GROMACS-CP2K Interface Tutorial

(Introduction to QM/MM simulations)

Dmitry Morozov

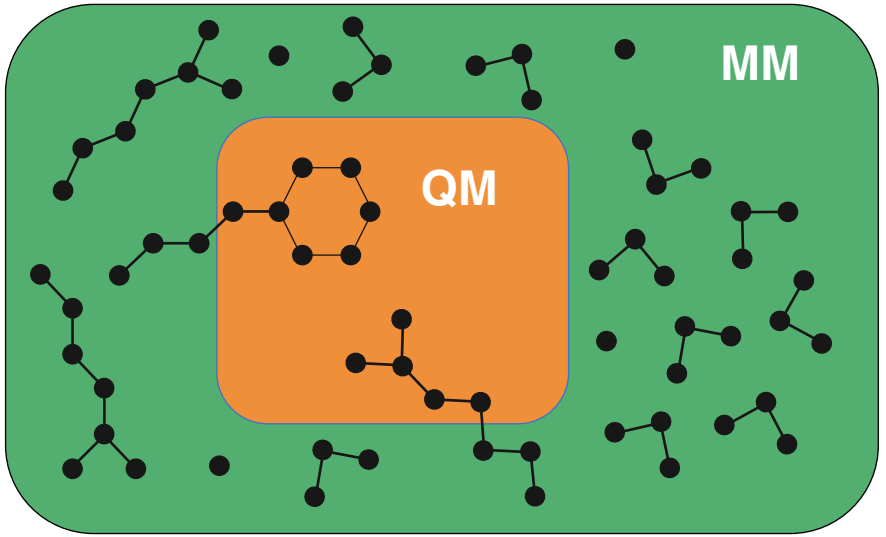
University of Jyväskylä, Finland

dmitry.morozov@jyu.fi

Horizon 2020
European Union Funding
for Research & Innovation

Practical: GROMACS + CP2K Part I

1. Lecture recap
2. Gromacs-CP2K interface for QM/MM
3. Setting up a QM/MM calculation
4. CP2K input and output



Lecture Recap: Forcefield (MM) - GROMACS

- Force field description of MM region

$$H = H_{MM} + H_{QM} + H_{QM/MM}$$

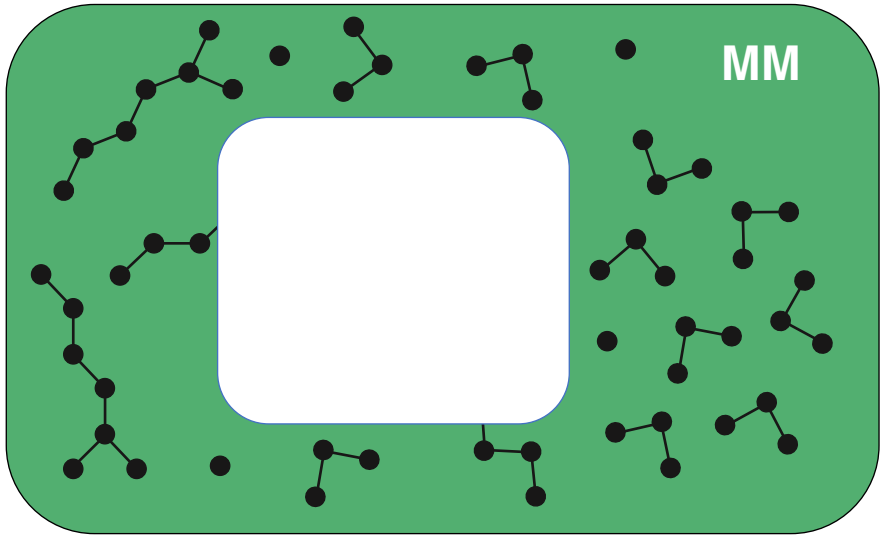
Forcefield Quickstep GEEP

$$V(r_1, r_2, \ldots r_N) = V_{bonded}(r_1, r_2, \ldots r_N) + V_{non-bonded}(r_1, r_2, \ldots r_N)$$

$$V_{bonded} = \sum_{bonds} \frac{1}{2} k_b (r - r_0)^2 + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 + \sum_{torsions} k_\xi (\xi - \xi_0)^2$$

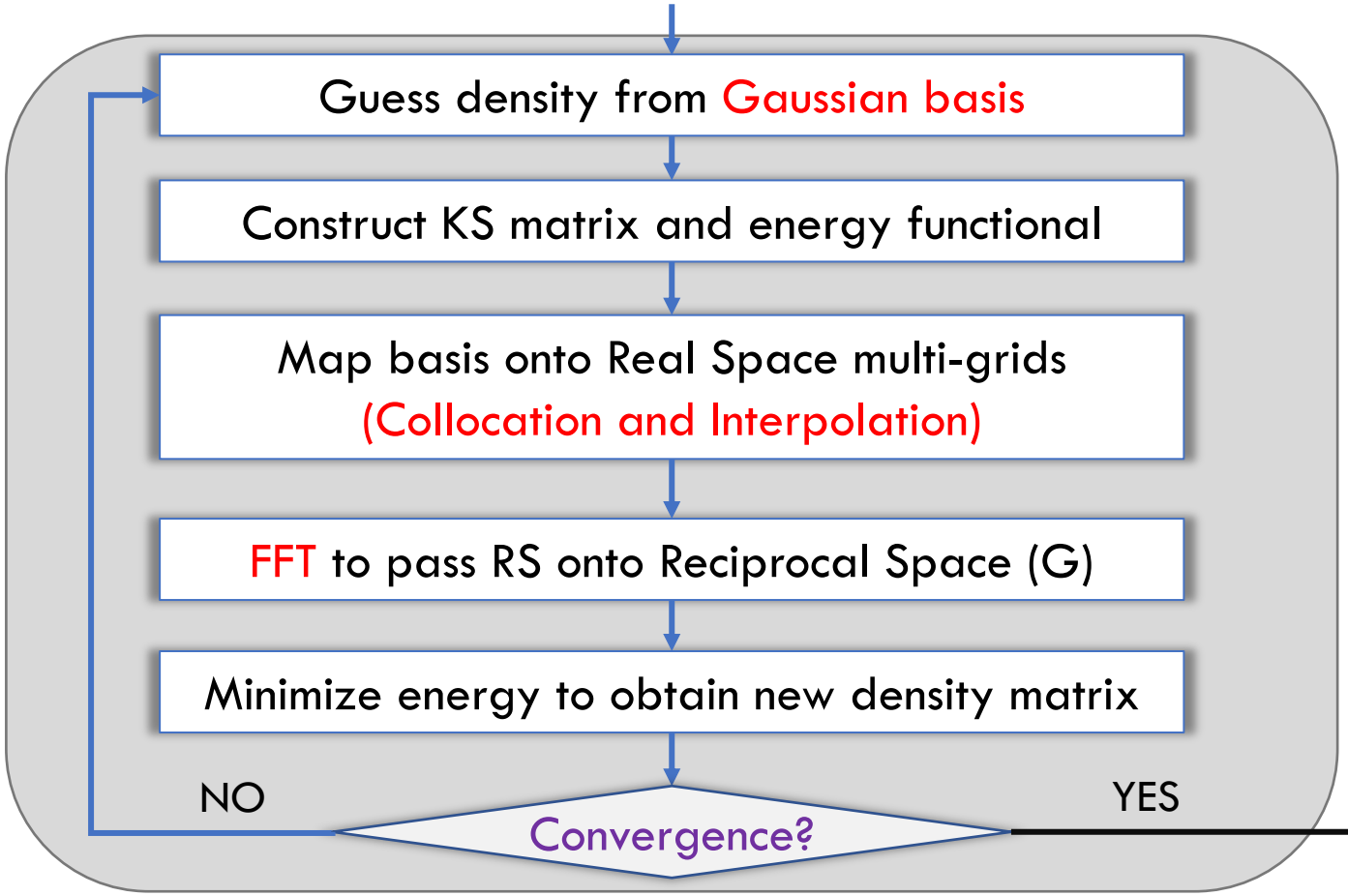
$$+ \sum_{torsions} \frac{1}{2} k_\phi [1 + \cos(n\phi - \phi_0)]$$

$$V_{non-bonded} = \sum_{LJ} 4\epsilon_{ij} \left(\frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^{6}} \right) + \sum_{Coul.} \frac{q_i q_j}{r_{ij}}$$



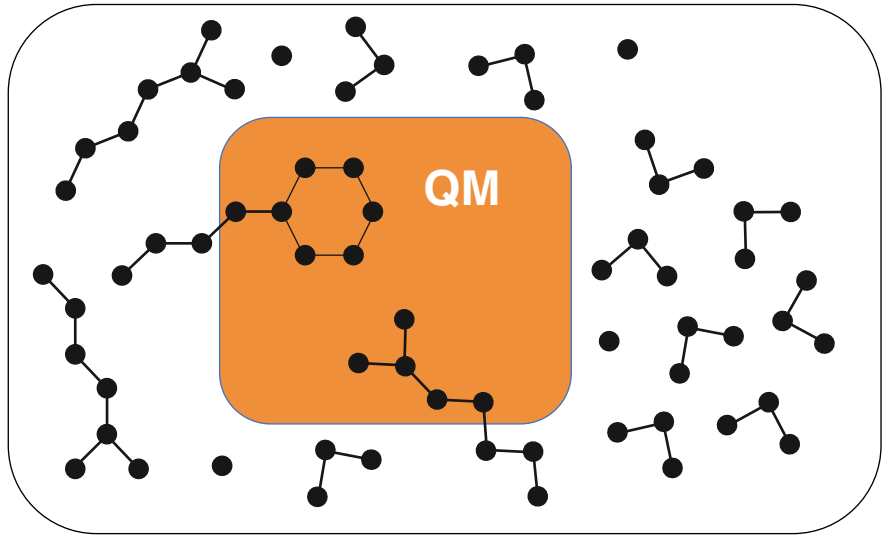
Lecture Recap: Quickstep (QM) - CP2K

QM region as CP2K input



$$H = H_{MM} + H_{QM} + H_{QM/MM}$$

Forcefield Quickstep GEEP



Energy, Forces and other properties

Practical: GROMACS + CP2K Part I

1. Lecture recap
2. Gromacs-CP2K interface for QM/MM
3. Setting up a QM/MM calculation
4. CP2K input and output

GROMACS-CP2K Interface

Quickstep: Mixed Gaussian and Plane wave basis implementation of Density Functional Theory

+

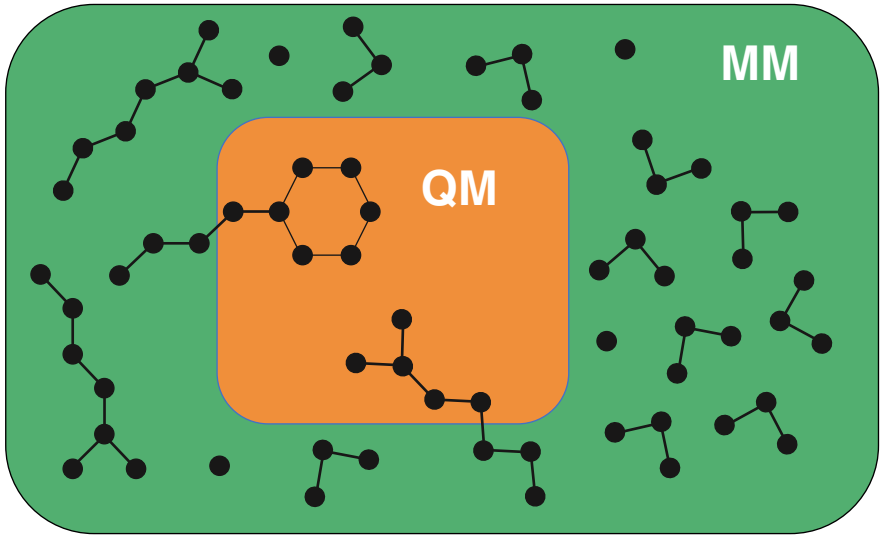
GEEP: **Gaussian Expansion of Electrostatic Potential (GEEP) to compute the QM/MM coupling**

Forcefield: Classical MM-MM interactions both bonded and non-bonded (PME)

