PC Calculations Using Gaussian for Windows: A Complement to Laboratory Measurements on HCl

3 AUTHORS, INCLUDING:

Darren L. Williams
Sam Houston State University
28 PUBLICATIONS **79** CITATIONS

SEE PROFILE

Joseph W. Nibler
Oregon State University
105 PUBLICATIONS **2,016** CITATIONS

SEE PROFILE

PC Calculations Using Gaussian for Windows

A Complement to Laboratory Measurements on HCl

Darren L. Williams, Philip R. Minarik, and Joseph W. Nibler
Department of Chemistry
Oregon State University
Corvallis, OR 97331-4003

Several articles in this *Computer Series* have illustrated the increasingly important role of computational chemistry in the undergraduate curriculum. For example, Pietro recently described the use of the program SPARTAN on a workstation to compute at the semiempirical AM1 level solvation effects on the deprotonation of nitromethane (*1*). Duke and O'Leary have discussed ab initio calculations on molecular hydrogen using the program Gaussian on a microVAX computer (*2*) and give explicit examples of the input statements necessary to run this program for a range of basis sets. Further examples of chemical applications of Gaussian are contained in the books by Foresman and Frisch (*3*) and by Hehre et al. (*4*).

The increased speed and memory capabilities of desktop PCs and the availability of sophisticated PC software at reasonable cost now allow even broader use of quantum mechanical programs. The purpose of this note is to show an application of the Gaussian for Windows program to a calculation of the structural and electronic properties of the diatomic molecule HCl. This molecule is particularly relevant to most physical chemistry laboratory courses, since these commonly include an experiment in which the vibrational–rotational infrared spectra of HCl and DCl are analyzed to obtain vibrational frequencies and the bond length (*5*). In addition, experiments are often done on HCl to obtain the dipole moment from a dielectric constant measurement and the polarizability from a refractive index measurement using a simple interferometer and a helium neon laser (*5, 6*). We have found that a combination of these measurements with a theoretical calculation of the same quantities using Gaussian works exceptionally well in illustrating the strengths (and weaknesses) of ab initio computations. The direct relation to

laboratory measurements heightens the student interest and motivation to learn more about the "innards" of the calculation. The fact that the exercise can now be done in reasonable times on local PC computers with which students are familiar (rather than on a remote main frame computer or expensive work station) makes the experience even more meaningful.

These calculations were done as a group student project this past year in our Junior/Senior integrated laboratory sequence at Oregon State University. The students synthesized DCl and analyzed its spectrum and combined the results with a similar analysis of HCl to obtain the equilibrium structural parameters R_e and ω_e. The analysis also gave the anharmonicity parameter $\omega_e x_e$, from which an estimate of the bond dissociation energy was made ($D_e = \frac{1}{4}\omega_e^2/\omega_e x_e$). Separate dielectric constant and refractive index measurements on HCl gave the dipole moment μ and the mean polarizability α_0. Although the latter quantities are really averages over a thermal distribution of states, rather than the equilibrium values obtained in the theoretical calculations, this difference is small and can be ignored in making comparisons.

Fifteen different calculations (Table 1) were done to explore possible improvement in calculated values of R_e, ω_e, m_e, and α_0 as the basis set size and level of the calculation were increased. The notation used for the Gaussian basis sets is "standard" and is defined in refs *3* and *4*. Briefly, the terminology STO-3G for the minimal basis set means that each of the hydrogenic Slater Type Orbitals (1s, 2s, $2p_x$, etc.) is replaced by three Gaussian functions. The latter have the advantage that the many integrations necessary in calculating molecular properties are done much faster for Gaussian functions of the form $\exp(-\zeta r^2)$ than is the case for Slater type orbitals of the form $\exp(-\zeta r)$.

To obtain an improved representation of the size of the electron distribution in the valence (outermost) orbitals, additional Gaussians with a different effective charge (zeta) can be added. The 3-21G notation indicates two zeta values (2s, 2s', $2p_x$, $2p_x'$, etc.) while the 6-31G label corresponds to six Gaussians for the 1s orbital and triple zeta values for the valence orbitals. Even larger (diffuse) orbitals are useful for systems with lone pair electrons, anions, and excited states; a "+" denotes addition of these orbitals to heavy atoms and "++" means that diffuse functions have been added to hydrogen atoms as well.

