

Introduction to Computational Chemistry Exercises Part 4

Eno Paenurk & Patrick Finkelstein & Felix Pultar

May 17, 2023

Goals:

- Load and run ORCA
- Perform optimization, frequency, and single point calculations for ground and transition states
- Read the relevant data from the output to construct a full reaction energy profile

Today, you will examine the reaction shown below, the thermal ring-opening of cyclobutene to form butadiene. You will combine all the different job types we mentioned in the lecture to construct a reaction energy profile.



1 Preparations

We will be using ORCA version 5.0.3. Make sure you have this version loaded by typing module list

Also make sure that you are using the icc-2023 working environment.

1.1 Installing suborca.sh on Euler or Grace

As with autodE we have made your life easier by providing a submission script. Update git and you will find suborca_grace.sh and suborca_euler.sh. As in Week 2, copy the version you need into your bin directory (reminder: this is in /cluster/home/username/ for Euler and /home/netid/ on Grace) and rename it to suborca.sh. Make the script executable. Now you will be able to use it from anywhere.

1.2 Preparing the workspace

We will run quite a few calculations today and it is good practice to have a clean folder setup. Navigate to the part4 folder in your workspace and make three new folders called starting_material, transition_state, and product.



2 Ground state structures

2.1 Generating coordinates

Within Chemdraw, draw the starting material (cyclobutene) and the product (butadiene). Transform them into 3D structures by using the smi2xyz script from week 3 and save them as sm_initial.xyz and prod_initial.xyz in their respective folders.

Important: Make sure the input xyz file does not have the same basename as the input file we will generate. Otherwise ORCA will try to overwrite your starting geometry that it also tries to read in, and that can cause problems. This is why we suggest you add the suffix _initial.

2.2 General information on the input file

The general structure of the input file is:

```
! Keywords
% Resources
% More detailed paramaters
* system information
```

The keyword section is preceded by an exclamation mark (!). Extra settings, including the resources you want to use, can be specified in blocks starting with the percentage sign (%). Finally, the compound information section starts with an asterisk (*).

The keywords can also appear in several lines:

```
! Keywords1
! Keywords2
% Resources
% More detailed paramaters
* system information
```

Anything that is preceded by a hash sign (#) is ignored and can be used to comment the input:

```
! Keywords1 # These are my first keywords
! Keywords2 # These are additional keywords
#yolo
% Resources
% More detailed paramaters
* system information
```

You can also have any amount of blank lines in the input:

```
! Keywords1 # These are my first keywords
! Keywords2 # These are additional keywords

% Resources
% More detailed paramaters
* system information
```

2.3 Generating the input file

As we mentioned in the lecture, there are countless combinations of functionals and basis sets to accomplish our task. To guide you a bit (and also make sure that the calculations do not take too long) we have chosen some parameters for you. We will be using the BP86 functional, together with the def2-SVP basis set. This should give us decent geometries, whilst only taking 1-2 minutes to optimize the structures and calculate the frequenices.

We will begin with the starting material. Navigate to the correct folder and make a new file called opt_freq.inp, either directly in the cluster with vim or with an external text editor. Type the following contents and save the file.



```
! BP86 def2-svp d3bj CPCM(Acetone) opt freq
%pal nprocs 4 end
%maxcore 2048
*xyzfile 0 1 sm_initial.xyz
```

Let's quickly go through all of what we just wrote:

- As we mentioned in the previous section, the first line is the keyword line, which starts with an
 exclamation mark. The order of the keywords is not important but we encourage you to stick to
 something similar, to make sure you don't forget anything and make it easier to compare different
 input files.
- BP86 refers to the functional we are using. There is no need to specify that we want to use DFT, as this is implied by asking for a functional.
- def2-svp is the basis set that we will use.
- d3bj refers to the London dispersion correction. We advise you to apply this to DFT calculations (unless you know that this is already accounted for).
- *CPCM(Acetone)* calls the implicit solvent model for acetone.
- opt freq are the two jobs we wish to run, a geometry optimization followed by frequency calculation. As we mentioned in the lecture it is **important** to run the frequency calculation with identical parameters as the geometry optimization, otherwise you have a mismatch between frequencies and related information of the optimized geometry. The easiest way to ensure this, is to do both job types in a single calculation. ORCA will know to first optimize the geometry and then run the frequency analysis.
- The second and third line, starting with the %'s, contain information on the resources we want ORCA to use. In this case we ask for 4 processors with 2048 MB memory each. The script we provide will read in these lines so it is important that they exist in this exact format. Note that memory refers to fast memory ORCA can access during the calculation and not space on the harddisk.
- Finally, the last line tells ORCA the system information. By envoking xyzfile we say that the coordinates can be found in an external file (called sm_initial with a charge of 0 (the first number) and a spin multiplicity of 1 (the second number).