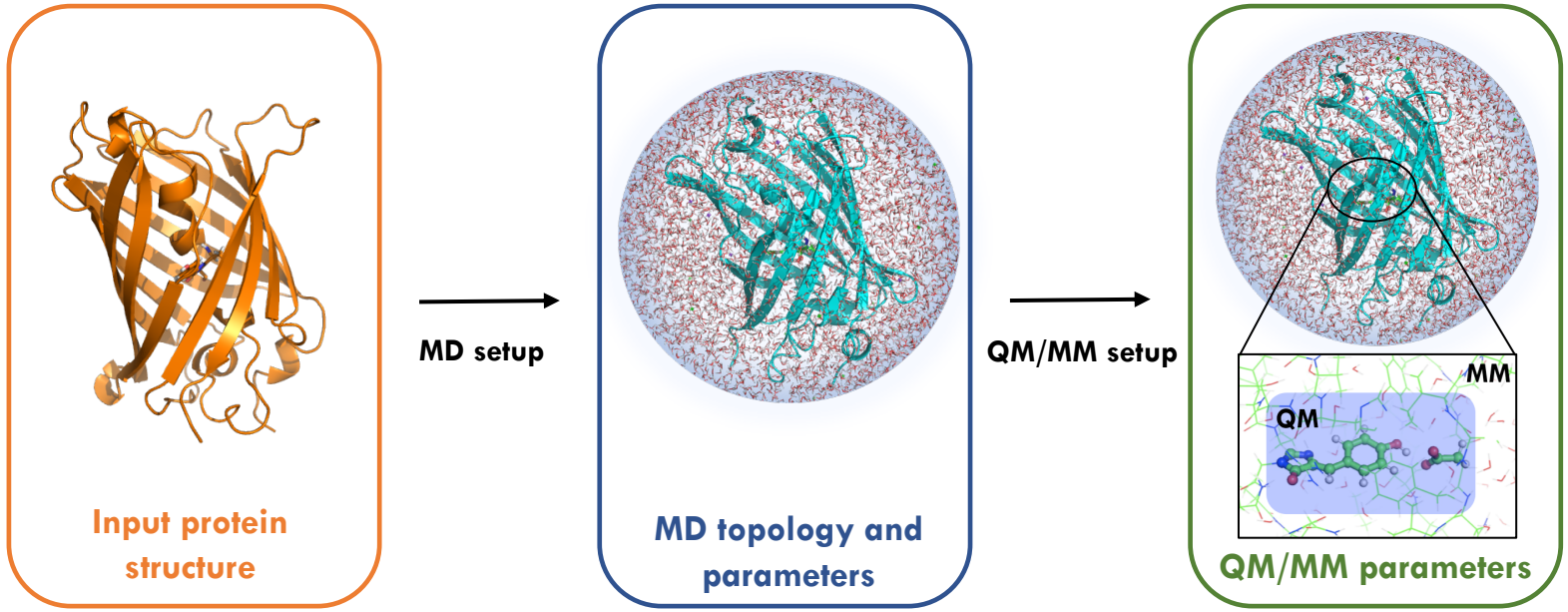
+

Integration: Classical MD using fully periodic QM/MM forces

$$H = \underbrace{H_{MM}}_{\text{Forcefield}} + \underbrace{H_{QM}}_{\text{Quickstep}} + \underbrace{H_{QM/MM}}_{\text{GEEP}}$$



Features of the Interface



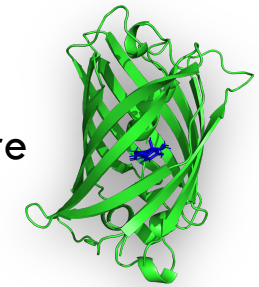
- Automatized topology conversion from classical MD to QM/MM: charges and bonds modifications, as well as link-atoms setup on the frontier
- Validated CP2K QM parameters setup for the biological systems
- Compatibility with the most simulation techniques available in Gromacs
- Compatibility with Gromacs tools and third-party software for analysis
- Supports highly parallelizable simulation methods, like umbrella sampling

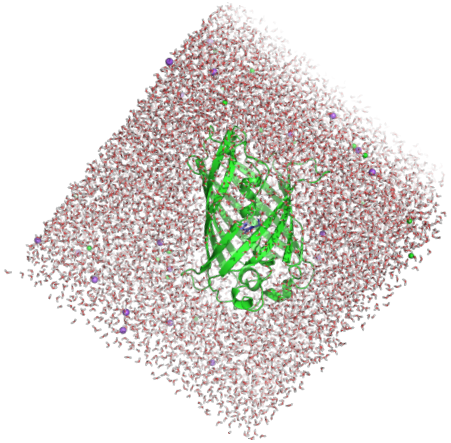
Practical: GROMACS + CP2K Part I

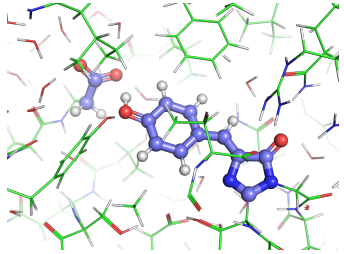
1. Lecture recap
2. Gromacs-CP2K interface for QM/MM
3. Setting up a QM/MM calculation
4. CP2K input and output

Setup a QM/MM calculation

Input structure







SYSTEM SETUP

- Missing residues
- Residue protonation
- pdb2gmx
- Define box
- Add solvent and ions
- ….

MD SIMULATIONS

- Prepare .mdp files
- Energy minimization
- Pressure & temperature equilibration
- Production run
- …

- Index file
- .mdp parameters for QM/MM
-External CP2K input (optional)

QM/MM Calculations

- Geometry optimization
- Atomic charges
- QM/MM MD
- Absorption spectra
- Umbrella Sampling
- ….

GROMACS-CP2K Tutorial files

Open "**Practical: GROMACS + CP2K Part I**" episode.

Open terminal window and finish "Setting up tutorial environment part"

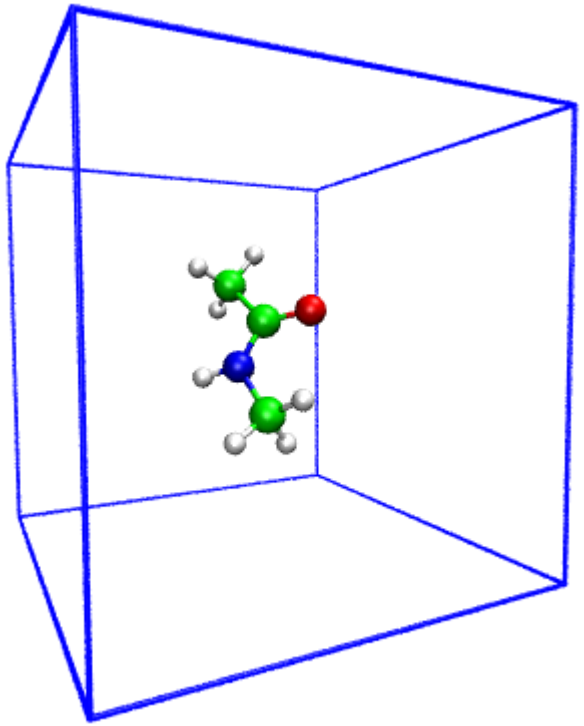
```
>> module load gromacs-cp2k

>> cd /work/ta025/ta025/<your login name>

>> git clone https://github.com/bioexcel/2021-04-22-gromacs-cp2k-tutorial.git tutorial

>> cd tutorial
```

Exercise 1: Setting up simple QM system



Objective: Make simple QM system with interface

QM subsystem : NMA molecule (12 atoms)

MM subsystem : No

QM charge: 0

QM multiplicity: 1

Functional: PBE

Do the steps (1)-(5) from the “Exercise 1”

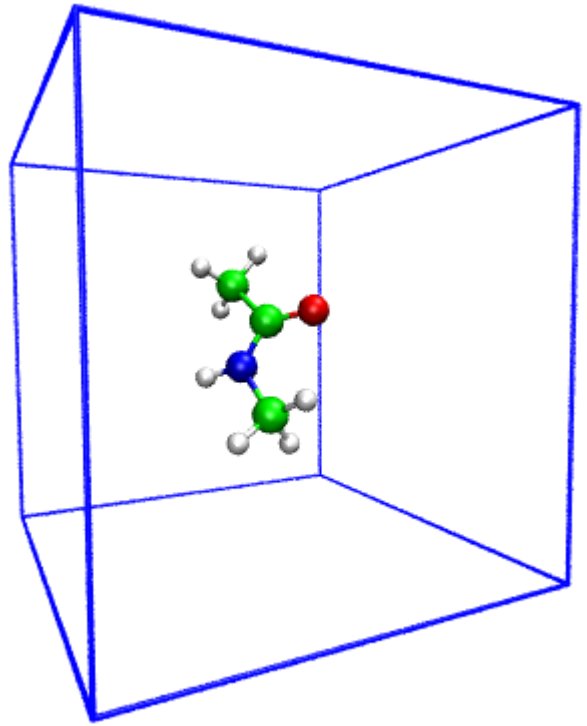
MDP Parameters for energy minimization

```
integrator    = steep ; Algorithm (steep = steepest descent minimization)
emtol         = 10.0  ; Stop minimization when the maximum force < 10.0 kJ/mol/nm
emstep        = 0.01  ; Energy step size
nsteps        = 100   ; Maximum number of (minimization) steps to perform

; Set output frequency to each step
nstxout               = 1 ; Coordinates to trr
nstlog                = 1 ; Energies to md.log
nstcalcenergy         = 1 ; Energies
nstenergy             = 1 ; Energies to ener.edr

; Set cut-offs
rlist           = 0.2 ; NB-search cut-off
rcoulomb        = 0.2 ; Short-range electrostatic cut-off
rvdw            = 0.2 ; Short-range Van der Waals cut-off

; CP2K QMMM parameters
qmmm-active             = true   ; Activate QMMM MdModule
qmmm-qmgroup            = System ; Index group of QM atoms
qmmm-qmmethod           = PBE    ; Method to use
qmmm-qmcharge           = 0      ; Charge of QM system
qmmm-qmmultiplicity     = 1      ; Multiplicity of QM system
```



CP2K: Basic Input Parameters

```
>> less nma-em.inp
```

Input Sections: GLOBAL

```
&GLOBAL
   PRINT_LEVEL LOW                 !HIGH/MEDIUM/LOW
   PROJECT GROMACS                 ! <projectname>
   RUN_TYPE ENERGY_FORCE           ! GEO_OPT/ENERGY_FORCE/BAND
&END GLOBAL
```

Practical: GROMACS + CP2K Part I

1. Lecture recap
2. Gromacs-CP2K interface for QM/MM
3. Setting up a QM/MM calculation
4. CP2K input and output

CP2K: Basic Input Parameters

Input Section: FORCE_EVAL

```
&FORCE_EVAL                    ! parameters for force evaluation
  METHOD QMMM                  ! method employed e.g. QMMM (Quickstep + external charges)
  &DFT                         ! DFT section - all QM
    .... contents of DFT section
  &END DFT
  &QMMM                        ! QMMM section - set up for QM box
    .... contents of QMMM section
  &END QMMM
  &MM                          ! MM section - MM point chrages, etc.
    .... contents of MM section
  &END MM
  &SUBSYS                      ! subsystem - coordinates, atom kinds etc.
    .... contents of SUBSYS section
  &SUBSYS
&END FORCE_EVAL
```

CP2K: Basic Input Parameters

Input Section: DFT

```
&FORCE_EVAL
  METHOD QMMM
  &DFT
    CHARGE 0
    MULTIPLICITY 1
    BASIS_SET_FILE_NAME  BASIS_MOLOPT        ! File with basis sets
    POTENTIAL_FILE_NAME  POTENTIAL           ! File with peudopotentials
     &MGRID
       NGRIDS 5                              ! Number of Grids
       CUTOFF 450                            ! Plane wave cutoff (Rydberg) for finest grid.
       REL_CUTOFF 50                         ! Cutoff to map product Gaussians onto the grids
       COMMENSURATE                          ! Align all the grids
     &END MGRID
     &SCF
      SCF_GUESS RESTART                      ! CP2K will search for existing *.wfn file
      EPS_SCF 5.0E-8                         ! Accuracy of SCF convergence
      …
     &END SCF
```