Finally, the shape of an orbital can be changed by adding orbitals of higher angular momentum, a procedure

Table 1. Gaussian 92 Times (Hours) to Calculate R_e, ω_e, μ_e, and α_o for HCl

Basis Set	Comments	# Functions	Level of Calculation			
			HF	MP2	MP3	QCISD(T)
STO-3G	Single zeta orbitals	10	0.4			
3-21G	Double zeta orbitals	15	0.4			
6-31G(d) ≡ 6-31G*	Adds d polarization to Cl	21	0.9	1.1		
6-311G(d,p) ≡ 6-311G**	Triple zeta, adds p polarization to H	32	1.5	2.9	2.5	
6-311+G(2d,2p)	More polarization, diffuse Cl orbitals	44	2.8	5.6	3.9	15.5
6-311++G(3df,3pd)	More polarization, diffuse Cl, H orbitals	65	8.7	33.8	11.9	74.7

said to "polarize" the orbital. To illustrate the notation, addition of two d and one f functions to heavy atom orbitals and two p and one d orbitals to hydrogen would be indicated by the (2df, 2pd) extension on the basis set description. The concept of polarization is analogous to mixing of orbitals to form sp³ hybrids from s and p orbitals in valence bond theory; both procedures result in a better description of the electron distributions and in a lowering of energy. The polarized basis set 6-31G(d), also known as 6-31G*, has become a common standard for calculations on medium-sized systems.

More details on basis set usage and notation can be found in refs 3 and 4. Also discussed there, and in other sources, are the levels of calculation. The most basic of these, the Hartree–Fock (HF) computation at the self consistent field (SCF) level gives the best possible orbital representation in which electrons are assigned individual orbitals. This gives reasonably good structures but its neglect of electron interactions (correlation) makes it inadequate for the accurate calculation of bond energies and many other properties. Configuration interaction (CI) methods endeavor to include the effect of electron correlation by adding excited orbital descriptions. As the basis set size becomes infinite, CI gives the exact solution to the nonrelativistic Schrödinger equation. In practical calculations, combinations of singly, doubly, and triply excited states are used, and one common approach for this is termed the QCISD(T) method. Finally, another approach to adding higher excitations to Hartree–Fock functions is Møller–Plesset theory. This involves the use of many-body perturbation theory to calculate second (MP2) and even higher order (MP3, MP4) corrections to the HF energies. MP methods are generally faster than CI calculations but, since they are not variational, can lead to ground state energies lower than the true value.

In our study of basis set and level choices, student pairs did three calculations each on HCl, with some redundancy to serve as a check. The results were then pooled for use in class discussion and in individual reports. Indicated in Table 1 are approximate times (in hours) for the calculations done on a 50 MHz 486 PC with 20 Mbyte of RAM memory and 100 Mbyte of hard disk scratch space. Comparison calculations on a 90 MHz Pentium computer gave times shorter by a factor of about 4. All times depend somewhat on the sequence of steps in the calculation and also upon whether analytical or numerical derivatives of the energy are utilized in the calculations. The latter accounts for the apparent time anomaly between MP2 and MP3 calculations. The times also depend upon the number of iterations necessary to find the minimum energy from the starting point for the calculation (initial guess at R_e), which, for consistency, was chosen to be the experimental value of 1.2747 Å (7).

The Gaussian input requirements to define a calculation are quite simple and consist of several sections, as shown below for a calculation of a potential energy curve displayed in Figure 1.

Route Section:	# UQCISD(T) 6-31G* SCAN
Title Section:	POTENTIAL ENERGY SCAN OF HCL
Charge & Multiplicity:	0 1
Molecule Specification Section:	H
	Cl 1 R
	<blank line>
	R=0.75 70 0.1