Run the same workflow for the product (don't forget to change the folder and modify the name of the xyzfile).

2.4 Submitting the job

Enter the starting_material folder where both the input file (opt_freq.inp) and the coordinates (sm_initial.xyz) files are located and type

```
suborca.sh opt_freq.inp
```

Verify with squeue that you job is in the queue. Don't forget to submit the calculation for the product molecule.

2.5 Checking if the job completed successfully

If you run the calculations on the cluster with the resources we suggest, the job should be done after a few minutes.

A fast way to check if an ORCA calculation terminated successfully without opening the file is by typing tail opt.out

This will give you the last 10 lines of the file. It's enough to only show the last two lines and you can do this with

```
tail -n 2 opt.out
```

If the job has terminated without errors you should see

```
****ORCA TERMINATED NORMALLY****
TOTAL RUN TIME: 0 days 0 hours 1 minutes 23 seconds 748 msec
```



2.6 Checking the generated files

As you can see, ORCA generate a lot of files. Some files are only generated for specific job types. A list of possible files for the job types considered in this session is given below.

File	Description	Occurrence
outfile	contains information and results of the calculation	always
basename.xyz	optimized xyz coordinates	opt
basename_trj.xyz	trajectory of the optimization	opt
basename.gbw	the wavefunction	always
basename.opt	numerical information about the optimization	opt
basename.engrad	the energy gradient of the optimized structure	opt
basename.cpcm	surface information for the solvent model	cpcm
basename.smd.out	information for the SMD solvent model	smd
basename.hess	the Hessian and vibrational spectrum information	freq
basename_property.txt	Values of some selected properties	always
basename.prop	Property file for further calculations	always

The first thing you should do is to check the .out file that was generated. As we have asked for a geometry optimization and frequency calculation, this should be full of information.

One of the important pieces of information is whether the calculated structure has the correct number of imaginary frequencies (this is a ground state structure so you should have 0). You can open the .out file in a text editor and search for "VIBRATIONAL FREQUENCIES" which is followed by the list of frequencies. The imaginary ones are listed with a minus sign and denoted as "imaginary".

You can also check this directly from the terminal, by using a new command, grep:

grep "VIB" -A 11 opt.out

grep will search for all the occurences of the string in the quotation marks and prints the lines where the string is found in the file. VIB only occurs in the context of VIBRATIONAL FREQUENCIES, so we can get away with not specifying the whole two words. In order for us to read the some of the lines under it, we can use the -A flag followed by the number of lines we wish to see. For our purpose, 11 is the ideal number as it prints up to the first frequency, so we can check if it is a positive or a negative value (which denotes an imaginary frequency).

You can also open the opt.out file with the visualization software that we've shown.

Chemcraft

You can open the output file with Chemcraft. To inspect the optimization steps, expand the "Optimization steps" list under "Optimization" on the left-hand side panel. You can visualize the different structures along the optimization trajectory by clicking on the entries in that list or navigating with the arrow keys.

You should see all the frequencies in the left side-bar, under the optimized geometry. Make sure you have the correct amount of imaginary frequencies, which, as this is a ground state structure, should be 0.

¹For large and floppy molecules, you might get small imaginary frequencies (< 50 cm⁻¹). These often stem from numerical errors and may sometimes be ignored (consult your local computationalist).



Avogadro

You can visualize the optimization steps by opening the "basename_trj.xyz" file, which contains the optimization trajectory. Select "View - Properties - Conformer properties"; this will open a window with the list of structures. You can click through the structures or navigate with the arrow keys to see how the optimization proceeded. Note that Avogadro does not update the bond visualization as the distances change, so it's less easy to spot bond-breaking and bond-making.

