

## **Lab 1**

### **Single point energy, geometry optimization and population analysis**

The purpose of this introductory project is to introduce you to the ab-initio methods and the Gaussian program and show you how to build an input file for Gaussian runs. In particular, you will learn to construct Cartesian and Z-matrices to describe the geometry, and to enter requests for particular calculations using particular basis sets.

We will calculate some molecular properties of the following atomic and molecular systems: hydrogen and fluoride atoms, hydrogen and hydrogen fluoride molecules (HF), water and ammonia.

In the course of doing these, you will see some of the features that you may have encountered in courses on simple bonding theory, molecular orbitals theory and other quantum chemical methods.

### **Theoretical and computational aspects of ab-initio procedures**

There are several procedures that can be carried out with any ab-initio program. This laboratory will focus on two main procedures: Single point calculations and Geometry optimization calculations.

### **Single point calculations**

This procedure simply calculates the energy, wave function and other requested properties at a single fixed geometry. It is usually done first at the beginning of a study on a new molecule to check out the nature of its wave function. It is also frequently carried out after a geometry optimization, but with a larger basis set or a more accurate method than is possible with the basis set and method used to optimize the geometry. Thus, for very large systems the geometry may be optimized at the Hartree-Fock level with the (relatively small) 3-21G basis set, but energy differences between isomers are then explored with more accurate methods and larger basis sets.

## Geometry optimization

Geometry optimization is a name for the procedure that attempts to find the configuration of minimum energy of the molecule. The procedure calculates the wave function and the energy at a starting geometry and then proceeds to search a new geometry of a lower energy. This is repeated until the lowest energy geometry is found. The procedure calculates the force on each atom by evaluating the gradient (first derivative) of the energy with respect to atomic positions. Sophisticated algorithms are then used at each step to select a new geometry, aiming for rapid convergence to the geometry of the lowest energy. In the final, minimum energy geometry the force on each atom is zero.

It is important to recognize that this procedure will not necessarily find the global minimum, i.e. the geometry with the lowest energy. By its nature, a successive search for a minimum finds a local minimum but not necessarily the lowest. In fact, the optimization procedure stops when it finds a stationary point, i.e. a point where forces on the atoms are zero, and this may be also a saddle point (i.e. a transition structure). This will occur particularly if we restrict the symmetry of the molecule and do not allow the program to search the full space of molecular configurational degrees of freedom. For example, if we restrict  $\text{NH}_3$  to be planar, we will get an optimized geometry for a planar structure. This is a minimum energy configuration within this restricted subset of configurations, however it is actually the transition structure involved in the “umbrella-like” flipping of the molecule from one pyramidal structure to the other.

It is always a good idea to start a procedure of a geometry optimization calculation with a small basis set and a relatively poor method before we move to the basis set and method of choice for a particular problem. You can then start the final geometry optimization from the geometry selected by the simpler and less accurate approach. We can even use the second derivatives of the energy with respect to atomic coordinates (obtained during the optimization process or from a frequency run<sup>1</sup>) to further improve the geometry optimization.

---

<sup>1</sup> A procedure to evaluate the vibrational normal modes of a molecule.

### Important notes about Gaussian calculations:

1. The format of the Gaussian input file is strict. You must follow the rules detailed in the Gaussian guide.
2. In Hartree-Fock calculations use the SCF=Tight option (for a tight convergence<sup>2</sup>).
3. After each run, make sure that the execution was accomplished successfully: The following line 'Normal termination of Gaussian' should appear at the end of the output file.
4. In geometry optimization calculations make sure that convergence has been accomplished.<sup>3</sup>

---

<sup>2</sup> Single point direct SCF calculations are run with **modest** convergence criteria automatically in the interest of speed. **The default for this case is sufficient for 0.1 kcal mole<sup>-1</sup> accuracy in the SCF energy and 3 decimal places in the density matrix-sufficient for population analysis, electrostatic potential derived charges, and the like. SCF=Tight requests full convergence for this case.**

<sup>3</sup> At the end of every optimization cycle and its SCF calculation, the following four parameters (see below) are checked, if the answer is YES for all of them the optimization has converged. For example,

Item	Value	Threshold	Converged
Maximum Force	0.000001	0.000450	YES
RMS Force	0.000001	0.000300	YES
Maximum Displacement	0.000000	0.001800	YES
RMS Displacement	0.000001	0.001200	YES
Predicted change in Energy=-2.172892D-13			
Optimization completed			

# 1. Hydrogen molecule

We start with the simplest possible molecule, the hydrogen molecule. The table below shows example of a Gaussian input file for this molecule.