The first line selects the calculation level (UQCISD(T)) and basis set (6-31G*), and SCAN indicates that the bond length variable will be scanned. The U on
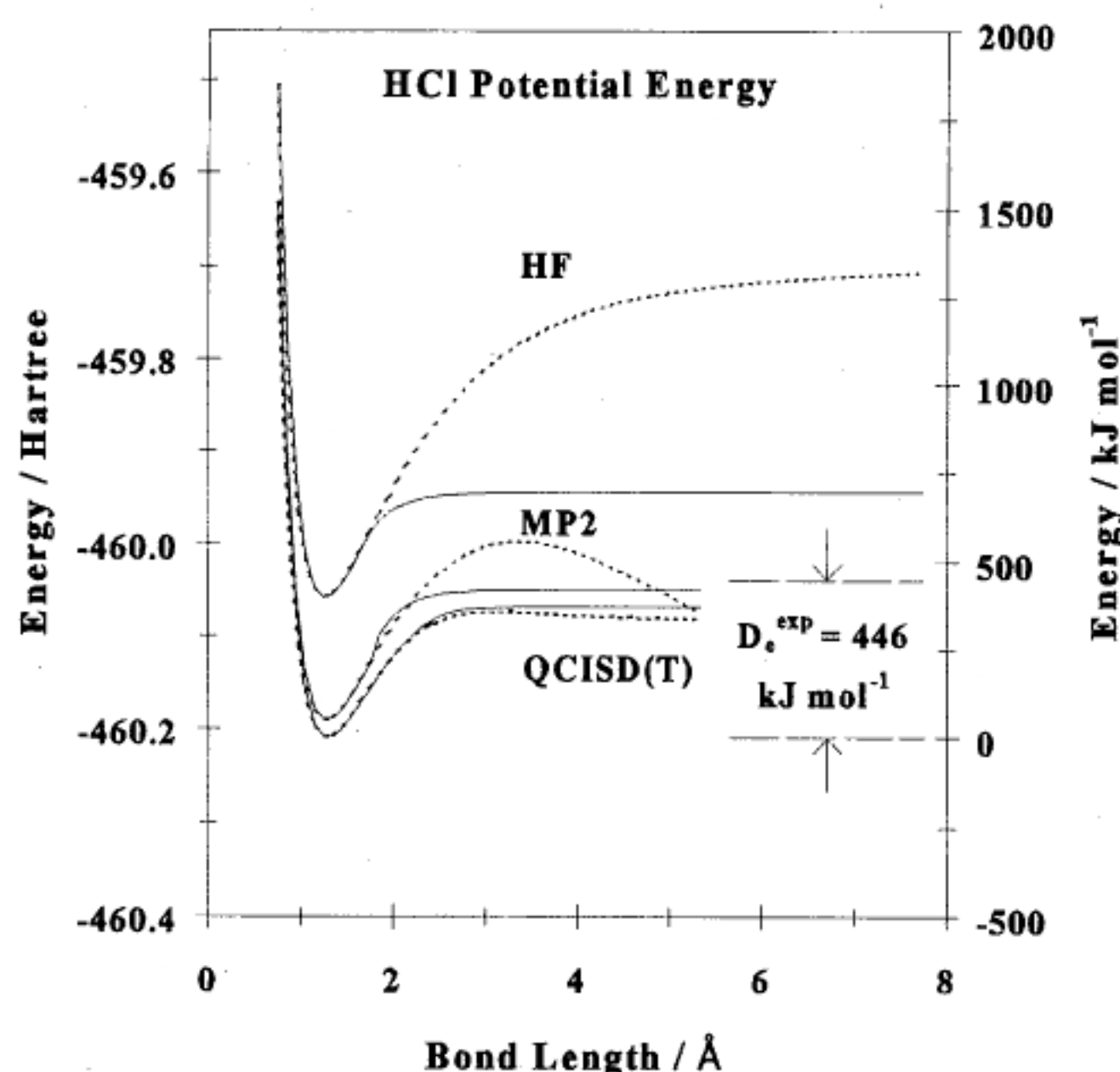

Figure 1. Gaussian 92 for Windows calculation of the HCl potential energy as a function of bond length using a 6-31G* basis set at the HF, MP2, and QCISD(T) levels. The solid curves are the open-shell, unrestricted results; the dashed curves are the closed-shell, restricted results.

the level selection means that the calculation is unrestricted (open-shell); that is, it is not required that the electrons remain paired in orbitals because this can lead to ionic, rather than neutral, products at dissociation. The next line is a title for the output and the third line gives the charge and spin multiplicity for the ground electronic state. Line 4 indicates atom 1 is H and line 5 identifies atom 2 as Cl attached to atom 1 at a bond distance R. The blank line terminates the atom definition section and the last line defines the starting value of R, the number of steps, and the increment for the scan. These parameters are readily entered in the Windows menu format of Gaussian or they can be separately prepared as an ASCII file.