To check if you have the correct amount of imaginary frequencies, you will have to inspect the "basename.out" file either in Chemcraft or Molden, or by opening the file in a text editor.

Molden

You can also open the output with Molden by typing

molden opt.out

Within the control panel (sometimes it opens in the background), you can open the "Geom. conv." window from the lower right corner that displays graphs about the convergence of the calculation, and choose the points you want to visualize by click on the respective points on the graphs.

You can check the list of frequencies by clicking in the top right on the "Norm. Mode" box. This will give you a IR spectrum as well as a list of all the frequencies.

Once you have verified that you indeed have the correct structure and vibrational frequencies, you can extract the correction term for the Gibbs free energy that we will use later on. This number can be found towards the end of the file:

```
For completeness - the Gibbs free energy minus the electronic energy G-E(el) ... 0.05801323 Eh 36.40 kcal/mol
```

You can extract this opening the file with a text editor and scrolling to the very end. Depending on the text editor it might be fast to just search for "G-E(el)".

You can open the file with vi and either scroll to the end use the key combination shift + g. Alternatively, if you do not want to open the file and just want the result printed in the terminal you can make use of grep:

```
grep "G-E(el)" opt.out
```

Note down these G-E(el) values for both the starting material and the product. We will add this to the electronic energy, which we will re-calaculate later on.

3 Transition state structure

In the following section we will show you how to perform the three different strategies we mentioned in the lecture. We recommend you try all three methods at some point as they each have their advantages.

3.1 Option 1: OptTS

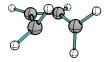
This method can work when you already have a good guess of the transition state structure, but will usually fail if your guess was too far away. As the reaction considered here is relatively simple, it is possible to make a good guess, based on chemical intuition and some additional tips from our side.

The input file, call it opt_freq.inp, is very similar to the normal geometry optimization, albeit with the optts keyword:

```
! BP86 def2-SVP D3BJ CPCM(Acetone) optts freq
%pal nprocs 4 end
%maxcore 2048
*xyzfile 0 1 ts_initial.xyz
```



For the ts_initial.xyz guess you could try to open the cyclobutene slightly, while also rotating both CH₂ groups clockwise, like in the figure:



You can do these structure modification in either Chemcraft or Avogadro (check the material of week 3). Submit it using the suborca.sh script.

3.2 Option 2: Scan

Any internal coordinate can be scanned with the **%geom** block. Together with the **ScanTS** keyword, the highest energy structure along the scan is selected for the TS optimization. For example, to scan a bond distance between two atoms numbered 1 and 2 from 3 Å to 1 Å in 10 steps, and optimize the TS along the pathway, you could use the following input (note that the indentation is not required):

```
! Method ScanTS
% resources
%geom
   Scan # perform a scan
   B 0 1 = 3.0, 1.0, 10 # scan the distance between atoms 1 (0) and 2 (1) in 10 steps
   end
end
*xyzfile 0 1 reactant.xyz # xyz file of the reactant structure
```

Note: Atom counting starts from 0 in ORCA, which means that the first atom is numbered 0, and so on. Make sure to pay attention to the numbering when preparing scan inputs.

The starting value of the scan parameter should be close to what it is in the starting structure (reactant.xyz in the example above). In the case of a successful calculation, the final optimized TS filename is saved in "basename.xyz".

Occasionally, the ScanTS method does not manage to optimize the TS. To make use of the scan, you can then still choose the structure along the scan that seems like a reasonable guess, and submit it to an OptTS calculation.

For more information about the scan features, see ORCA 5.0.3 manual Sections 8.3.8 and 8.3.10.1.

For our exercise, we suggest that you start from the product and scan "backwards" to the starting material. For this, generate the s-cis isomer of butadiene and scan the bond between the two terminal carbons, starting from the distance in the open form to the distance you got out for the optimized cyclobutene. You can easily generate the correct isomer with some creative SMILES: as you may have noticed, drawing butadiene will typically lead to the s-trans isomer. To force the correct geometry around the C-C bond you can simply move the double bond to be between the two central carbons, and add positive charges to the terminal positions.