Read the *Gaussian guide* and create the following Gaussian input file:

%	
Route section	# RHF 3-21G Pop=Regular SCF=Tight
Title section	Hydrogen molecule, 3-21G, R=0.734
Charge and multiplicity	0 1
Molecular specification	H H 1 0.734

Make sure you understand each line and each flag in this input.

## # RHF 3-21G Pop=Regular SCF=Tight

This line is the *route section* of this input file. **RHF** is a flag for Restricted Hartree-Fock calculation, **3-21G** denotes the basis set to be used in this calculation. **Pop=Regular** is a request about how the program should provide output about populated and non-populated levels (pop=Regular implies that five highest occupied and five lowest virtual orbitals are printed, along with the density matrices and a full orbital by orbital and atom by atom Mulliken population analysis). The absence of the flag **OPT** implies request for a single point calculation (the molecular system is taken at a fixed geometry). **SCF=Tight** is a flag requesting tight convergence criteria for the self-consistent field calculation.

## Hydrogen molecule, 3-21G, R=0.734

This line is the *title section* of this Gaussian input file. You can write your own title, however by convention it contains the essential information about the calculation – here the molecule name, the basis used and the bond length.

```
0 1  
H 0.0 0.0 0.0  
H 0.734 0.0 0.0
```

This section specifies the *molecular structure*. **0** is the total charge of the molecule, zero in this case. **1** stands for the spin-multiplicity of the molecule, which is  $2S+1=1$  (a singlet state) in the present case. The 2<sup>nd</sup> and 3<sup>rd</sup> lines give the molecular geometry,

In this example we use the Cartesian coordinates (Angstrom units given in floating point format) form, and the number given specify that one hydrogen atom is taken to be at the origin and the other is a distance 0.734 along the x axis. The atomic Cartesian coordinates are given in floating point (0.0), not as an integer (0), notation.

An alternative way to provide the molecular configuration is by the Z-matrix. The input for the hydrogen molecule will now look like:

%	
Route section	# RHF 3-21G Pop=Regular SCF=Tight
Title section	Hydrogen molecule, 3-21G, R=0.734
Charge and multiplicity	0 1
Molecular specification	H H 1 R  R=0.734

Here, a variable R appears in the Z-matrix<sup>4</sup>. Its value is provided after the configuration section (*separated from it by a blank line*). This is useful if the same parameter appears several times in the Z-matrix (and later when a molecular parameter is to be optimized).

*Perform the Gaussian calculation using the input above. Examine the **output** file and answer the following questions:*

- 1.1 Has the run terminated successfully?
- 1.2 What is the total energy of the hydrogen molecule? What does this total energy include (kinetic energy of electrons? Nuclei? Electron electron potential energy? Nuclear-nuclear interaction? Correlation energy? Etc).
- 1.3 How is the zero energy defined in this calculation?
- 1.4 How many atomic orbitals were used in the above calculation?

*Now perform the SP calculation above using the **6-31G(d)** basis set.*

---

<sup>4</sup> Read more about the molecular specification in the Gaussian guide.

- 1.5 What is the total energy of the hydrogen molecule? How many atomic orbitals are used in this case? How many molecular orbitals are there? How many of these molecular orbitals are occupied?
- 1.6 Which are the atomic orbitals that make substantial contribution to the populated molecular orbitals (derived from the SP calculation with 6-31G(d) basis set)? List only atomic orbitals with substantial weight (amplitude above 0.1 in absolute values).
- 1.7 Report the molecular orbitals energies and the total energy derived from both calculations (with 3-21G and 6-31G(d) basis sets). Express energies in Hartrees and in electron-volts.
- 1.8 How many basis functions are used with each one of the basis sets above? Explain the difference in MO energies and total energies of H<sub>2</sub> in the two calculations: If the number of basis functions affected this difference, discuss how it did so, and if not explain what it was that made the difference.
- 1.9 Given that the exact ground state energy of the hydrogen atom is -0.50000 Hartree (neglecting the relativistic effects, interactions of the electron and nuclear spin, and the effect of the finite nuclear mass), is the molecule bonded?
- 1.10 What is the predicted bond energy for H<sub>2</sub> with RHF method and 3-21G basis set?
- 1.11 What is the predicted bond energy for H<sub>2</sub> with RHF method and 6-31G(d) basis set?
- 1.12 Compare the calculated bond energies in 1.10 and 1.11 to the experimental value. Why are they different? Discuss the validity of your calculations (1.8 and 1.9).
- 1.13 Discuss the effect of the number of basis functions on the calculated molecular orbital energies and binding energy.

