

Electronic Structure calculations in *Gaussian*

It is imperative to preoptimize any geometry using semi-empirical methods (PM3 etc before submitting to *ab initio* calculations. In cases where PM3 arrives at unreasonable minima, it may be useful to submit directly to *ab initio* calculations. The easiest way to set up calculations in *Gaussian* is with the aid of the graphical user interface (GUI) *WebMO*. The cartesian file generated in *MOPAC*, *Gaussview* or any other text editing program can be uploaded into *WebMO*. Details on using *WebMO* are not described here as they are described in the *WebMO* manual:

http://www.webmo.net/download/WebMO_Users_Guide.pdf

Keywords are specified with dropdown menus and can be further modified in the *Generate* window.

1. Geometry optimization

The following keyword input is typical for a ground state geometry optimization:

```
#N B3LYP/6-31G(d) Opt Freq Test temperature=195
```

How much output is generated during a geometry optimization (and actually all other types of *Gaussian* calculations as well) depends on the beginning of the keyword line:

#P will provide somewhat more detailed output

(#N) will provide normal output (normal is the default)

#T will provide somewhat less output (terse)

A geometry optimization (keyword **opt**) can be performed in three different types of coordinate systems. The way the geometry is defined in the input file is actually separate from deciding in which coordinate system the optimization will be performed:

opt=Z-Matrix will optimize the geometry in internal coordinates (as provided in the input file)

opt or **opt=Redundant** will optimize the geometry in redundant internal coordinates (chosen as default coordinate system)

opt=Cartesian will optimize the geometry in Cartesian coordinates

The level of theory in this example is the DFT method **B3LYP**. It is recommended over **HF** (Hartree Fock). The basis set is **6-31G(d)**. This is adequate for first and second row elements.

2. Frequency calculations

The keyword **Freq** calls for a second calculation that computes the frequencies and with it the energy of the molecule. For all methods used in *Gaussian*, the energy will be given in atomic units (au). The atomic unit of energy is called *Hartree* and equates to kcal/mol as follows:

$$1 \text{ Hartree} = 627.509391 \text{ kcal/mol}$$

The energies of molecular systems as calculated by most ab initio methods are calculated relative to separate electrons and nuclei. Energies for molecular systems are therefore very large and negative. The energy to be reported is the *Free Energy* (and not the B3LYP energy, ZPE or Internal Energy).

3. Specifying temperature and Isotopes:

If no temperature is specified, the energy is computed at room temperature of 298K. The temperature, however, can be specified as in the above example. It must be reported in Kelvin (unit must not explicitly be reported). Note, that the geometry is independent of temperature. Therefore, specifying a temperature for a plain geometry optimization (without the **Freq** keyword) is nonsensical. For the same reason, an isotopic substitution will not change the geometry. To calculate the energy of an isotopically substituted structure, the keyword **ReadIsotopes** must be included inside the **Freq** command

```
#N B3LYP/6-31G(d) Freq=ReadIsotopes
```

```
0, 1
```

```
[coordinates]
```

```
195 1.0 0.9135
```

```
[isotopic masses]
```

blank line

temperature, pressure, scaling factor

rounded to integer values and in same order as coordinates above. An example is shown below:

```
12
```

```
2 (deuterium)
```

```
1
```

```
1
```

```
7
```

```
...
```

4. Multiprocessor Computing (Parallel Processing)

If calculations are run on multi-core processors, the calculation can be run as a parallel job. In addition, memory must be allocated to each core. The default in Gaussian is 6 MW (1 MW=2 MB). In the following example, a job is run over eight cores each being allocated 1024 MW. Multicore processing should not be confused with parallel processing over separate computers. For this the additional Gaussian software *Linda* is required, which we currently do not have access to.

```
%Mem=1024MW
%NProcShared=8
#N B3LYP/6-31G(d) Opt Freq Test temperature=195
```

5. A Deeper Understanding of optimizations

Here are a few important concepts of geometry optimizations for the wary practitioner:

After each iteration of the geometry optimization, the output file contains a summary of the current stage of the optimization:

	Item	Value	Threshold	Converged?
Maximum	Force	0.021672	0.000450	NO
RMS	Force	0.018596	0.000300	NO
Maximum	Displacement	0.038954	0.001800	NO
RMS	Displacement	0.033876	0.001200	NO
Predicted change in Energy=-1.250480D-03				