CP2K: Basic Input Parameters

Input Section: DFT

```
&DFT
  …
 &XC
    DENSITY_CUTOFF     1.0E-12        ! DFT Precision parameters
    GRADIENT_CUTOFF    1.0E-12
    TAU_CUTOFF         1.0E-12
    &XC_FUNCTIONAL PBE                ! Choice of DFT functional
    &END XC_FUNCTIONAL
  &END XC
  &QS
   METHOD GPW                         ! Mixed Gaussian/Plane-wave method
   EPS_DEFAULT 1.0E-10                ! Accuracy of SCF energies
   EXTRAPOLATION ASPC                 ! Extrapolation of wavefunction from previous calculation
   EXTRAPOLATION_ORDER  4
  &END QS
&END DFT
```

CP2K: Basic Input Parameters

Input Section: SUBSYS

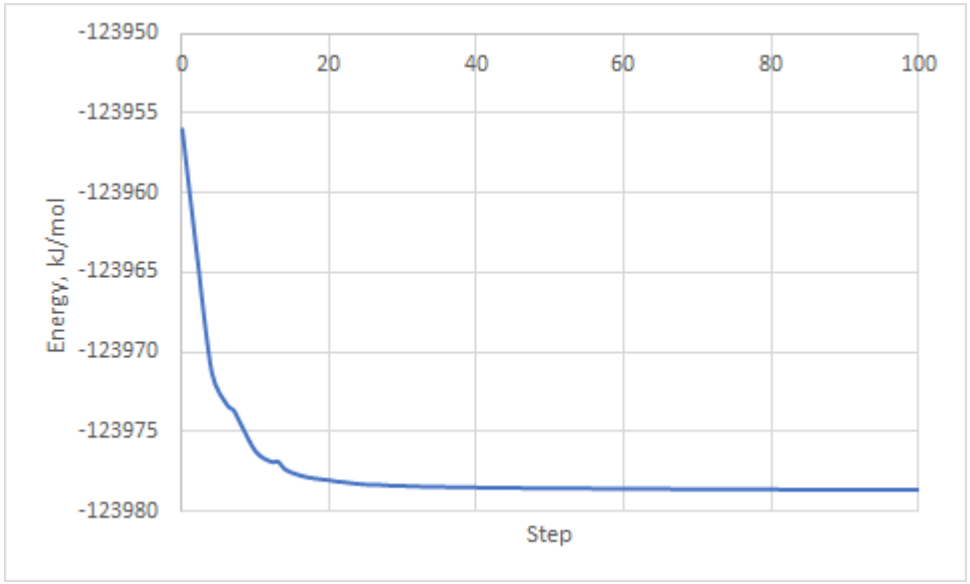
```
&FORCE_EVAL
  …
 &SUBSYS                    ! specifies information of the system: coordinates, topology, molecules & full cell
   &CELL                    ! Full system box size (will be the same as in Gromacs)
    A 10.000 0.000 0.000    ! Defined with three vectors A, B, C (in Angstroms)
    B 0.000 10.000 0.000
    C 0.000 0.000 10.000
    PERIODIC XYZ            ! Fully periodic cell
  &END CELL
   …
```

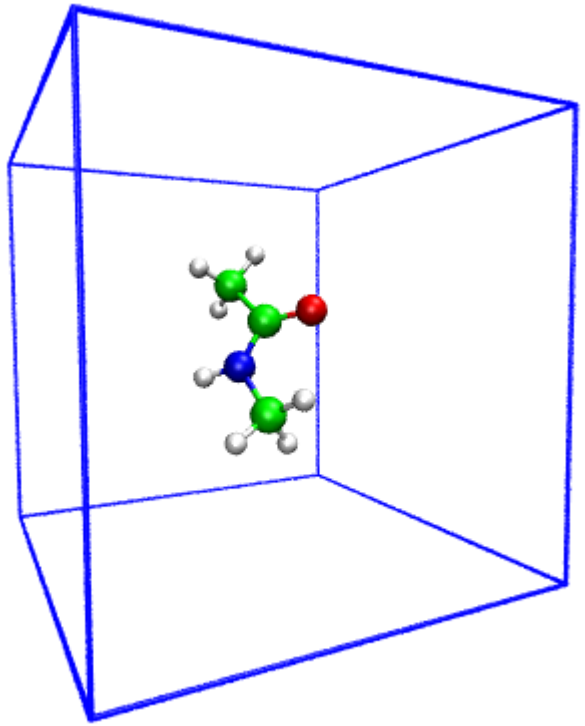
CP2K: Basic Input Parameters

Input Section: KIND

```
&FORCE_EVAL
  &SUBSYS
   …
    &KIND H                          ! Each kind of QM atoms should have basis and PP assigned
      ELEMENT H
      BASIS_SET DZVP-MOLOPT-GTH      ! Gaussian Basis set to be used for Hydrogens
      POTENTIAL GTH-PBE              ! Make sure Basis and PP match
    &END KIND
    &KIND C                           ! Each kind of QM atoms should have basis and PP assigned
      …
    &END KIND
    …
  &END SUBSYS
&END FORCE_EVAL
```

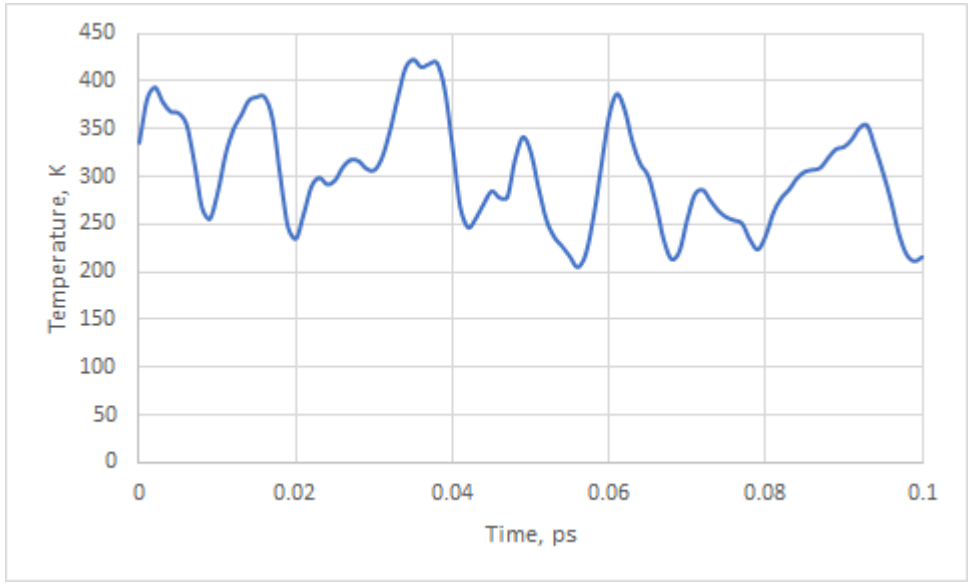
Result of the energy minimization

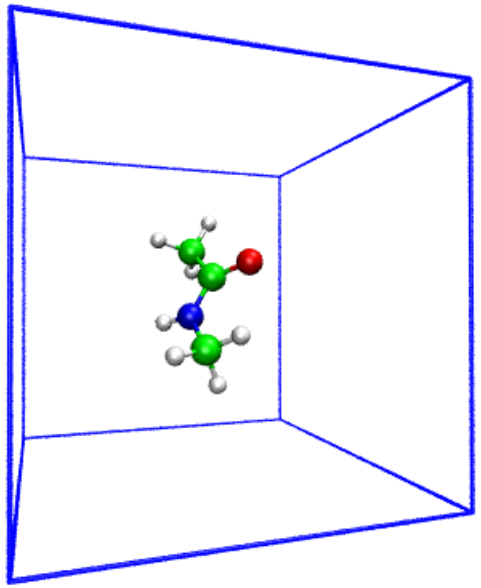




No do the steps (8)-(11) from the “Exercise 1”

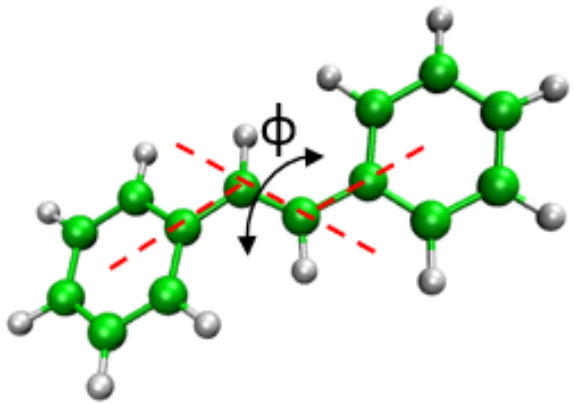
Result of the molecular dynamics with QM forces





Congratulations, you have done first QM simulation with GROMACS-CP2K Interface!

Exercise 2: Stilbene isomerization



Objective: Make isomerization Free-energy profile

QM subsystem : Stilbene (26 atoms)

MM subsystem : No

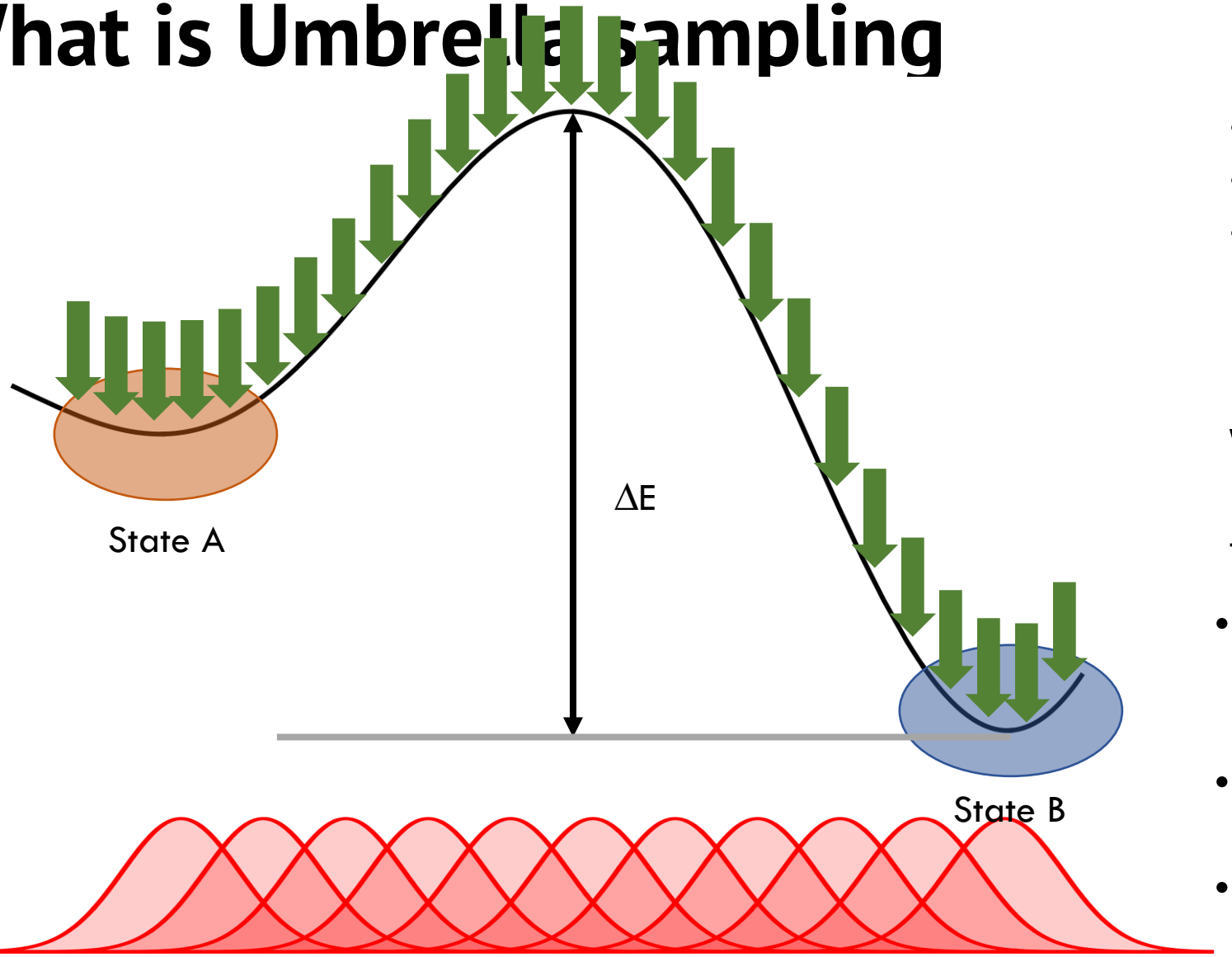
QM charge: 0

QM multiplicity: 1

Functional: PBE

Do the steps (1)-(7) from the “Exercise 2”

