DFT reaction energy calculations

Introduction

In this practical we are going to use density functional theory (DFT) to calculate the activation and reaction free energies ($\Delta G^{\ddagger}$ and ΔG^0, respectively) for the uncatalyzed transformation of chorismate to prephenate:
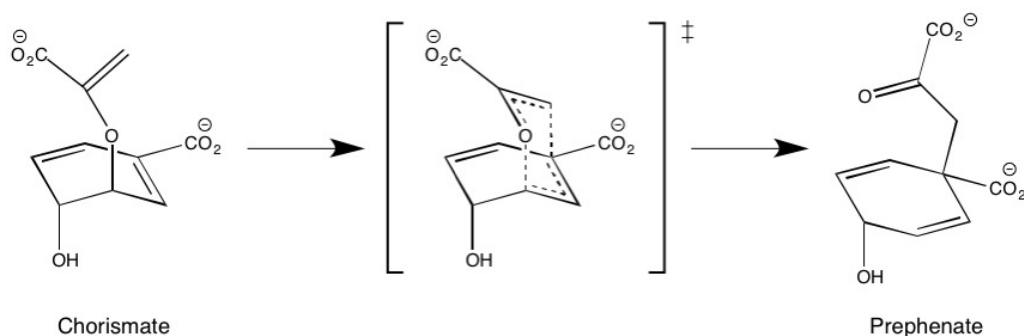

Figure 1: The unimolecular transformation of chorismate to prephenate.

These energies will later be used in the EVB reference reaction to calibrate the EVB Hamiltonian.

The reaction energies for the uncatalyzed transformation of chorismate to prephenate (Figure 1) in water are going to be calculated with DFT using Gaussian 16[1]. Structure optimization and frequency calculations at the reactant (chorismate), transition and product (prephenate) states are to be computed with the B3LYP functional[2] and the 6-31G(d,p) basis set. Dispersion effects must be included in all calculations using Grimme's B3LYP-D3 method[3,4]. Intrinsic reaction coordinate (IRC) calculations are to be performed in both directions from the TS to verify that the correct minima are connected (reactant and product state). The electronic energies will be obtained from single-point calculations on the optimized geometries (RS, TS and PS) at the b3lyp/6-311G+(2d,2p) level of theory. The final energies to be reported and used further for calibrating the EVB Hamiltonian must be corrected for zero-point energy (ZPE), dispersion and solvation effects that are calculated with the SMD model[5] using water as solvent (eps=80). While the actual DFT computations are performed by Gaussian, we will use the software REACT (https://github.com/chemREACT/REACT) to prepare inputfiles and to analyse the the output files generated by Gaussian. Figure 2 shows a screenshot of the main window of REACT and we will refer to this figure troughout the tutorial.
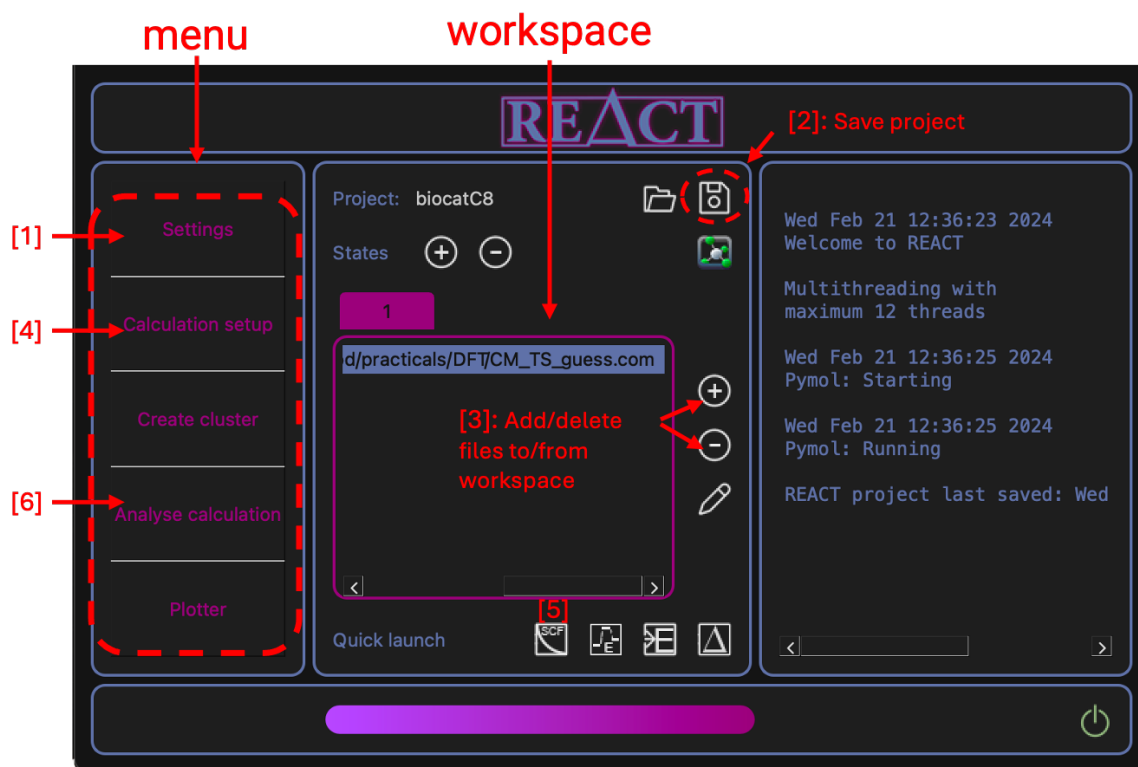
Recommended literature