We also suggest to change from scants to a simple scan (we show the input below). This will skip the last part (optimizing the highest energy structure as a transition state). Quite often the scans are not ideal, as



the start or the end structure can end up being higher in energy than any potential transition state, due to unreasonable bond lengths. This way, the calculation is finished faster and you can manually choose the structure you think is reasonable for a subsequent optts calculation.

Your input file will look something like:

```
! BP86 def2-SVP D3BJ CPCM(Acetone) opt
%pal nprocs 4 end
%maxcore 2048
%geom
    Scan
    B 2 3 = 3.02, 1.57, 15
    end
end
*xyzfile 0 1 product.xyz
```

Remember to change the Scan part, so you ask for the right atoms. You can look up the atom numbering in Chemcraft or Avogadro (and subtract 1 due to the numbering of ORCA!). As our system is relatively small and we only ask for 15 steps, the calculation should be done reasonably fast. You can, of course, also choose a different internal coordinate to perform the scan.

The calculation will generate a lot of files (around 10 plus 2 per step you have requested). The interesting one for now will be the .out file. Open it with your favorite visualizer.

Chemcraft

On the left hand side you will see all the steps of the scan. In the bottom you have the option to view the scan graph. Click on this and select Energy vs scan step. From here, you can check which scan step is highest in energy and then use those coordinate as the input for a optts calculation.

Avogadro

Go to View- Properties - Conformer Properties and you can see which scan step is the maxima. Remember the number and use the .xyz file for that structure as the input for a optts calculation.

Molden

Molden cannot visualize the .out file from a scan job, but you can convert the .allxyz file (which has an ORCA-specific format) into a regular trajectory file with a simple cryptic command:

```
sed '/>/d' basename.allxyz > basename_all.xyz
```

where basename_all.xyz now contains the coordinates for each optimized step. You can open this file with Molden or any other visualization program.

3.3 Option 3: NEB

If the TS is hard to guess or your geometry preparation skills are failing you, you can try the NEB method. It is a fairly robust tool for finding TS structures, albeit often slow (we will be using DFT for this exercise, but for larger systems it might make more sense to use xtb). You need to provide the calculation with the starting structure and the final structure. There are two important things to note:

- If you have several reactants or products, you need to combine them into one xyz file. You also should usually take care to have them in a reasonable relative orientation for the reaction. As our reaction is intramolecular, this is luckily the case.
- Make sure that the atoms in the reactants and the products are numbered consistently NEB interpolates the coordinates of each atom according to its number and will give nonsense results if the atom numbers change.

For both of those points, the easiest solution is usually to open the product (or reactant, whichever seems easier) structure with Chemcraft/Avogadro, manually break it into the necessary components using the structure manipulation tools we showed in Part 2, and reoptimize the resulting structure with xtb. The



relative orientation and atom numbering should then be suitable for NEB calculations.

Important: As you may have noticed, the product, butadiene most likely optimized to the s-trans confomer, which is to be expected as it is the thermodynamically more stable isomer. However, it makes everything a lot easier if we construct the s-cis isomer, as this is the one that is formed directly after the transition state. Refer to Section 3.2 to generate the other isomer.

We will be using the BP86/def2-svp combination again. The minimal input for our NEB calculation is:

```
! BP86 def2-svp d3bj CPCM(Acetone) neb-ts freq
%pal nprocs 4 end
%maxcore 2048
%neb
   NEB_END_XYZFile "end.xyz" # xyz file of the final structure
end
*xyzfile 0 1 start.xyz # xyz file of the reactant structure
```

This input will first calculate the regular NEB trajectory connecting reactant and product. In a second step it will select the highest-energy structure in the calculation and "drive" it to the maximum along the pathway before performing a TS optimization on that point. We added the freq keyword so that we also obtain the frequencies of the suggested transition state.

There is a list of additional options you can invoke in the NEB module. Some are exemplified below:

```
%neb
   NEB_END_XYZFile "product.xyz" # xyz file of the final structure
   PreOpt_Ends True # optimize the reactant and product before NEB (default: false)
   NImages 8 # number of structures along the trajectory (default: 8)
end
```

The NEB calculation will produce quite a lot of files. The most important files are the following xyz files:

File	Description
basename_MEP_trj.xyz	xyz of the final trajectory
basename_NEB-CI_converged.xyz	xyz of the NEB-CI result (highest-energy image – HEI)
basename_NEB-TS_converged.xyz	xyz of the NEB-TS result (the optimized TS)
basename_TSOpt.xyz	same as above

These xyz files can be visualized with Chemcraft, Avogadro, or Molden. To view trajectories in Avogadro, go to "View-Properties-Conformer Properties". This will give you a list that you can navigate to visualize the steps along the trajectory.