What is Umbrella sampling



- System is stable in state A
- System is stable in state B
- The transitions between states are possible

-

We want to know what is the barrier ΔE and states relative free-energies ΔG

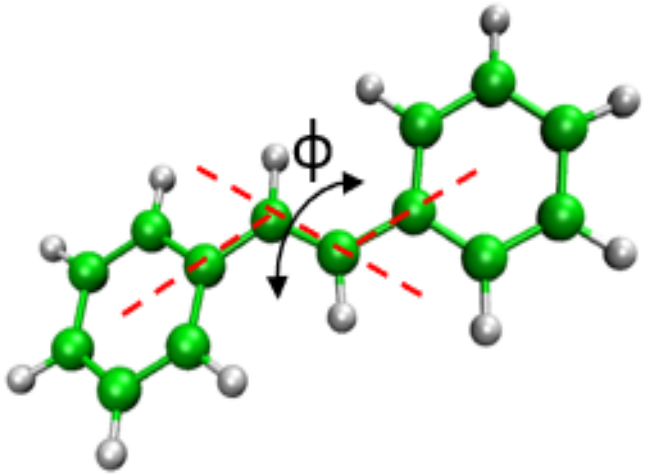
- Energy profile integrated from the coordinate distribution in each window
- Sufficient overlap between windows needed
- Gromacs has tool **gmx wham** to perform integration

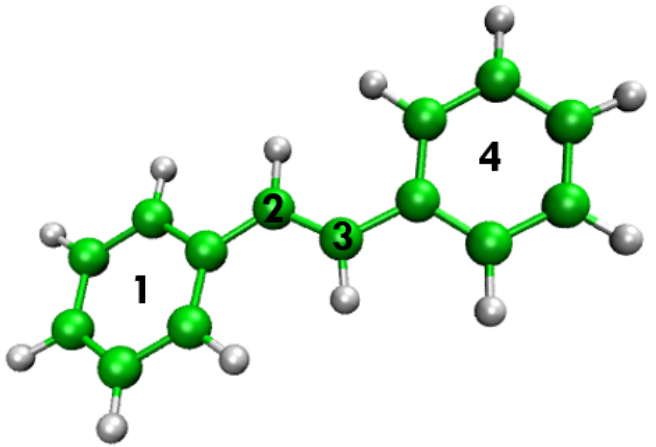
For further information please follow Umbrella sampling tutorial:
http://www.mdtutorials.com/gmx/umbrella/index.html

MDP Parameters: umbrella sampling

>> less **qmmm_md_umbrella.mdp**

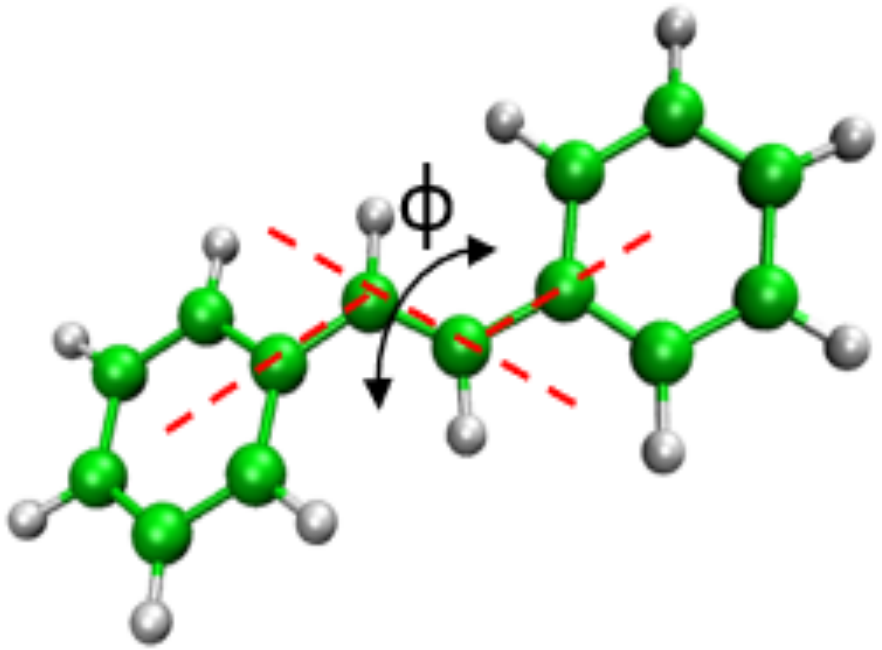
```
pull                    = yes
pull_ncoords            = 1         ; only one reaction coordinate
pull_ngroups            = 4         ; four groups defining one reaction coordinate
pull_group1_name        = group1 ; groups are defined in index file
pull_group2_name        = group2
pull_group3_name        = group3
pull_group4_name        = group4
pull_coord1_type        = umbrella
pull_coord1_geometry    = dihedral
pull_coord1_dim         = Y Y Y
pull_coord1_groups      = 1 2 2 3 3 4
pull-coord1-init        = -180          ; this is your angle value
pull_coord1_rate        = 0.00          ; restrain in place
pull_coord1_k           = 418.4         ; kJ mol^-1 nm^-2
pull-nstxout            = 1             ; output pulling coordinate each step
pull-nstfout            = 1             ; output pulling force each step
```





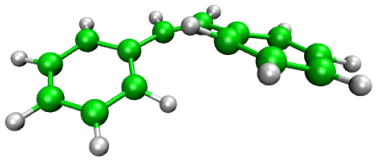
MDP Parameters: QM/MM

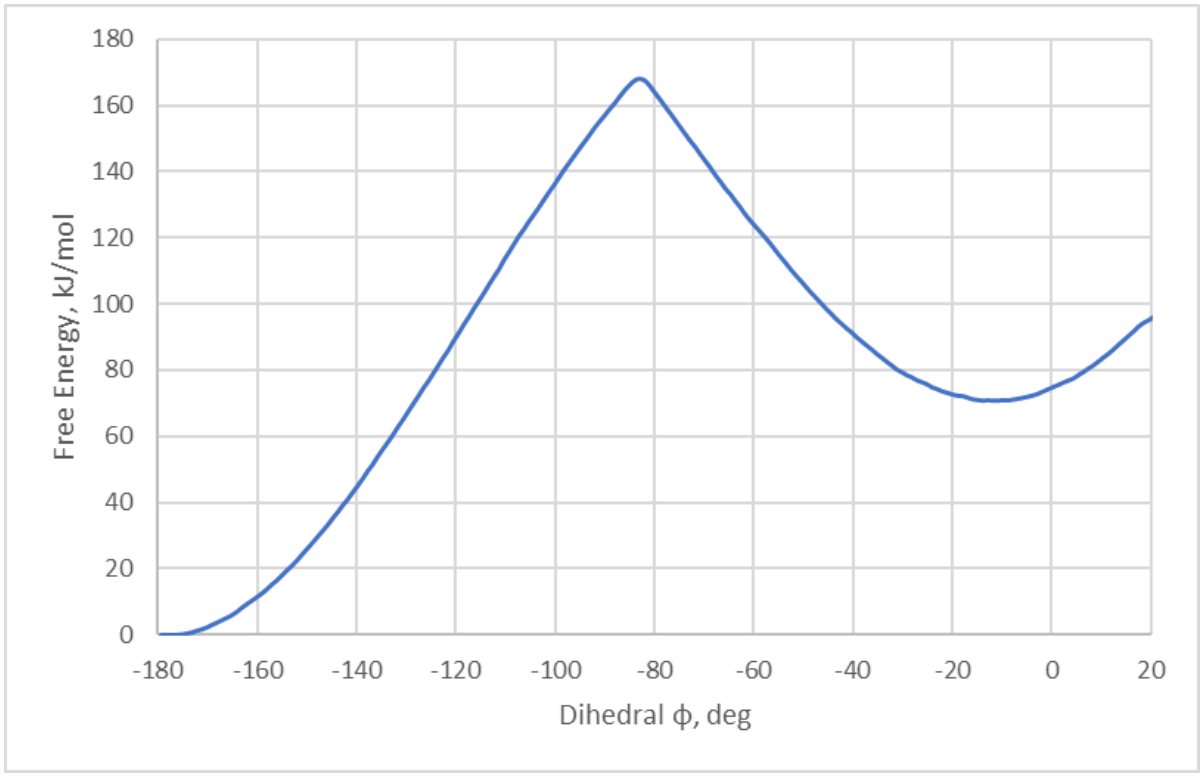
```
qmmm-active            = true
qmmm-qmgroup           = QMatoms
qmmm-qmmethod          = PBE
qmmm-qmcharge          = 0
qmmm-qmmultiplicity    = 1
```



Isomerization free energy with MM forcefield

Amber14 Forcefield, Gromacs simulation, 1ns each window



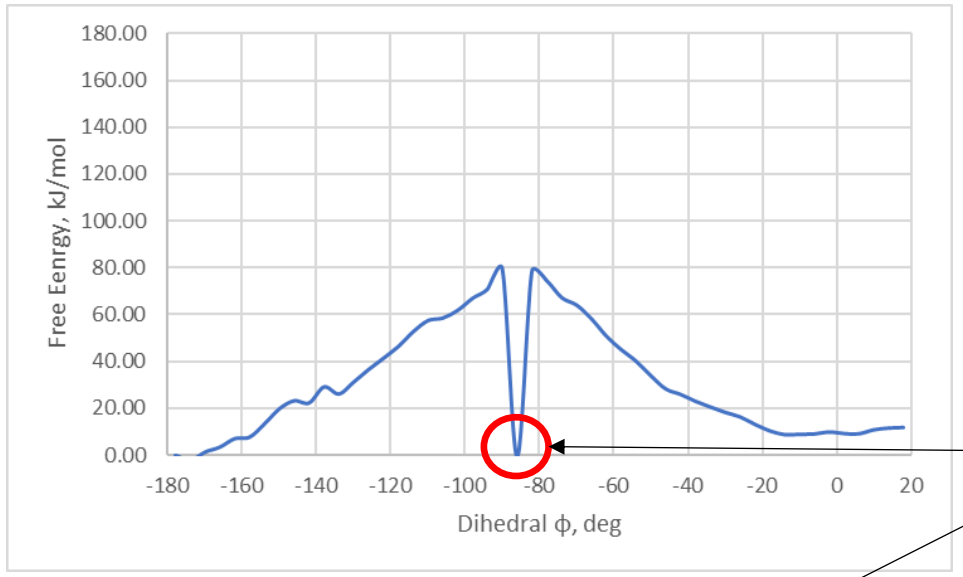


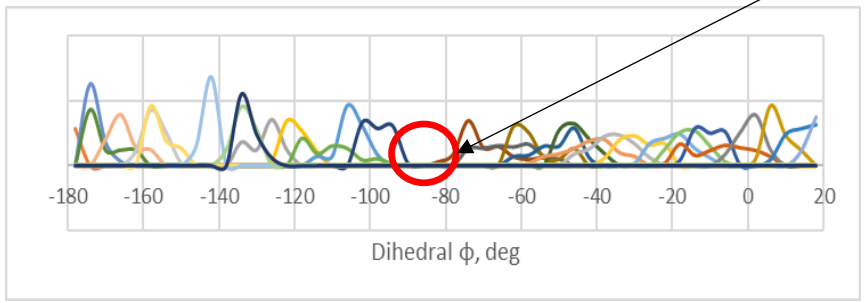
>160 KJ/mol isomerization barrier. Lets see how it changes if we will go for QM simulation!