The potential energy curves displayed for HCl in Figure 1 were obtained using the basis set 6-31G* at three levels of calculation for both restricted (dashed curves) and unrestricted (solid curves) cases. The MP2 and QCISD(T) levels, which include electron correlation corrections, yield curves significantly lower in energy than those obtained at the HF level. The predicted HF values of R_e and the bond energy D_e are also in poorer accord with the experimental values of 1.2747 Å and 446 kJ/mol, respectively. For D_e, computed as $E(H) + E(Cl) - E(HCl)$ from separate calculations for isolated species, values of 298, 370, and 373 kJ/mol were obtained for UHF, UMP2, and UQCISD(T), respectively. Values closer to the experimental result are obtained if larger basis sets are used. The large disparity in the RHF and UHF curves occurs because the former extrapolates to $H^- + Cl^+$ dissociation products, whereas the latter correctly yields neutral atoms. Note also that, since the MP2 calculation is a perturbation treatment, it is not subject to the variational constraint, and the energy curves at large separations are not well represented.

A detailed display of the effect of basis set size and calculation level on the R_e, ω_e, μ_e, and α_o parameters is offered in Figure 2, with experimental values indicated with the horizontal lines. Since the interest was in prop-
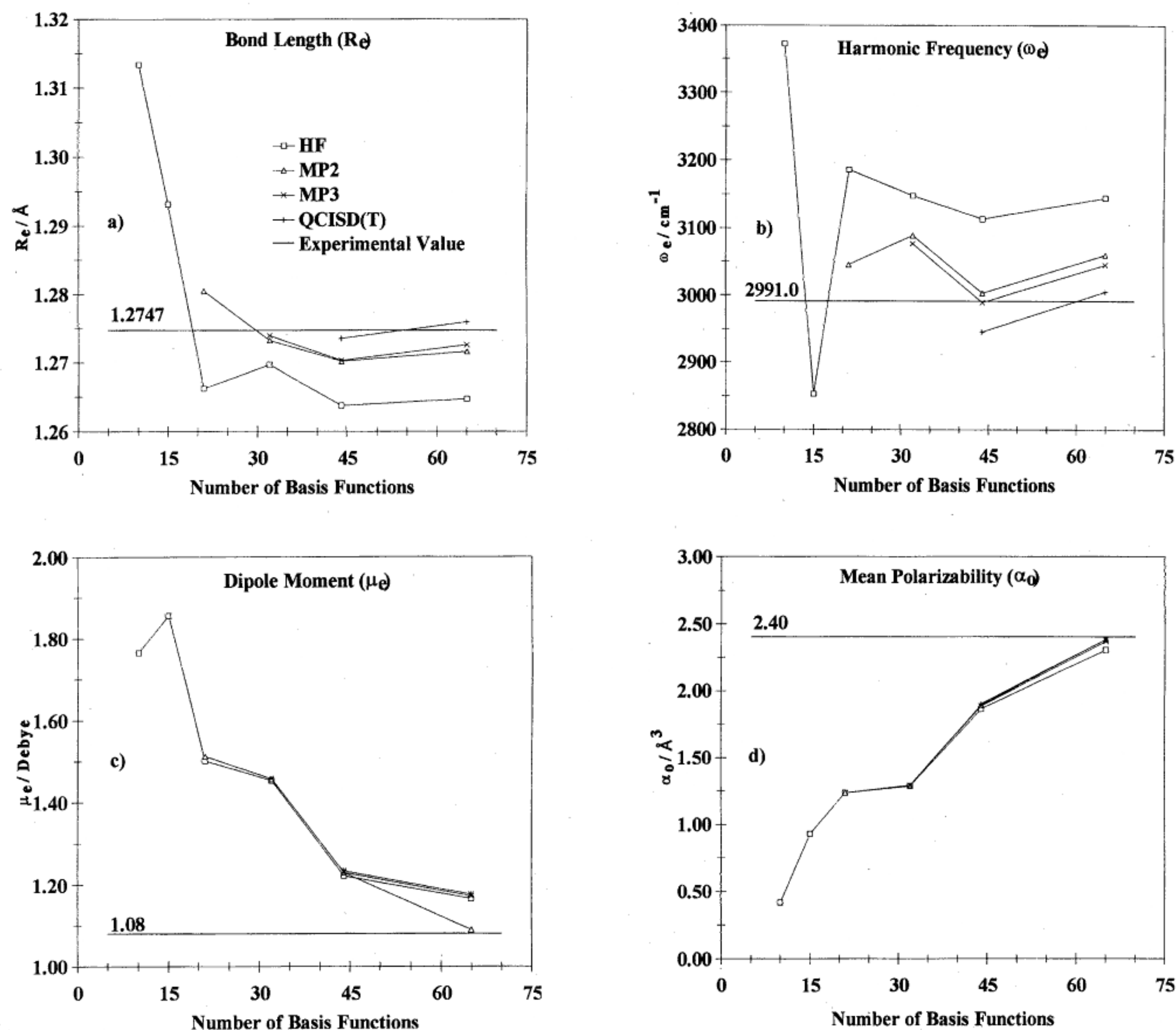
Figure 2. Calculated properties of HCl versus the number of basis functions at various calculational levels. The experimental values are indicated with the solid line. (a) Equilibrium bond length. (b) Harmonic vibrational frequency. (c) Dipole moment. (d) Mean polarizability.

