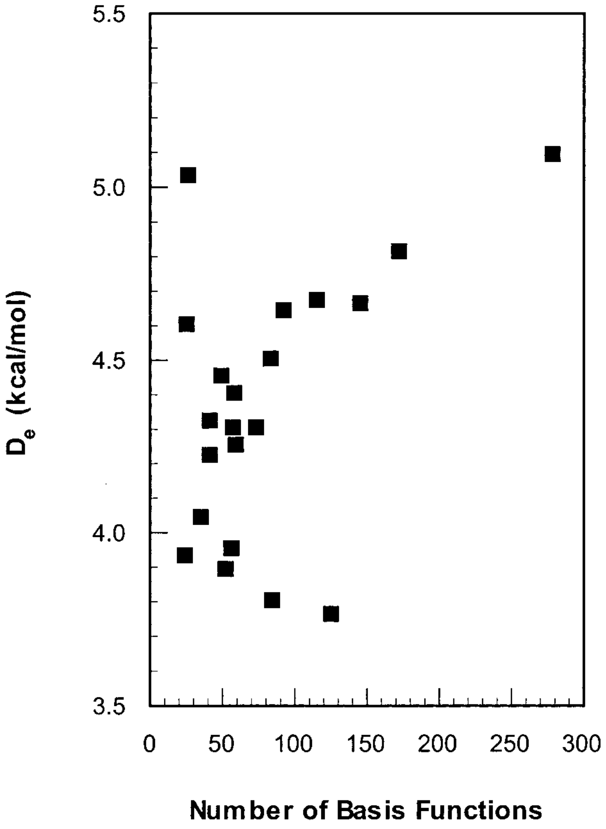

Figure 2. Values of D_e for the water dimer computed with the MP2 method and a number of popular basis sets. Calculated values have been corrected for basis set superposition error using the counterpoise correction.

solution of the CCSD equations plus an N^7 step for the perturbative calculation of the triple excitations.

4. Correlation Consistent Basis Sets

Since the first application of quantum mechanics to chemistry by Heitler and London,[42] basis sets have been used to convert the electronic Schrödinger equation into an algebraic equation. At first, the basis sets used were simple representations of the atomic orbitals and the method was referred to as the "linear combination of atomic orbitals," or LCAO, method. Later, as theoretical chemists began to employ electronic computers to solve the resulting equations, it became clear that a simple set of atomic orbitals did not provide the desired accuracy—a molecule, although composed of atoms, is composed of "deformed" atoms. More flexible atomic sets as well as higher angular momentum functions are required to represent the deformation and correlation effects. As a result, modern basis sets contain several functions of a given angular momentum, i.e., several *s*-functions, several *p*-functions, and so on, as well as functions of high angular momentum, e.g., the most accurate sets available today include up to *i*-functions.

The main problem with most of the basis sets in common use today is that they do not provide a well defined path to the complete basis set (CBS) limit. Although the total energy always decreases with the addition of functions to a basis set, many other properties, including molecular binding energies, do not show continuous, monotonic improvement. The general situation is not unlike that represented in Figure 2, which plots the binding energy of water dimer calculated with the MP2 method as a function of the basis set (labeled by the number of functions in the set). Although not a scatter plot, no clear convergence pattern is evident in the figure. Without a well-defined path to the CBS limit, it is not be possible to separate the contributions to the readily observed calculational errors. A hierarchy of basis sets, which provide a systematic approach to a complete basis set, would allow us to cleanly separate the electronic structure method error (ΔD_e^{M}) from the basis set convergence error [$\Delta D_e^{bs}(n)$] and enable us to better understand the source of errors in molecular calculations. The correlation consistent basis sets provide such a family.

4.1. Construction of Correlation Consistent Basis Sets.[43] Although the importance of the basis set in limiting the accuracy of electronic structure calculations on molecules was well recognized by the end of the 1960s, the principles involved in constructing basis sets for use in correlated molecular calculations were not understood. As a result, the sets used in most early correlated calculations were based on experience gained from Hartree−Fock calculations. The first advance came in 1987 when Almlöf and Taylor[44] showed that basis sets constructed from atomic natural orbitals (ANOs) provided accurate solutions of the molecular Schrödinger equation. Further, Almlöf and Taylor found that the occupation numbers of the ANOs fell into clearly defined groups. Using this fact, they constructed a hierarchy of contracted basis sets by adding groups of secondary natural orbitals with nearly equal occupation numbers to the principal natural orbitals. The accuracy of the calculations systematically improved as larger and larger ANO sets were used, up to the limit imposed by the set of primitive Gaussian functions used to represent the natural orbitals. This concept was extended by Widmark et al.[45] to atoms in the second row of the periodic chart and by Bauschlicher and co-workers to the third row,[46] including the transition metals[47] (see also Pou-Amerigo et al.[48]) The major problem with the ANO sets is efficiency. Any calculation that uses the sets has to compute integrals over all of the primitive functions included in the set. Since very large primitive sets are used in constructing the ANO sets, the cost of integral calculations can be prohibitive for all but the smallest ANO sets.

At the time that Almlöf and Taylor were exploring the use of ANO sets, Dunning[49] was carrying out a systematic study of the basis set requirements for describing correlation effects in atoms. Whereas Almlöf and Taylor chose the occupation numbers of the natural orbitals as a measure of the importance of ANOs to the wave function, Dunning chose the contribution that a basis function makes to the correlation energy. This choice had a significant side benefit: it moved the criterion from one based on orbitals, which are linear combinations of basis functions, to one based on the basis functions themselves, allowing a significant reduction in the number of primitive functions included in the sets. For the first row atoms Dunning found that the computed energy lowerings fell into distinct groups, e.g.,

$$\Delta E_{1,0}(sp) \approx \Delta E_{1,0}(d) \tag{11a}$$

$$\Delta E_{2,1}(sp) \approx \Delta E_{2,1}(d) \approx \Delta E_{1,0}(f) \tag{11b}$$

$$\Delta E_{3,2}(sp) \approx \Delta E_{3,2}(d) \approx \Delta E_{2,1}(f) \approx \Delta E_{1,0}(g) \tag{11c}$$

where $\Delta E_{m+1,m}(l)$ is the energy lowering associated with increasing the number of functions of angular momentum, l, from m to $m+1$ (the s and p functions were considered a single set). Further, the contributions of the three groups to the atomic correlation energy were well separated, e.g., for the oxygen atom the first group averaged (with negative and positive ranges as subscripts and superscripts, respectively) $-62.0_{0.0}^{0.0}$ mE_h, the

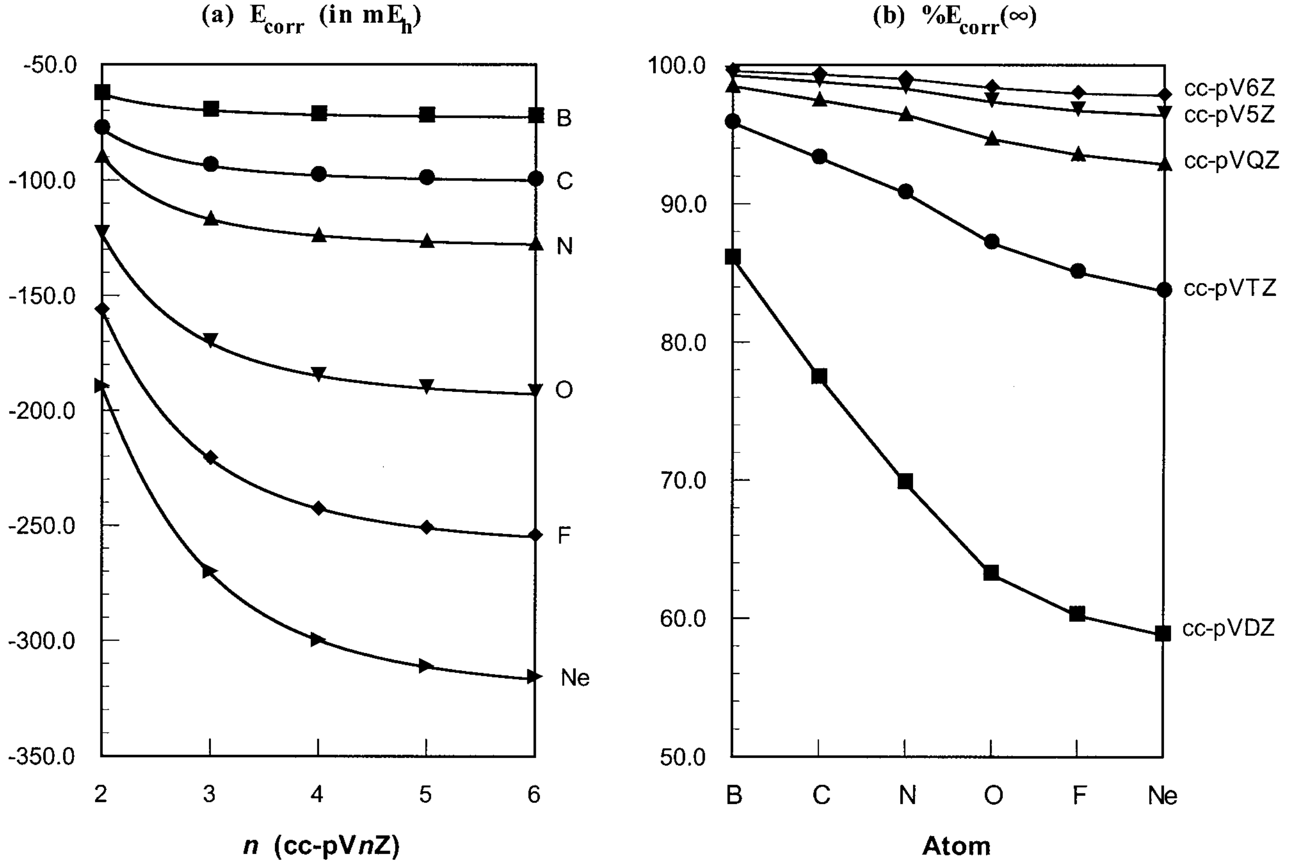

Figure 3. Calculation of the valence-electron correlation energy of the first-row atoms, B−Ne, using the CCSD(T) or RCCSD(T) method and the correlation-consistent basis sets. (a) Total valence-electron correlation energy (in millihartrees). (b) Fraction of the correlation energy recovered for each atom and basis set, i.e., $E_{corr}(n)/E_{corr}(\infty)$ (in percent).

second group averaged $-15.0_{-0.8}^{+0.5}$ mE_h, and the third group averaged $-3.5_{-0.6}^{+0.9}$ mE_h.

With the above information, the schema for constructing correlation consistent basis sets is clear. The smallest basis set, the cc-pVDZ set, is formed from the 1s, 2*s*, and 2*p* atomic HF orbitals plus a single *s*, *p*, and *d* primitive Gaussian function. This yields a [3*s*2*p*1*d*] set. The next basis set, the cc-pVTZ set, is constructed by adding an optimized set of two primitive *s*-, *p*-, and *d*-functions plus a single primitive *f*-function to the atomic HF orbitals, yielding a [4*s*3*p*2*d*1*f*] set, and so on. This is the same *aufbau* scheme found by Almlöf and Taylor,[44] except that primitive basis functions are added to the atomic Hartree−Fock orbitals, rather than secondary natural orbitals to the primary natural orbitals. This pattern results in a very systematic improvement of the basis set. The correlation consistent basis sets are collectively designated cc-pV*n*Z, where $n = 2$ represents the cc-pVDZ set; $n = 3$ the cc-pVTZ set; $n = 4$ the cc-pVQZ set; and so on.

Although the correlation consistent basis sets constructed in the above manner provide an excellent description of the neutral and singly positive ions of atoms, they do not provide an adequate description of negative ions. The wave functions for negative ions are substantially more diffuse than those for the corresponding neutral atoms, and basis sets must be tailored to describe anions. Kendall, Dunning, and Harrison[50] investigated the addition of extra functions to the correlation consistent sets for calculations on anions, explicitly optimizing the exponents of the additional functions for the negative ions. Based on these results, Kendall et al. constructed *augmented* correlation consistent basis sets, denoted aug-cc-pV*n*Z, by adding a single diffuse function to each angular momentum present in the standard set, i.e., a diffuse *s*-, *p*-, and *d*-function to the cc-pVDZ set, a diffuse *s*-, *p*-, *d*-, and *f*-function to the cc-pVTZ set, etc. In later studies of the electrical properties of molecules, Woon and Dunning[51] found that additional diffuse functions were required to describe the higher order electrical properties of atoms and molecules (polarizabilities, hyperpolarizabilities, etc.). They defined d(oubly)-aug-cc-pV*n*Z and t(riply)-aug-cc-pV*n*Z sets that were derived from the aug-cc-pV*n*Z sets by adding a second (and third) diffuse function for each angular symmetry present in the original sets. The designations for the correlation consistent sets are often abbreviated to v*nz* (cc-pV*n*Z), av*nz* (aug-cc-pV*n*Z), dav*nz* (d-aug-cc-pV*n*Z), etc.

Dunning and co-workers have used the above approach to construct correlation consistent basis sets up to $n = 6$, including basis sets for all-electron calculations, for the first-row atoms (boron through neon)[52] and hydrogen and helium, as well as similar sets for the second row atoms (aluminum through argon).[53] Correlation consistent sets have recently been published for the third-row atoms gallium through krypton.[54] Work has been essentially completed on sets for the alkali metals and is underway on the first-row transition metal atoms.

4.2. Extrapolation to the Complete Basis Set Limit. To illustrate the convergence behavior of the correlation consistent basis sets, consider use of the sets to calculate the valence correlation energies of the first-row atoms, B through Ne. To the extent that molecules are collections of perturbed atoms, these calculations provide invaluable insights into the convergence behavior of the correlation consistent sets in molecular calculations.