For much more information on the NEB method, consult the ORCA 5.0.3 manual section 8.3.17.

3.4 Confirming the transition state

There are two additional steps we will do to confirm that the transition state we calculated is a true transition state connecting starting materials and product.

- 1. Check the imaginary frequency
- 2. Run an IRC calculation

3.4.1 Visual inspection of the imaginary frequency

You can check the imaginary frequency using either Chemcraft, Avogadro, or Molden.

Chemcraft

Chemcraft is the most straightforward. Simply open the .out file of the transition state optimization and you will directly see all the frquencies in the left side bar. Select the imaginary one (it will have a minus sign in front of it) and you will be presented with an animation. In the bottom left corner you have some options to play around with.



Avogadro

Unfortunately, neither Avogadro nor Avogadro 2 can directly show the vibrations from the files that are newer than Orca 4.2.1. However there is a workaround, that works well if you only want to check the imaginary frequency.

Navigate to the folder where the .out file is located and type

```
orca_pltvib optts.hess 6
```

This command will save the 7th frequency (remember, ORCA starts counting at 0) and as the first 6 entires are rotations and translations, the 7th will be the lowest frequency, which will be the imaginary one. You should now have a new file called optts.hess.v006.xyz. You can now open this in Avogadro. Under the Menu option Extensions select Animations and you can animate the imaginary frequency.

Note that in Avogadro2, you should instead click on the little clock widget on the top to access these options.

Molden

Navigate to the folder where the .out file is located and type

```
orca_pltvib optts.hess 6
```

This command will save the 7th frequency (remember, ORCA starts counting at 0) and as the first 6 entires are rotations and translations, the 7th will be the lowest frequency, which will be the imaginary one. You should now have a new file called optts.hess.v006.xyz.

Type

```
molden optts.hess.v006.xyz
```

On the molden control panel you can now click on Movie, which should be in the top-left panel. The animation will most likely be too fast, so you can slow it down by clicking on the hourglass in the top-left panel, typing in 100 and clicking apply.

Note: ORCA has a lot of these command line utilities such as orca_pltvib that make it possible to plot all sorts of properties (amongst other things) and we will make more use of these tools in the coming weeks.

3.4.2 IRC calculations

To verify that the TS you have located really connects the two intermediates (or reactant-product) that you think it does, you can use the Intrinsic Reaction Coordinate (IRC) calculation. You need to provide the calculation with the structure of the TS, and it will follow the largest imaginary mode forward and backwards to the first local minima, which should correspond to the intermediates/reactants/products. The minimal input for our case would be:

```
! BP86 def2-svp d3bj CPCM(Acetone) IRC
%pal nprocs 4 end
%maxcore 2048
*xyzfile 0 1 ts.xyz # xyz file of the TS structure
```

The IRC calculation also produces several xyz files:

File	Description
basename_IRC_Full_trj.xyz	xyz of the final trajectory
$basename_IRC_F_trj.xyz$	xyz of the "forward" trajectory
basename_IRC_B_trj.xyz	xyz of the "backward" trajectory
$basename_IRC_F.xyz$	xyz of the "forward" minimum
basename_IRC_B.xyz	xyz of the "backward" minimum



where forward and backward represent the arbitrary directions from the TS to the next minimum. The full final trajectory file provides a good visualization of the reaction from the reactant to the product, and the last two files given the list should correspond to the intermediates/reactants/products. These xyz files can also be visualized with Chemcraft, Avogadro, or Molden.

3.5 Finalizing

Whichever approach you took, you should have an output file that contains the frequencies and the thermodynamic corrections. Extract the "G-E(el)" value from that output as explained above (this value will be used to correct the barrier free energy).

