MOLECULAR MODELING AND ELECTRONIC STRUCTURE CALCULATIONS

George Schatz, Baudilio Tejerina, Shelby Hatch and Jennifer Roden
Department of Chemistry, Northwestern University, Evanston, Illinois 60208-3113

This laboratory is designed to use the program GAMESS (General Atomic Molecular Electronic Structure System, developed in Gordon research group at Iowa State) through a website called nanoHUB (www.nanoHUB.org) to determine the geometric and electronic properties of numerous small molecules. GAMESS uses *ab initio* and semi-empirical calculations to determine these properties. *Ab initio* ("from first principles") calculations solve the Schrödinger equation using the exact computational expression for the energy of the electrons.[1] The particular *ab initio* method that we will use for this lab is called Hartree-Fock (HF). HF uses an **approximate** wavefunction to solve Schrödinger, so the resulting molecular properties are approximate, but for many applications the accuracy is adequate for interpreting experiments. Semi-empirical calculations use an approximate energy expression for the electrons, but solve for the exact wavefunction associated with this expression. Usually the energy expression uses empirical parameters (found experimentally) to match molecular properties, but the resulting properties are still approximations to the correct values. The semi-empirical method that we used in this lab is called PM3. This stands for "parameterized model 3", which was the third (and best) method that the original authors of the method developed.

The underlying theory for GAMESS will be described in the lecture. In brief, GAMESS self-consistently solves the Schrödinger equation. The self-consistent method is an iterative approach that minimizes the energy by adjusting the wave functions of the molecules. Further information can be obtained from the GAMESS user guide (http://www.msg.ameslab.gov/gamess/documentation.html).

Pre-Lab Information

GAMESS is located in the QC-Lab tool (https://nanohub.org/tools/qclab) on the nanoHUB website. To access this tool you must first create a user account (see Appendix A for directions).

Make sure to include the results that you obtain (bond distances, bond angles, energy values, and charge per atom) in your lab notebooks.

The units on these results should be the same as the units used in GAMESS (distances in Angstroms (Å), angles in degrees, energy in Hartrees).

The results should be presented neatly, preferably in a tabular form with ***some experimental data obtained from literature*** also recorded. Literature values can be located through the National Institute of Standards and Technology (NIST) Computational Chemistry Comparison and Benchmark Database (CCCBD) at http://cccbdb.nist.gov

It is important to note that calculated energy values cannot be directly compared to experimental energy values.[2]

---

[1] The time-independent Schrödinger equation is $H\psi=E\psi$ where ψ is the wavefunction representing atomic/ electronic positions, E is the energy and H is the Hamiltonian, an operator which takes the wavefunction and finds the energy.

[2] The calculated energy values are absolute energy values while the experimental values obtained are typically energy differences between states. Comparison between experimental and computational energy values can be made. How?

PROBLEM 1: Warm-up and Practice

In this problem, the properties of three small linear molecules (CO, H_2, and N_2) will be calculated using a semi-empirical and an *ab initio* method. The calculation for CO will be shown in full detail and the remaining two molecules (H_2 and N_2) will be left for the reader to perform individually.

Carbon Monoxide Walk through:

1.) (In lab notebook) Determine the input coordinates
  a. Draw the Lewis Dot structure to determine basic bonding and lone pairs.

:C≡O:

  b. Redraw the molecule using the appropriate Valence Shell Electron Pair Repulsion (VSEPR) model to determine basic structure. *Linear*
  c. Place molecule on Cartesian coordinates using the average bond lengths given in the textbook.
    i. Make x-axis the bonding axis

C	0.0	0.0	0.0
O	1.12	0.0	0.0

    ii. Reassign the origin of the coordinates in order to achieve the highest symmetry possible. This will reduce number of calculations and speed up the process (trivial for small linear molecules but helpful for larger non-linear molecules).

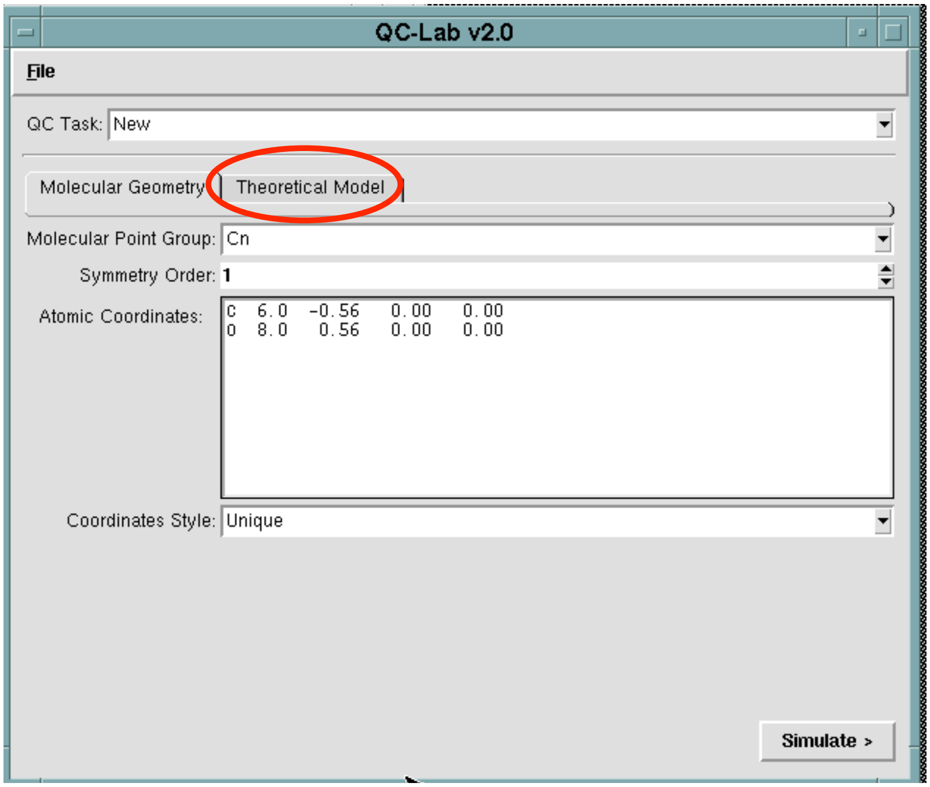
C	-0.56	0.0	0.0
O	0.56	0.0	0.0

  d. Be sure to show all of your work in your notebook

2.) (In nanoHUB) Perform semi-empirical calculations
  a. After setting up a nanoHUB account and launching the QC-lab tool, select new from the 'QC task' pull-down menu
  b. Delete the text in the atomic coordinates box and enter the input coordinates (X, Y, Z) that you determined, being sure to first put the atomic symbol and atomic number. Be sure to note the syntax (spacing, using the decimal, etc.) See screenshot on next page.
  c. Leave the 'Molecular Point Group,' 'Symmetry Order,' and 'Coordinate Style' at the default setting (C_n, 1, unique)[3]

---

[3] These three selections have to do with taking advantage of symmetry to determine the input geometry with the different point groups dictating what symmetry operations (mirror of plane, rotation, inversion, etc.) need to be performed on a minimal number of input atoms to obtain the whole molecule. For water you would need to enter C_{nv}, 2, and unique with the input of one O and one H. The second H would be found through the performed symmetry operations.



QC Task: New

Molecular Point Group: Cn

Symmetry Order: 1

Atomic Coordinates:

```
C  6.0  -0.56   0.00   0.00
O  8.0   0.56   0.00   0.00
```

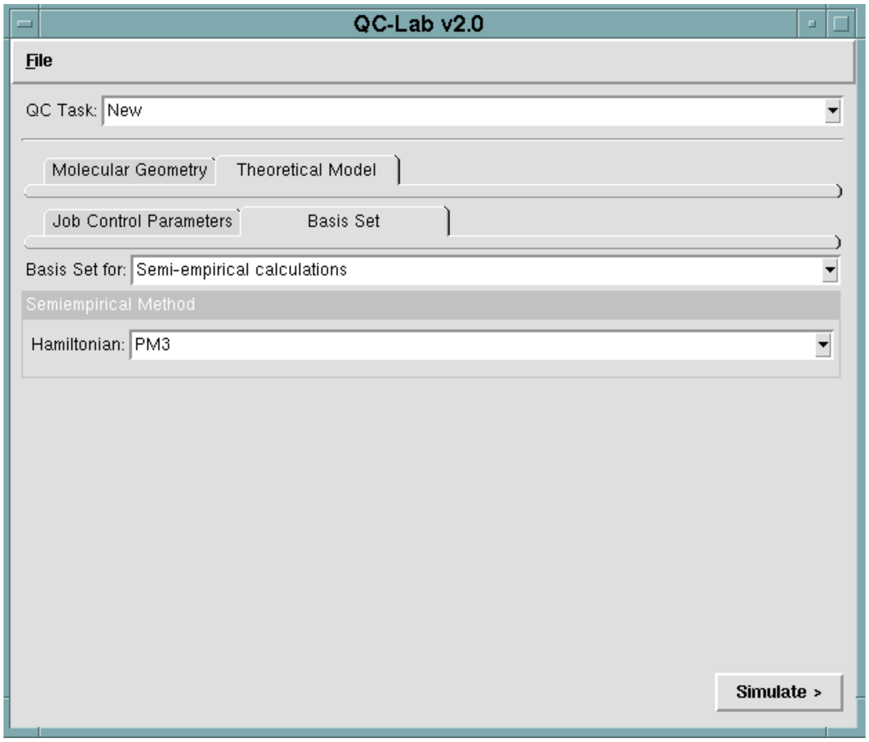
Coordinates Style: Unique

d. Click on the ‘Theoretical Model’ tab (circled in previous image)
e. Leave the default ‘Job Control Parameters’ settings for now (Run, Geometry Optimization, Restricted Hartree-Fock Calculation, 0 and 1).[4]
f. Under the ‘Basis Set’ tab, set the ‘Basis Set for’ pull down menu to ‘Semi-empirical calculations.’ This action will refresh your screen and bring you back to the ‘Job Control Parameters’ tab so click back into the ‘Basis Set’ tab (you might notice other tabs have vanished; this is okay).
g. In the ‘Hamiltonian’ pull-down, select PM3.

---

[4] These setting are very important in telling the software what calculations to run.

Execution Type	Tells the software to run the calculation or check the input
Run Type	Tells the software what type of calculation to run (geometry optimization = find ‘lowest’ energy geometry, Hessian = determines more information about the potential energy surface and can be used to find vibration information, single energy = finds energy of input geometry)
SCF Type	Provides computational details, to be discussed later
Molecular Charge	Indicates if the molecules in an ion (positively or negatively charged)
Spin Multiplicity	Gives the software information on how many electrons are spin up or spin down (which has to do with the number of unpaired electrons)

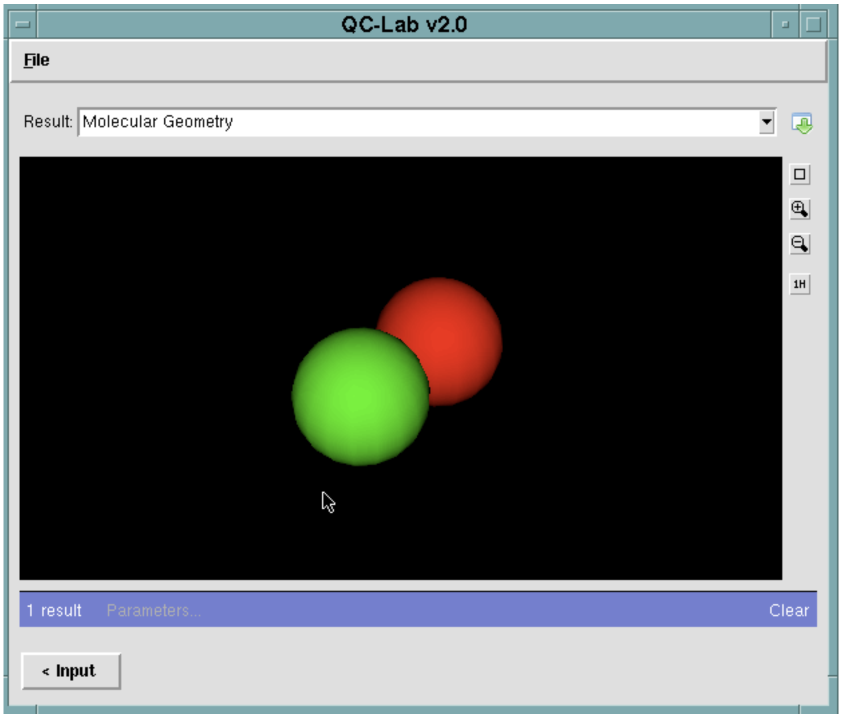

QC-Lab v2.0
File
Molecular Geometry
Theoretical Model
Job Control Parameters
Basis Set
Semiempirical Method
Simulate >

QC Task: New

Basis Set for: Semi-empirical calculations

Hamiltonian: PM3

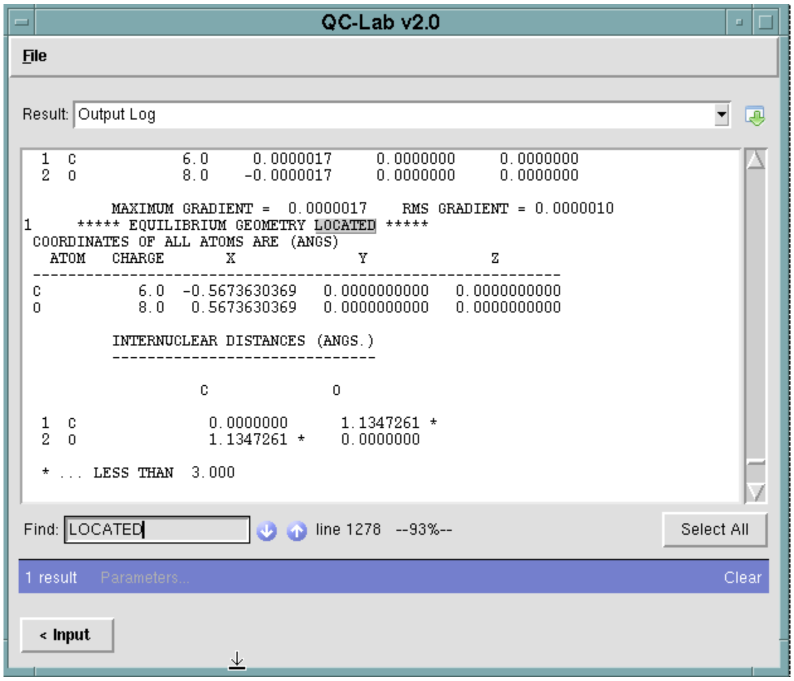
h. Click the ‘Simulate’ button in the lower right.
i. After the job has finished running, an image of the molecule will appear in the window.


QC-Lab v2.0
File
1H
1 result
Parameters...
Clear
< Input

Result: Molecular Geometry

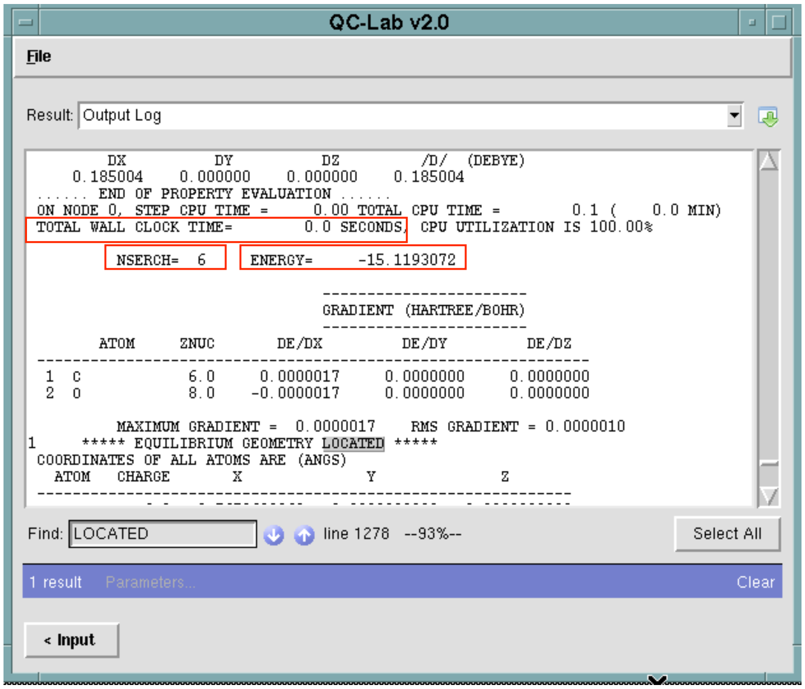
j. In order to obtain the necessary information, select the ‘Output Log’ from the ‘Results’ pull-down. The output contains all the results and information about the calculation.

k. The following key words can be found using the ‘Find’ feature to locate the desired information:
   i. LOCATED = the location in the output with the optimized coordinates and bond distances are printed


QC-Lab v2.0
File
Result: Output Log
Find: LOCATED
line 1278 --93%--
Select All
1 result
Parameters...
Clear
< Input

```
  1  C            6.0     0.0000017     0.0000000     0.0000000
  2  O            8.0    -0.0000017     0.0000000     0.0000000

         MAXIMUM GRADIENT =  0.0000017    RMS GRADIENT = 0.0000010
1     ***** EQUILIBRIUM GEOMETRY LOCATED *****
 COORDINATES OF ALL ATOMS ARE (ANGS)
   ATOM   CHARGE       X              Y              Z
 ------------------------------------------------------------
 C           6.0  -0.5673630369   0.0000000000   0.0000000000
 O           8.0   0.5673630369   0.0000000000   0.0000000000

         INTERNUCLEAR DISTANCES (ANGS.)
         ------------------------------

                   C              O

  1  C             0.0000000      1.1347261 *
  2  O             1.1347261 *    0.0000000

  * ... LESS THAN  3.000
```

   ii. Slightly above the word LOCATED will be a value for TOTAL WALL CLOCK TIME, NSERCH and ENERGY, which gives the time the calculation took, the number of steps taken by the computer to obtain an optimized geometry and the energy for the optimized structure.


QC-Lab v2.0
File
Result: Output Log
Find: LOCATED
line 1278 --93%--
Select All
1 result
Parameters...
Clear
< Input

```
          DX          DY          DZ         /D/  (DEBYE)
     0.185004    0.000000    0.000000    0.185004
 ...... END OF PROPERTY EVALUATION ......
 ON NODE 0, STEP CPU TIME =     0.00 TOTAL CPU TIME =        0.1 (    0.0 MIN)
 TOTAL WALL CLOCK TIME=        0.0 SECONDS, CPU UTILIZATION IS 100.00%

         NSERCH=  6      ENERGY=      -15.1193072

                                ------------------------
                                GRADIENT (HARTREE/BOHR)
                                ------------------------

       ATOM     ZNUC       DE/DX         DE/DY         DE/DZ
 --------------------------------------------------------------
  1  C            6.0     0.0000017     0.0000000     0.0000000
  2  O            8.0    -0.0000017     0.0000000     0.0000000

         MAXIMUM GRADIENT =  0.0000017    RMS GRADIENT = 0.0000010
1     ***** EQUILIBRIUM GEOMETRY LOCATED *****
 COORDINATES OF ALL ATOMS ARE (ANGS)
   ATOM   CHARGE       X              Y              Z
 ------------------------------------------------------------
```

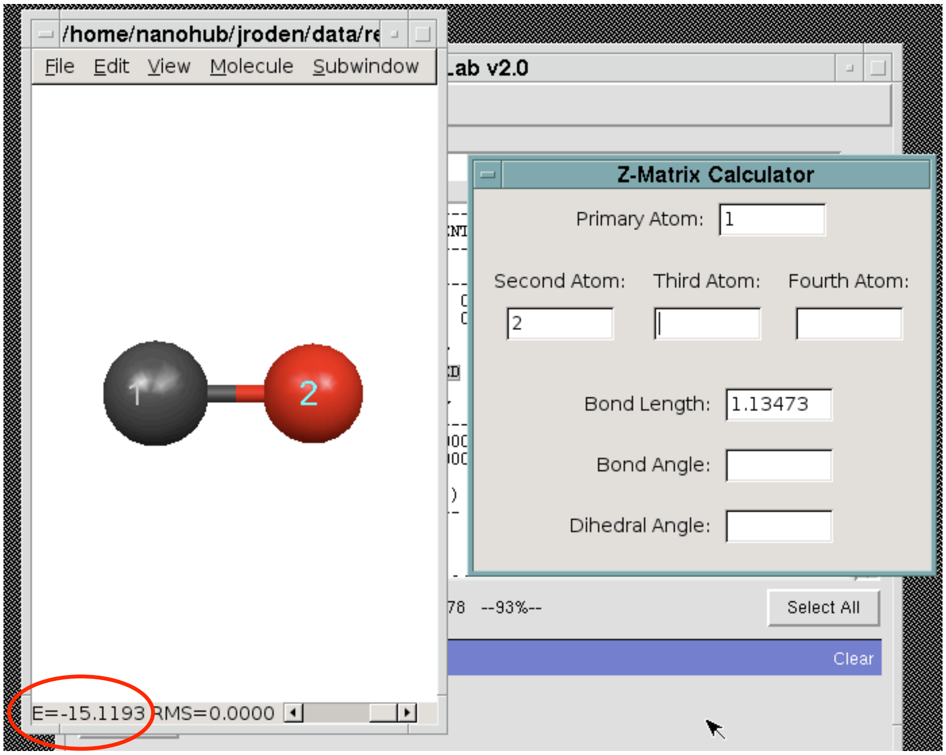
    iii. MOPAC CHARGES = this is located below the LOCATED and gives the charge for each atom.

l. In addition to searching the output file, results can also be obtained through MacMolPlot. Clicking on the launchmolviewer tab at the bottom of the screen and using the pull down menu to select MacMolPlot will access this software.

    i. Note the nanoHUB session number, this will appear in the URL - Select 'Open' from the 'File' menu. Open the folder with your username. Open data. Open sessions. Open your session folder. Select the proper file ending in .inp and click open.

    ii. The molecule should appear in the window with the energy written in the bottom left corner.

    iii. The bond length (and angle) can be found by using the 'Z-Matrix Calculator' from the 'Subwindow' menu. The atoms will be numbered based on the order their coordinates were input (in this case C will be 1 and O will be 2). The assigned atom numbers can be displayed on the molecule image by selecting 'Atom Number' under the 'Atom Labels' menu in the 'View' pull-down.



Primary Atom: 1

Second Atom: 2

Third Atom:

Fourth Atom:

Bond Length: 1.13473

Bond Angle:

Dihedral Angle:

E=-15.1193 RMS=0.0000

    iv. Close out of the 'Z-Matrix Calculator' using the hyphen bar but DO NOT close the MacMolPlot window. Click back into the QC-Lab v2.0 window.

    v. Note: MacMolPlot cannot give you atomic charge information.

3.) Perform *Ab Initio* Calculations

a. In the QC-Lab v2.0 window, click on the 'INPUT' button in the bottom left corner.

b. Click on the 'Molecular Geometry' tab and verify that your input geometry is the same as you entered for the semi-empirical calculation. Alternately, you can replace this input geometry with your optimized geometry obtained using the semi-empirical method. That may reduce the number of steps (nserch) the computer will have to use. This is a common technique for more complex structures; however, is not necessary for simple structures.

c. Click on the ‘Theoretical Model’ tab and select the ‘Basis Set’ tab. In the ‘Basis set for’ pull-down select ‘All-electron calculation.’ This will refresh your session and switch the screen back to the ‘Job Control Parameters’ tab; in addition it will add more tabs.
d. Click on the ‘Basis Set’ tab again and select 6-31G from the ‘Basis Set’ pull down.
e. Press the Simulate button.
f. When your job has finished running, an image of the molecule will be present on the screen. The output log can be accessed in the same way it was in the semi-empirical calculation. However, you need to search for slightly different words using the ‘Find’ function.
    i. LOCATED = will take you to the region in the output where you can easily find the optimized coordinates, bond lengths, nserch, and total wall clock time.
    ii. TOTAL MULLIKEN = will give you the charge on the atoms, but be careful because the TOTAL MULLIKEN is printed off for each nserch and the only charge that matters is the one corresponding to the optimized geometry at the bottom of the output.
g. You could also use the MacMolPlot GUI to process the results, keeping in mind that you cannot get atomic charge data from it.

H_2 and N_2 Practice:

Perform the same calculations (semi-empirical PM3 and *ab initio* 6-31G) except replace CO with H_2 and N_2. Report the same results.

Problem 1 Question:

1.) What method of calculation compared the best with experimental results?
2.) Which method of calculation took a longer amount of time?
3.) Keeping 1 and 2 in mind, why would it be more beneficial to take the geometry generated from the semi-empirical calculation and set it as an input for the *ab initio* calculation?

PROBLEM 2: Formaldehyde

Perform the same calculations on formaldehyde ($O=CH_2$) and obtain the same data as in Problem 1.

Problem 2 Questions:

1.) Which method gave results closer to the literature values?
2.) Is formaldehyde a polar molecule? Why?
3.) What are the major differences between the input geometry, based on average bond lengths and VSEPR, and the optimized geometry (aka: How do bond lengths, angles, etc. change from what you were expecting)? What accounts for these differences?

PROBLEM 3: Ozone

Oxygen is a paramagnetic compound, which means it has unpaired electrons in degenerate orbitals, despite having an even number of electrons. Ozone (O_3) has 18 valence electrons and could be diamagnetic or paramagnetic. In order to determine its most stable state a slightly different method needs to be applied because restricted HF (what we have used up till now) cannot handle unpaired electrons.[5] We will use restricted open shell HF (ROHF) in order to determine if the diamagnetic or paramagnetic compound is the most stable for ozone. In addition, determine which geometry (triangular or bent) is most stable for ozone (Hint: not all combinations of geometry and spin will be possible). The lower the energy of the compound, the more stable it is.

Diamagnetic Directions:

1.) In the QC-Lab v2.0 window enter the starting geometry for ozone and then open the 'Job Control Parameters' under the 'Theoretical Method' tab.
2.) Change the 'SCF Type' pull-down menu to 'Restricted Open Shell Hartree-Fock Calculation.'
3.) Make sure the correct input geometry has been entered and the basis set 6-31G is selected. Then click the 'Simulate' button.
4.) Obtain data as before.

Paramagnetic Directions:

1.) In the QC-Lab v2.0 window enter the starting geometry for ozone and then open the 'Job Control Parameters' under the 'Theoretical Method' tab.
2.) Change the 'SCF Type' pull-down menu to 'Restricted Open Shell Hartree-Fock Calculation.' Also, type '3' in for the 'Spin Multiplicity.'
3.) Make sure the correct input geometry has been entered and the basis set 6-31G is selected. Then click the 'Simulate' button.
4.) Obtain data as before.

Problem 3 Questions:

1.) Is ozone a polar molecule? Why?
2.) Based on your calculations is ozone predicted to be paramagnetic or diamagnetic?
In truth ozone has two weakly coupled electrons meaning formally it is diamagnetic but the energy difference is small so at times it appears paramagnetic.
3.) How do your calculations compare to reality (did they give the correct magnetism, was the energy difference small)? Note: room temperature can excite molecules with an energy difference of approximately 0.6 kcal/ mol.
4.) Which geometry was lower in energy for the diamagnetic species? Paramagnetic? Do these results agree with the literature results?

---

[5] This comes from the way the orbitals are represented in the programming of the calculations. In RHF, an orbital must get filled with two electrons before the next orbital can be filled. In UHF and ROHF, the orbitals are separated by spin so each orbital is filled with one electron. This means that UHF has twice the number of orbitals (aka instead of one S orbital with two electrons there are two S orbitals, one for each spin.

DISCUSSION QUESTIONS:

1.) How would you, using a theoretical (model/ method), calculate the change in enthalpy for the following reaction:

$$H_2 + CO \longrightarrow H_2CO$$

Be sure to describe all the calculations you would run. (Hint: Assume delta H equals delta E. Connection Questions: Is this a safe assumption? When does this assumption break down?)

2.) If you were to model a protein, using one of the computational approaches that we have considered here, which type of calculation would you use (semi-empirical or *ab initio*)? Why?

3.) In the literature, computational-based papers often have a section devoted to method justification where computations are performed on well studied molecules/ systems and compared to the experimental data. Why are these comparisons important? (Be sure to relate to the results you obtained.)