Before tutorial

Everything, which is written on the **green background**, is a command, that should be executed in the terminal window, string-by-string, each following with the ENTER button.

Please note that **<...>** in the commands means, that everything, including **<>** symbols, must be replaced with your own specific information. Be careful!

Some exercises will require usage of **less** Linux tool for looking up into the content of files. In case you are not familiar with it, here is a short list of hotkeys, which could be used inside LESS editor:

- q – exit
- / – search for a pattern which will take you to the next occurrence.
- n – for next match in forward
- N (SHIFT+n) – for previous match in backward
- g – go to the start of file
- G (SHIFT+g) – go to the end of file

All exercises will require you to submit job for computing using **sbatch run.sh** command. To check status of your job following commands would be useful.

- **squeue -u <your username>** - **c**hecks status of all your jobs. Output will look like that:

```
             JOBID PARTITION     NAME     USER ST       TIME  NODES NODELIST(REA
SON)
            279859      test gromacs- dmorozov  R       0:00      1 cl101
```

Job ID — Status — Run time

- **scancel <JOBID>** will remove the job, if you occasionally submitted it.

Setting up tutorial environment

Let's start the tutorial with the following steps

1) Execute commands in the terminal:

cd /scratch/project_2003487/training<your number>

module load git

module load python-env

module load cp2k/8.1-gmx

git clone https://github.com/bioexcel/gromacs_cp2k_tutorial.git tutorial

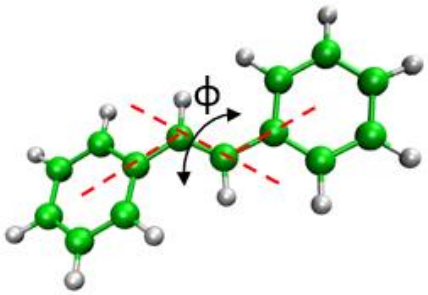
2) Go to the copied tutorial directory

cd tutorial

Exercise 1: Umbrella sampling simulation with QM/MM

1) Go to stilbene_vacuum directory:

```
cd stilbene_vacuum
```



2) Look up in the table and pick-up starting structure and dihedral angle value (anyone you want)

Structure	Dihedral angle, ϕ
1	-180
2	-173
3	-166
4	-159
5	-152
6	-145
7	-138
8	-131
9	-124
10	-117
11	-110
12	-103
13	-96
14	-89
15	-82
16	-75
17	-68
18	-61
19	-54
20	-47
21	-40
22	-33
23	-26
24	-19
25	-12
26	-5
27	2
28	9
29	16
30	23

3) Copy chosen starting structure:

```
cp eq_gro/md-equilb<your number>.gro ./conf.gro
```

4) Modify Gromacs input file **qmmm_md_umbrella.mdp** with value of your chosen Dihedral angle:

```
sed -i "s/@umbr@/<your dihedral angle>/" qmmm_md_umbrella.mdp
```

You can also modify **pull-coord1-init** option in the **qmmm_md_umbrella.mdp** file with **vim** or any other editor.

5) Add group of atoms which will be treated with QM to the index file (in that case all atoms are QM):

```
gmx_cp2k make_ndx -f conf.gro -n index.ndx
```

```
> 0
```

```
> name 7 QMatoms
```

```
> q
```

6) Generate Gromacs-CP2K simulation file:

```
gmx_cp2k grompp -f qmmm_md_umbrella.mdp -p topol.top -c conf.gro -n index.ndx -o stilbene.tpr
```

files **stilbene.tpr**, **stilbene.inp** and **stilbene.pdb** should appear in the directory

7) Run QMMM simulation:

```
sbatch run.sh
```

8) While job is running you can check the content of **stilbene.inp**

```
less stilbene.inp
```

and of **qmmm_md_umbrella.mdp**

```
less qmmm_md_umbrella.mdp
```

9) At the end of the job inspect **pullx.xvg** file appeared in the directory:

```
less pullx.xvg
```

