Lab #5: Computational Chemistry

Introduction

In this investigation we will apply the techniques of computational chemistry to several of the molecular systems that we have investigated experimentally. As with the use of any tool, instrumental or otherwise, we need to understand the underlying machinery before we can use the results to forward our investigation of a particular molecule or reaction. With computational chemistry it is even more critical that we understand the principles so that we can correctly interpret the results and not be mislead by erroneous results.

We will focus our attention on *ab initio*, from the beginning, calculations that are based solely on the underlying quantum mechanical treatment of electrons and nuclei within molecules and in principle require no experimental values to calibrate the results of the calculations nor parameters based on experiment. These methods are even applicable to chemical compounds, which do not yet exist, or which exist for such a short time that their properties can not be determined experimentally. Among these short lived states are transition states. These computations are also important for the calculation of molecular geometry, electronic properties, and other molecular properties of stable molecules.

You have likely already done some computational chemistry in other courses using WebMO. This lab assignment takes a closer look at the how computations are set-up in Gaussian and what is being calculated. For this reason, we will spend one week on an introduction to computational chemistry looking at simple molecules like H_2 and HCl, and a follow-up week on more complex molecules (chosen by you). We will not use the graphical interface, WebMO, for analysis of the simple molecules, but you have the option of using it for analysis of the complex molecules.

As a starting point we will examine the approximate description of the radial (away from the nucleus) distribution of electrons. The description of the radial distribution of electrons is key to understanding any chemistry and properties of atomic and molecular systems. Most quantum chemical programs use Gaussian functions to simulate this distribution and thus we will start at this point.

Overview of Theory

$$H\Psi = E\Psi \qquad (1)$$

In the quantum mechanical description of molecular electronic structure it is necessary to formulate the Schrödinger equation in terms of a Hamiltonian with several approximations. First, since nuclei are much more massive than electrons and therefore electron motion is much faster than that of nuclei we can remove the nuclear kinetic energy terms and arrive at a separate molecular electronic Hamiltonian. This

approximation is known as the Born-Oppenheimer approximation. In the potential energy part of the Hamiltonian we must treat the potential term corresponding to electron-electron repulsion in an average way rather than having the motions of every pair of electrons correlated.

With the exception of the case of the hydrogen atom we can not formulate the exact wavefunctions which are solutions to the Schrödinger equation. Molecules can present even more difficulty than multielectron atoms in terms of formulating the electronic wavefunctions. An orbital is a wavefunction for an individual electron. Wavefunctions for electrons in molecules are known as molecular orbitals. Since we don't know the exact molecular wavefunctions or orbitals we can approximate them by expanding them in a set of known functions each with coefficients corresponding to their importance in the overall wavefunction. This is possible because any function can be described by expansion in an infinite set of functions. When this set becomes finite the description may not be as good but an approximate function may be reached.

$$\Phi(r) = \sum_{i=1}^{\infty} c_i \Psi_i(r) \qquad (2)$$

$$\Phi(r) \cong \sum_{i=1}^{n} c_i \Psi_i(r) \qquad (3)$$

Many computational programs model molecular electronic structure by forming molecular orbitals that are linear combinations of atomic orbitals (LCAO-MO). These atomic orbitals are approximated by using terms consisting of several Gaussian functions with varying exponents. Atomic orbitals are described by hydrogenic wavefunctions with quantum numbers n, l, m_l. (see Ch. 6 McQuarrie). It is most convenient to express these wavefunctions in spherical coordinates as:

$$\Psi_{nlm}(r,\theta,\phi) = R_{nl}(r)Y_l^m(\theta,\phi) \qquad (4)$$

where $Y_l^m(\theta,\phi)$ is a spherical harmonic.