4 Refining the electronic energy

It is customary to calculate single point energies on the optimized geometry with more expensive (and thereby hopefully more accurate) methods. This step is under the assumption that stationary points identified with a lower level method are also stationary points at the higher level methods, which is typically a reasonable approximation. We will thus refine the electronic energy by moving to a hybrid functional and employing a larger basis set. First, go to the folder with the starting material and make a new file called singlepoint_b3lyp.inp with the following contents:

```
! B3LYP def2-tzvp d3bj CPCM(Acetone)
%pal nprocs 4 end
%maxcore 2048
*xyzfile 0 1 opt.xyz
```

Let's again quickly go through all of what we just wrote:

- We decided to switch to a hybrid functional to obtain a higher accuracy and have chosen B3LYP.
- def2-tzvp is the (now larger) basis set.
- d3bj refers to the London dispersion correction.
- *CPCM(Acetone)* refers to the implicit solvent model for acetone.
- We do not need a specific keyword to request a single point calculation, as this is the default
- The second and third line, starting with the %'s, again contain information on the resources we want ORCA to use.
- For the last system information, remember to specify the file containing the coordinates of the optimized structure in this case we can find them in opt.xyz

Submit the calculation and then run the same job for the product (again don't forget to change the folder and modify the name of the xyzfile you are calling).

Once the calculation has finished (this should be within a minute), you can extract the electronic energy from the output file. Like before, you have multiple options to get to the information which is towards the end of the file and should look something like this.

```
FINAL SINGLE POINT ENERGY -155.851712819613
```

Open the singlepoint_b3lyp.out file and scroll to close to the very bottom, or search for "FINAL SINGLE POINT ENERGY".

If you want to open the .out file, you can use the key-combination shift-g to get to the bottom of the file. You will then have to scroll up a tiny bit.

Alternatively, you can use the grep command again.



```
grep "FINAL SINGLE POINT ENERGY" singlepoint_b3lyp.out
```

5 Constructing the reaction energy profile

The final ΔG value is calculated:

$$\Delta G = \Delta E_{\rm el}({\rm B3LYP/TZVP}) + \Delta G_{\rm corr}({\rm BP86/SVP})$$

Add the correction term from the opt.out files to the single point energy from the XXX.out files. Now, choose the 0 point, such as the starting material and calculate the relative energies of the transition state and the product with respect to the product.² The energy units you typically obtain from ORCA are in Hartree, so you can use the table below to convert this into other units you might be more familiar with.

Unit	Hartree	eV	$\rm kcal/mol$	$\mathrm{kJ/mol}$
Hartree	1	27.211	627.509	2625.499
eV	3.675×10^{-2}	1	23.061	96.485
kcal/mol	1.594×10^{-3}	4.336×10^{-2}	1	4.184
kJ/mol	3.809×10^{-4}	1.036×10^{-2}	0.2390	1

5.1 Changing reaction temperatures

Whenever you run a frequency calculation, ORCA will automatically print the thermochemical information at 298.15 K at the end of the output. You can find it under the "THERMOCHEMISTRY AT" section. It also lists all the approximations made in the calculation of the thermochemical corrections. Then the enthalpy, the entropy, and the Gibbs free energy calculations and results are printed, in that order.

It is cheap to calculate this information, so there is an option to easily do this once you already have ran the optimization and frequency calculations. You can request the thermochemistry calculation with the following input:

```
! PrintThermoChem
%geom
InHessName "dummy.hess" # the .hess file from the frequency calculation
end
%freq
Temp 283.15, 293.15, 303.15 # the temperature(s) for the thermochemistry calculations
end
*xyzfile 0 1 dummy.xyz # xyz of the structure
```

You can list several temperatures and the calculation will be performed for each of the temperatures. Note that you do not need to specify any methods. In addition, this is a very cheap calculation, so there is no need to ask for any extra resources (running this interactively or even on the login node with the default 1 core is enough).

6 Final thoughts

If you reached the end of the exercise, we invite you to try to get the reaction energy profile of this reaction with autodE, to get a sense of how much faster the automated process can be – if it works. Knowing how each of the individual steps work can allow you to tackle the cases where automated processes like autodE do not work and a more hands-on approach is required.

In addition, the skills you learned in this session will also be essential for other computations besides mechanistic ones. We will learn more about those in the coming weeks.

²The ΔG of the reaction should be around -17 kcal/mol and the ΔG^{\ddagger} should be around 31 kcal/mol