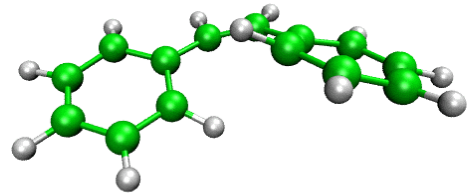
Do the part (9) from the "Exercise 2"

Isomerization free energy with QM

PBE, Gromacs-CP2K simulation for ~100fs (100 steps) for each frame



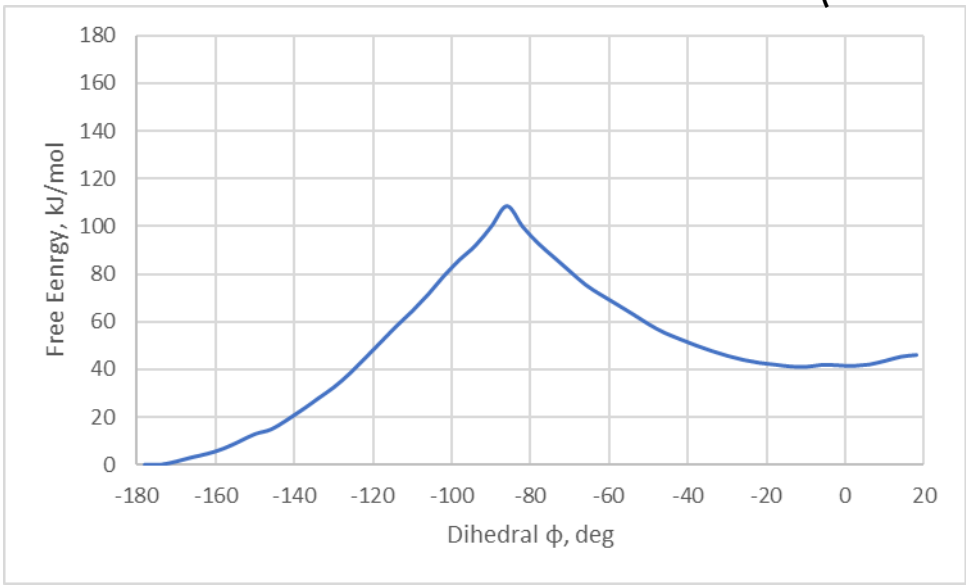


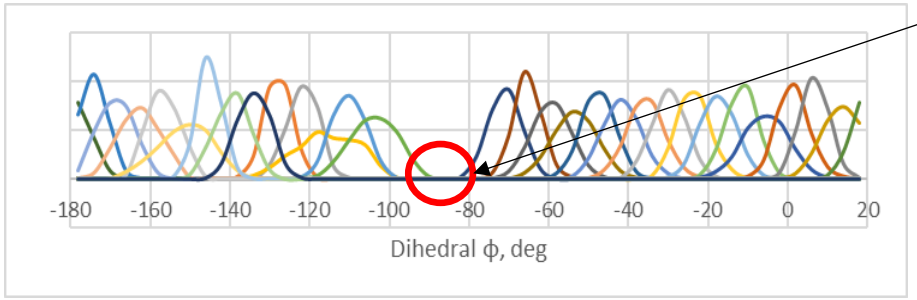


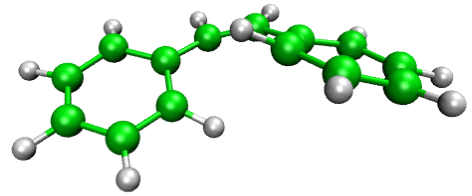
Heavily Under-sampled region

Isomerization free energy with QM

PBE, Gromacs-CP2K simulation extended to 10 ps
(10000 steps)



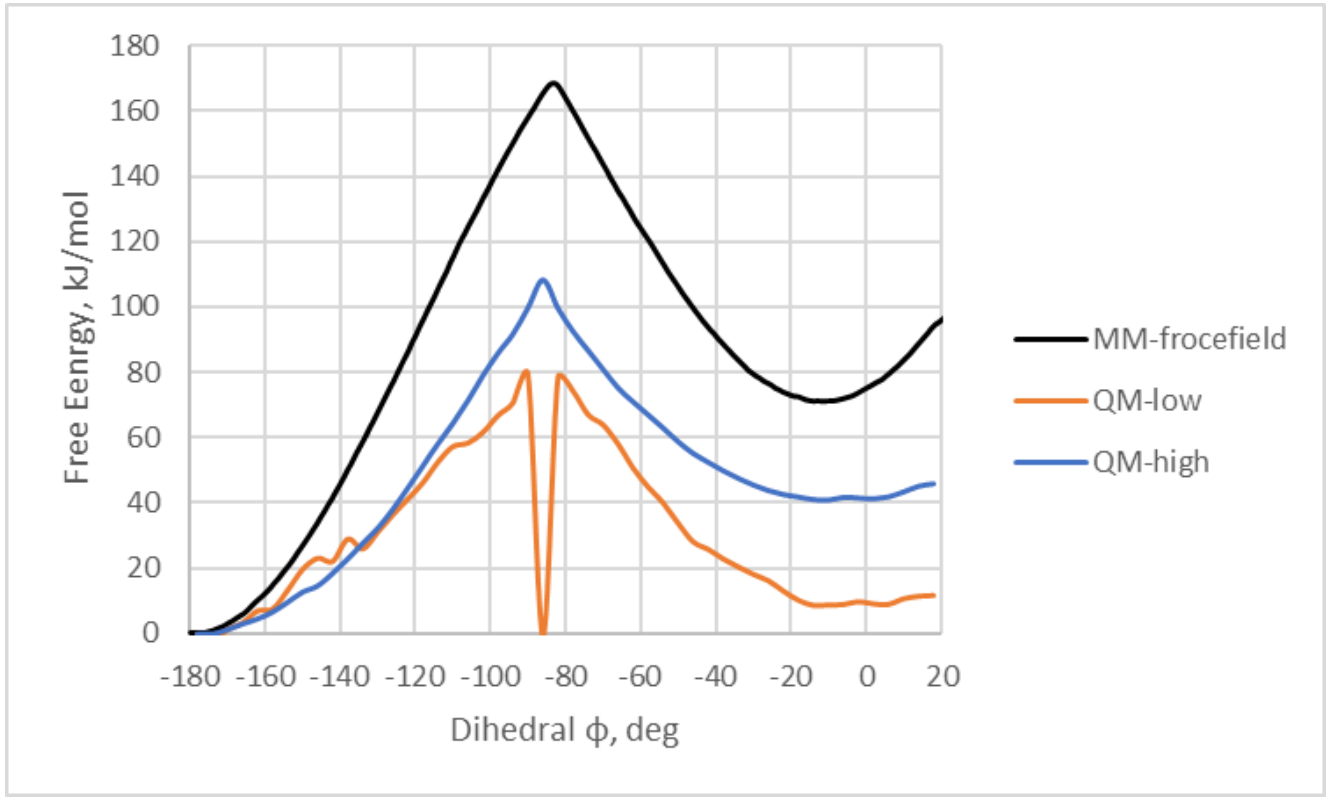




Still under-sampled

For QM/MM suggested amount of sampling is at least 30-50 ps per window after extensive (1-10 ns) MM equilibration!

Isomerization free energy with QM



MM forcefield gives artificially high barrier

Questions?

End of the practical: GROMACS + CP2K Part I

Practical: GROMACS + CP2K Part II

1. Lecture recap (QM/MM, GEEP, PBC)
2. Setting up a QM/MM calculation with solvent
3. CP2K input and output
4. Large protein system setup

Lecture Recap: GEEP for QM/MM Coupling - CP2K

- **QM polarization due to the MM part included.**

$$H = H_{MM} + H_{QM} + H_{QM/MM}$$

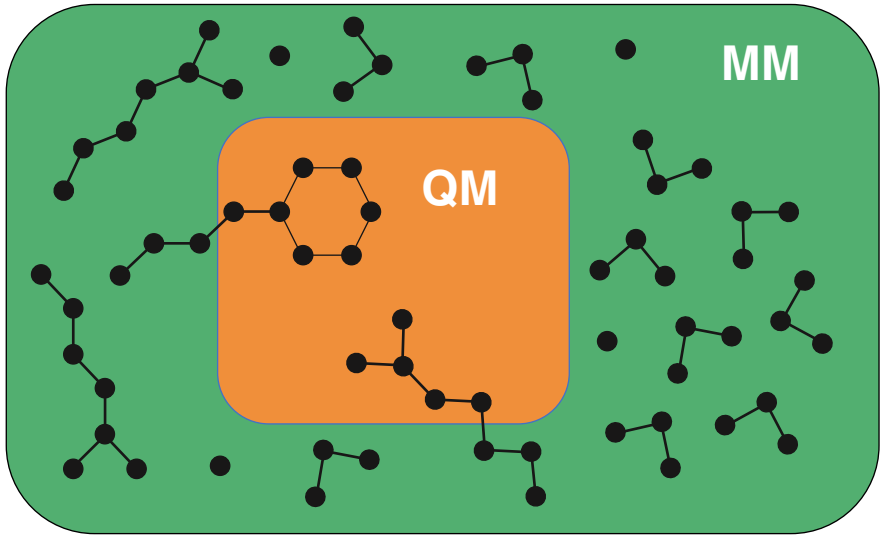
Forcefield Quickstep GEEP

$$E^{QM-MM}_{electrostatic} = \sum_{I \in MM} q_I \int \frac{\rho(\mathbf{r})\, v_I^{smear}(|\mathbf{r}_i - \mathbf{R}_I|)}{|\mathbf{r}_i - \mathbf{R}_I|} d\mathbf{r}$$

$$q_I v_I^{smear}(|\mathbf{r}_i - \mathbf{R}_I|) = \sum_{N_g} A_g e^{-(|\mathbf{r}_i - \mathbf{R}_I|/G_g)^2} + R_{low}(|\mathbf{r}_i - \mathbf{R}_I|)$$

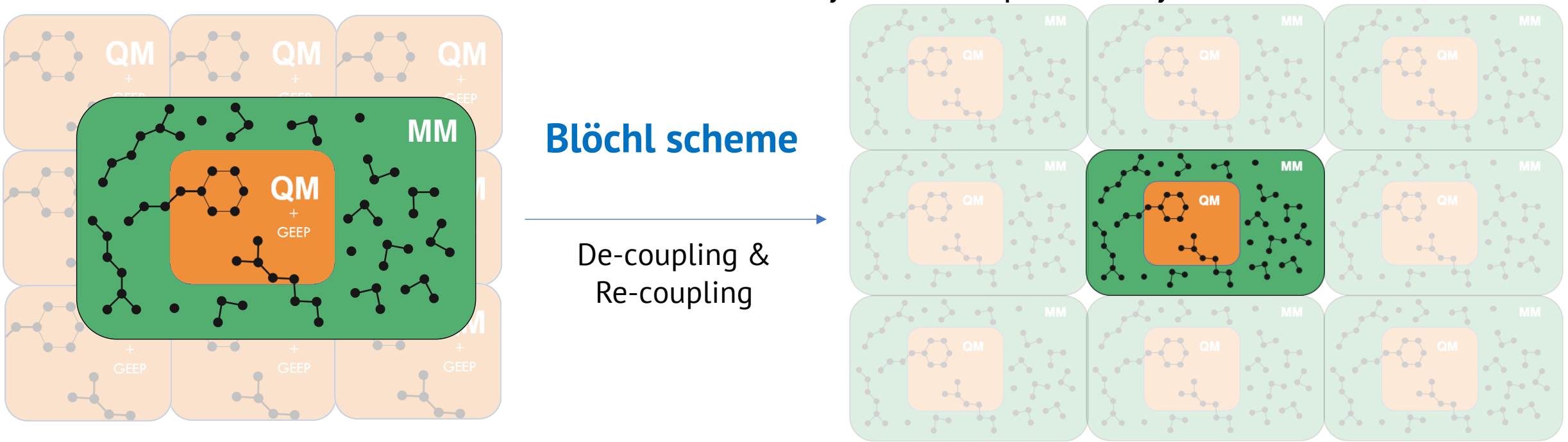
+

Real Space multi-grid approach



Fully periodic QM/MM

- GEEP projects electrostatic potential from point charges onto the multi-grid of QM box
- QM-QM periodic interactions are threated efficiently with Quickstep
- Unless the QM and MM box have same dimensions the QM images over PBC will have incorrect periodicity
- Blöchl scheme is used in CP2K to restore full system box periodicity