## 2. Hydrogen atom

A possible form of an input file for atomic input can be:

%	
Route section	# ROHF 3-21G Pop=Regular SCF=Tight
Title section	H atom – 3-21G Cartesian
Charge and multiplicity	0 2
Molecular specification	H 0.0 0.0 0.0

*Perform an energy calculation of a hydrogen atom (use the **ROHF** – Restricted Open Shell Hartree-Fock method and the **6-31G(d)** basis set).*

2.1 Explain the input file<sup>5</sup>.

The atomic hydrogen is an open shell system and we can use the ROHF (Restricted Open Shell Hartree-Fock) method in order to perform this kind of calculation. Note that the spin multiplicity of a hydrogen atom is 2 (a doublet state).

2.2 Use 1.5 to calculate the H<sub>2</sub> bond energy, compare to the experiment and to the bond energy calculated in 1.9.

*Perform an energy calculation of a hydrogen atom (this time use the **UHF** – Unrestricted Hartree-Fock method and the **6-31G(d)** basis set).*

2.3 Compare the calculated energy of hydrogen atom using both methods above. Explain any differences or similarities.

2.4 Why are the calculated and the experimental (given in 1.7) hydrogen atom energies different? Explain this regarding the fact that there is no correlation in this type of calculation.

---

<sup>5</sup> Use the Gaussian guide for that purpose.

### 3. Hydrogen fluoride molecule and the fluorine atom

*Calculate the electronic structure of the hydrogen and fluoride atoms using the ROHF and UHF methods with the 3-21G basis set.*

- 3.1 What are the energies of the fluoride atom in both methods? Are they the different? Why?
- 3.2 Would an 'Opt' type calculation of the fluoride atom be any different from the single point calculation?

An input file for the hydrogen fluoride molecule is:

%	
Route section	# RHF 3-21G Pop=Regular SCF=Tight
Title section	Hydrogen fluoride molecule – 3-21G Cartesian
Charge and multiplicity	0 1
Molecular specification	H 0.0 0.0 0.0 F 0.9374 0.0 0.0

*Calculate the electronic structure of the hydrogen fluoride molecule using the above input file.*

- 3.3 Report the total energy of the hydrogen fluoride molecule. Compare the atomic charges and the dipole moment of the H<sub>2</sub> and the hydrogen fluoride molecules.
- 3.4 Use the atomic energies in 3.1 to calculate the bond energies of HF. (Use the calculated atomic energies, not the experimental atomic energies. Why?). Which of the atomic energies in 3.1 is valid for this calculation?
- 3.5 Use the atomic energies in 3.1 and the energy in 1.2 to calculate the bond energies of H<sub>2</sub>.
- 3.6 Compare between the bond energies of HF and H<sub>2</sub> and to the literature. Is it valid to compare between the bond energies of HF and H<sub>2</sub>?
- 3.7 Compare the bond energy of H<sub>2</sub> calculated in 3.5 to that calculated in 2.2. Which is more close to experiment? What is the effect of the basis set?

3.8 Which are the atomic orbitals that contribute substantially to the occupied molecular orbitals considered? (List only atomic orbitals with substantial weight, larger than 0.1).