A more convenient function for computational study is the Slater type orbital which has no nodes and has an adjustable parameter, ξ, in the exponent.

$$S_{nlm}(r,\theta,\phi) = \frac{(2\xi)^{n+1/2}}{[(2n)!]^{1/2}} r^{n-1} e^{-\xi r} Y_l^m(\theta,\phi) \qquad (5)$$

However, certain integrals are much more efficiently evaluated within a computational program if we use Gaussian functions.

$$G_{nlm}(r,\theta,\phi) = N_n r^{n-1} e^{-\alpha r^2} Y_l^m(\theta,\phi) \qquad (6)$$

In order for these Gaussian functions to approximate the Slater type functions it is useful to sum up several Gaussian functions with the different exponents, for example:

$$\Phi_{1s}^{STO}(r) = \sum_{i=1}^{3} d_{1si} \Phi_{1s}^{GF}(r, \alpha_{1si}) \qquad (7)$$

where:

$$\Phi_{1s}^{GF}(r, \alpha) = \left(\frac{2\alpha}{\pi}\right)^{3/4} e^{-\alpha r^2} = N_\alpha e^{-\alpha r^2} \qquad (8)$$

The above example illustrates the STO–3G (i.e., Slater-type orbitals-3 Gaussians) basis set since the wavefunction for the atomic orbital is formulated from the sum of 3 Gaussians. Basis sets are groups of functions with particular parameters that are used to approximate the atomic orbitals that make up the molecular wavefunction. The more functions and flexibility built into the basis set the better the description of molecular electronic properties but the more time the calculation will require. As you will soon see even fairly small molecules will take substantial time to compute molecular properties for.

TABLE 2. Exponents and coefficients for 1s STO-3G

i	α	d
1	0.1688	0.4446
2	0.6239	0.5353
3	3.425	0.1543

Analyzing approximate radial functions:

Our description of the molecular orbitals is thus in terms of a sum of atomic orbitals each of which is described by a sum(s) of Gaussian functions. That is the ith molecular orbital, Ψ_i, is described as:

$$\Psi_i = \sum_{j=1}^{M} c_{ji} \Phi_j \qquad (9)$$

where Φ_j is one of the M atomic orbitals used.

The STO–3G basis set is useful for some calculations but is not flexible enough to describe the molecular properties of most molecules with reasonable accuracy. This basis set describes all atomic orbitals of a particular atom in an identical manner. This means that the three p–orbitals on carbon are treated identically. Thus those atomic orbitals participating in π bonding have the same spatial extent as those atomic orbitals

involved in σ bonding. The atomic orbital involved in π bonding should be much more diffuse. This difficulty is overcome by the use of basis sets with Gaussian terms describing Slater orbitals with two different exponents and thus one of smaller size and one of larger extent. The computational program then optimizes the mixture of these two components to develop an orbital with a more appropriate size. These basis sets are called double zeta (ζ) because of the two different exponents on the Slater type orbitals. The 6−31G basis set uses six Gaussian functions to describe the core orbitals and the valence orbitals are each made up of combination of a smaller Slater type orbital (larger exponent, ζ, equation 5) described by three Gaussian functions and a larger Slater type orbital described by only a single Gaussian function.

Determine the number and type of basis functions (each Slater type orbital) used to describe each atom in H_2O using the 6−31G basis set. How many total Gaussian functions?

Answer: Each hydrogen has one smaller orbital described by a linear combination of three Gaussian functions and a second larger orbital described by a single Gaussian. The oxygen 1s orbital is described by a linear combination of six Gaussians. The 2s and the three 2p oxygen orbitals are each described by the 31G double zeta basis set.

TABLE 3. Number of Basis Functions and Gaussians for 6−31G H_2O

Atom	Basis Functions	Gaussians
H	2	4
O	9	22
Total	13	30

The time required for a calculation is a steep function of the number of basis functions so careful choice must be made for each problem.

There are two other types of functions that may be added to a basis set to better describe molecular electronic structure. The first is polarization terms that amount to the addition of orbitals with higher angular momentum (d orbitals for second row elements). These orbitals provide flexibility in the spatial distribution of the electron. This addition is denoted by a "(d)" or * added to the basis set specification (*i.e.* 6−31G(d) or 6-31G*) and amounts to the addition of 6 basis functions each described by a single Gaussian. Additional very diffuse functions (small ζ) can also be added by adding a "+" sign or a second * to the designation (*i.e.* 6-31+G(d) or 6-31G**).

These calculations fail to account for the correlated movement of individual electrons within an atom. The electrons only "see" each other through an average effect within the approximations of this theory. This unquantified effect is known as electron correlation and can be accounted for approximately. Calculations which attempt to

include correlation effects require even more computational time but often lead to more accurate electronic properties. Several computational models which include some treatment of electron correlation will be explored.