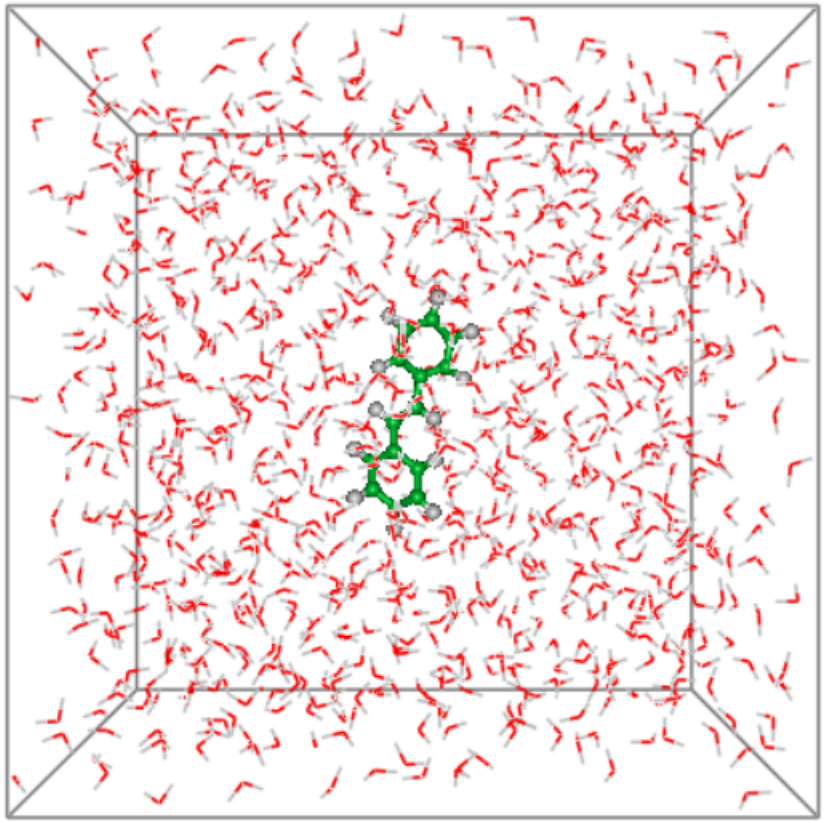


Practical: GROMACS + CP2K Part II

1. Lecture recap (QM/MM, GEEP, PBC)
2. Setting up a QM/MM calculation with solvent
3. CP2K input and output
4. Large protein system setup

Exercise 3: Energy minimization with QM/MM

```
>> cd ../stilbene_water
```



Objective: Optimize system with QM/MM

QM subsystem : Stilbene

MM subsystem : 1001 waters with TIP3P parameters

QM charge: 0

QM multiplicity: 1

Functional: PBE

You can download and open stilbene-sol.pdb with PyMOL

Do the steps (1)-(4) from the "Exercise 3"

Practical: GROMACS + CP2K Part II

1. Lecture recap (QM/MM, GEEP, PBC)
2. Setting up a QM/MM calculation with solvent
3. CP2K input and output
4. Large protein system setup

Exercise 3: CP2K input files

Input Section: QMMM

```
&FORCE_EVAL
...
  &QMMM                        !Definition of QM region and QM-MM coupling
    &CELL                      ! QM Cell
      A 9.720 0.000 0.000      ! Defined with three vectors A, B, C (in Angstroms)
      B 0.000 21.740 0.000
      C 0.000 0.000 7.500
      PERIODIC XYZ             ! Fully periodic cell
    &END CELL
    ECOUPL GAUSS               ! QM-MM coupling method (GEEP)
    USE_GEEP_LIB 12            ! Number of gaussian functions used in GEEP
...
```

Exercise 3: CP2K input files

Input Section: QMMM

```
&FORCE_EVAL
  &QMMM                          !Definition of QM region and QM-MM coupling
    ...
    &PERIODIC                    ! Treating periodic QM-MM
      GMAX      1.0E+00
      &MULTIPOLE ON              ! Use Blöchl scheme (decoupling & re-coupling)
        RCUT      1.0E+01
        EWALD_PRECISION     1.0E-06
      &END
    &END PERIODIC
    &QM_KIND H                   ! Hydrogen, which should be treated as QM atoms
      MM_INDEX 2 4 6 9 11 13 15 18 20 22 24 26          ! Indexes of atoms starting from 1
    &END QM_KIND
    ...
  &END QMMM
```

Exercise 3: CP2K input files

Input Section: MM

```
&FORCE_EVAL
...
  &MM                                  ! MM region treatment
    &FORCEFIELD
      DO_NONBONDED FALSE               ! Do NOT do MM-MM point charges and VdW interactions
    &END FORCEFIELD
    &POISSON
      &EWALD
        EWALD_TYPE NONE                ! Do NOT do MM-MM periodic interactions
      &END EWALD
    &END POISSON
  &END MM
...
```

GROMACS handles the description of the MM region!

Exercise 3: CP2K input files

Input Section: TOPOLOGY

```
&SUBSYS
   ...
   &TOPOLOGY                          ! grompp will generate pdb with atomic charges for CP2K
     COORD_FILE_NAME stilbene.pdb     ! Make sure that files exists
     COORD_FILE_FORMAT PDB
     CHARGE_EXTENDED TRUE             ! Read charges from PDB Extended Beta field (starting from column 81)
     CONNECTIVITY OFF                 ! Do not read or generate bonds (MM treated by Gromacs)
     &GENERATE
       &ISOLATED_ATOMS                ! Generate topology consisting of isolated atoms
         LIST 1..26
       &END
     &END GENERATE
   &END TOPOLOGY
   ...
```

Exercise 3: CP2K input files

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

```
&FORCE_EVAL
...
 &SUBSYS
    ...
    &QM_KIND H
      MM_INDEX 2 4 6 9 11 13 15 18 20 22 24 26
    &END QM_KIND
    &QM_KIND C
      MM_INDEX 1 3 5 7 8 10 12 14 16 17 19 21 23 25
    &END QM_KIND
    ...
 &END SUBSYS
...
&END FORCE_EVAL
```

Only stilbene atoms marked as QM

Questions?

Exercise 3: CP2K input files

```
>> less stilbene-sol-opt.pdb
```

```
....
ATOM     22  C    QM     1      18.947  14.687  16.609  1.00  0.00          C   0.000000
ATOM     23  H    QM     1      19.399  13.704  16.735  1.00  0.00          H   0.000000
ATOM     24  C    QM     1      17.563  14.803  16.517  1.00  0.00          C   0.000000
ATOM     25  H    QM     1      16.951  13.902  16.556  1.00  0.00          H   0.000000
ATOM     26  O    MM     2       1.816   6.680   1.359  1.00  0.00          O  -0.834000
ATOM     27  H    MM     2       0.966   6.696   1.800  1.00  0.00          H   0.417000
ATOM     28  H    MM     2       1.615   6.408   0.463  1.00  0.00          H   0.417000
ATOM     29  O    MM     2       1.559   2.257  10.377  1.00  0.00          O  -0.834000
ATOM     30  H    MM     2       1.900   2.175  11.268  1.00  0.00          H   0.417000
ATOM     31  H    MM     2       1.127   1.420  10.208  1.00  0.00          H   0.417000
ATOM     32  O    MM     2      31.142   4.832   6.637  1.00  0.00          O  -0.834000
ATOM     33  H    MM     2      30.197   4.943   6.736  1.00  0.00          H   0.417000
ATOM     34  H    MM     2      31.238   4.242   5.889  1.00  0.00          H   0.417000
....
```

TIP3P water

Exercise 3: results of energy minimization

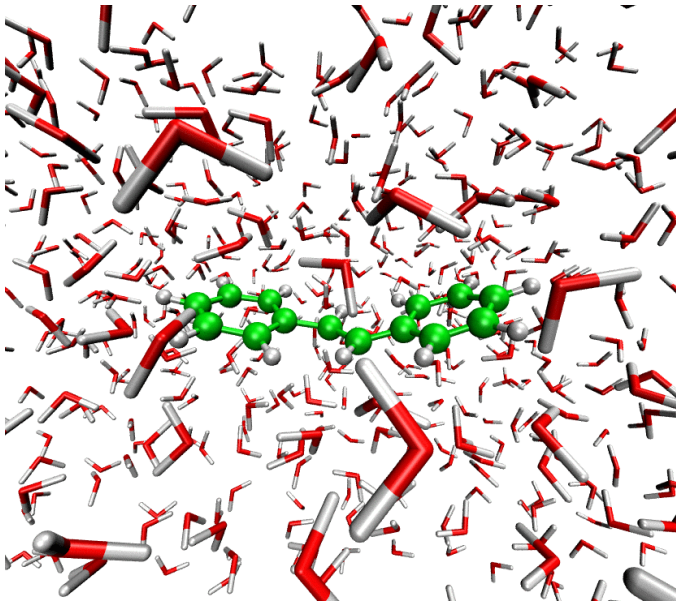
Do step (6) of "Exercise 3".

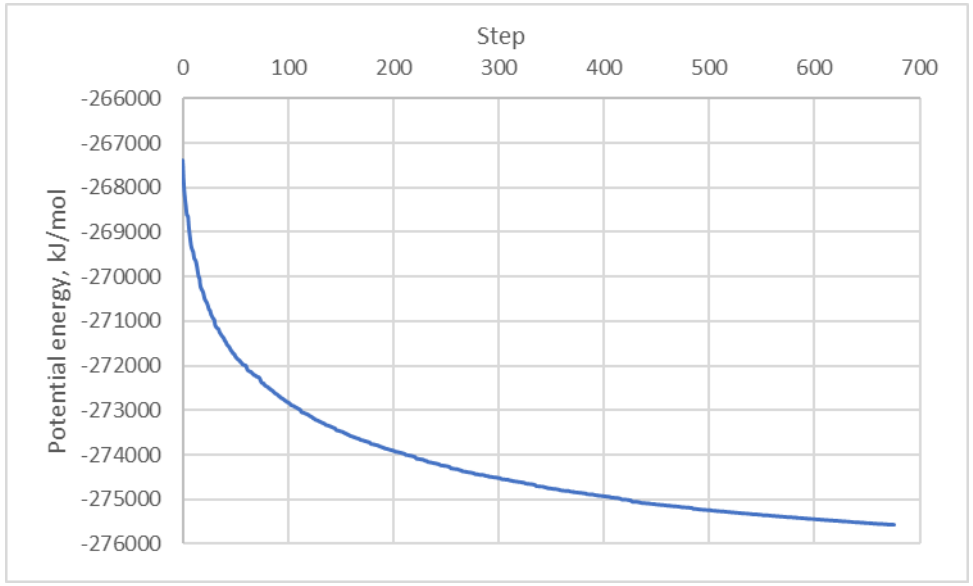
>> gmx_cp2k energy (reads data from ener.edr file)

....