3.9 Which are the atomic orbitals that contribute most substantially to the bonding? How do they affect the bond type?

## 4. Water and Ammonia

Next we explore the polyatomic molecules water and ammonia. An example of an input file for a water molecule computation is:

%	
Route section	# RHF 3-21G Opt SCF=Tight
Title section	Water 3-21G RHF
Charge and multiplicity	0 1
Molecular specification	O H 1 R H 1 R 2 A  R=1.0 A=105.0

Explanation:

# RHF 3-21G Opt SCF=Tight

In this line **RHF** is the method, **3-21G** is the basis set, **Opt** is a request for a geometry optimization (the default uses the Berny algorithm) and **SCF=Tight** is a request for a tight convergence.

O  
H 1 R  
H 1 R 2 A

R=1.0  
A=105.0

The first assigns first atom to be an oxygen atom. In the second line, the distance between the second atom, H, and the first atom, O, is defined by the yet unspecified variable R. The third line sets the distance between the third atom, H, and the first atom, O, to be also R. It also denotes the angle between the line connecting atoms 1 and 2 and the line connecting atoms 1 and 3 (the angle 2-1-3) as A. Then comes a blank line and following that, the parameter R is set to 1.0 (Angstrom) and the parameter A is set to 105.0 (degrees).

An example of a more complex polyatomic molecule is Ammonia. In this case it is easier to define the Z-matrix using a dummy (ghost) atom. Such an atom is not part of the structure and certainly not part of the calculation. It is only used as a reference point in the specification of the molecular structure. In the present case we put such a dummy atom on the  $C_3$  rotational axis ( $C_3$  is a symmetry operation that rotates the system in  $120^\circ$ ) where we normally consider the nitrogen lone pair to be. The Z-matrix data is:

```
=====
N
X 1 1.0
H 1 R 2 A
H 1 R 2 A 3 120.0
H 1 R 2 A 3 -120.0

R=1.0
A=115.0
=====
```

The dummy atom, always called X, introduced here as atom 2, is used here to define the molecular axis. Its distance from atom 1 (N) is arbitrary (chosen above as  $1\text{\AA}$ ). The other atoms, 3, 4, 5 are hydrogen atoms, their distance from atom 1 is R, and the angles HNX (for H=3, 4, 5) are all equal and denoted A. Now, we need to specify what are called dihedral angles. Consider for example the H atom 4. It is linked to atom 1 at a distance R with an angle A between the line 4-1 and the line 2-1. The same is true for the H atom 3. To distinguish between them we need to specify another angle, and we take it to be (for atom 4) the angle between the planes 4-1-2 and 1-2-3. This angle should be  $120.0^\circ$ , and for the H atom 5 it is taken  $-120^\circ$  in order to give the 3-fold rotational symmetry. A. The use of dihedral angles requires some practice, but you will soon get used to it. Now run the Gaussian code for this molecule.

If you are familiar with the group theory basis of molecular symmetry, you will recognize that the ammonia belongs to the  $C_{3v}$  group and water belongs to the  $C_{2v}$  group. The program should have discovered and printed this information if you have written the Z-matrix correctly.

When Gaussian is executed using the above input files, it will do a geometry optimization. Note that Gaussian varies all the variables distances and angles. In this case, the experimental values are<sup>6</sup>:

Water: OH = 0.958 Å, H-O-H = 104.45°.

Ammonia: NH = 1.008 Å, H-N-H = 107.3°.

*Use Gaussian to find optimal geometries of H<sub>2</sub>O and NH<sub>3</sub> using the 6-31G(d), 6-31G(d, p) and 6-311G(d, p) basis sets. Note that in the ammonia case the angle A is the H-N-X angle. Answer the following questions<sup>7</sup>*

- 4.1 How can you deduce from your output that your calculations have converged?
- 4.2 For each of molecule (water and ammonia), make a table of the results for the optimal geometry and total energy obtained with each basis set. Compare your results with experiment. Discuss the effect of the basis set on the results.
- 4.3 What is the dipole moment along each axis of the water and ammonia molecules in the optimized structures? What is the total dipole moment of each molecule? Clearly define your convention for the axes, discuss the obtained results and compare to literature.

---

<sup>6</sup> These values are taken from the CRC handbook of chemistry and physics website (<http://www.hbcpnetbase.com>)

<sup>7</sup> The information about the energy of the optimized molecule (in an output file) begins with the last 'SCF Done' before the 'Optimized Parameters' table. For more information, use the annotated output files on the web page.