- Siegbahn & Himo (2011) "The quantum chemical approach for modeling enzyme reactions"[6]
- Himo (2017) "Recent trend in quantum chemical modeling of enzymatic reactions"[7]

Practical step by step

0. Setup

1. Download course material: git clone https://github.com/benteb/biomolmod
2. Move to the Activate the virtual environment: source venv/bin/activate
3. Open REACT: python REACT.py
4. Open the settings window ([1] in Figure 2). In DFT settings, set the defaults:

   Functional: B3LYP
   Basis set: 6-31G (d,p)
   Additional keywords: empiricaldispersion=gd3

   In the uppermost right corner, change working directory to biomolmod/practicals/DFT. Close the window by clicking "Save".

5. Make a new REACT session by clicking [2] in Figure 2, choose a project name, and click save in the file handler window. As you progress in your project, remember to save you session once in a while by clicking the [2] button.
6. Open a new terminal window and log into Saga.
7. Create or move to the folder you want to work and get the full path using this command: pwd
8. Switch back to the local window were the course material is (/DFT). Open the file rsync_saga.sh and change the variables remote_user and remote_dir. We will use this script to move files to/from Saga from you local computer.
   Move files TO Saga: ./rsync_saga.sh to
   Move files FROM Saga: ./rsync_saga.sh from
   During this practical, if you see a file with this symbol '~' behind it, it means you have (accidentally) overwritten a file. This is a backup for the file that was overwritten. If it was not an accident, you can delete this file.
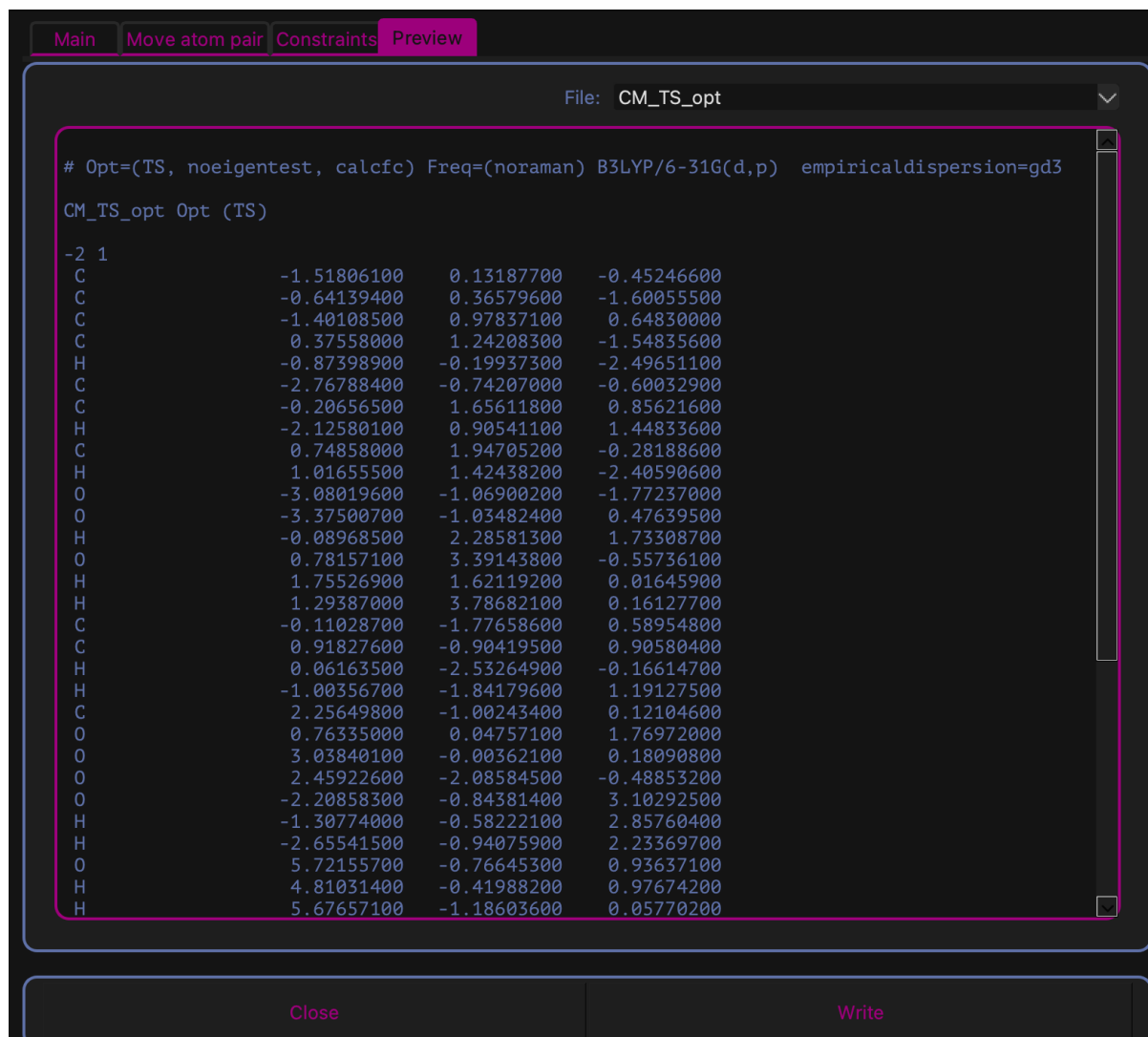
1. Optimizing the transition state