In this investigation we will compare different levels of theory to our recent experimental results on HCl and we will then attempt to apply our new found working knowledge of this type of calculation to other problems of interest.

Computational study of H$_2$ and beyond: Methodology

We will use the computational program suite Gaussian 03 with the graphical interface WebMO. All of these programs are resident on Slater, the LINUX computational server. (See Gustavus Computational Server and Gustavus Computational Resources for helpful information and links.) For each computation we must define a computational model in terms of the basis set we will use and the level of theory we will use. The energies resulting from these calculations are all expressed in Hartrees, E$_h$. E$_h$ = 2.1947463 x 10^5 cm^{-1} = 4.359748 x 10^{-18} J.

Gaussian 03 example input file for acetylene

```
#HF 3-21G GFPRINT OPT Pop=Reg

 Acetylene

 0 1

  C

  C     1 rcc

  H     1 rhc     2 a1

  H     2 rhc     1 a1     3 dihed

Variables:

rcc= 1.2024

rhc= 1.0625

a1= 180.00

dihed= 0.00
```

The first line is known as the route. It describes the calculations to be performed. In this case we are performing a Hartree-Fock (HF; SCF-LCAO) calculation with the 3-21G basis set. We are requesting that the coefficients and exponents of the Gaussian functions we are using be printed (GFPRINT). We want the geometry set out below to be optimized within this model chemistry. The final keyword requests that the coefficients on the atomic orbitals making up the molecular orbitals be printed in the output. A blank line is left followed by the title which we chose. After another blank line the molecular charge (0 in this case) followed by the spin multiplicity of the state (1 in this case to denote singlet meaning that all electrons are paired). The very next line begins the definition of the molecular geometry. The first atom listed is numbered 1 and the rest of the geometry is in Z-matrix form which uniquely defines the geometry in terms of variables specifying bond distances, bond angles, and dihedral angles. At the end these variables are given initial values. Each bond length is defined relative to other atoms. For hydrogen atom #3, the bond distance is defined relative to carbon atom 1 which it is connected to. The bond angle is defined relative to carbon atom #1 and carbon atom #2.

For an optimization (OPT) these values will be optimized whereas in a single point calculation (SP) the energy is determined at the geometry specified. In a SCAN type of calculation all variables are fixed except those that are given a starting point, the # of points to calculate, and an increment. All distances in Gaussian are given in Angstrom and all angles are specified in degrees.

Systems Under Investigation

1. H_2

As an initial investigation, we will explore the H_2 molecule using Gaussian 03. For H_2 we know the energy of the separated atoms ($r = \infty$). The experimental parameters we will compare our calculation to are listed in the table below.

TABLE 4. Experimental properties of H_2

r_e	0.7416×10^{-10} m
$\bar{\nu}$	4401 cm^{-1}
B_e	1822.7381 GHz
Binding Energy	0.174 E_h
Energy of H atom	$-1\ E_h/2 = -109737.315$ cm^{-1}

We will use Gaussian 03 on the Linux computational server. For more detailed information about the specifics of carrying out these calculations on this computer system see: Gustavus Gaussian Computational Server Gaussian Web Site

In lab:
1. In your notebook, prepare an (handwritten) input file for H_2.
2. In WebMO, perform a geometry optimization using the HF/STO-3G basis set. Does your input file match the one you wrote in your notebook?

3. While you are examining the output, you should start another calculation in which you use the optimized bond length from above in a frequency calculation (FREQ keyword) and the same computational model. This calculation will give you the vibrational frequency which you can compare with the above as well as thermodynamic properties.

4. For your first calculation, how do the results of this calculation compare with the above table of data? How large is the binding energy predicted by the above calculation? Examine the molecular orbitals and the contributing atomic orbitals. Record the time required for each calculation by examining the end of the output.

5. As a third calculation, do a single point (SP) calculation at a large internuclear separation and compare the SCF energy with that of two infinitely separated hydrogen atoms.

6. Do a fourth calculation in which you use a larger basis set to optimized the geometry and calculate the frequency. Compare these results. Discuss the form of the basis functions.

2. HCl
a)
We will next examine $H^{35}Cl$ and make comparisons to our experimental data. HCl is also a diatomic molecule like H_2 but the larger chlorine molecule makes the calculation of molecular properties both more time consuming and increases the demands on the computational model. We will explore several models and how well they reproduce the experimental data on HCl. Start with a small basis set HF model of your choice. Again, write out an example input file in your notebook. Set up your calculation and Z–matrix within Gaussian 03 (WebMO). How many basis functions and Gaussians does your model use? (Note in notebook.) You want to optimize the geometry and calculate the vibrational frequency at this geometry. Compare the rotational constant, internuclear separation, and vibrational frequency with your experimental value. Record the calculation time from the end of your output. You will do a second calculation on HCl in which you will select a larger basis set and different model. Choose your model from the table posted in the lab so that we will get a variety of results to compare. Once you complete these two calculations, post your results in the spreadsheet posted in the lab. These overall results should be discussed in your report (due after computational labs are **all** done). A plot of calculation time versus accuracy should be constructed based on this data. Which model is most time effective for handling this problem?

b)
We will next examine the internuclear potential of HCl. We can roughly approximate this potential as a harmonic potential. We will examine the intermolecular potential with Gaussian 03 and then compare this result with a simple harmonic potential with a force constant, k, based on your experimental data. In making this comparison you can arbitrarily set the minimum of the calculated potential at zero. Within Gaussian the SCAN keyword can be used to examine the energy as a function of internuclear separation. Start with the HF/3–21G(d) model and create the appropriate Z–matrix within Gaussian 03. When using the SCAN keyword the internuclear separation parameter must have a starting point, number of points at which we want the energy

calculated, and the increment. About 30 total points from about 0.75 angstroms to about 4.5 angstroms should make the form of the internuclear potential clear (This calculation may take an hour or more). See the next paragraph for more details on the input file. A table of the energies and distances is included at the end of the Gaussian 03 output and can be put into another text file and then imported into Sigmaplot or Excel for plotting.

The command for the scan job is "OPT=z-matrix". At the end of the input file you need a line that gives the bond to be changed (e.g., B1), the starting position (e.g., 0.75), the command S (i.e., S), the number of steps (e.g., 10), and the increment (e.g, 0.25). Note that the numbers in parenthesis can be changed to match the directions above for this problem. The starting position and increment must be in decimal form and the number of steps must be an integer. In my example, my line would be:

<div align="center">B1 0.75 S 10 0.25</div>

Note the spaces between each entry, and always remember that the final entry in your input file must be a return.

Week 2. *Modeling Polyatomic Systems*

WebMO coupled to Gaussian 03 will be used to investigate the properties of a system of your own choice. Since computational time can go up as the number of electrons to the 4^{th}(!) power, it is best to confine our study to molecules of 20 atoms or less. Before beginning your calculations include a picture of the molecule in its optimized ground state geometry in your notebook. You should choose an interesting chemical "problem" to explore… perhaps you want to better understand the effect of solvating a molecule, perhaps you want to explore the more stable form of an aromatic compound (i.e., meta, para, ortho), perhaps you want to verify some computational data from the literature (i.e, verify that one of a given set of structures has the lowest energy or verify the bond angles, lengths, etc at a couple levels of theory), etc… Summarize your goals and results clearly in your report. Some systems explored last year are listed here, but you can be creative in your choice of project.

Some Possible Systems:
1. Water clusters are examined in some detail by Gordon *et al.* See:
   Gordon, Mark S.; Freitag, Mark A.; et al. The Effective Fragment Potential Method: A QM-Based MM Approach to Modeling Environmental Effects in Chemistry. Journal of Physical Chemistry A . 2001; 105:293.
2. Cysteine exhibits an interesting series of equilibria for deprotonation. The relative values of these equilibrium constants are not fully quantified.
3. 2-naphthol exhibits a large change in pKa upon electronic excitation. See:
   Richard Knochenmuss; Volker Karbach; Claudia Wickleder; Stephan Graf, and Samuel Leutwyler. Vibrational-Energy Redistribution and Vibronic Coupling in 1-Naphthol·Water Complexes . Journal of Physical Chemistry A. 1998; **102** :1935-1944.

References

D. A. McQuarrie and J. D. Simon, Physical Chemistry: A Molecular Approach, University Science Books, Sausalito, CA, Chapter 11. (1997).

D.P. Shoemaker, C.W. Garland, and J.P. Nibler, Experiments in Physical Chemistry, 6th ed., McGraw-Hill, pp. 397-404 (1996).