> 6 (potential energy)

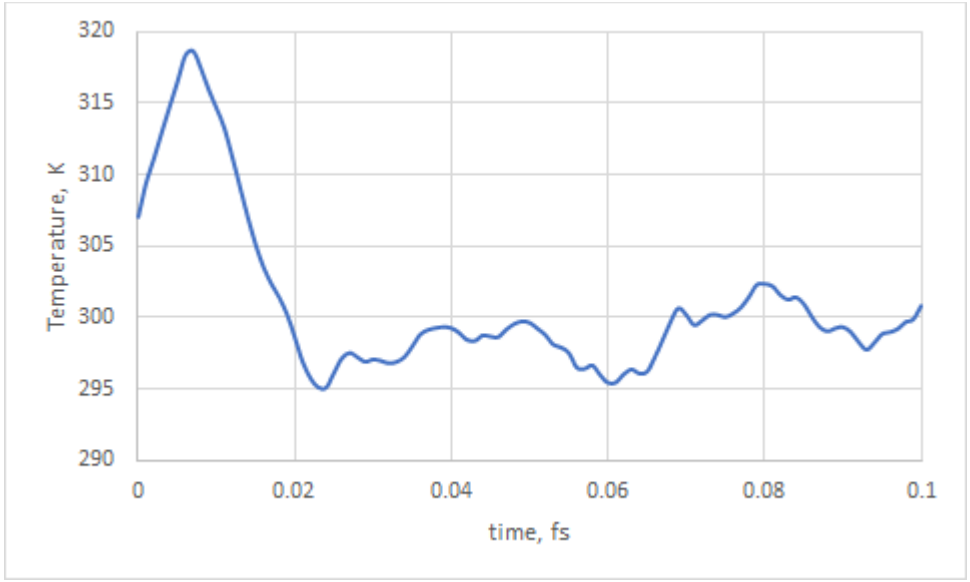
Download and open energy.xvg your need Grace to open file or copy data from file to Excel by columns

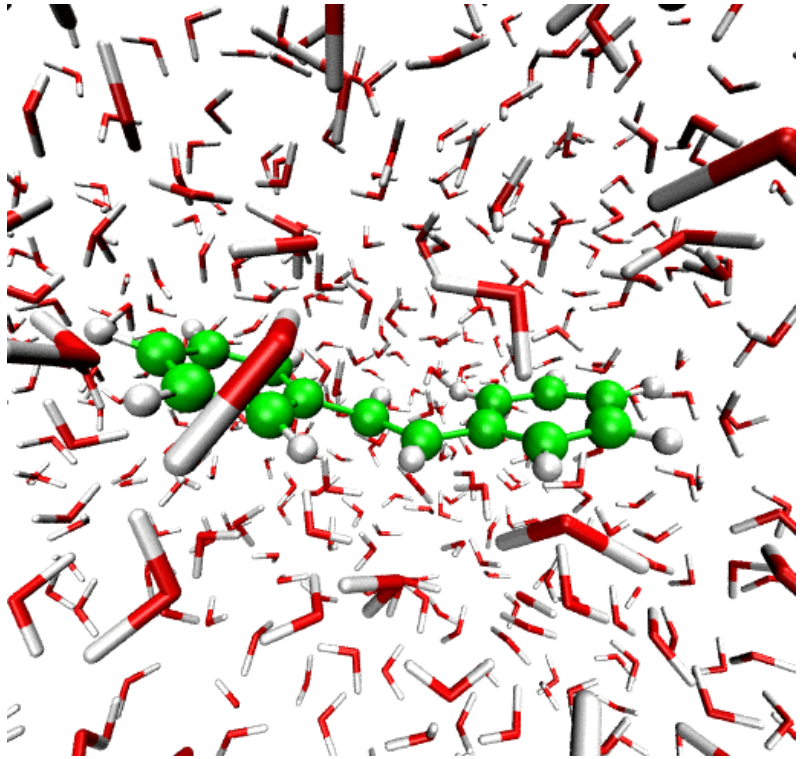




Exercise 3: MD simulations

Do steps (7)-(9) of “Exercise 3”



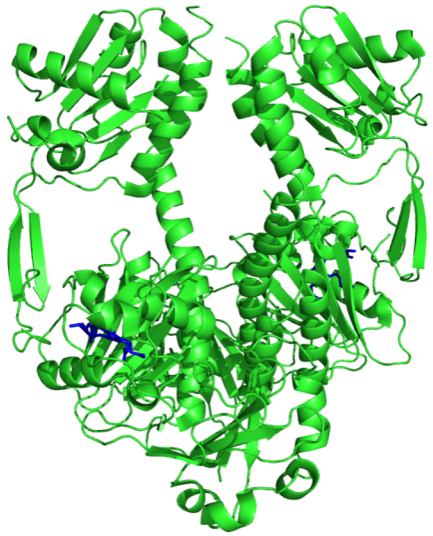


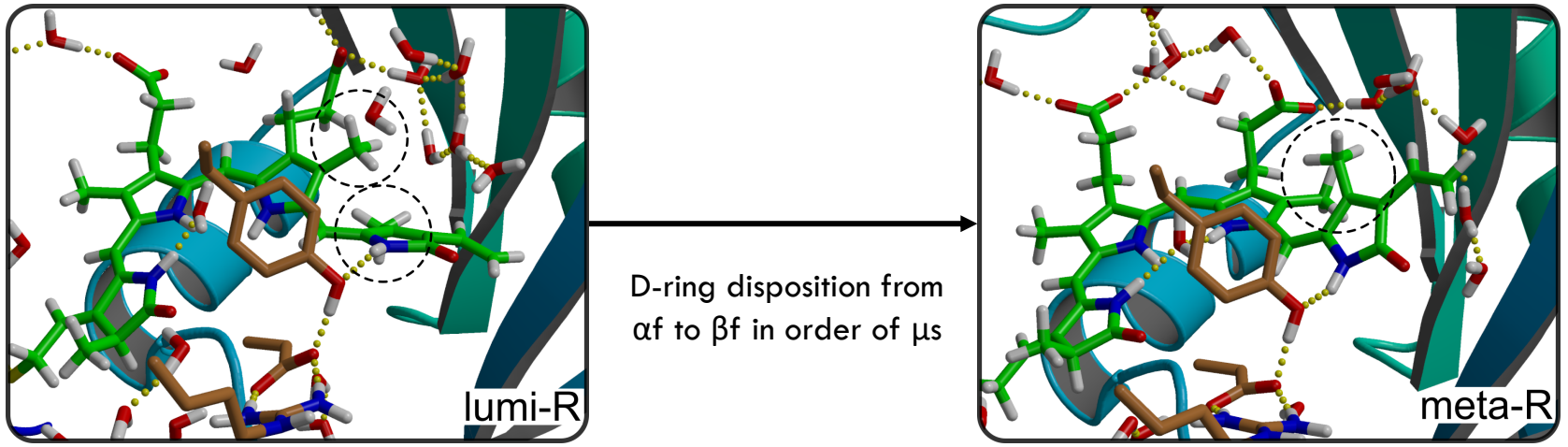
Practical: GROMACS + CP2K Part II

1. Lecture recap (QM/MM, GEEP, PBC)
2. Setting up a QM/MM calculation with solvent
3. CP2K input and output
4. Large protein system setup

Exercise 4: Protein simulations

>> cd ../phytochrome





Objective:

D-ring disposition energy barrier (α_f —> β_f)

NEB + umbrella sampling simulations

QM part - Chromophore

QM method - PBE/DZVP-MOLOPT-GTH

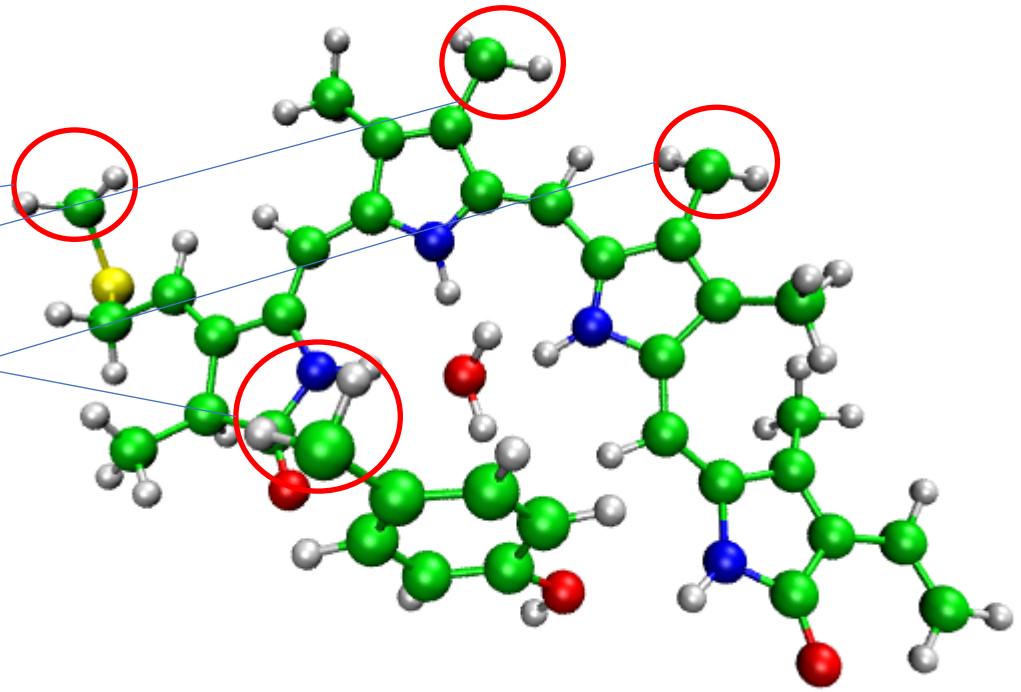
MM Frocefield - Amber03

Do the steps (1)-(3) from the "Exercise 4"