The first two lines contain the maximum remaining force on an atom in the system as well as the average (RMS, root mean square) force on all atoms. In any case of doubt, this information is given in atomic units (here: Hartrees/Bohr and Hartrees/Radians). Together with the actual value for the current structure appears the **Threshold** value. The third and fourth convergence criteria are the maximum displacement, that is, the maximum structural change of one coordinate as well as the average (RMS) change over all structural parameters in the last two iterations. Once the current values of all four criteria fall below the threshold, the optimization is complete. The convergence criteria can be changed in the following way:

Using the opt keyword

opt	without any additional information sets the RMS force criterion to 3×10^{-4}
opt=tight	will set the RMS force criterion to 1×10^{-5} and scale the other three criteria accordingly

`opt=verytight` will set the RMS force criterion to 1×10^{-6} and scale the other three criteria accordingly

Choosing tighter convergence criteria will, of course, give improved results but will also need more computer time. The default settings are appropriate for small systems. Especially for large structures, however, convergence of the last two criteria can be very slow and it is sometimes advisable to stop optimizations before all four criteria are fulfilled. The maximum number of optimization cycles depends on the size of the system and is automatically adjusted by *Gaussian*. If a particular setting of optimization cycles is desired, however, this can be specified using

`opt=(maxcycles=n)` (default maxcycle in Gaussian is n=20)

The default optimization algorithm included in *Gaussian* is the "Bery algorithm" developed by Bernhard Schlegel. This algorithm uses the forces acting on the atoms of a given structure together with the second derivative matrix (called the Hessian matrix) to predict energetically more favorable structures and thus optimize the molecular structure towards the next local minimum on the potential energy surface. An explicit calculation of the second derivative matrix is quite costly, the Bery algorithm constructs an approximate Hessian at the beginning of the optimization procedure through application of a simple valence force field, and then uses the energies and first derivatives calculated along the optimization pathway to update this approximate Hessian matrix. The success of the optimization procedure therefore depends to some degree on how well the approximate Hessian represents the true situation at a given point. For many "normal" systems, the approximate Hessians work quite well, but in a few cases a better Hessian has to be used. Often it is sufficient to calculate the Hessian matrix explicitly once at the beginning of the calculation and then use the standard updating scheme of the Bery algorithm. This is specified using the

`opt=calcfc`

keyword. In some very rare cases, the Hessian changes considerably between optimization steps and must then be recomputed after each optimization step using the

`opt=calcall`

keyword. In case a number of different options are to be specified for a geometry optimization, the options must be given in parenthesis:

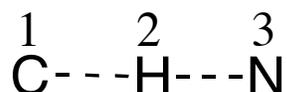
`opt=(Z-Matrix,calcfc,tight,maxcycles=25)`

6. Transition State optimization

Two methods of locating transition state geometries will be described:

(a) As in the case of ground state optimizations, it is essential to preoptimize the transition state geometry semi-empirically in *MOPAC*. The output of this calculation is then submitted to the *ab initio* calculation in *Gaussian*.

Transition state calculations can be tricky and require tricks specific to each case. Let us consider the transition state of an amide mediated deprotonation:



For any transition state, the imaginary vibrational frequency must be associated with the motion over the saddle point (i.e. the motion of the proton from carbon to nitrogen). In other words, the final optimized transition state should exhibit one negative (imaginary) frequency that corresponds to the translational motion of the proton from carbon to nitrogen. This principle serves as a final check for the validity of the transition state. For deprotonations the magnitude of the imaginary frequency should be around 1000 cm^{-1} . For heavier atoms this frequency will be in the range of $100\text{-}300 \text{ cm}^{-1}$. An excellent treatment of transition state theory can be found in "Chemical Kinetics and Reaction Dynamics" by Paul L. Houston (Cornell University, Dover Publications, Inc. Mineola New York, 2006).

We make use of the above stated principle and restrict the C-H and H-N bond lengths to a value close to ones previously established or chemical intuition: $r(\text{C-H}) \sim 1.35 \text{ \AA}$ and $r(\text{H-N}) \sim 1.42 \text{ \AA}$. These bond lengths are unaltered during the transition state calculation and allow for a more expedient calculation. In a second transition state calculation we will release this constraint.