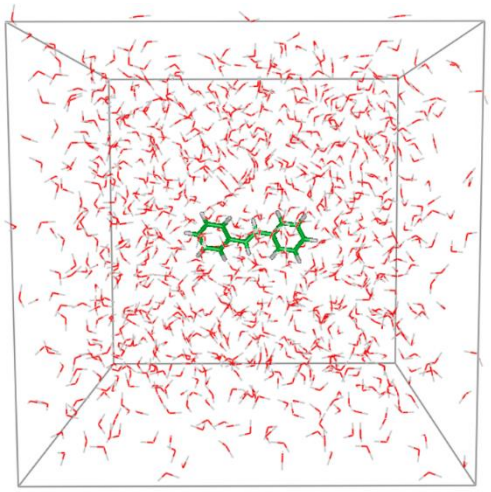
It contains information about chosen coordinate dynamics over the simulation trajectory. By performing that sampling over the many points along reaction coordinate and gathering all ***.tpr** and **pullx.xvg** files you could produce free-energy profile of the reaction with **gmx wham** tool.

10) Check the free energy profiles generated from 100 steps (100fs) and 10000 steps (10ps) of QMMM MD simulation for each frame from the **eq_gro** directory: **profile-100fs.xvg** and **profile-10ps.xvg.**

You can open them with Grace or copy data into any other software (i.e. Excel).

Notice, how important is long sampling! For biological systems at least 30 ps of QMMM simulation per frame is recommended.

Exercise 2: Energy minimization with QM/MM



1) Go to stilbene_water directory:

```
cd ../stilbene_water
```

2) Generate QM/MM index file:

```
gmx_cp2k make_ndx -f stilbene-sol.pdb
> 2
> name 6 QMatoms
> q
```

file **index.ndx** should appear in the directory

3) Generate Gromacs-CP2K simulation file:

```
gmx_cp2k grompp -f em-qmmm.mdp -p topol.top -c stilbene-sol.pdb -n index.ndx -o stilbene-sol-opt.tpr
```

files **stilbene-sol-opt.tpr**, **stilbene-sol-opt.inp** and **stilbene-sol-opt.pdb** should appear in the directory

4) Run QMMM minimization

```
sbatch run.sh
```

5) While job is running you can check the content of **stilbene-sol-opt.inp**

```
less stilbene-sol-opt.inp
```

and of **em-qmmm.mdp**

```
less em-qmmm.mdp
```

6) At the end of the job you can print out energy at each step with the following command:

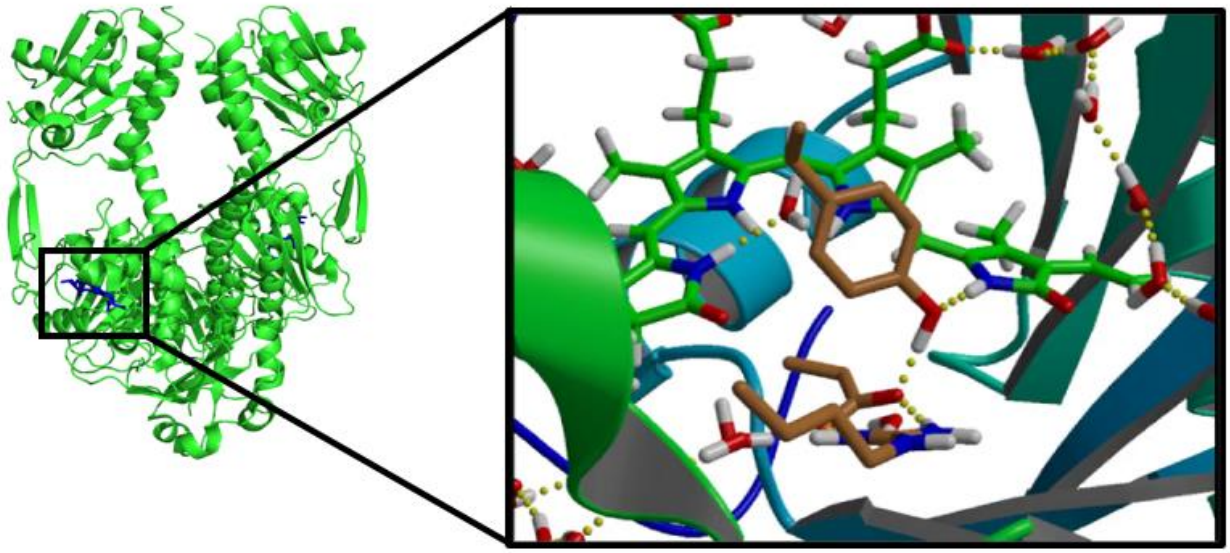
```
gmx_cp2k energy
6
<!Press ENTER button second time!>
```

energy.xvg file should appear in the directory.