In Figure 3a, we plot the absolute values of the valence correlation energies of the first-row atoms, boron through neon, from CCSD(T) or RCCSD(T)[55] calculations with the cc-pVnZ basis sets. Note the smooth, monotonic increase in the magnitude of $E_{\mathrm{corr}}(n)$ with increasing basis set size. For the first three basis sets ($n = 2-4$), the magnitude of the correlation energy increases nearly exponentially with increasing n. This led early users of the correlation consistent sets to use a simple exponential function to extrapolate the results for $n = 2-4$ to the complete basis set limit.[49,56] However, this function tends to underestimate the contribution of the sets with $n > 4$. Drawing an analogy with the convergence of the correlation energy of the helium atom with angular momentum, Martin[39d] proposed the use of inverse powers of the maximum angular momentum in the set to estimate the error in the correlation energy caused by truncation of the basis set expansion (note that for the first-row atoms considered here $l_{\max} = n$ for the correlation consistent sets). In fact, the smooth curves plotted in Figure 3a are based on the expansion

$$E_{\mathrm{corr}}(n) = E_{\mathrm{a}} + E_{\mathrm{b}}/n^3 + E_{\mathrm{c}}/n^5 \tag{12}$$

The root-mean-square deviations of the solid line from the points are just 0.02 (B), 0.03 (C), 0.05 (N), 0.10 (O), 0.13 (F), and 0.15 mE_{h} (Ne). The predicted valence correlation energies at the CBS limit are −73.1 (B), −100.9 (C), −129.3 (N), −196.2 (O), −260.5 (F), and −324.4 mE_{h} (Ne). These latter values may be compared to estimates of the valence correlation energy of C, O, and Ne of −101.2, −195.5, and −322.8 mE_{h}, respectively, an early illustration of the accuracy of the CCSD(T) method. More recently, Halkier and co-workers[40] have reported good success with a simple $^1/_{n^3}$ extrapolation based on results from calculations with two successive basis sets, provided that cc-pVQZ or larger sets are employed, and Truhlar and co-workers[57] have shown that quite useful results are obtained with such a two point scheme based on the cc-pVDZ and cc-pVTZ sets. Despite the successes achieved in extrapolating calculated results to the complete basis set limit, none of the existing functions has proven to be clearly superior to all of the other functions (for a discussion of the issues, see ref 58).

In Figure 3b, we plot the percentage of E_{corr}[CCSD(T);∞] recovered by the cc-pVnZ sets for each of the first-row atoms, B through Ne. The poor performance of the cc-pVDZ set is obvious. Although the double-zeta set recovers over 85% of the correlation energy of the boron atom, the fraction of the correlation energy recovered drops dramatically along the row, becoming less than 60% for the neon atom. Thus, use of the cc-pVDZ set to calculate atomic and molecular properties that are sensitive to electron correlation effects is questionable, especially for molecules involving atoms from the latter half of the row. This behavior also compromises trends that involve changes in the number of electrons predicted using the cc-pVDZ basis set. Thus, despite its popularity, calculations with the double-zeta set must be viewed with caution.

The fraction of the correlation energy recovered with the cc-pVTZ set is not only much higher than that for the cc-pVDZ set, but its variation along the row from boron to neon is dramatically less—from 96% (B) to 86% (Ne). Thus, one would expect the triple-zeta set to provide a far more consistent description of the electronic structure of molecules and this is observed to be the case. In fact, the cc-pVTZ set should be considered the "minimal basis set" to be used in correlated calculations. Further expanding the basis set, one finds that the cc-pV5Z set nearly quantitatively recovers the valence correlation energy of the first row atoms, yielding 99% of the estimated CBS limit for boron to 96% of the limit for neon.

5. Electronic Structure Method Errors for Molecular Binding Energies

The electronic structure method error, or intrinsic error, for a molecular binding energy, ΔD_{e}(M), is given by eq 2: D_{e}(M;∞) − D_{e}(expt'l). Thus, to determine the error for electronic structure method "M," we must first establish the value of the binding energy at the complete basis set limit ($n = \infty$). As noted in the previous section, it is possible to extrapolate binding energies calculated with a sequence of correlation consistent basis sets to the CBS limit. The extrapolations can be quite accurate if sufficiently large basis sets are used. In the current study, we limit our considerations to small molecules for which calculations could be carried out with basis sets as large as d-aug-cc-pV6Z. Although some generality is lost by this limitation, our studies, as well as those of others, indicate that the conclusions drawn here are broadly applicable. Using this approach, it is expected that the error in the extrapolated bond energies is on the order of 0.1−0.2%, significantly smaller than the intrinsic errors in the methods themselves.

In the current study we limit our considerations to valence electron calculations. Valence electron calculations are generally capable of predicting molecular binding energies to an accuracy of 0.5−1%. However, in discussing the intrinsic errors in the calculated binding energies, we will, whenever possible, report the results of calculations that include all of the electrons to better understand the nature of the residual error in the valence electron calculations. It should be noted that in carrying out all-electron calculations it is important to use basis sets designed for such calculations (see, e.g., ref 52). Use of standard basis sets in all-electron calculations recovers only a small, often nonsystematic fraction of the core and core-valence correlation energy. Because of this, all-electron calculations with standard basis sets can produce misleading results.

In predicting the binding energies, calculations on all but the "chemically bound molecules" use the counterpoise method[59] to correct for basis set superposition error (BSSE[60]). This is critical. If this correction is not made, the "observed" basis set convergence error is actually a combination of the "true" basis set convergence error and the basis set superposition error; a point that is illustrated in Figure 4. Once BSSE has been eliminated, the remaining error is the basis set convergence error.[61] This is of more than conceptual importance. If BSSE is not eliminated, $\Delta D_{\mathrm{e}}^{\mathrm{bs}}(n)$ can be erratic and it may not be possible to extrapolate the results to the CBS limit.[62] Of course, in the limit of a complete basis set, BSSE is identically zero and, in fact, we find that the magnitude of the counterpoise correction decreases steadily with increasing n. It is interesting to note that van Mourik et al.[62] found that use of the counterpoise correction often led to a smoother dependence of D_{e} on n even for strongly bound molecules.

It is quite possible, and even probable, that binding energies computed without the counterpoise correction are closer to the complete basis set limit than the uncorrected values. This frustrating (or lucky) situation (depending on your proclivities) is due to the fact that BSSE corrections and basis set convergence errors are often of opposite sign.

5.1. Chemically bound Molecules. In Table 1, we list the experimental bond energies[63] and the intrinsic errors for the perturbation theory and coupled cluster methods for a prototypical set of first-row, diatomic molecules.[38,52] The set includes CH, a covalent, singly bonded molecule; HF, a very ionic, singly bonded molecule; N_2, a covalent, triply bonded, diatomic

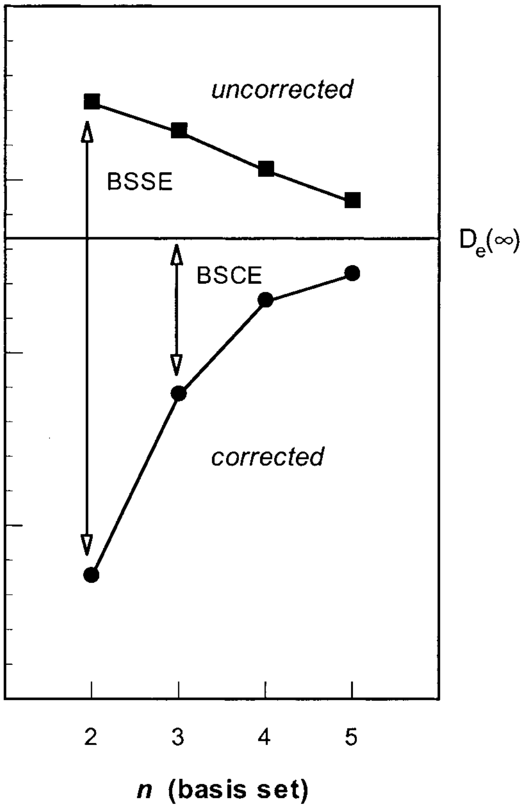

Figure 4. Comparison of basis set superposition error and basis set convergence error in the calculation of binding energies. Note the improvement in the convergence behavior of the binding energies after correction for BSSE.

TABLE 1: Intrinsic Errors in D_e for a Prototypical Set of First-row, Diatomic Molecules: ΔD_e (kcal/mol)[a]

	ΔD_e[CH]	ΔD_e[HF]	ΔD_e[N_2]	ΔD_e[CO]
D_e(expt'l)[b]	83.9	141.6	228.4	259.3
Δ(cv)	−0.2	−0.2	−0.8	−0.9
CCSD	−1.0	−2.2	−10.7	−8.4
CCSD(T)	−0.2	−0.1	−1.1	−0.8
CCSDT[c]		−0.2	−1.7	−1.2
MP2	−2.9	4.2	11.6	12.7
MP3	−1.4	−3.1	−12.6	−8.8
MP4	−0.6	1.1	3.4	5.0
MP5		−1.0		

[a] Δ(cv) refers to the core-valence correction to the experimental values given in the first row of the table. There is perfect agreement between the measured and calculated values when the intrinsic error is the same as the core-valence correction. [b] For comparison to the current nonrelativistic calculations, the experimental dissociation energies have been corrected for atomic spin−orbit effects. [c] See ref 87.

molecule; and CO, an ionic, multiply bonded, diatomic molecule. Although a more extensive set of molecules could be considered, the current set illustrates the main points to be made here.

From the intrinsic errors for the bond energies listed in Table 1, we see that the (valence electron) CCSD(T) method predicts the bond energies to an accuracy of better than 1 kcal/mol, save for N_2, where the error is −1.1 kcal/mol (this still corresponds to 99.6% of the experimental N_2 bond energy). Note that the intrinsic errors are all negative, i.e., the calculated values underestimate the experimental values. This is as it should be: the correlation energy should increase with decreasing R as should the error. For the singly bonded species, including the very ionic HF molecule, the errors are just a few tenths of a kcal/mol. The errors for the multiply bonded molecules are much larger than those for the singly bonded molecules (as are the bond energies). The fact that the intrinsic errors for the CCSD(T) method are larger for the multiply bonded species is partially due to the fact that the CCSD(T) method is based on a Hartree−Fock wave function and Hartree−Fock wave functions do not provide a good zero-order description of π bonds. Inclusion of core-valence correlation effects[38] reduces the intrinsic errors in D_e[CCSD(T)] to just a few tenths of a kcal/mol for all species considered.

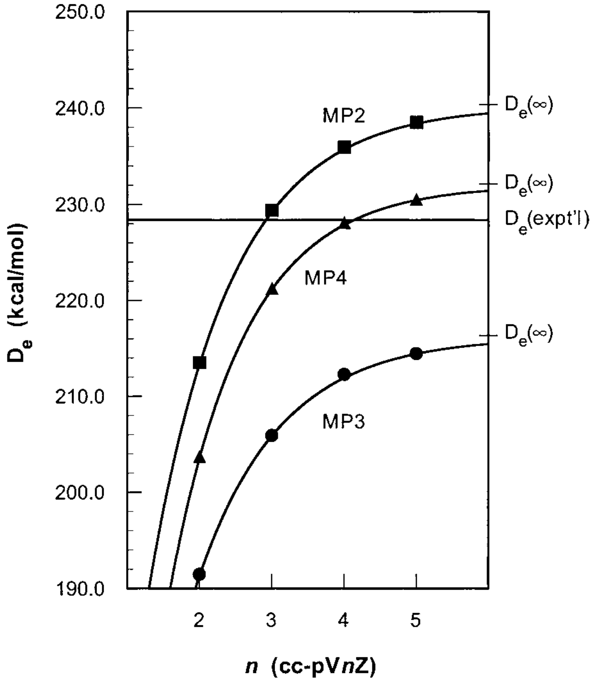

Figure 5. $D_e(N_2)$ calculated with MP2, MP3 and MP4 methods and the correlation-consistent basis sets. $D_e(\infty)$ refers to the dissociation energies at the complete basis set limit and D_e(expt'l) to the experimental dissociation energy.

Omission of the perturbative triples correction in the CCSD(T) method substantially increases the error in the coupled cluster method: from −0.2 to −1.0 kcal/mol for CH, from −0.1 to −2.2 kcal/mol for HF, from −1.1 to −10.7 kcal/mol for N_2, and from −0.8 to −8.4 kcal/mol for CO. Thus, *connected* triple excitations (t_3) are critical for accurate predictions of bond energies with the coupled cluster method. As expected, the triples correction is far more important for multiply bonded molecules than for the singly bonded molecules (compare CH and HF to N_2 and CO) as well as for molecules with more electrons (compare CH to HF).

The MP2 method is clearly of limited use in computing the bond energies of chemically bound molecules. The errors in the calculated D_e can exceed 10 kcal/mol and vary dramatically from molecule to molecule, underestimating D_e(expt'l) by 2.9 kcal/mol for CH and overestimating D_e(expt'l) by 4.2, 11.6, and 12.7 kcal/mol for HF, N_2 and CO, a range of more than 15 kcal/mol. In fact, the behavior observed in Table 1, the systematic over-prediction of the bond energies of many, but not all, molecules, has been a major source of confusion in the use of perturbation theory to compute the bond energies of molecules. To illustrate this point, in Figure 5, we plot the bond energy of N_2 as a function of basis set for the MP2−4 methods.[38b] As can be seen, as the basis set increases in size (and completeness), the calculated bond energies also increase. In fact, the curves of $D_e(n)$ for both the MP2 and MP4 methods pass through the experimental value of the bond energy on their way to intrinsic errors of +11.6 kcal/mol (MP2) and +3.4 kcal/mol (MP4). At $n \approx 3$, D_e(MP2;n) crosses the experimental value and the calculational error is essentially zero. Unless one was aware of the dependence of D_e(MP2) on n, it would be tempting to conclude that the MP2 method with the cc-pVTZ set gives a very accurate description of the wave function of N_2. This is clearly an incorrect conclusion: the intrinsic error in D_e(MP2) is nearly 12 kcal/mol. Similar behavior is observed for the MP4 method where use of a cc-pVQZ basis set results in near cancellation of the basis set convergence error and the electronic structure method error and, thus, a false sense of confidence in the calculation. This is not an uncommon occurrence in perturbation theory calculations.[21,22,64]

The intrinsic errors for the MP4 method are substantially smaller than those for the MP2 method. However, even the MP4 method does not approach chemical accuracy, i.e., errors of less than 1 kcal/mol. In addition, the relative errors vary from −0.6 kcal/mol for CH to +5.0 kcal/mol for CO. Thus, with the MP4 method, it is not even possible to reliably predict the *relative* bond energies among a heterogeneous set of molecules such as those listed in Table 1. Similar arguments hold for the MP3 method.