The *Gaussian* input for the first calculation is the following:

```
#N B3LYP/6-31G(d) Opt=(TS,EstmFC,noeigentest,ModRedundant) Test
```

```
0,1  
[coordinates]
```

```
1 2 1.35 F  
2 3 1.42 F
```

1. We specify the optimization (**Opt**) to be a transition state (**TS**) optimization.
2. **EstmFC** means that the force constants (FC) are computed at the beginning of

the calculation by constructing an approximate Hessian (see section 2).

3. By default, the Berny optimization program checks the curvature (number of negative eigenvalues) of the approximate second derivative matrix at each step of a transition state optimization. If the number is not exactly one negative eigenvalue, the job is aborted. The test can be suppressed with the `NoEigenTest` option. It appears that Gaussian never finds a transition state unless the `NoEigenTest` keyword is included.

4. `ModRedundant` allows us to explicitly freeze (F) variables during the optimization. In this example we freeze the bondlength between atom number 1 and 2 and atom number 2 and 3. The bondlength is given in Ångstroms. Note that there must be a space between the final line of the coordinates and the `ModRedundant` input.

5. A note on the level of theory: DFT (B3LYP) has the great advantage of treating electron correlation quite well whereas HF does not treat it at all. Electron correlation is especially relevant for transition state calculations. It is therefore recommended to use DFT methods for such calculations.

The calculation must now be rerun without any coordinate restrictions and is followed by a frequency calculation to arrive at the thermally corrected energy of the transition state. The imaginary frequency should be verified at this point and can be visualized directly in *WebMO*.

The input for the second transition state calculation is the following:

```
#N B3LYP/6-31G(d) Opt=(TS, CalcFC, noeigentest) Freq Test temperature=195
```

1. The Hessian is calculated exactly with `CalcFC`.
2. The optimization can now freely adjust the bondlengths between atoms 1, 2 and 3.
3. The optimization is followed by a frequency calculation at 195 Kelvin.

(b) The second method of locating transition state geometries is the 'scanning' technique. Identify the bondlength, bondangle or dihedral angle that most closely describes the imaginary frequency of the transition state and perform a scan of this variable. An example is given below:

```
#N B3LYP/6-31G(d) Opt=(ModRedundant) SCF=(Tight,NoSymm) Test
```

```
0,1  
[coordinates]
```

```
1 2 3 125.6 S 10 +2.5
```

The ModRedundant keyword is used to perform a scan. This scanning method is also called a *Relaxed Potential Energy Scan*. In the above example the angle between atom number 1, 2 and 3 is first optimized at 125.6 and once completed increased by 2.5 degrees. Once this geometry is optimized the angle is again increased by the same amount. The process is repeated ten times. In *WebMO* the B3LYP energies can be viewed by clicking on the watchglass at the top of the geometry sequence box. Alternatively, the output file can be opened by *Gaussview* and the reaction coordinate can be viewed by selecting *Scan* under the *Results* menu. Ideally, one will find that the B3LYP energy during this sequence rises, reaches a maximum and subsequently decreases. Choose the geometry of maximum energy and submit this to a restricted transition state calculation as described in point (a). Selecting the appropriate scanning variable can be tricky. Sometimes scanning a bondangle instead of a bondlength or a dihedral angle instead of a bondangle may work better. To choose the ideal variable, attempt to understand what the key geometrical changes are that occur at the transition state.

(c) Alternate ways of locating transition states are to use so called saddle-point calculations: QST2 and QST3. In the former one submits the reactant and product structure and the program interpolates between the two structures to find a reasonable starting point for a transition state calculation. It is absolutely necessary for the atom numbering to be consistent between both structures. This can be achieved in *Gaussview* using the Connection Editor. QST3 requires in addition a best estimate for the transition state geometry. Thus far, I (ach) have not been able to use this method to my advantage and therefore limit this section to simply mentioning it.

7. IRC calculations

Transition state geometries may be connected to ground state geometries with an intrinsic reaction coordinate (IRC) calculation. In the process one follows the path downhill from the transition state in either the *forward* or *reverse* direction. The direction does not correspond to the user's understanding of forward and

reverse for his/her particular chemical system. The calculation is set up in the following way:

```
#N B3LYP/6-31G(d) IRC=(CalcFC,Forward)
```

or

```
#N B3LYP/6-31G(d) IRC=(CalcFC,Reverse)
```

An IRC calculation will not reach the bottom of the energy well. Therefore, submit the output of the IRC to a geometry optimization to reach the ground state.