It contains data with Potential energy (kJ/mol) against optimization step. You can open it in Grace or copy data into any other software (i.e. Excel).

7) You can also check geometry changes by downloading **stilbene-sol.pdb** (coordinates) and **traj.trr** (trajectory) and open them with PyMOL or VMD

Exercise 3: Large protein system setup with QM/MM



1) Go to phytochrome directory:

```
cd ../phytochrome
```

2) Generate Gromacs-CP2K simulation file:

```
gmx_cp2k grompp -f md-qmmm.mdp -p topol-sol.top -c lumi-R-conf.gro -n index.ndx -o phytochrome.tpr
```

files **phytochrome.tpr**, **phytochrome.inp** and **phytochrome.pdb** should appear in the directory

3) Run QMMM MD simulation:

```
sbatch run.sh
```

4) While job is running you can check the content of **phytochrome.inp**

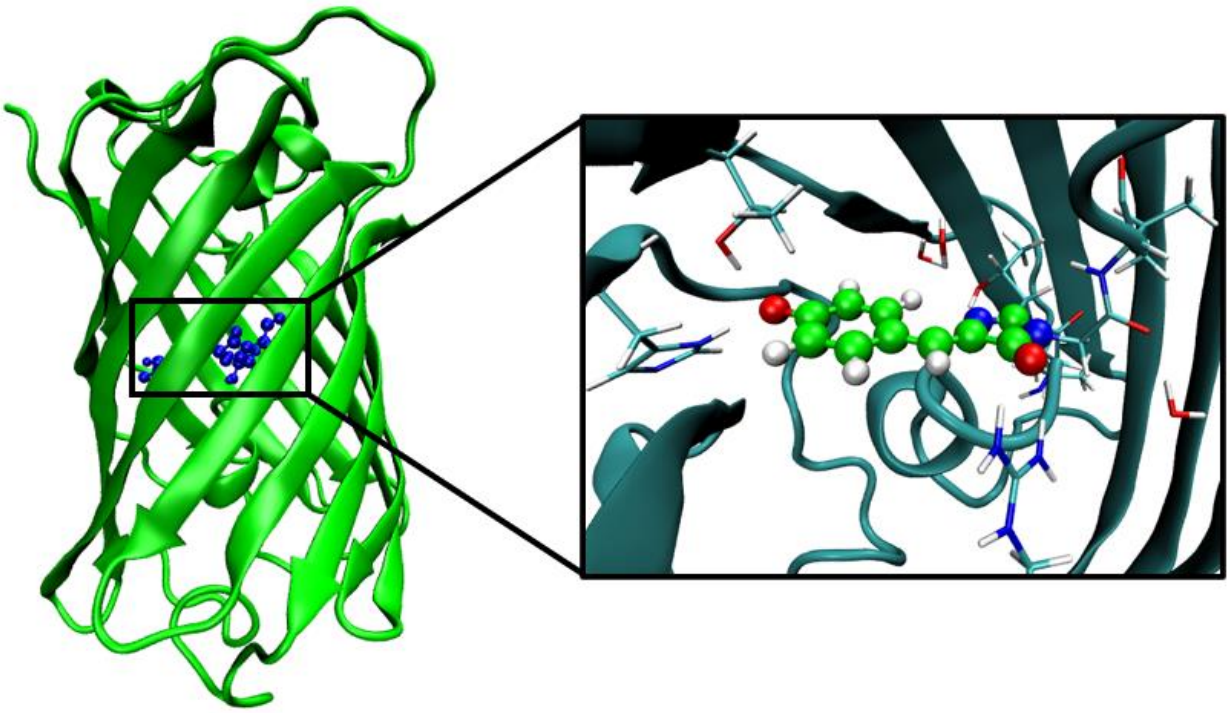
```
less phytochrome.inp
```

and of **md-qmmm.mdp**

```
less md-qmmm.mdp
```

5) Notice how CP2K input file has changed. Four **&LINK** sections appeared at the end of **&QMMM** part. They are notifying CP2K that broken chemical bond should be treated between MM and QM atoms. From the side of QM part that that result in adding “fake” Hydrogen Link-atom between QM and MM atoms. This is automatically done inside the CP2K.