1. Normally you would now use you chemical intuition to generate a guess for the transition state. We have already done this for you, so you can upload "CM_TS_guess.xyz" in REACT ([3] in Figure 2).
2. Now we will create the inputfile for the transition state optimization calculation. Select the transition state guess structure from the REACT workspace ("CM_TS_guess.xyz") and click "Calculation setup" ([4]) from the menu.

   In the 'Calculation setup' window:

   a. Change the charge from 0 to -2 (in the top right square titled 'molecule')
   b. Change filename to "CM_TS_opt".
   c. Make changes in the window such that the following calculations details are set:

      - Functional: B3LYP
      - Basis set: 6-31G (d,p)
      - Additional keywords: empiricaldispersion=gd3
      - Job type: Opt (TS)
      - Freq ON, Raman OFF
      - Additional job details: noeigentest, calcfc
      - SCRF OFF (ignore rest of the options on the same line)
      - Memory OFF

- Chk OFF

3. Move to the "Preview" tab, and inspect the input file before clicking the "write" button. If any of the text differ from Figure 3, go back to the "Main" tab and adjust the setup options accordingly.

```
Main   Move atom pair  Constraints   Preview

                              File:  CM_TS_opt

# Opt=(TS, noeigentest, calcfc) Freq=(noraman) B3LYP/6-31G(d,p)  empiricaldispersion=gd3

CM_TS_opt Opt (TS)

-2 1
 C                -1.51806100    0.13187700   -0.45246600
 C                -0.64139400    0.36579600   -1.60055500
 C                -1.40108500    0.97837100    0.64830000
 C                 0.37558000    1.24208300   -1.54835600
 H                -0.87398900   -0.19937300   -2.49651100
 C                -2.76788400   -0.74207000   -0.60032900
 C                -0.20656500    1.65611800    0.85621600
 H                -2.12580100    0.90541100    1.44833600
 C                 0.74858000    1.94705200   -0.28188600
 H                 1.01655500    1.42438200   -2.40590600
 O                -3.08019600   -1.06900200   -1.77237000
 O                -3.37500700   -1.03482400    0.47639500
 H                -0.08968500    2.28581300    1.73308700
 O                 0.78157100    3.39143800   -0.55736100
 H                 1.75526900    1.62119200    0.01645900
 H                 1.29387000    3.78682100    0.16127700
 C                -0.11028700   -1.77658600    0.58954800
 C                 0.91827600   -0.90419500    0.90580400
 H                 0.06163500   -2.53264900   -0.16614700
 H                -1.00356700   -1.84179600    1.19127500
 C                 2.25649800   -1.00243400    0.12104600
 O                 0.76335000    0.04757100    1.76972000
 O                 3.03840100   -0.00362100    0.18090800
 O                 2.45922600   -2.08584500   -0.48853200
 O                -2.20858300   -0.84381400    3.10292500
 H                -1.30774000   -0.58222100    2.85760400
 H                -2.65541500   -0.94075900    2.23369700
 O                 5.72155700   -0.76645300    0.93637100
 H                 4.81031400   -0.41988200    0.97674200
 H                 5.67657100   -1.18603600    0.05770200

              Close                              Write
```

(opt=noeigentest suppresses testing the curvature in Berny optimizations. The test is on by default only for transition states in internal (Z-matrix) or Cartesian coordinates, for which it is recommended. Occasionally, transition state optimizations converge even if the test is not passed, but **NoEigenTest** is only recommended for those with large computing budgets, which is not a problem in our case.)

4. Back in the terminal, you will find a new file called "CM_TS_opt.com" (unless you chose another filename). Copy this file to Saga by using the command:
   `./rsync_saga.sh to`

5. On Saga, to submit the DFT calculation, we use the script g16run.sh. The script takes 3 arguments: filename <mark>without extention</mark>, number of nodes and number of hours. To use the script with arguments node=1 and hour=1 (appropriate for this calcualtion), type this in the command line:

   `./g16run.sh CM_TS_opt 1 1`

   The job takes approximately 15 minutes to finish. You can view your jobs in the queue with the following command: `squeue -u "you_username"`

2. Analyzing the TS geometry optimzation

1. When the calculation has finished, upload the result ("CM_TS_opt.out") to REACT. If the filename appears in the color green, it has sucessfully converged. If not (red color), you need read the file to see if you find any error messages. The slurm output file can also be useful to read. Sometimes, it is useful to start a new optimization from one of the last frames in the failed optimization. The frames can be inspected by selecting the failed optimization and clicking the SCF button ([5]) from the "Quick launch" menu. This upload all frames from the optimization jobs to PyMOL. In PyMOL, you can save any of the frames as a new .xyz file.

2. It is always important to visually inspect your geometries. Inspect the optimized transition state in PyMOL. Is it similar to the expected transition state (Figure 1)?

3. With the output file selected in the workspace, click "Analyse calculation" ([6]) from the menu. Switch to the "Frequencies" tab and evaluate the frequencies. How many, if any, imaginary (listed as "negative" frequencies) frequencies are present? How

many imaginary frequencies are expected for a transition state?

3. Intrinsic reaction coordinate (IRC) calculation

1. You are now going to do a intrinsic reaction coordinate calculation from the transition state to locate and verify the connected reactant (chorismate) and product (prephenate). With the optimized TS file selected, open "Calculation setup" ([4]). First, change the job type to "IRC". Proceed by adding the following computational details:

   - Functional: B3LYP
   - Basis set: 6-31G (d,p)
   - Additional job details: maxpoints=100, stepsize=5, calcfc
   - SCRF OFF (ignore rest of the options on the same line)
   - Memory OFF
   - Chk OFF
   - OldChk OFF

   Move to "Preview" tab and inspect the two files prepared before clicking "Write".

2. Submit both input files using the g16run.sh script (run the script separately for each file = two times).

   ```
   ./g16run.sh CM_TS_opt_rev 1 2
   ```

```
./g16run.sh CM_TS_opt_frwd 1 2
```

While this is running (for approximately 1 hour), feel free to jump to step 5 before performing step 4. As you allerady have the optimized TS geometry, you can start with this one in step 5.

4. Optimizing RS and PS

Up to this point, we have been working with 1 state only, the TS. We are going to organize our REACT workspace by introducing 2 new states, one for the RS and one for the PS.