MP5 calculations have been reported on HF.[21] The magnitude of the error in D_e(MP5) is nearly identical to that in D_e(MP4), although of opposite sign. This is an example of the convergence problem in perturbation expansions noted earlier. The results on HF combined with many other results[20,21] caution against the use of perturbation theory in molecular calculations. Although one might argue that the MP2 method appears to improve upon the Hartree−Fock method, one must keep in mind that the MP2 method is the first term in a slowly or nonconverging series and, therefore, its use should also be viewed with caution.

5.2. Hydrogen-Bonded Molecules. The hydrogen bond is the most important of the weak molecular interactions. The concept of the hydrogen bond was introduced into chemistry by Pauling in the 1939 edition of his book *Nature of the Chemical Bond*,[2] although others were also involved in unraveling the anomalous physical and thermodynamic properties of matter resulting from the presence of hydrogen bonds. It is now widely recognized that hydrogen bonds play a critical role in many chemical processes, especially biochemical processes. Hydrogen bond strengths range from less than 4 kcal/mol to over 40 kcal/mol, depending on the nature of the molecular species involved, although most fall within the range of 4−10 kcal/mol. Unfortunately, few hydrogen-bonded species have been characterized to the extent that they can serve as reliable benchmarks for theoretical calculations.

The HF dimer has been the subject of a number of detailed studies. Of particular interest are the modeling and experimental studies of Quack, Suhm, and co-workers[65,66] and the experimental studies of Miller et al.,[67] Nesbitt and co-workers,[68] and Klemperer and co-workers.[69] This combination of studies has allowed a reliable value of D_e to be extracted from the measurements of D_0. Klopper, Quack, and Suhm[65f] report that $D_e[(HF)_2] = 4.56 \pm 0.05$ kcal/mol. Unfortunately, comparable data does not exist for the prototype hydrogen-bonded molecule, $(H_2O)_2$. The estimate of the equilibrium binding energy of the water dimer from experimental studies is 5.4 ± 0.7 kcal/mol,[70,71] an unacceptably large error bound for the present purpose.

The HF−HF binding energy derived from the studies of Klopper, Quack, and Suhm[65f] is listed in Table 2 along with the intrinsic errors for the coupled cluster and perturbation theory methods.[72−74] With the exception of the CCSD and MP2 methods, the errors in the calculated binding energies for $(HF)_2$ fall within the stated error bounds, ±0.05 kcal/mol. The CCSD(T), MP4, and MP3 binding energies are just 0.02−0.03 kcal/mol smaller than the experimental value. For the CCSD and MP2 methods, the intrinsic errors are −0.15 and −0.09 kcal/mol, respectively. So, even the CCSD and MP2 methods predict more than 96% of the $(HF)_2$ binding energy.

TABLE 2: Intrinsic Errors in D_e for the Hydrogen Fluoride Dimer: ΔD_e (kcal/mol).

	$\Delta D_e[(HF)_2]$
D_e(expt'l)[a]	4.56 ± 0.05
CCSD	−0.16
CCSD(T)	−0.02
MP2	−0.09
MP3	−0.03
MP4	−0.02

[a] Reference 65f.

All-electron calculations have not yet been reported for $(HF)_2$. However, Schütz et al.[75] and Klopper and Luthi[73b] have carried out all-electron and valence-electron MP2 calculations on the water dimer and reported a core-valence correction of 0.04 kcal/mol. The core-valence correction for $(HF)_2$ is expected to be somewhat smaller than this value (because of the larger atomic charge of fluorine). Addition of the core-valence correction to the CCSD(T), MP4 and MP3 values for $D_e[(HF)_2]$ should bring these results into almost perfect agreement with the experimentally derived value.

In summary, the CCSD(T) method provides unparalleled accuracy for the hydrogen-bond energy of $(HF)_2$. In contrast to chemically bound molecules, third-order and fourth-order perturbation theory also provide accurate predictions of D_e, and the intrinsic error for the MP2 method is less that 0.1 kcal/mol. The available evidence suggests that these conclusions also hold for the water dimer.[56,73] Indeed, these conclusions may be generally valid if the two molecules involved in the hydrogen bond are well described by a Hartree−Fock wave function. It may not hold if the wave function for one or more of the molecules (or in the case of intramolecular hydrogen bonding, molecule) have substantial multireference character. In this case, the coupled cluster methods are expected to be more reliable.

5.3. Weakly Bound Molecules. Weakly bound molecular complexes, in which the binding is derived mainly from the electrostatic interactions between molecules, have always posed a serious challenge for the computational quantum chemist. For these species, the binding energies fall in the range of a few tenths of a kcal/mol to a few kcal/mol. In fact, it is common to quote the binding energies for weakly bound molecules in wavenumbers, where 1 kcal/mol = 349.76 cm^{-1}, and that will be done here. In the past, the binding energies for these species were usually underestimated, often by substantial amounts. Good agreement was sometimes obtained when the BSSE nearly canceled the basis set convergence error, but these instances are clearly fortuitous.

Within the past decade, substantial experimental and modeling efforts have focused on three weakly bound molecules: N_2−HF,[76] Ar−HF,[77] and Ar−HCl.[78] As for $(HF)_2$, this combination has led to a detailed characterization of the potential energy surfaces of these species. Because of these investments, N_2−HF, Ar−HF, and Ar−HCl provide excellent benchmarks for assessing the accuracy of perturbation theory and coupled cluster methods for describing the interactions of weakly bound molecules. The values of D_e, derived from the experimentally based modeling studies are listed in Table 3 along with

TABLE 3: Intrinsic Errors in D_e for the N_2−HF, Ar−HF, and Ar−HCl Weakly Bound Molecules: ΔD_e (cm^{-1})

ΔD_e	ΔD_e[N_2−HF]	ΔD_e[Ar−HF]	ΔD_e[Ar−FH]	ΔD_e[Ar−HCl]	ΔD_e[Ar−ClH]
D_e(expt'l)	776 ± 30	211 ± 4	109 ± 10	176 ± 5	148 ± 10
CCSD	−52	−45	−36		
CCSD(T)	17	0	−15	0	−1
MP2	35	−10	−16	31	33
MP3	−36	−31	−31		
MP4	38	7	−10	10	7

the intrinsic errors for the CCSD, CCSD(T), and MP2−4 methods.[38c,f,h] In this table, information on both the ArHX and ArXH isomers is given, although the errors in the "experimental" D_e are substantially larger for the ArXH isomers than that for the ArHX isomers (probably even larger than the uncertainties quoted by the author[77,78]).

From Table 3 we see that, save for the Ar−FH isomer, the intrinsic errors for the CCSD(T) method fall well within the stated error bounds for the experimentally derived binding energies. The intrinsic error in D_e [CCSD(T)] is essentially zero for Ar−HF, Ar−HCl, and Ar−ClH. It is 17 cm^{-1} in N_2−HF and −15 cm^{-1} in Ar−FH, while the estimated error bounds in these two cases are ±30 and ±10 cm^{-1}, respectively (as noted above, the latter uncertainty may be too small). Omission of the triples correction in the CCSD(T) method substantially increases the intrinsic errors in D_e. The CCSD method underestimates the binding energy of N_2−HF by 52 cm^{-1}, Ar−HF by 45 cm^{-1}, and Ar−FH by 36 cm^{-1}.

Given its simplicity, the MP2 method does a surprisingly good job of predicting the properties of weakly bound molecules (as it did for hydrogen-bonded molecules). The intrinsic errors for D_e range from +35 cm^{-1} in N_2−HF to −16 cm^{-1} in Ar−FH, although the percentage error in ArFH (14.7%), ArHCl (17.6%), and ArClH (22.2%) can hardly be considered exceptionally good. In ArFH, the error for the MP2 method is comparable to that for the CCSD(T) method. The MP2 method is nearly as accurate as the MP4 method. Only for Ar−HCl and Ar−ClH are significant improvements realized by using the MP4 method. The errors for the MP3 method tend to be larger than the errors in either the MP2 or MP4 methods, although smaller than for the CCSD method.

5.4. Very Weakly Bound Molecules. There is no more demanding problem than the calculation of the binding energies of molecules whose binding arises only from dispersion interactions. Dispersion interactions are very weak, depending on $^1/_{r^6}$ (or higher powers of $^1/_r$). These forces have no classical counterpart, being due to the internal, quantum mechanical structure of the two (or more) systems involved in the complex. Binding energies arising from dispersion forces range from a few hundredths to a few tenths of a kcal/mol (a few wavenumbers to a few hundred wavenumbers).

The rare gas dimers have been well characterized by a combination of experimental and modeling studies and provide accurate benchmarks for the electronic structure methods discussed here. The experimentally derived binding energies for He_2, Ne_2, and Ar_2 are listed in Table 4[79] along with the corresponding intrinsic errors for both coupled cluster and perturbation theory methods.[81,32,38j] The CCSD method predicts 85% of the binding energy in He_2, 77% of the binding energy in Ne_2, and 72% of the binding energy in Ar_2. The decreased percentage for Ne_2 and Ar_2 reflects the increased importance of connected triple excitations in these latter two species. Including the triples correction, CCSD(T), reduces the error in D_e to just 3% for all three dimers. Thus, the CCSD(T) method provides an accurate, consistent description of the electronic structure of the rare gas dimers.

TABLE 4: Intrinsic Errors in D_e for the Rare Gas Dimers, He_2, Ne_2, and Ar_2: ΔD_e (cm^{-1})[a]

	ΔD_e(He_2)	ΔD_e(Ne_2)	ΔD_e(Ar_2)
D_e(expt'l)[b]	7.59	29.4	99.6
Δ(cv)		0.05	−0.8
CCSD	−1.1	−6.8	−27.6
CCSD(T)	−0.2	−1.0	−2.6
CCSDT	0.0		
MP2	−2.7	−10.5	13.4
MP3	−1.1	−7.1	−17.6
MP4	−0.5	−1.9	0.4
MP5	−0.2		

[a] Δ(cv) refers to the core-valence correction to the experimental values given in the first row of the table. There is perfect agreement between the measured and calculated values when the intrinsic error is the same as the core-valence correction. [b] Δ(cv) Refers to the core−valence correction to the experimental values given in the first row of the table. There is perfect agreement between the measured and calculated values when the intrinsic error is the same as the core−valence correction.

Van Mourik, Wilson, and Dunning[38i] have reported core-valence calculations on Ne_2 and Ar_2 (the K-shell electrons were not included in the Ar_2 calculations). They found the core-valence correction to be very small in Ne_2 (0.05 cm^{-1}). Oddly enough, the correction was positive, i.e., the core-valence contribution to the correlation energy is larger for two separate neon atoms than for Ne_2. For Ar_2, the core-valence correction at the CBS limit was predicted to be −0.8 cm^{-1}. Thus, the core-valence correction accounts for approximately one-third of the error in D_e[(CCSD(T)].

Full inclusion of connected triple excitations in the CCSDT method essentially eliminates the error in the He_2 binding energy. Van Mourik and Dunning[32] have shown that the intrinsic error in D_e(CCSDT) is nearly zero: connected quadruple excitations contribute only 0.01−0.015 cm^{-1} to the binding energy of this dimer; see also Burda et al.[81] Burda et al. also reported CCSDT calculations on Ne_2, but only with the aug-cc-pVTZ basis set. These calculations predict that inclusion of full triple excitations will increase the Ne_2 binding energy by 0.25 cm^{-1}, or 25% of the remaining error in D_e[CCSD(T)]. However, this is a small basis set and, until calculations with larger basis sets are reported, we will not be able to definitively establish the intrinsic error in the all-electron CCSDT method for Ne_2.

As expected, neither the MP2 or MP3 methods provide accurate descriptions of the binding in the rare gas dimers. These methods do not include triple excitations, and it is known that triple excitations are important for describing dispersion interactions.[82] The MP4 method, which does include triple excitations, predicts 93% of D_e(He_2) and D_e(Ne_2), and 100.5% of D_e(Ar_2). The accuracy of D_e(Ar_2) is somewhat surprising. However, given the other terms in the perturbation expansion, this result may be fortuitous. For He_2, MP5 calculations have also been carried out. The intrinsic error for the MP5 method is −0.2 cm^{-1}, or 40% of that for the MP4 method. In fact, the error in D_e(MP5) is the same as that in D_e[CCSD(T)]. Thus, the perturbation theory series for He_2 is converging very slowly. This is

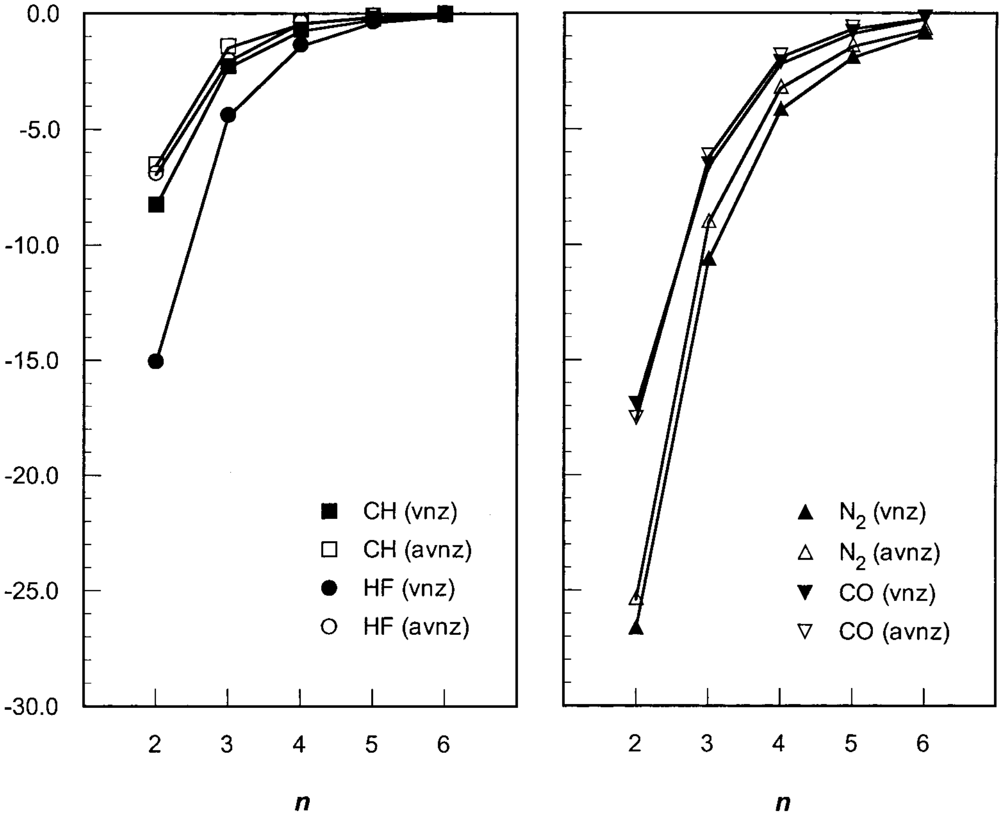

Figure 6. Basis set convergence errors for CCSD(T) or RCCSD(T) calculations of D_e for (a) CH and HF, and (b) N_2 and CO with the standard (v*n*z) and augmented (av*n*z) correlation-consistent basis sets.

unfortunate since the cost of MP5 calculations is comparable to that of CCSDT calculations, both scale as N^8, (although the CCSDT equations must be solved iteratively), and the intrinsic error in the CCSDT method is nearly zero for He_2.

6. Basis Set Convergence Errors

Having established the electronic structure method errors for perturbation theory and coupled cluster methods, we may now turn our attention to the basis set convergence error in the calculation of molecular binding energies, $\Delta D_e^{bs}(M;n) = D_e(M;n) - D_e(M;\infty)$. We will again examine the basis set convergence errors separately for the various types of interactions considered here: chemically bound molecules, hydrogen-bonded molecules, weakly bound molecules, and very weakly bound molecules. For these studies, we will use the CCSD(T) method because, as shown in the last section, the CCSD(T) method provides an accurate description of the binding energies for all of the molecules considered here. However, the convergence behavior of perturbation theory and CCSD methods are similar to that of the CCSD(T) method, a fact used to advantage by Dunning and Peterson to approximate CCSD(T) calculations with very large basis sets.[83]

6.1. Chemically Bound Molecules. The basis set convergence errors for calculation of the bond energies of the chemically bound molecules (CH, HF, N_2, and CO) are plotted in Figure 6. Consider first the results of calculations with the cc-pV*n*Z basis sets. With the cc-pVDZ basis set, D_e(CH) is 8.3 kcal/mol from the complete basis set (CBS) limit, and the bond energies of HF, N_2, and CO are 15.1, 26.7, and 17.5 kcal/mol, respectively, from the limit. In percentages, these correspond to 10%, 11%, 12%, and 7% of D_e for CH, HF, N_2, and CO. For the cc-pVTZ set, the convergence errors range from 2.3 (CH) to 10.7 kcal/mol (N_2), or from 2.8% (CO) to 4.7% (N_2). Even for the cc-pV5Z set, the variation in the convergence error is 1.7 kcal/mol. Thus, very large basis sets will be required to calculate either the absolute bond energies or the relative bond energies of this diverse set of molecules to chemical accuracy, i.e., with an error of less than 1 kcal/mol.

The above clearly illustrates the difficulties caused by the varying rates of convergence for single, double and triple bonds, and ionic and covalent molecules. For a given bond type, on the other hand, the situation is not so grim. Peterson and Dunning[38e] have shown that the rates of convergence of the CH bond energies in the CH_m and C_2H_m($m = 1-4$) series are very similar. With the cc-pVDZ set the spread in the convergence errors is just 4.5 kcal/mol (vs 18.4 kcal/mol above), with the cc-pVTZ set it is just 1.4 kcal/mol (vs 8.4 kcal/mol above), and with the cc-pVQZ set the spread drops to just 0.5 kcal/mol (vs 3.4 kcal/mol above). Because of this consistency, if one corrects the D_e(CH) obtained with the cc-pVTZ set by adding approximately 2 kcal/mol to the calculated values, the resulting bond energies agree with the experimental values to within ±0.7 kcal/mol.

But even for similar bonds one must be careful. $\Delta D_e^{bs}(n)$ for the CH bond formed by addition of a hydrogen atom to acetylene is far different than those of the other bonds in the series.[38e] The $H-C_2H_2$ bond is formed by breaking one of the acetylene π-bonds as the CH σ-bond is formed. Calculations show that there is little dependence of the energy of this process on basis set. In fact, the convergence error is positive for this bond, i.e., the CBS limit is approached from above, not from below as is the case for the other CH bonds. The other CH bonds are formed by simply pairing a radical orbital on the CH_{m-1} or C_2H_{m-1} fragment with the 1s hydrogen orbital. This has the "standard" type of dependence on basis set illustrated

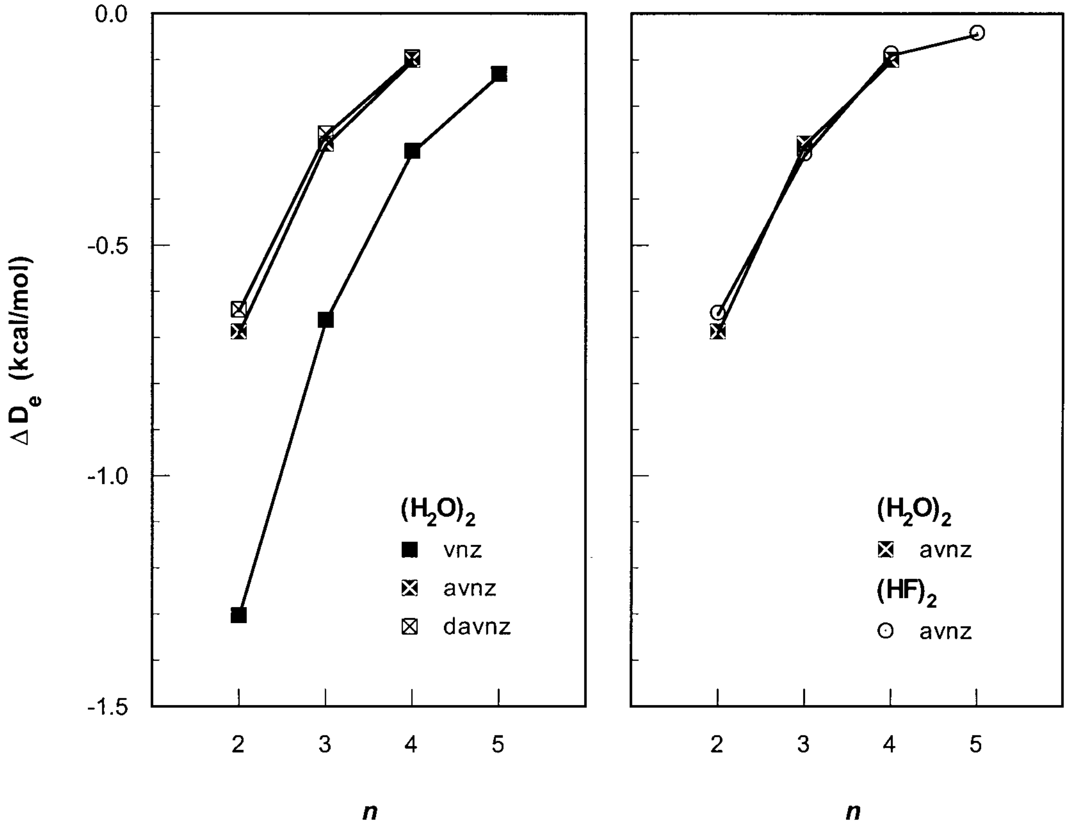

Figure 7. Basis set convergence errors for CCSD(T) calculations of D_e for the water and hydrogen fluoride dimers with the correlation-consistent basis sets. (a) Calculations on $(H_2O)_2$ with the standard and augmented basis sets. (b) Comparison of the convergence errors for $(H_2O)_2$ and $(HF)_2$ with the augmented basis sets.

in Figure 6 for CH. Thus, the simple correction given above cannot be applied to the H−C_2H_2 bond.

Cancellation of basis set convergence errors is also responsible for the success of the isodesmic reaction scheme[84] as a means of computing molecular bond energies. In an isodesmic reaction

$$A + B \rightarrow C + D \qquad (13)$$

the numbers of chemical bonds of each type (e.g., single C−H bonds, double C═O bonds, etc.) are the same on each side of the reaction. To the extent that the convergence rates of the C−H, C═O, etc. bonds on the left and right sides of eq 13 are similar, the calculated ΔE_{rxn} will be only weakly dependent on the basis set. This fact can be used to compute the binding energy of one of the molecules involved in eq 13 if accurate values of the remaining three binding energies are known (either from calculation or experiment). For example, to compute the energy of the C−Cl bond in C_6H_5Cl, one could use the reaction

$$C_6H_5Cl + CH_4 \rightarrow C_6H_6 + CH_3Cl \qquad (14a)$$

and

$$D_e(C_6H_5{-}Cl) = D_e(C_6H_5{-}H) + D_e(CH_3{-}Cl) - D_e(CH_3{-}H) - \Delta E_{rxn} \qquad (14b)$$

Using eq 14b, a rather accurate value of the C−Cl bond energy can be computed without ever performing calculations on benzyl chloride with a large basis set.

Comparing the D_e's calculated with the aug-cc-pV*n*Z sets with those calculated with the cc-pV*n*Z sets, we see a marked difference in convergence behavior only for HF. For HF the basis set convergence error is reduced from 15.1 to 7.0 kcal/mol upon augmentation of the cc-pVDZ set. Even for the cc-pVTZ and cc-pVQZ sets, augmentation reduces the error by 2.4 and 1.0 kcal/mol, respectively. This is characteristic of highly ionic molecules such as HF. Because of the anion character in the wave function, such molecules are much better described by the augmented sets. Note that with the augmented sets, the convergence behavior of D_e is essentially the same for HF as for CH. Use of the augmented sets tends to reduce the differences in the basis set convergence behavior for ionic and covalent molecules and, thus, can be considered more "universal" sets. This concept will be further extended in the following two sections.

6.2. Hydrogen-Bonded Molecules. The basis set convergence errors for the water dimer, $(H_2O)_2$, with the standard and augmented sets are plotted in Figure 7a.[40e] The importance of the extra diffuse functions in the augmented sets, a fact first noted by Feller,[56] is clear. The convergence errors for the aug-cc-pV*n*Z sets are essentially identical to those for the cc-pV(n+1)Z sets. Since an aug-cc-pV*n*Z set for oxygen is smaller than a cc-pV(n+1)Z set by (2n+3) functions, it is computationally more efficient to use the augmented sets rather than the standard sets of equivalent accuracy. Use of the doubly augmented sets, on the other hand, offers only minor improvements over the singly augmented sets and need not be considered further. This conclusion applies to all hydrogen-bonded systems investigated to date.[85]

For $(H_2O)_2$, calculations with the aug-cc-pV*n*Z sets yield 86%, 94%, and 98% of D_e[CCSD(T)] for n = 2−4. The absolute errors are 0.69, 0.28, and 0.10 kcal/mol. Thus, the aug-cc-pVQZ basis set yields a hydrogen bond energy for $(H_2O)_2$ within 0.1 kcal/mol of the complete basis set limit. The basis set convergence errors for $(H_2O)_2$ and $(HF)_2$[40k] calculated with the

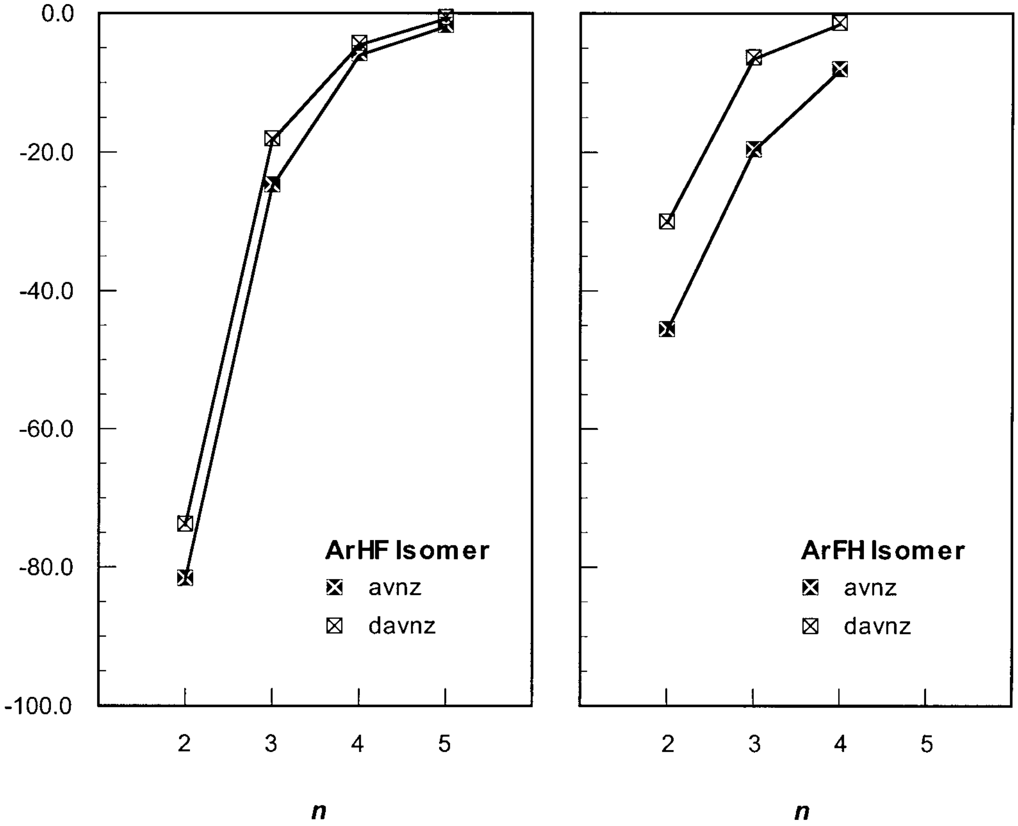

Figure 8. Basis set convergence errors for CCSD(T) calculations of D_e for the two isomers of Ar–HF with the augmented correlation-consistent basis sets. (a) ArHF isomer. (b) ArFH isomer.

augmented sets are compared in Figure 7b. As can be seen, $\Delta D_e^{bs}(n)$ for $(H_2O)_2$ and $(HF)_2$ are very similar. This means that it would be possible to use $\Delta D_e^{bs}(n)$ for $(HF)_2$ to correct for the basis set convergence error in $(H_2O)_2$. It would be interesting to know how universal this finding is.

6.3. Weakly Bound Molecules. The dependence of $\Delta D_e^{bs}(n)$ of the two isomers of Ar–HF (ArHF and ArFH) on basis set, for both the aug-cc-pV*n*Z and d-aug-cc-pV*n*Z sets, is plotted in Figure 8. Both classical electrostatic (e.g., dipole–induced dipole) and dispersion interactions contribute to D_e for Ar–HF, although the former interactions are dominant. As was the case for the hydrogen-bonded molecules, $(H_2O)_2$ and $(HF)_2$, augmented sets are required to properly describe the binding in the Ar–HF complex (which may also be considered a hydrogen-bonded molecule, although a very weakly bound one). There is a strong dependence of $\Delta D_e^{bs}(n)$ on basis set. The aug-cc-pVDZ basis set yields just 61% of D_e(ArHF). The fraction of D_e recovered increases to 87% for the aug-cc-pVTZ set and to nearly 96% for the aug-cc-pVQZ set. This can be contrasted to $(H_2O)_2$ where the aug-cc-pVDZ set yields 86% of D_e(CBS) and the aug-cc-pVTZ and aug-cc-pVQZ sets yield 94% and 98%, respectively. Thus, D_e for weakly bound molecules converges more slowly than D_e for hydrogen-bonded molecules; a clear indication of the increased difficulty in describing weakly bound complexes.

As noted in section 5, the binding energy of the ArFH isomer is only about one-half that of the ArHF isomer (109 cm^{-1}). As can be seen in Figure 8b, use of the doubly augmented sets significantly improves the convergence behavior of ΔD_e(ArFH). In fact, the improvement is of sufficient magnitude to argue for use of a d-aug-cc-pV*n*Z set rather than an aug-cc-pV*n*Z set for calculations on ArFH. This means that doubly augmented sets could be needed if the entire Ar–HF potential energy surface is of interest. However, it should also be noted that the convergence errors are more comparable for the two isomers when the aug-cc-pV*n*Z sets are used. As we shall see in the next subsection, d-aug-cc-pV*n*Z sets must be used to obtain an adequate description of dispersion interactions. Thus, the behavior of $\Delta D_e^{bs}(n)$ for the ArFH isomer suggests that dispersion interactions are more important for the more weakly bound ArFH isomer than for the ArHF isomer.

6.4. Very Weakly Bound Molecules. The basis set convergence errors for $D_e(Ne_2)$ are plotted in Figure 9a. The binding in Ne_2 is due entirely to dispersion forces, which have long been known to place heavy demands on the basis set. These demands are evident in the figure. In contrast to the description of weakly bound molecules, we now find that two sets of diffuse functions, i.e., the d-aug-cc-pV*n*Z basis sets, are required to adequately describe the wave function of Ne_2. Although the aug-cc-pV(n+1) sets yield convergence errors that are not too much larger than those of the d-aug-cc-pV*n*Z sets, the d-aug-cc-pV*n*Z sets have 2(2n+3) fewer basis functions than the aug-cc-pV(n+1)Z sets and thus are recommended for calculations on very weakly bound molecules. Use of the triply augmented sets, on the other hand, leads to only a minor improvement in the calculated $\Delta D_e(n)$ and need not be considered further.

The d-aug-cc-pVDZ set recovers 62% of $D_e(\infty)$. Use of the d-aug-cc-pVTZ set increases this fraction to 87%, while use of the d-aug-cc-pVQZ set recovers 96% of $D_e(\infty)$. Interestingly enough, these numbers are essentially identical to those from the calculations on ArHF, but with the aug-cc-pV*n*Z sets.

In Figure 8(b) $\Delta D_e^{bs}(n)$ for Ne_2 is plotted along with $f\Delta D_e^{bs}(n)$ for Ar_2, where $f = 0.225$ [a scaling factor that minimizes the difference between the Ne_2 and Ar_2 convergence errors]. As can be seen, the $\Delta D_e^{bs}(n)$ curve for Ne_2 is very similar to the "reduced" $\Delta D_e(n)$ curve for Ar_2. This similarity is not totally surprising given the fact that the binding of both of these species

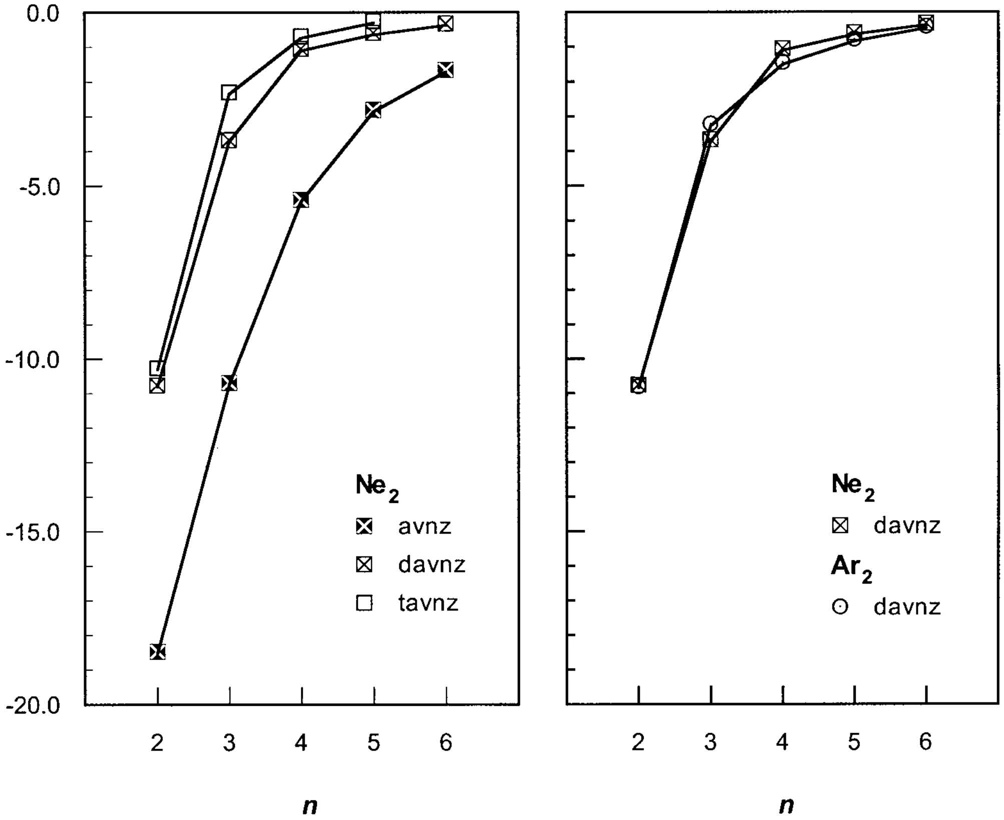

Figure 9. Basis set convergence errors for CCSD(T) calculations of D_e for Ne_2 with the augmented correlation-consistent basis sets. (a) Ne_2. (b) Comparison of the errors for Ne_2 and Ar_2; the Ar_2 points have been scaled by $f = 0.225$.

are determined by long-range dispersion forces, which are largely determined by the polarizability of the Ne and Ar atoms.

7. Path Forward

From the above it is clear that a major breakthrough in the calculation of molecular binding energies was achieved in the 1990s. The level of accuracy that is now obtainable was all but unimaginable at the beginning of that decade. But, what if the desired accuracy is higher than that achievable by the CCSD(T) method? What about the next step in refining the calculated binding energies? Also, as noted in section 3.2, CCSD(T) calculations scale as N^7. Does this mean that the CCSD(T) method is an illusion, prohibitively expensive for molecules of any real interest to practicing chemists? Finally, the accuracy of the CCSD(T) method is dependent on extrapolation of the calculated binding energies to the complete basis set limit. What can be done about the slow convergence of binding energies with basis set? In this section we will briefly touch on each of these issues.

7.1. More Accurate Coupled Cluster Calculations on Molecules. There are two sources of error in the CCSD(T) method. The first is due to truncation of the coupled cluster expansion; the second is due to the use of the HF wave function as the zero-order wave function. The equations for the CCSDT and CCSDTQ methods have been derived by Bartlett and coworkers.[86] However, use of these methods poses significant computational problems as CCSDT calculations scale as N^8, while CCSDTQ calculations scale as N^{10}! Nevertheless, the CCSDT method has been implemented in the ACES II program[9] and has been used by Feller,[87] Feller and Sordo,[88] and Bak et al.[89] to systematically investigate the differences between the perturbative treatment of connected triple excitations in the CCSD(T) method and the full inclusion of these excitations in the CCSDT method. For a range of molecules, these authors found that the differences were usually small, from 0.1 to 0.7 kcal/mol, with D_e(CCSDT) being smaller than D_e[CCSD(T)], see Table 1. The two exceptions to this trend were $D_e(C_2)$ and D_e(CN). For C_2, D_e(CCSDT) was 2.1 kcal/mol less than D_e[CCSD(T)]; for CN, D_e(CCSDT) is 0.8 kcal/mol greater than D_e[CCSD(T)]. In general, a complete treatment of triple excitations did not improve agreement with experiment. Thus, a cancellation of errors is at least partially responsible for the good performance of the CCSD(T) method. No CCSDTQ calculations with a basis set large enough to be meaningful (cc-pVTZ or beyond) have yet been reported. So, further work will be required to understand the convergence of the coupled cluster expansion.

Both the C_2 and CN molecules are poorly described by the HF wave function. If these errors are due to the inadequacy of the HF zero-order wave function, the obvious solution to the problem is to use a multiconfiguration wave function for Ψ_0. Unfortunately, the use of multireference wave functions in eq 8 leads to mathematical difficulties in the solution of the resulting coupled cluster equations. Several approaches have been developed to address such problems, including the state-universal multireference coupled cluster method of Kucharski and Bartlet[90] (based on an ansatz originally proposed by Jeriorski and Monkhorst[91]) and the state-specific multireference coupled cluster method of Li and Paldus.[92] Both of these methods have met with some success. However, neither have been benchmarked as described herein. Additional mathematical and/or computational work remains to be done to establish an efficient, theoretically sound multireference coupled cluster method.

7.2. CCSD(T) Calculations on Large Molecules. Since CCSD(T) calculations scale as N^7, doubling the size of a molecule increases the computing requirements by over a factor of 100. Given the cost of CCSD(T) calculations on triatomic and tetratomic molecules, this steep dependence on the number of basis functions would seem to rule out calculations on large molecules. However, advances in both computing technology and algorithms for coupled cluster calculations are pushing this approach to larger and larger molecules. Using massive parallelism, computers capable of 1 trillion arithmetic operations per second (TFLOPS) are already in place, e.g., at the San Diego Supercomputer Center, and computers capable of 100 TFLOPS are expected to become available within the next four years. Although few computational chemistry codes can exploit this level of parallelism, the high-performance computational chemistry group at the Pacific Northwest National Laboratory, working with a worldwide group of collaborators, has recently released a new code, NWChem,[93] which is capable of performing coupled cluster calculations on terascale computers. With NWChem, it should be possible to routinely carry out CCSD(T) calculations on molecules containing 20 or more atoms with a cc-pVQZ basis set on the 3 $^1/_2$-TFLOPS computer to be installed at DOE's National Energy Research Scientific Computing Center (Lawrence Berkeley National Laboratory) later this year.

The scaling laws given above assume that the calculations are carried out in terms of the canonical, delocalized Hartree–Fock orbitals. Although an N^5 step is required to transform the two-electron integrals over atomic orbitals (AOs) to integrals over molecular orbitals (MOs), this approach has a number of advantages. For example, in terms of the canonical HF orbitals, the number of double excitation amplitudes that must be calculated in the CCSD equations is $^1/_2(n_o n_v(n_o n_v+1))$, where n_o is the number of occupied orbitals and n_v is the number of virtual orbitals ($N = n_o + n_v$). If the equations were instead written in terms of the AOs, there would be $^1/_2(N^2(N^2+1))$ amplitudes. If we assume that $n_v \gg n_o$, there will be $(n_v/n_o)^2$ more amplitudes in the AO case than in the MO case. For the aug-cc-pVTZ set, $n_v \approx 10n_o$, so there will be 100 times more AO than MO amplitudes. This argument, however, neglects the efficiencies that could be realized by screening. Screening has long been used in HF calculations and has been found to reduce the formal N^4 scaling to $N^2 - N^3$ (with a few more tricks, linear scaling is even possible for sufficiently large molecule[94]). Just as in the HF case, screening the contributions to the correlation energy can reduce the formal scaling of correlated calculations to far less than those predicted by the scaling laws.

To take maximum advantage of screening, the equations for the coupled cluster methods must be rewritten in terms of AOs rather than MOs. It is in the localized AO basis, rather than the delocalized MO basis, that the decay of the interaction integrals with distance can be most readily exploited. This has recently been done for the CCD method by Scuseria and Ayala.[95] By examining a number of test cases, these authors showed that, for sufficiently large molecules, the CCD equations could be solved just as efficiently in the AO basis as in the MO basis. Because the effectiveness of the screening procedure will increase with molecule size, the question is not whether the use of AOs is more efficient than the use of MOs, but where the crossover point occurs. The preliminary studies of Scuseria and Ayala[95] suggest that the crossover point for extended molecules may occur around 10–20 atoms (the exact crossover point depends on a number of factors, including the basis set, the shape of the molecule, etc.).

It is also possible to base correlated calculations on localized molecular orbitals instead of the delocalized, canonical HF orbitals. This approach, which was first put forward by Pulay and co-workers,[96] has recently been applied to coupled cluster methods by Werner, and co-workers.[97] Their algorithm uses a set of orbitals obtained by combining localized HF MOs with AOs projected onto the HF orbitals. Because of the localized nature of the MOs, correlation between electrons in orbitals that are widely separated can be neglected. This method has been shown to lead to much improved scaling with molecular size. However, it involves additional approximations relative to the algorithm put forward by Scuseria and Ayala.[95]

Save for the HF and density functional methods, the impact of reduced scaling algorithms has yet to be felt in chemistry. As the above discussion shows, reduced scaling algorithms for correlated molecular calculations are imminent. When these algorithmic advances are combined with the advances in computing technology wrought by massive parallelism, the impact will be truly revolutionary. Problems that currently seem intractable will not only become doable, they will become routine.

7.3. Slow Convergence of Basis Set Expansions. The slow convergence of molecular binding energies with basis set is a result of the fundamental inability of basis set expansions to represent the wave function as the interelectronic distance (r_{ij}) approaches zero. For a two-electron system, it can be shown that (un-normalized)[98]

$$\mathrm{Lim}_{r_{12}\to 0}\, \Psi(r_{12}) = 1 + {}^1/_2 r_{12} \tag{15}$$

This is a result of the singularity in the electron–electron interaction ($1/r_{12}$) when $r_{12} \rightarrow 0$ and is analogous to the electron–nuclear cusp condition. Expansions in a basis set of one-electron functions cannot reproduce this behavior. Although it is beyond the scope of the present paper to discuss many-electron approaches to the calculation of molecular binding energies (see the review by Klopper[99]), it is worthwhile to note that Kutzelnigg, Klopper, and co-workers[100] have developed an efficient scheme to include terms linear in r_{12} in the wave function expansion. Inclusion of linear r_{12} terms ensures that eq 15 is satisfied and substantially improves the rate of convergence of the calculations. Closure relations are used to eliminate the many-electron integrals that arise in r_{12}-dependent wave functions.

The coupled cluster equations in the linear r_{12} formalism, CC–R12, have been derived by Noga and Kutzelnigg[100j] and implemented in DIRCCR12 by Noga and Klopper[101] [see also ref 100k]. These authors have shown that CCSD(T)-R12 calculations can provide essentially converged results.[100,102] However, large basis sets, slightly larger than the aug-cc-pV5Z basis set, must be used to ensure the accuracy of the closure approximation. Nonetheless, the CCSD(T)-R12 method is a significant advance when high accuracy is sought. For example, for the four chemically bound molecules considered here, the convergence errors in calculations of D_e[CCSD(T)] with the aug-cc-pV5Z set are −0.18 (CH), −0.15 (HF), −0.70 (CO), and −1.45 kcal/mol (N_2). While the errors in D_e(CH) and D_e(HF) may be acceptable, those in D_eCO) and $D_e(N_2)$ are not.

8. Conclusions

The work presented in the previous sections clearly establishes that quantum chemists achieved a major advance in the ab initio calculation of molecular binding energies (D_e) in the 1990s. It is now possible to calculate accurate binding energies for a broad range of molecules, from those bound by very weak dispersion interactions to those bound by the strongest chemical interactions. This advance is a result of the development of new electronic structure methods (the coupled cluster methods) and basis sets (the correlation consistent basis sets) for solving the Schrödinger equation as well as a detailed understanding of the errors that arise in the calculation of binding energies. In this article we discussed the sources of the errors and then systematically analyzed the errors in the calculation of molecular binding energies for chemical interactions ("chemically bound molecules"), hydrogen-bonded interactions ("hydrogen-bonded molecules"), electrostatic interactions ("weakly bound molecules"), and dispersion interactions ("very weakly bound molecules"). The binding energies associated with these interactions range from hundreds of kcal/mol to a few hundredths of a kcal/mol—a variation of 4 orders of magnitude! This analysis leads to clear guidelines for the choice of electronic structure method and basis set to use in the calculation of molecular binding energies. Let us recap our findings.

The *electronic structure method error* in the binding energy is the error inherent in a given electronic structure method, i.e., it is the difference between the value of the binding energy computed with a complete basis set and the experimental binding energy. This error is also referred to as the intrinsic error for the method. In the current study we limited our considerations to valence-electron calculations; such calculations are generally capable of predicting molecular binding energies to an accuracy of 1/2 to 1%. The results from all-electron calculations were noted whenever they were available to quantify the magnitude of the core-valence correction. From the work presented here and elsewhere[38−41] we can draw the following conclusions.

Coupled Cluster Methods. The CCSD(T) method provides an accurate description of the electronic structure of all of the molecular systems considered. The calculated binding energies are in excellent agreement with the experimental values as follows. (a) For chemically bound molecules, hydrogen-bonded molecules, and weakly bound molecules, the intrinsic errors in binding energies (D_e) computed with the valence-electron CCSD(T) method are less than 1% (and often less than 0.5%). For chemically bound molecules, this is usually close to chemical accuracy, i.e., ±1 kcal/mol. (b) For very weakly bound molecules the error is somewhat larger; it is approximately 3% for the rare gas dimers. (c) Inclusion of core-valence and relativistic effects in calculations on strongly bound molecules reduces the intrinsic error in the CCSD(T) method to well under one kcal/mol.[38,46,103−105]

Inclusion of full triple excitations, the CCSDT method, in calculations of D_e for strongly bound molecules leads to slightly larger errors than does the CCSD(T) method.[87−89] Thus, the excellent agreement achieved for the CCSD(T) method is partly a result of compensating of errors. However, the observed differences are usually just a few tenths of a kcal/mol, although they can be larger for molecules that are poorly described by the HF wave function.

For weakly bound molecules, specifically He_2, the intrinsic error in the CCSDT method is only on the order of 0.1% of D_e.

Omission of the perturbative triples correction in the CCSD method substantially increases the error relative to that for the CCSD(T) method. The CCSD method does not provide accurate predictions of molecular binding energies, although it clearly provides semiquantitative trends.

Perturbation Theory Methods. Perturbation theory methods are unsatisfactory for describing strongly bound molecules as the perturbation series converges slowly or even diverges and does not yield chemical accuracy (intrinsic errors of less than 1 kcal/mol) even at fourth (or fifth) order perturbation theory. Perturbation theory methods appear to be useful for describing hydrogen-bonded molecules and may be useful for treating some weakly bound molecules (electrostatic interactions); even the MP2 method performs satisfactorily in these cases. Perturbation theory methods are marginally useful for describing the binding in very weakly bound molecules. The series converges very slowly, e.g., for He_2 the MP5 method, an N^8 method, is no more accurate that the CCSD(T) method, an N^7 method.

The *basis set convergence error* is the error in the molecular property calculated with a given basis set relative to that obtained with a complete basis set. The functional form of the error depends on the correlation method used. From the work presented here and elsewhere,[36,38−41] we can draw the following conclusions.

Correlation Consistent Basis Sets. Correlation consistent basis sets, cc-pV*n*Z, provide a hierarchy of basis sets with accuracy (i.e., completeness) systematically increasing with *n*. Binding energies computed with the sets appear to converge smoothly to the complete basis set limit and can be extrapolated to that limit with reasonable error bounds.

The basis set convergence error for the bond energies of different types of chemical bonds (covalent, ionic, single, double, triple, ...) differ significantly. Thus, for a general set of molecules, large basis sets will be required to achieve either high *absolute* or high *relative* accuracy. The convergence behavior of similar bonds in different molecules differ far less, and more computationally tractable basis sets can provide high relative accuracy for these bonds. The isodesmic reaction scheme can also be used to decrease the dependence on basis set.

Different families of correlation consistent basis sets are required to describe different types of molecular interactions. (a) Standard sets provide an excellent description of covalent molecules. (b) Augmented sets are required to efficiently describe highly ionic molecules, hydrogen-bonded molecules, and weakly bound molecules (i.e., those bound by electrostatic interactions). (c) Doubly augmented sets are required to describe very weakly bound molecules where the binding is mainly due to dispersion interactions.

Convergence to the complete basis set limit is fastest for strongly bound molecules with the aug-cc-pVTZ set yielding 96−98% of $D_e(\infty)$, and slightly slower for hydrogen-bonded molecules with the aug-cc-pVTZ set yielding 93−94% of $D_e(\infty)$. For weakly bound molecules an aug-cc-pVTZ set yields just 85−87% of $D_e(\infty)$, while a d-aug-cc-pVTZ set is required to achieve 85−87% of $D_e(\infty)$ for very weakly bound molecules.

A major source of confusion in assessing the accuracy of calculations was identified.[38b] The *calculational error*, which is the error for a given electronic structure method and basis set, is the sum of the electronic structure method error and the basis set convergence error. If the latter two quantities are of different sign, the errors may nearly cancel for some choice of basis set (*n*). This is often the case for MP2 and MP4 perturbation theory calculations. These methods overestimate D_e for many strongly bound molecules (three of the four considered here), and the resulting cancellation of the convergence error and the method error could lead the unsuspecting researcher to erroneously conclude that a given calculation is very accurate.

The approach outlined here for analyzing the requirements for the calculation of molecular binding energies can be readily extended to other molecular properties. In addition, it can be readily extended to electronic structure methods not considered here. For an excellent example of the application of these concepts to the calculation of atomization energies and reaction enthalpies, the reader is referred to a recent article by Bak et al.[106]

With the advent of parallel supercomputers capable of trillions of arithmetic operations per second and the implementation of new algorithms that take advantage of "reduced scaling" techniques, it will be possible to extend coupled cluster calculations to a broad range of molecules important in combustion chemistry, environmental chemistry, chemical vapor deposition, and industrial processing. This will extend the revolution begun in the 1990s to the real world of chemistry.

Acknowledgment. This work benefited from the efforts of many colleagues at the Pacific Northwest National Laboratory. In particular, I wish to thank Drs. David E. Woon and Kirk A. Peterson, who were involved in this work almost from the beginning and helped shape its evolution. Other colleagues who contributed significantly to the work described herein include Drs. Tanja van Mourik, Angela K. Wilson, David Feller, Ricky A. Kendall, and Robert J. Harrison. Discussions with Drs. David A. Dixon, Michel Dupuis, and Lawrence B. Harding are gratefully acknowledged. The manuscript benefited from the comments of Drs. Feller, Peterson, and Karl K. Irikura as well as the two referees. This work was supported by the Division of Chemical Sciences in the Office of Basis Energy Sciences of the U.S. Department of Energy at Pacific Northwest National Laboratory, a multiprogram national laboratory operated by Battelle Memorial Institute, under Contract No. DE-AC06-76RLO 1830. Computing resources were provided by the Division of Chemical Sciences and by the Division of Mathematical, Information, and Computational Sciences at the National Energy Research Scientific Computing Center (NERSC) at Lawrence Berkeley National Laboratory.

References and Notes

(1) P. A. M. Dirac, *Proc. R. Soc. A* **1929**, *123*, 714.

(2) Pauling, L. *The Nature of the Chemical Bond*; Cornell University Press: Ithaca, NY; 1939.

(3) (a) Luo, F.; McBane, G. C.; Kim, G.; Giese, C. F.; Gentry, W. R. *J. Chem. Phys.* **1993**, *98*, 3564. (b) Luo, F.; Giese, C. F.; Gentry, W. R. *J. Chem. Phys.* **1996**, *104*, 1151.

(4) Schöllkopf, W.; Toennies, J. P. *J. Chem. Phys.* **1996**, *104*, 1155.

(5) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98 (Revision A.9)*; Gaussian, Inc.: Pittsburgh, PA. For further information, see: http://www.Gaussian.com/index.htm.

(6) Amos, R. D.; Alberts, I. L.; Andrews, J. S.; Colwell, S. M.; Handy, N. C.; Jayatilaka, D.; Knowles, P. J.; Kobayashi, R.; Laidig, K. E.; Laming, G.; Lee, A. M.; Maslen, P. E.; Murray, C. W.; Rice, J. E.; Simandiras, E. D.; Stone, A. J.; Su, M.-D.; Tozer, D. J. *CADPAC: The Cambridge Analytic Derivatives Package Issue 6, Cambridge*, 1995. For further information, see: http://ket.ch.cam.ac.uk/software/cadpac.html.

(7) Schmidt, M. A.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347. For further information, see: http://www.msg.ameslab.gov/GAMESS/GAMESS.html.

(8) Werner, H.-J.; Knowles, P. J.; Amos, R. D.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Leininger, T.; Lindh, R.; Lloyd, A. W.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Peterson, K. A.; Pitzer, R. M.; Pulay, P.; Rauhut, G.; Schütz, M.; Stoll, H.; Stone, A. J.; Thorsteinsson, T. *MOLPRO Quantum Chemistry Package 2000.1*, **2000**. For further information, see: http://www.tc.bham.ac.uk/molpro/.

(9) Stanton, J. F.; Gauss, J.; Watts, J. D.; Nooijen, M.; Oliphant, N.; Perera, S. A.; Szalay, P. G.; Lauderdale, W. J.; Kucharski, S. A.; Gwaltney, S. R.; Beck, S.; Balková, A.; Bernholdt, D. E.; Baeck, K. K.; Rozyczko, P.; Sekino, H.; Hober, C.; Bartlett, R. J. Integral packages included are VMOL (Almlöf, J.; Taylor, P. R.); VPROPS (Taylor, P. R.), ABACUS; (Helgaker, T.; Aa. Jensen, H. J.; Jørgensen, P.; Olsen, J.; Taylor, P. R.). *ACES II*. For further information, see: http://www.qtp.ufl.edu/Aces2/.

(10) Helgaker, T.; Aa. Jensen, H. J.; Jørgensen, P.; Olsen, J.; Ruud, K.; Ågren, H.; Andersen, T.; Bak, K. L.; Bakken, V.; Christiansen, O.; Dahle, P.; Dalskov, E. K.; Enevoldsen, T.; Fernandez, B.; Heiberg, H.; Hettema, H.; Jonsson, D.; Kirpekar, S.; Kobayashi, R.; Koch, H.; Mikkelsen, K. V.; Norman, P.; Packer, M. J.; Saue, T.; Taylor, P. R.; Vahtras, O. *DALTON, An Ab Initio Electronic Structure Program, Release 1.0.*, 1997. For further information, see: http://www.kjemi.uio.no/software/dalton/dalton.html.

(11) The author and his colleagues have studied a number of energy related properties (ionization potentials, electron affinities, proton affinities, etc.) as well as spectroscopic properties (bond lengths, vibrational frequencies, etc.) using this approach. A closely related study of molecular dipole moments has recently been reported in Halkier, A.; Klopper, W.; Helgaker, T.; Jørgensen, P.; Taylor, P. R. *J. Chem. Phys.* **1999**, *111*, 4424, although only BH and HF were explicitly considered in the paper.

(12) A third error type, in which the computed value starts above or below the experimental value at $n = 2$ and then converges away from the experimental result, exists but is rarely observed.

(13) Pople, J. A.; Binkley, J. S.; Seeger, R. *Int. J. Quantum Chem. Symp.* **1976**, *10*, 1.

(14) Bartlett, R. J.; Purvis, G. D. *Int. J. Quantum Chem.* **1978**, *14*, 561.

(15) (a) Liu, B.; McLean, A. D. *J. Chem. Phys.* **1980**, *72*, 3418. (b) Lengsfield, B. H.; McLean, A. D.; Yoshimine, M.; Liu, B. *J. Chem. Phys.* **1983**, *79*, 1891. (c) Liu, B.; McLean, A. D. *J. Chem. Phys.* **1989**, *91*, 2348.

(16) The author extended the approach of Liu and McLean to chemically bound species in the early 1980s. In hindsight, this generalization was quite successful, although it was not recognized as such at the time because of the magnitude of the basis set convergence error.

(17) Cremer, D.; He, Z. *J. Phys. Chem.* **1996**, *100*, 6173.

(18) (a) Knowles, P. J.; Somasundram, K.; Handy, N. C. *Chem. Phys. Lett.* **1985**, *113*, 8. (b) Handy, N. C.; Knowles, P. J.; Somasundram, K. *Theor. Chim. Acta* **1985**, *68*, 87.

(19) Gill, P. M. W.; Radom, L. *Chem. Phys. Lett.* **1986**, *132*, 16.

(20) (a) Olsen, J.; Christiansen, O.; Kock, H.; Jørgensen, P. *J. Chem. Phys.* **1996**, *105*, 5082. (b) Christiansen, O.; Olsen, J.; Jorgensen, P.; Koch, H.; Malmqvist, P.-A. *Chem. Phys. Lett.* **1996**, *261*, 369.

(21) Dunning, T. H., Jr.; Peterson, K. A. *J. Chem. Phys.* **1998**, *108*, 4761.

(22) Leininger, M. L.; Allen, W. D.; Schaefer, H. F., III.; Sherrill, C. D. *J. Chem. Phys.* **2000**, *112*, 9213.

(23) Stillinger, F. H. *J. Chem. Phys.* **2000**, *112*, 9711.

(24) (a) Coseter, F. *Nucl. Phys.* **1958**, *7*, 421. (b) Coester, F.; Kümmel, H. *Nucl. Phys.* **1960**, *17*, 477. (c) Kümmel, H. *Nucl. Phys.* **1969**, *22*, 177.

(25) (a) Cizek, J. *J. Chem. Phys.* **1966**, *45*, 4256. (b) Cizek, J. *Adv. Chem. Phys.* **1969**, *14*, 35. (c) Paldus, J.; Cizek, J.; Shavitt, I. *Phys. Rev.* **1972**, *A5*, 50.

(26) (a) Bartlett, R. J.; Purvis, G. D. *Int. J. Quantum Chem.* **1978**, *14*, 561. (b) Bartlett, R. J.; Purvis, G. D. *Phys. Scr.* **1980**, *21*, 251. (c)

(27) Purvis, G. D.; Bartlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910.

(28) (a) Lee, Y. S.; Bartlett, R. J. *J. Chem. Phys.* **1984**, *80*, 4371. (b) Lee, Y. S.; Kucharski, S. A.; Bartlett, R. J. *J. Chem. Phys.* **1984**, *81*, 5906. (c) Lee, Y. S.; Kucharski, S. A.; Bartlett, R. J. *J. Chem. Phys.* **1985**, *82*, 5761. (d) Urban, M.; Noga, J.; Cole, S. J.; Bartlett, R. J. *J. Chem. Phys.* **1985**, *83*, 4041. (e) Kacharski, S. A.; Bartlett, R. J. *Adv. Quantum Chem.* **1986**, *18*, 281. (f) Noga, J.; Bartlett, R. J. *J. Chem. Phys.* **1987**, *86*, 7041; Erratum, *J. Chem. Phys.* **1988**, *89*, 3401.

(29) (a) Noga, J.; Bartlett, R. J. *J. Chem. Phys.* **1987**, *86*, 7041. (b) Noga, J.; Bartlett, R. J. *J. Chem. Phys.* **1988**, *89*, 3401.

(30) Kucharski, S. A.; Bartlett, R. J. *J. Chem. Phys.* **1992**, *97*, 4282.

(31) (a) Sinanoglu, O. *J. Chem. Phys.* **1962**, *36*, 706. (b) Sinanoglu, O. *J. Chem. Phys.* **1962**, *36*, 3198. (c) Sinanoglu, O. *Adv. Chem. Phys.* **1964**, *6*, 315.

(32) van Mourik, T.; Dunning, T. H., Jr. *J. Chem. Phys.* **1999**, *111*, 9248

(33) Despite the computational cost of CCSDT calculations, D. Feller and co-workers have reported such calculations on molecules as large as propane using up to triple-ζ basis sets; Feller, D., private communication.

(34) (a) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479. (b) Raghavachari, K.; Pople, J. A.; Replogle, E. S.; Head-Gordon, M. *J. Phys. Chem.* **1990**, *94*, 5579.

(35) Bartlett, R. J. *J. Phys. Chem.* **1989**, *93*, 697.

(36) (a) Scuseria, G. E.; Lee, T. J. *J. Chem. Phys.* **1990**, *93*, 5851. (b) Lee, T. J.; Scuseria, G. E. In *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy*; Langhoff, S. R. Ed.; Kluwer Academic Publishers: Dordrecht, 1997.

(37) Stanton, J. F. *Chem. Phys. Lett.* **1997**, *281*, 130.

(38) (a) Peterson, K. A.; Dunning, T. H., Jr. *J. Chem. Phys.* **1995**, *102*, 2032. (b) Peterson, K. A.; Dunning, T. H., Jr. *J. Phys. Chem.* **1995**, *99*, 3898. (c) Woon, D. E.; Dunning, T. H., Jr.; Peterson, K. A. *J. Chem. Phys.* **1996**, *104*, 5883. (d) Xantheas, S. S.; Dunning, T. H., Jr.; Mavridis, A. *J. Chem. Phys.* **1997**, *106*, 3280. (e) Peterson, K. A.; Dunning, T. H., Jr. *J. Chem. Phys.* **1997**, *106*, 4119. (f) Peterson, K. A.; Dunning, T. H., Jr. *J. Mol. Struct. (Theochem)* **1997**, *400*, 93. (g) van Mourik, T.; Dunning, T. H., Jr. *J. Chem. Phys.* **1997**, *107*, 2451. (h) Peterson, K. A.; Wilson, A. K.; Woon, D. E.; Dunning, T. H., Jr. *Theor. Chem. Acc.* **1997**, *97*, 251. (i) Woon, D. E.; Peterson, K. A.; Dunning, T. H., Jr. *J. Chem. Phys.* **1998**, *109*, 2233. (j) van Mourik, T.; Wilson, A. K.; Dunning, T. H., Jr. *Mol. Phys.* **1999**, *96*, 529.

(39) (a) Martin, J. M. L.; Taylor, P. R. *Chem. Phys. Lett.* **1996**, *248*, 336. (b) Martin, J. M. L.; Lee, T. J. *Chem. Phys. Lett.* **1996**, *258*, 129. (c) Martin, J. M. L.; Lee, T. J. *Chem. Phys. Lett.* **1996**, *258*, 136. (d) Martin, J. M. L. *Chem. Phys. Lett.* **1996**, *259*, 669. (e) Martin, J. M. L. *Chem. Phys. Lett.* **1996**, *259*, 679. (f) Martin, J. M. L. In *Computational Thermochemistry. Prediction and Estimation of Molecular Thermodynamics*; Irikura, K. K., Frurip, D. J., Eds; American Chemical Society: Washington, DC, 1996. (g) Martin, J. M. L.; Taylor, P. R. *J. Chem. Phys.* **1997**, *106*, 8620. (h) Martin, J. M. L. *Theor. Chem. Acc.* **1997**, *97*, 227. (i) Martin, J. M. L. *Chem. Phys. Lett.* **1997**, *273*, 98. (j) Martin, J. M. L. *Chem. Phys. Lett.* **1998**, *292*, 411. (k) Martin, J. M. L.; Lee, T. J.; Taylor, P. R. *J. Chem. Phys.* **1998**, *108*, 676.

(40) (a) Halkier, A.; Christiansen, O.; Sundholm, D.; Pyykkoo, P. *Chem. Phys. Lett.* **1997**, *271*, 273. (b) Halkier, A.; Jørgensen, P.; Gauss, J.; Helgaker, T. *Chem. Phys. Lett.* **1997**, *274*, 235. (c) Helgaker, T.; Gauss, J.; Jørgensen, P.; Olsen, J. *J. Chem. Phys.* **1997**, *106*, 6430. (d) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J. Chem. Phys.* **1997**, *106*, 9639. (e) Halkier, A.; Koch, H.; Jørgensen, P.; Christiansen, O.; Beck Nielsen, I. M.; Helgaker, T. *Theor. Chem. Acc.* **1997**, *97*, 150. (f) Halkier, A.; Taylor, P. R. *Chem. Phys. Lett.* **1998**, *285*, 133. (g) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243. (h) Halkier, A.; Coriani, S.; Jørgensen, P. *Chem. Phys. Lett.* **1998**, *294*, 292. (i) Halkier, A.; Larsen, H.; Olsen, J.; Jørgensen, P. *J. Chem. Phys.* **1999**, *110*, 734. (j) Halkier, A.; Coriani, S. *Chem. Phys. Lett.* **1999**, *303*, 408. (k) Halkier, A.; Klopper, W.; Helgaker, T.; Jørgensen, P.; Taylor, P. R. *J. Chem. Phys.* **1999**, *111*, 9157.

(41) Feller, D.; Peterson, K. A. *J. Chem. Phys.* **1998**, *108*, 154.

(42) Heitler H.; London, F. *Z. Phys.* **1927**, *44*, 455.

(43) For a more complete discussion of correlation consistent basis sets see: Dunning, T. H., Jr.; Peterson, K. A.; Woon, D. E. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed; John Wiley & Sons Ltd.: New York, 1998.

(44) (a) Almlöf, J.; Taylor, P. R. *J. Chem. Phys.* **1987**, *86*, 4070. (b) Almlöf, J.; Taylor, P. R. *J. Chem. Phys.* **1990**, *92*, 551.

(45) (a) Widmark, P.-O.; Malmqvist, P.-A.; Roos, B. O. *Theor. Chim. Acta* **1990**, *77*, 291. (b) Widmark, P.-O.; Persson, B. J.; Roos, B. O. *Theor. Chim. Acta* **1991**, *79*, 419.

(46) Bauschlicher, C. W., Jr. *J. Phys. Chem. A* **1998**, *102*, 10424.

(47) (a) Bauschlicher, C. W., Jr.; Taylor, P. R. *Theor. Chim. Acta* **1993**, *86*, 13. (b) Bauschlicher, C. W., Jr. *Theor. Chim. Acta* **1995**, *92*, 183.

(48) Pou-Amerigo, R.; Merchan, M.; Nebot-Gil, I.; Widmark, P.-O.; Roos, B. O. *Theor. Chim. Act* **1995**, *92*, 149.

(49) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.

(50) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.

(51) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1994**, *100*, 2975.

(52) (a) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1995**, *103*, 4572. (b) Wilson, A. K.; van Mourik, T.; Dunning, T. H., Jr. *J. Mol. Struct. (THEOCHEM)* **1996**, *388*, 339.

(53) (a) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1993**, *98*, 1358. (b) van Mourik, T.; Dunning, T. H., Jr. *Int. J. Quantum Chem.* **2000**, *76*, 205.