Exercise 4: Using user-defined QM input file



1) Go to egfp_spectra directory:

```
cd ../egfp_spectra
```

2) Generate Gromacs-CP2K simulation file:

```
gmx_cp2k grompp -f md-qmmm.mdp -p topol.top -c conf.gro -n index.ndx -o egfp-spec.tpr
```

file **egfp-spec.tpr** should appear in the directory, while **egfp-spec.inp** and **egfp-spec.pdb** was already there before tpr file generation.

3) Run QMMM MD simulation:

```
sbatch run.sh
```

4) While job is running you can check the content of **egfp-spec.inp**

```
less egfp-spec.inp
```

and of **md-qmmm.mdp**

```
less md-qmmm.mdp
```

5) Notice that in the **md-qmmm.mdp** file:

qmmm-qmmethod = INPUT

qmmm-qminputfile = egfp-spec.inp

And CP2K input file **egfp-spec.inp** has and additional section **&PROPERTIES.** It requests CP2K to perform additional TDDFT calculation for the lowest 5 excitations at each MD step.

6) After job is finished, we need to gather information about excitation energies over the calculated trajectory:

```
grep " TDDFPT|" egfp-spec.out | awk '{ print $3 " " $7 }' > excitations
```

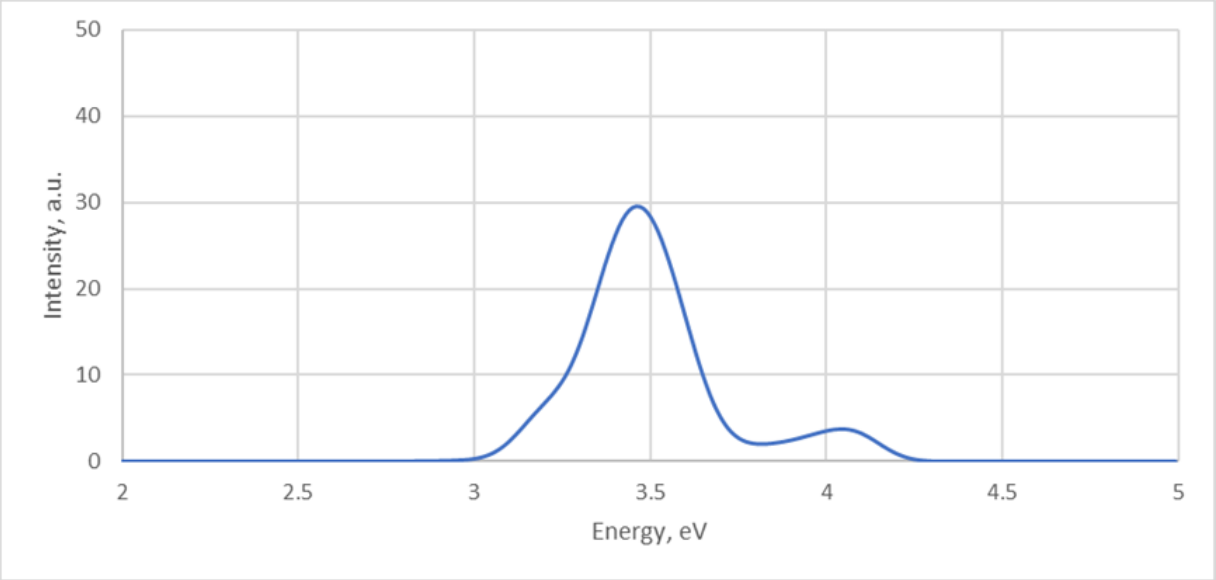
The **excitations** file should appear in the directory, it will consist out of two columns. First column is an excitation energy (in eV) and second is an oscillator strength (in a.u.) for each excitation computed by CP2K. Final absorption spectra could be convolved by representing each excitation with gaussian function and sum up over all of them.

7) Convolve the spectra using:

```
./conv.py excitations 0.1 2 5
```

File **spec.xvg** should appear in the directory. You can open it in Grace or copy data into any other software (i.e. Excel).

As an example, convolved spectra with 0.1 eV half-width gaussians over 100fs (100 steps) trajectory:



8) Spectra collected over 3 ps (3000 MD steps) will look like that: