9062 *J. Phys. Chem. A* **2000,** *104,* 9062−9080

# 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

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

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

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 $\Delta D_e^{\mathrm{bs}}(\mathrm{M};n)$ 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

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

where $D_e(\mathrm{M};\infty)$ is defined above and $D_e(\mathrm{expt'l})$ is the experimental value of the binding energy. $\Delta D_e^{\mathrm{M}}$ 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

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

We sometimes refer to $\Delta D_e^{\mathrm{calc}}(\mathrm{M};n)$ 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:

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

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 $\Delta D_e^{\mathrm{calc}}(\mathrm{M};n)$ is always greater than $\Delta D_e^{\mathrm{M}}$. For Type II, on the other hand, the calculated $D_e$ overshoots the experimental $D_e$ for sufficiently large $n$. In this case, $\Delta D_e^{\mathrm{calc}}(\mathrm{M};n)$ can be less than $\Delta D_e^{\mathrm{M}}$ for some $n$. In fact, in the figure $\Delta D_e^{\mathrm{calc}}(\mathrm{M};n) \approx 0$ 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.

## 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.

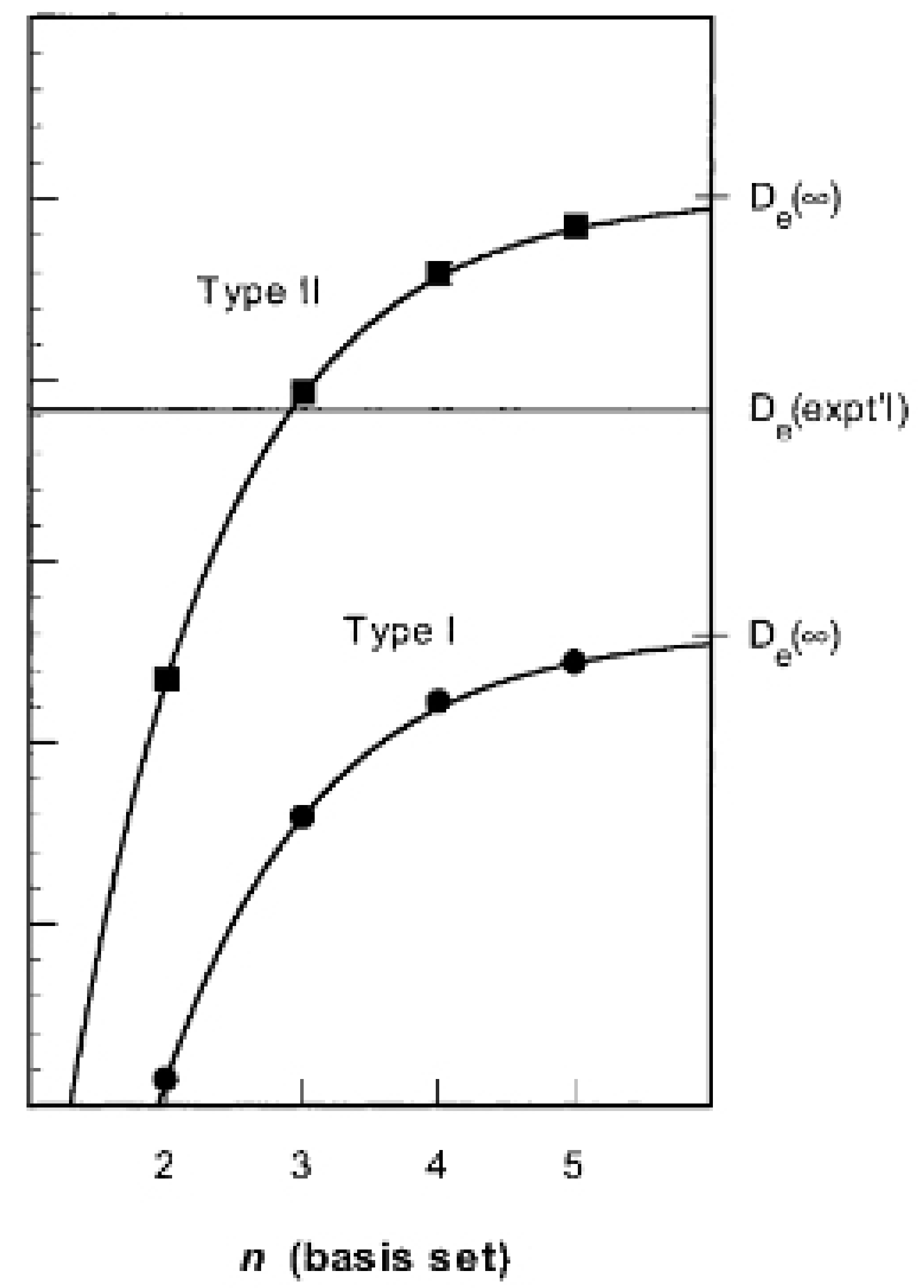

**Figure 1.** Type I and II errors for the calculation of molecular binding energies. $D_e(\infty)$ refers to the binding energy at the complete basis set limit.

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

$$\mathrm{H} = \mathrm{H}_0 + \lambda \mathrm{H}_1 \qquad (5)$$

where $\mathrm{H}_0$ is the Hartree−Fock Hamiltonian and $\lambda \mathrm{H}_1$, 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 $\lambda$

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

$$\mathrm{E} = \mathrm{E}_0 + \lambda \mathrm{E}_1 + \lambda^2 \mathrm{E}_2 + \lambda^3 \mathrm{E}_3 + \lambda^4 \mathrm{E}_4 + \ldots \qquad (7)$$

and the wave function and energy are computed order by order. For sufficiently small perturbations, only $\Psi_0$ (the Hartree−Fock wave function) and $\Psi_1$ (the first-order wave function) are important. For larger perturbations, $\Psi_2$, $\Psi_3$, ... must also be taken into account.

A wave function through $n$th order in perturbation theory is sufficient to calculate the energy to $(2n+1)$th order. Thus, $\mathrm{E}_0 + \mathrm{E}_1$ 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, $\Psi_0$ and $\Psi_1$ determine both $\mathrm{E}_2$(MP2) and $\mathrm{E}_3$(MP3); $\Psi_0$, $\Psi_1$, and $\Psi_2$