erties at the minimum energy configuration, where the electrons are paired, the calculations were at the restricted level. In these computations, three steps were performed:

a. geometry optimization (keyword = FOPT, in place of SCAN in the above example)
b. frequency determination (keyword = FREQ)
c. dipole moment and polarizability determination (keyword = POLAR).

More details on the specifics of these commands are given in the Gaussian guide (ref 3). The three steps were generally linked to form one job, and the times indicated in Table 1 are totals for all three jobs. The Gaussian output typically contains a large amount of detailed information about each iteration of the calculation, but the final optimized values of interest are readily extracted from a convenient archive summary at the end of the output. It should be noted that the individual polarizability tensor elements α_{xx}, α_{yy}, and α_{zz} are given in the output and the students must calculate the mean of these, $\alpha_o = \frac{1}{3} (\alpha_{xx} + \alpha_{yy} + \alpha_{zz})$, since it is this mean polarizability that is determined in the refractive index experiment.

The results shown in the figures serve as a basis for student discussion and an assessment of the quality of the calculations as a function of time (= cost). Comparisons can also be made with the quality of their lab measurements. It is seen that for basis sizes of about 20 or larger, R_e can be calculated to an accuracy nearly comparable to that achieved in the lab. This is not true for ω_e, although the QCISD(T) values are within 2% of the experimental result.

In contrast to the trends observed for R_e and ω_e, for μ_e and α_o the basis set size (rather than calculation level) is seen to be most critical in obtaining accurate values. For the highest basis set size, the agreement with experiment is comparable to that typically obtained by students in the laboratory determination of these quantities (5–10%). Characterization of the basis sets solely by size is a simplification of course. Particular improvement occurs for the last two basis sets, which have p-, d-, and f-type "polarization" orbitals added to the basic valence sets, along with functional changes (+ and ++) which make the orbitals larger (diffuse) to better approximate the lone pair electrons, and thereby to give a more accurate charge dis-

tribution description. In discussion of such trends, the instructor will find refs 3 and 4 useful. In particular, generalizations and extensions to larger molecules can be made using the many tables and graphs of ref 4.

We have found that the easy access to such calculations on a laboratory PC encourages students to explore further the effects of different orbital choices and calculational levels, and to apply Gaussian to other laboratory experiments. The Windows interface of Gaussian is familiar to students and there is no need to set up a university computing account, learn UNIX, VMS, or FTP as is often the case in using Gaussian on a campus mainframe computer. Moreover the shorter calculation times of such large computers are frequently offset by the long batch queue that exists if many users are requesting service. With Pentium and faster computer processor units to come, calculations such as those described here can be expected to become common and an integral part of the undergraduate training of chemistry majors.

Acknowledgment

Literature Cited

1. Pietro, W. J. *J. Chem Educ.* **1994**, *71*, 416.
2. Duke, B. J.; O'Leary, B. *J. Chem. Educ.* **1992**, *69*, 529.
3. Foresman, J. B.; Frisch, A. *Exploring Chemistry with Electronic Structure Methods: A Guide to Using Gaussian*; Gaussian: Pittsburgh, PA, 1993.
4. Hehre, W. J.; Radom, L.; Schleyer, P. v.R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.
5. Shoemaker, D. P.; Garland, C. W.; Nibler, J. W. *Experiments in Physical Chemistry*, 5th ed.; McGraw–Hill: New York, 1989; Expts 32 and 38.
6. Coe, D. A.; Nibler, J. W. *J. Chem. Educ.* **1973**, *50*, 82.
7. Huber, K. P.; Herzberg, G. *Constants of Diatomic Molecules*; Van Nostrand Reinhold: New York, 1979.

QUNLAB: A Program for the Quantitative Analysis Laboratory

William E. L. Grossman
Department of Chemistry
Hunter College of CUNY
695 Park Avenue
New York, NY 10021

In the Chemistry Department at Hunter College, as in many other chemistry departments, the laboratory in quantitative analysis is graded differently from laboratories in other courses. Assuming that the student's lab book passes muster, the grade on an experiment is based entirely on the accuracy of the student's analytical result.

No written report is required. The student submits the result of the required analysis and a brief calculation of the estimated error. The submitted result is compared to the known analysis of the sample, and the grade is determined by reference to an error curve for the particular analysis: a certain error range, say 8.2–9.8 parts per thousand, yields a certain grade, say 90%. As in real life, evaluation of the reported result does not depend on the source of the error. That is, any error in the reported result causes a lowered grade. This definitely includes calculation errors.

At first glance this may seem unfair. An erroneous analysis in a commercial or industrial situation may have important financial or other consequences, even more serious than a lowered grade. However, the industrial analyst will probably be protected from calculation errors by

instrument or laboratory software. Why shouldn't our students have the same resources to protect them?

Of course the reason is that we want the students to learn how to do the calculation, and so we have to make them responsible for it. What we need is a way of telling the students when their method of calculation is correct, without revealing either the correct analytical result or a correct calculation method for their sample.

Our solution to this problem is embodied in a program called QUNLAB, which runs on a computer in the quantitative analysis laboratory. Students are strongly encouraged to check their analytical results with QUNLAB before reporting them to the instructor. QUNLAB asks for the student's analytical data (sample weights and titration volumes, for example) and final calculated result.

QUNLAB first correctly calculates the analytical result from those data. It does not give the student this result, nor does it show how it did the calculation. The program then compares the student's calculated result with the one it has calculated, without revealing the latter. If the student has done the calculation correctly, the two results will be the same. We allow a small margin of error around the correct result for rounding differences, etc. QUNLAB only reports to the student that the results either agree or disagree. If the two results agree, the student can be confident that he or she has not made a calculation error. The student knows that the lab grade will now be based entirely on analytical technique. A disagreement between the two results constitutes a warning to the student that something is amiss with the calculation. QUNLAB uses the student's experimental results to do the analytical calculations the way the student *should* do them; it does not know the actual analysis of the student's sample, which is a carefully guarded secret. Thus even the most skillful hacker cannot crack the program and find the right answer. The student learns that there is an error, without ever knowing the actual composition of the analyzed unknown. Most calculation errors are substantial and result in a failing grade on the experiment. Thus there is a strong incentive to find the problem.

The benefit to the students has been considerable. It has been my experience over the years that more than 80% of the students make a calculation error on the first experiment. Thanks to the program, the grade distribution has changed from a (large) majority of F's to a majority of A's.

From the instructor's point of view, the benefits are twofold. The student who has made a mistake understands the error and has learned the right way to do the calculation. The immediate feedback that the student gets from QUNLAB gives a terrific boost to the learning process. Second, the student's grade reflects his or her experimental technique. The latter point is also an important one, since one of the major purposes of the laboratory is to teach and reward good laboratory techniques.

For historical, financial, and religious reasons, QUNLAB is written in Digital Equipment Corporation FORTRAN IV. It runs under RT-11, DEC's real-time operating system for the LSI-11 family of computers. It is efficiently implemented in our quantitative analysis laboratory on an LSI-11/23 running a time-shared executive (TSX-11) on top of the operating system. We can thus make the program available simultaneously on as many terminals as are needed by the class, at substantially lower cost than the equivalent number of PCs would be. There is no reason, however, why the program could not be implemented on any hardware that was available, using any convenient language.