1. Add 2 new states by clicking on button [6] two times. State 1 with all the TS files should now be labelled as state 2 instead, and this is easily done by draging state 1 (tab 1) inbetween state 2 (tab 2) and state 3 (tab 3).
2. Upload CM_TS_opt_frwd.out and CM_TS_opt_rev.out separately to one of the empty states (1 and 3). It is not given that the forward (reverse) IRC analysis will yield the PS (RS) state, therefore it does not matter which output file is put into with state for the moment.
3. Select the output file in state 1 and click the SCF button ([5]). In the PyMOL window, click play (right bottom corner) to visualize the IRC calculation. Does it converge to the RS or the PS state? If it converges to the PS state, switch the placement of state 1 (tab 1) and state 3 (tab 3), as we want the order of states in the workplace to follow the order of the reaction (RS->TS->PS). Move to the next state and verify that it converges to the expected state.
4. Next, we are optimizing the RS and PS structure. For both IRC reverse and forward output file, open "Calculation setup", change filename to "CM_RS_opt" ("CM_PS_opt") and set up jobs with the following calucaltion parameters:

   - Functional: B3LYP
   - Basis set: 6-31G (d,p)
   - Additional keywords: empiricaldispersion=gd3
   - Job type: Opt
   - Freq ON, Raman OFF
   - Additional job details: calcfc
   - SCRF OFF (ignore rest of the options on the same line)
   - Memory OFF
   - Chk OFF

5. Run calculations with g16run.sh. 1 node and 1 hour should be enough.

5. Calculating energies and correction for RS, TS and PS; big basis and solvation

Do the follwing steps for all three of your <u>converged</u> structures:

1. The goal here is to obtain improved electronic energies by performing a single point calculation using a bigger basis set on the already optimized structure. With the "CM_*S_opt.out" selected, open "Calculation setup". Change the filename to "CM_*S_opt_bbs" Change the job type to "Single point" and add the following computational details:

   - Functional: B3LYP
   - Basis set: 6-311G+(2d,2p)
   - Additional keywords: empiricaldispersion=gd3
   - SCRF OFF (ignore rest of the options on the same line)
   - Memory OFF
   - Chk OFF

   Move to the "Preview" tab and save the file by clicking "Write".

1. Move back to the "Main tab" (If you have already closed the "Calculation setup" window, open it again). Change the name to "CM_*S_opt_solv". Change the job type to "Single point" and add the following computational details:
   - Functional: B3LYP
   - Basis set: 6-31G(d,p)
   - Additional keywords: empiricaldispersion=gd3
   - SCRF ON, method=SMD, solvent: water
   - Memory ON = 16GB
   - Chk ON
2. Submit all 6 jobs with the "g16run.sh" script. 1 node and 1 hour should be enough.

6. Putting it all together

You should now have a *opt.out (small basis set in gas phase + frequency calculation), *opt_bbs.out and a *opt_solv.out for the RS (chorismate), TS and PS (prephenate). Upload any missing output files to REACT. Confirm that they all are colored green, indicating successful convergence. If any are red, open and investigate the output file.

In all three states, select the "CM_*S_opt.out" file and open "Analyse calculation" from the

REACT menu. In the upper left section, add missing files. For example for solvent: 1) click the button titled "solvent", 2) back in the REACT workspace, select the "CM_*S_opt_solv.out" file, and 3) back in the analyse window, click the round "+" button. Use the two arrows to switch between states. With all four files added for all three states.

The total activation and reaction Gibbs free energies are in the rightmost column. Without closing the "Analyse window", open the "Plotter" tool from REACT's main window. Make a DFT energy diagram.

For the rapport?

- Figures of optimized structures with key distances of the 3 stationary points (RS, TS, PS). In REACT, outputfiles can be converted to pdb or xyz files using the 'Pen tool' in workspace, located under the plus and minus buttons. Feel free to use any program to prepare your figures. If you wish to save figures from the REACT's pymol window, you can combine these two commands directly in the pymol window (substituting '/my/path/' and 'filename.png' with actual path on your computer and prefered filename):

Bg_color white

png /my/path/filename.png

- Construct a DFT energy diagram

References

1. 1 Frisch, M. J. *et al.* (Wallingford CT, 2009).
2. 2 Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys* **98**, 5648-5652, doi:10.1063/1.464913 (1993).
3. 3 Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J Chem Phys* **132**, 154104, doi:10.1063/1.3382344 (2010).
4. 4 Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* **32**, 1456-1465, doi:10.1002/jcc.21759 (2011).
5. 5 Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J Phys Chem B* **113**, 6378-6396, doi:10.1021/jp810292n (2009).
6. 6 Siegbahn, P. E. M. & Himo, F. The quantum chemical cluster approach for modeling enzyme reactions. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1**, 323-336, doi:10.1002/wcms.13 (2011).
7. 7 Himo, F. Recent Trends in Quantum Chemical Modeling of Enzymatic Reactions. *J Am Chem Soc* **139**, 6780-6786, doi:10.1021/jacs.7b02671 (2017).