(54) Wilson, A. K.; Woon, D. E.; Peterson, K. A.; Dunning, T. H., Jr. *J. Chem. Phys.* **1999**, *110*, 7667.

(55) Knowles, P. J.; Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **1994**, *99*, 5219.

(56) (a) Feller, D. *J. Chem. Phys.* **1992**, *96*, 6104. (b) Feyereisen, M. W.; Feller, D.; Dixon, D. A. *J. Phys. Chem.* **1996**, *100*, 2993.

(57) Fast, P. L.; Sanchez, L.; Truhlar, D. G. *J. Chem. Phys.* **1999**, *111*, 2921.

(58) Wilson, A. K.; Dunning, T. H., Jr. *J. Chem. Phys.* **1997**, *107*, 8718.

(59) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(60) Liu, B.; McLean, A. D. *J. Chem. Phys.* **1973**, *59*, 4557.

(61) There have been extensive discussions in the literature about the ability of the counterpoise correction to correct for basis set superposition error. Be this as it may, it is clear that use of the counterpoise correction significantly improves the convergence behavior of molecular properties computed with the correlation consistent basis sets.[62]

(62) van Mourik, T.; Wilson, A. K.; Peterson, K. A.; Woon, D. E.; Dunning, T. H., Jr. In *Advances in Quantum Chemistry. Quantum Systems in Chemistry and Physics, Part I*; Löwdin, P.-O., Ed.; Academic Press: New York, 1999.

(63) Huber, K. P.; Herzberg, G. *Molecular Spectra and Molecular Structure IV. Constants of Diatomic Molecules*; Van Nostrand: Princeton, 1979.

(64) Dunning, T. H., Jr.; Peterson, K. A., to be published.

(65) (a) Quack, M.; Suhm, M. A. *Mol. Phys.* **1990**, *69*, 791. (b) Quack, M.; Suhm, M. A. *Chem. Phys. Lett.* **1991**, *183*, 187. (c) Quack, M.; Suhm, M. A. *J. Chem. Phys.* **1991**, *95*, 28. (d) Klopper, W.; Quack, M.; Suhm, M. A. *Chem. Phys. Lett.* **1996**, *261*, 35. (e) Quack, M.; Suhm, M. A. *Theor. Chim. Acta* **1996**, *93*, 61. (f) Klopper, W.; Quack, M.; Suhm, M. A. *J. Chem. Phys.* **1998**, *108*, 10096.

(66) (a) Von Puttkamer, K.; Quack, M. *Mol. Phys.* **1987**, *62*, 1047. (b) Von Puttkamer, K.; Quack, M.; Suhm, M. A. *Mol. Phys.* **1988**, *65*, 1025. (c) Von Puttkamer, K.; Quack, M.; Suhm, M. A. *Infrared Phys.* **1989**, *29*, 535. (d) Von Puttkamer, K.; Quack, M. *Chem. Phys.* **1989**, *139*, 31. (e) Quack, M.; Suhm, M. A. *Chem. Phys. Lett.* **1990**, *171*, 517.

(67) Cayton, D. C.; Jucks, K. W.; Miller, R. E. *J. Chem. Phys.* **1989**, *90*, 2631.

(68) (a) Anderson, D. T.; Davis, S.; Nesbitt, D. J. *J. Chem. Phys.* **1996**, *104*, 6225. (b) Davis, S.; Anderson, D. T.; Farrell, J. T., Jr.; Nesbitt, D. J. *J. Chem. Phys.* **1996**, *104*, 8197. (c) Farrell, J. T.; Suhm, M. A.; Nesbitt, D. J. *J. Chem. Phys.* **1996**, *104*, 9313. (d) Anderson, D. T.; Davis, S.; Nesbitt, D. J. *J. Chem. Phys.* **1996**, *105*, 4488. (e) Davis, S.; Anderson, D. T.; Nesbitt, D. J. *J. Chem. Phys.* **1996**, *105*, 6645.

(69) (a) Howard, B. J.; Dyke, T. R.; Klemperer, W. *J. Chem. Phys.* **1984**, *81*, 5417. (b) Chang, H.-C.; Klemperer, W. *J. Chem. Phys.* **1993**, *98*, 9266. (c) Chang, H.-C.; Klemperer, W. *J. Chem. Phys.* **1994**, *100*, 1. (d) Chang, H.-C.; Klemperer, W. *J. Chem. Phys.* **1996**, *104*, 7830.

(70) (a) Curtiss, L. A.; Frurip, D. J.; Blander, M. *J. Chem. Phys.* **1979**, *71*, 2703.

(71) Reimers; Watts, R.; Klein, M. *Chem. Phys.* **1982**, *64*, 95.

(72) Peterson, K. A.; Dunning, T. H., Jr. *J. Chem. Phys.* **1995**, *102*, 2032.

(73) (a) Halkier, A.; Klopper, W.; Helgaker, T.; Jorgensen, P.; Taylor, P. R. *J. Chem. Phys.* **1999**, *111*, 9157. (b) Klopper, W.; Luthi, H. P. *Mol. Phys.* **1999**, *96*, 559.

(74) Klopper, W.; Quack, M.; Suhm, M. A. *Mol. Phys.*, in press.

(75) Schütz, M.; Brdarski, S.; Widmark, P.-O.; Lindh, R.; Karlström, G. *J. Chem. Phys.* **1997**, *107*, 4597.

(76) Soper, P. D.; Legon, A. C.; Read, W. G.; Flygare, W. H. *J. Chem. Phys.* **1982**, *76*, 292. Kolenbrander, K. D.; Lisy, J. M. *J. Chem. Phys.* **1985**, *85*, 2463. Kucks, K. W.; Huang, Z. S.; Miller, R. E. *J. Chem. Phys.* **1987**, *86*, 1098. Lovejoy, C. M.; Nesbitt, D. J. *J. Chem. Phys.* **1987**, *86*, 3151. Nesbitt, D. J.; Child, M. S. *J. Chem. Phys.* **1993**, *98*, 478. Nesbitt, D. J.; Lindeman, T. G.; Farrell, J. T., Jr.; Lovejoy, C. M. *J. Chem. Phys.* **1994**, *100*, 775. Bemish, R. J.; Bohac, E. J.; Wu, M.; Miller, R. E. *J. Chem. Phys.* **1994**, *101*, 9457. Farrell, J. T.; Sneh, O.; Nesbitt, D. J. *J. Phys. Chem.* **1994**, *98*, 6068. Tang, S. N.; Chang, H.-C.; Klemperer, W. *J. Phys. Chem.* **1994**, *98*, 7313.

(77) Hutson, J. M. *J. Chem. Phys.* **1992**, *96*, 6752 and references therein.

(78) Hutson, J. M. *J. Chem. Phys.* **1988**, *89*, 4550. Hutson, J. M. *J. Phys. Chem.* **1992**, *96*, 4237 and references therein.

(79) (a) He_2: Aziz, R. A.; Slaman, M. J. *J. Chem. Phys.* **1991**, *94*, 8047. Aziz, R. A.; Janzen, A. R.; Moldover, R. *Phys. Rev. Lett.* **1995**, *74*, 1586. (b) Ne_2: Aziz, R. A.; Meath, W. J.; Allnatt, A. R. *Chem. Phys.* **1983**, *78*, 295. Aziz, R. A.; Slaman, M. J. *Chem. Phys.* **1989**, *130*, 187. (c) Ar_2: Aziz, R. A.; Slaman, M. J. *Mol. Phys.* **1986**, *58*, 679. Aziz, R. A. *J. Chem. Phys.* **1993**, *99*, 4518.

(80) Woon, D. E. *J. Chem. Phys.* **1994**, *100*, 2838; Woon, D. E. *Chem. Phys. Lett.* **1993**, *204*, 29.

(81) Burda, J. V.; Zahradnik, R.; Hobza, P.; Urban, M. *Mol. Phys.* **1996**, *89*, 425.

(82) Briggs, M. P.; Murrell, J. N.; Stamper, J. G. *Mol. Phys.* **1969**, *17*, 381. Riemenschneider, B. M.; Kestner, N. R. *Chem. Phys.* **1974**, *3*, 193. Maeder, F.; Kutzelnigg, W. *Chem. Phys. Lett.* **1976**, *37*, 285. Kutzelnigg, W. *Faraday Discuss. Chem. Soc.* **1977**, *62*, 185. Meyer, W.; Hariharan, P. C.; Kutzelnigg, W. *J. Chem. Phys.* **1980**, *73*, 1880.

(83) Dunning, T. H., Jr.; Peterson, K. A. *J. Chem. Phys.*, submitted.

(84) (a) Hehre, W. J.; Ditchfield, R.; Radom, L.; Pople, J. A. *J. Am. Chem. Soc.* **1970**, *92*, 4796. (b) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(85) See: Del Bene, J. E.; Shavitt, I. In *Molecular Interactions: From van der Waals to Strongly Bound Complexes*; S. Scheiner, Ed.; Wiley: Chichester, 1997.

(86) (a) CCSDT: Urban, M.; Noga, J.; Cole, S. J.; Bartlett, R. J. *J. Chem. Phys.* **1985**, *83*, 4041. (b) CCSDTQ: Kucharski, S. A.; Bartlett, R. J. *J. Chem. Phys.* **1992**, *97*, 4282.

(87) Feller, D. *J. Chem. Phys.* **1999**, *111*, 4373.

(88) (a) Feller, D.; Sordo, J. A. *J. Chem. Phys.* **2000**, *112*, 5604. (b) Feller, D.; Sordo, J. A. *J. Chem. Phys.* **2000**, *113*, 485.

(89) Bak, K. L.; Jørgensen, P.; Olsen, J.; Helgaker, T.; Gauss, J. *Chem. Phys. Lett.* **2000**, *317*, 116.

(90) Kucharski, S. A.; Bartlett, R. J. *J. Chem. Phys.* **1991**, *95*, 8227.

(91) Jeriorski, B.; Monkhorst, H. *J. Phys. Rev. A* **1981**, *24*, 1668.

(92) Li, X.; Paldus, J. *J. Chem. Phys.* **1997**, *107*, 6257.

(93) Kendall, R. A.; Aprà, E.; Bernholdt, D. E.; Bylaska, E. J.; Dupuis, M.; Fann, G. I.; Harrison, R. J.; Ju, J.; Nichols, J. A.; Nieplocha, J.; Straatsma, T. P.; Windus, T. L.; Wong, A. T. *Comput. Phys. Commun.* **2000**, *128*, 260 and references therein. For further information, see: http://www.emsl.pnl.gov: 2080/docs/nwchem/nwchem.html.

(94) Ochsenfeld, C.; White, C. A.; Head-Gordon, M. *J. Chem. Phys.* **1998**, *109*, 1663.

(95) Scuseria, G.; Ayala, P. Y. *J. Chem. Phys.* **1999**, *111*, 8330.

(96) For a review of local approaches, see: Saebo, S.; Pulay, P. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213.

(97) (a) Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **1996**, *104*, 6286. (b) Schütz, M.; Werner, H.-J. *Chem. Phys. Lett.* **2000**, *318*, 370.

(98) Kato, T. *Commun. Pure Appl. Math.* **1957**, *10*, 151.

(99) Klopper, W. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Screiner, P. R. Eds; Wiley: Chichester; 1998.

(100) (a) Klopper, W.; Kutzelnigg, W. *Chem. Phys. Lett.* **1987**, *134*, 17. (b) Klopper, W.; Kutzelnigg, W. *J. Phys. Chem.* **1990**, *94*, 5625. (c) Termath, V.; Klopper, W.; Kutzelnigg, W. *J. Chem. Phys.* **1991**, *94*, 2002. (d) Klopper, W.; Kutzelnigg, W. *J. Chem. Phys.* **1991**, *94*, 2020. (e) Klopper, W.; Rohse, R.; Kutzelnigg, W. *Chem. Phys. Lett.* **1991**, *178*, 455. (f) Klopper, W. *Chem. Phys. Lett.* **1991**, *186*, 583. (g) Klopper, W.; Almlof, J. *J. Chem. Phys.* **1993**, *99*, 5167. (h) Klopper, W. *J. Chem. Phys.* **1995**, *102*, 6168. (i) Noga, J.; Klopper, W.; Kutzelnigg, W. *Chem. Phys. Lett.* **1992**, *199*, 497. (j) Noga, J.; Kutzelnigg, W. *J. Chem. Phys.* **1994**, *101*, 7738. (k) Noga, J.; Klopper, W.; Kutzelnigg, W. In *Modern Ideas in Coupled-Cluster Methods*; Bartlett, R. J., Ed.; World Scientific: Singapore; 1997.

(101) Noga, J. A.; Klopper, W. *DIRCCR12* **1995**, unpublished.

(102) Klopper, W.; Bak, K. L.; Jørgensen, J.; Olsen, J.; Helgaker, T. *J. Phys. B* **1999**, *32*, R103.

(103) (a) Feller, D.; Glendening, E. D.; de Jong, W. A. *J. Chem. Phys.* **1999**, *110*, 1475. (b) Feller, D.; Peterson, K. A. *J. Chem. Phys.* **1999**, *110*, 8384. (c) Dixon, D. A.; Feller, D.; Sandrone, G. *J. Phys. Chem. A* **1999**, *103*, 4744. (d) Feller, D.; Dixon, D. A. *J. Phys. Chem. A* **1999**, *103*, 6413.

(104) (a) Ricca, A.; Bauschlicher, C. W., Jr. *J. Phys. Chem. A* **1999**, *103*, 11121. (b) Bauschlicher, C. W., Jr. *J. Phys. Chem. A* **1999**, *103*, 11126.

(105) Bauschlicher, C. W., Jr.; Martin, J. M. L.; Taylor, P. R. *J. Phys. Chem. A* **1999**, *103*, 7715.

(106) Bak, K. L.; Jørgensen, P.; Helgaker, T.; Klopper, W. *J. Chem. Phys.* **2000**, *112*, 9229.