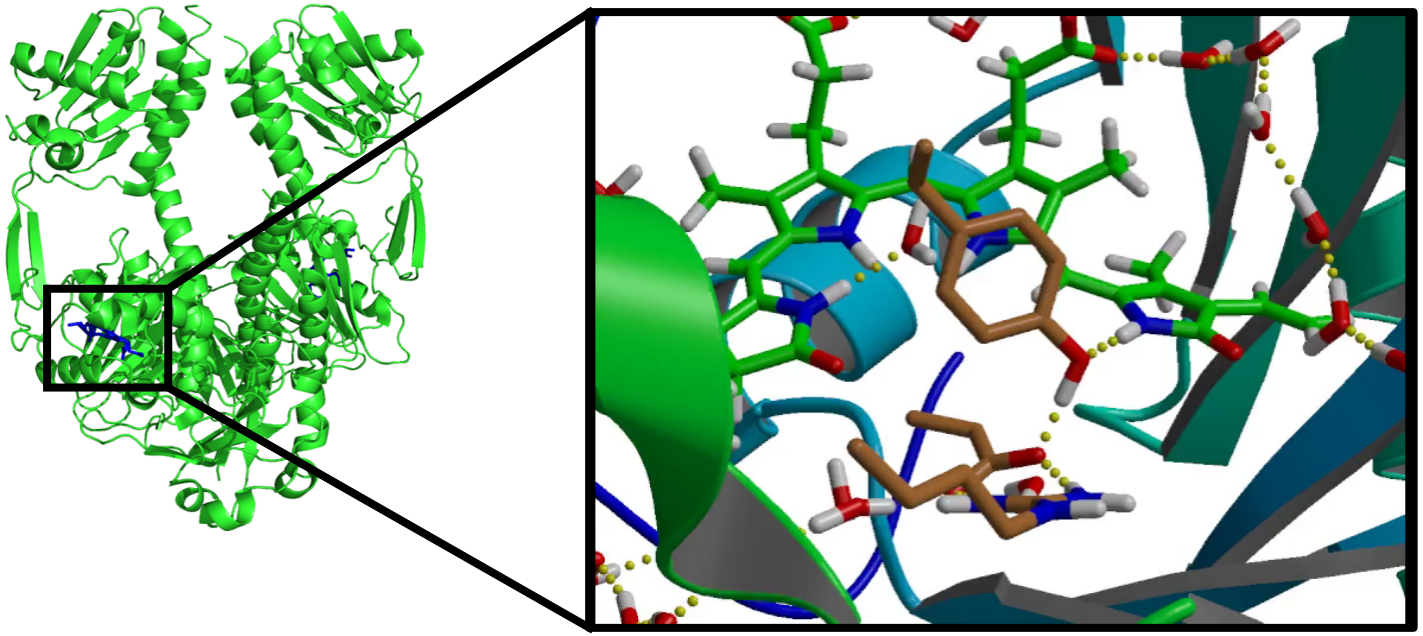
Exercise 4: Protein simulations

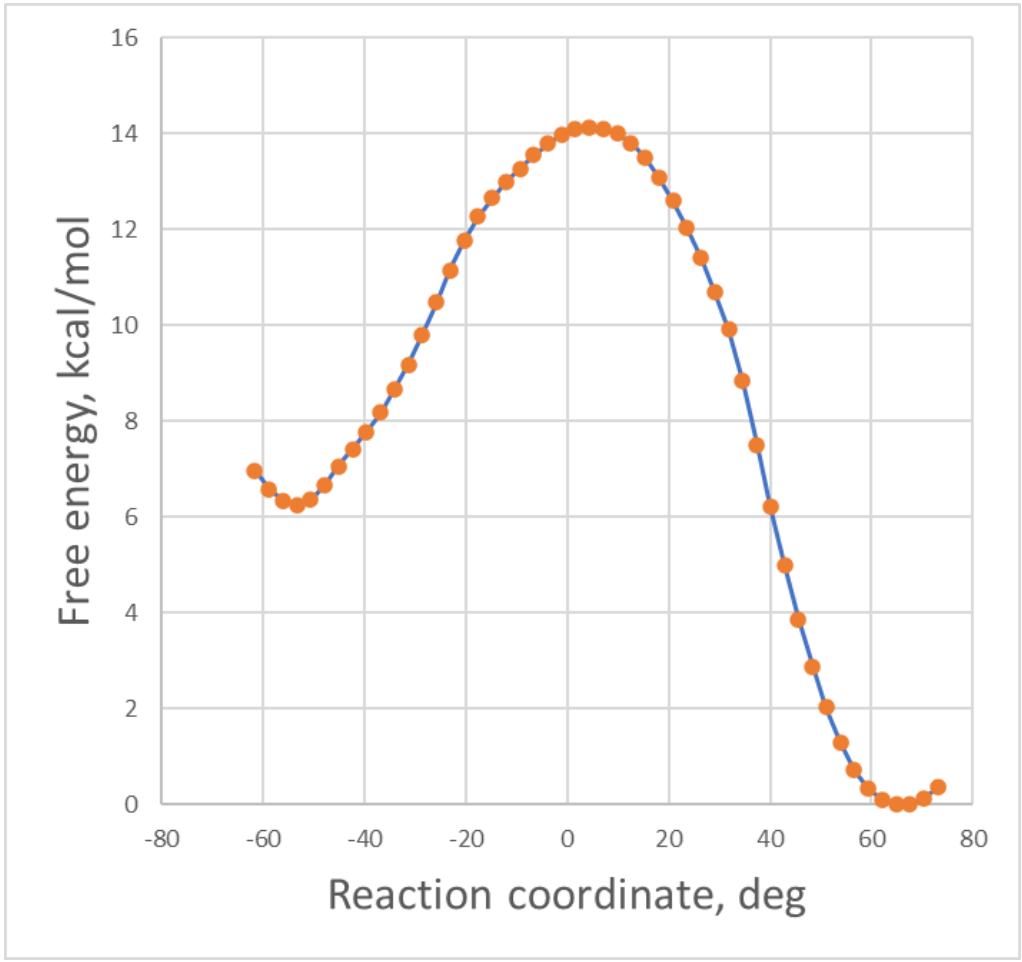
>> less phytochrome.inp

```
&FORCE_EVAL
...
 &QMMM
    ...
 &LINK
      QM_INDEX 547
      MM_INDEX 545
    &END LINK
    &LINK
      QM_INDEX 4162
      MM_INDEX 4160
    &END LINK
    &LINK
      QM_INDEX 7987
      MM_INDEX 7972
    &END LINK
    &LINK
      QM_INDEX 7984
      MM_INDEX 7978
    &END LINK    ...
 &END QMMM
...
&END FORCE_EVAL
```



Protein simulations: umbrella sampling





End of the practical: GROMACS + CP2K Part II

BioExcel Partners

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

NORMAN CONSULTING

EMBL-EBI

KTH VETENSKAP OCH KONST

THE UNIVERSITY OF EDINBURGH

JÜLICH FORSCHUNGSZENTRUM

IRB BARCELONA
INSTITUTO DE INVESTIGACIÓN BIOMÉDICA

Utrecht University

acrosslimits

Ian Harrow Consulting

MAX-PLANCK-GESELLSCHAFT

MANCHESTER 1824

UNIVERSITY OF JYVÄSKYLA

NBD NOSTRUM BIODISCOVERY

Horizon 2020
European Union Funding
for Research & Innovation

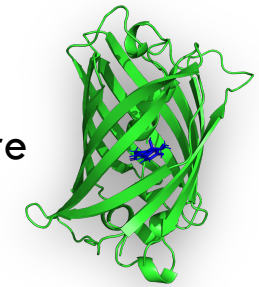
BioExcel is funded by the European Union Horizon 2020 program under grant agreements 675728 and 823830.

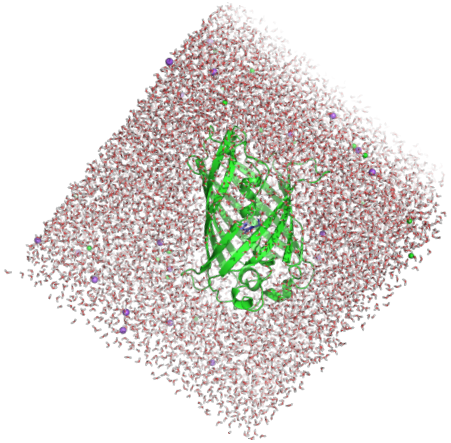
Practical: GROMACS + CP2K Part III

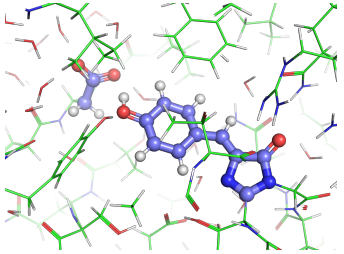
1. Make protein QMMM system starting from the PDB structure
2. Usage of non-standard CP2K input parameters
3. Calculation of the absorption spectra for your system

Build protein system from pdb file

Input structure







SYSTEM SETUP

- Missing residues
- Residue protonation
- pdb2gmx
- Define box
- Add solvent and ions
- ….

MD SIMULATIONS

- Prepare .mdp files
- Energy minimization
- Pressure & temperature equilibration
- Production run
- …

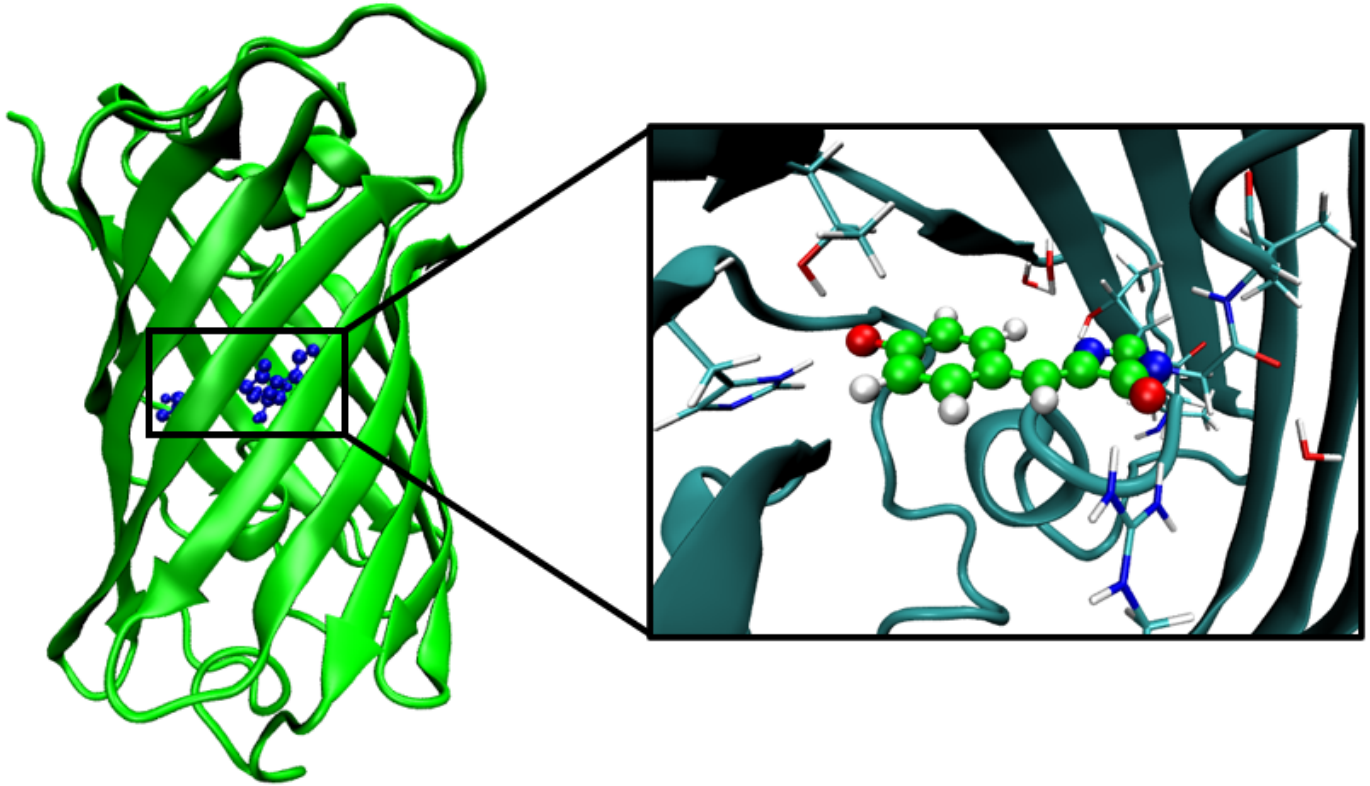
- Index file
- .mdp parameters for QM/MM
-External CP2K input (optional)

QM/MM Calculations

- Geometry optimization
- Atomic charges
- QM/MM MD
- Absorption spectra
- Umbrella Sampling
- ….

Exercise 5: build protein system from pdb file

```
>> cd egfp
```



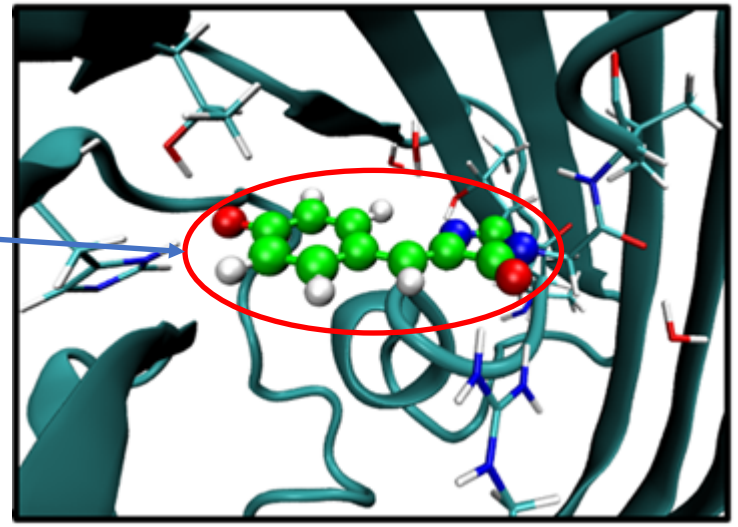
Objective:
Make QM/MM model of EGFP protein and perform MD simulation

System:
QM part - Chromophore
QM method - PBE/DZVP-MOLOPT-GTH
MM Forcefield - Amber03

Do the steps (1)-(5) from the "Exercise 5" they are a pure MM simulations

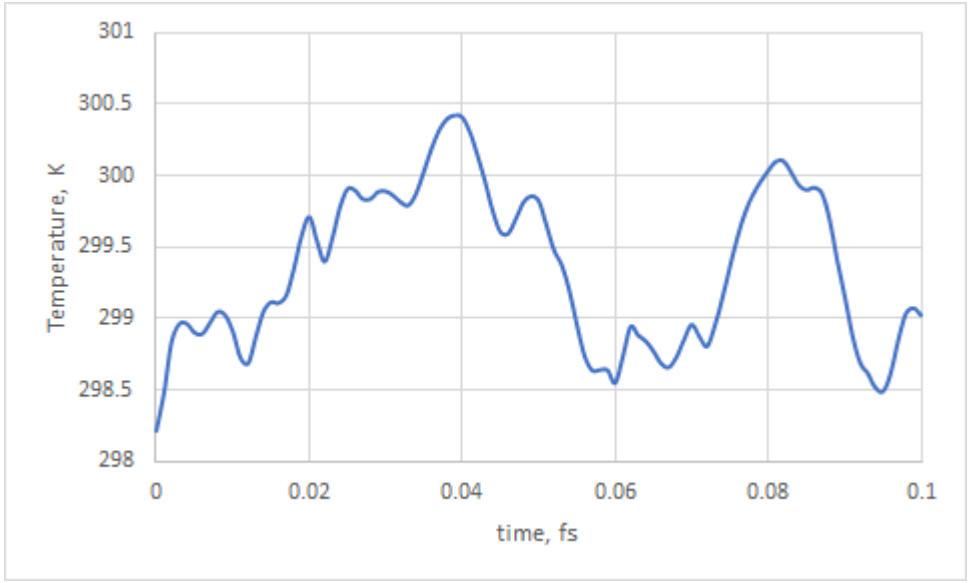
MDP Parameters: QM/MM

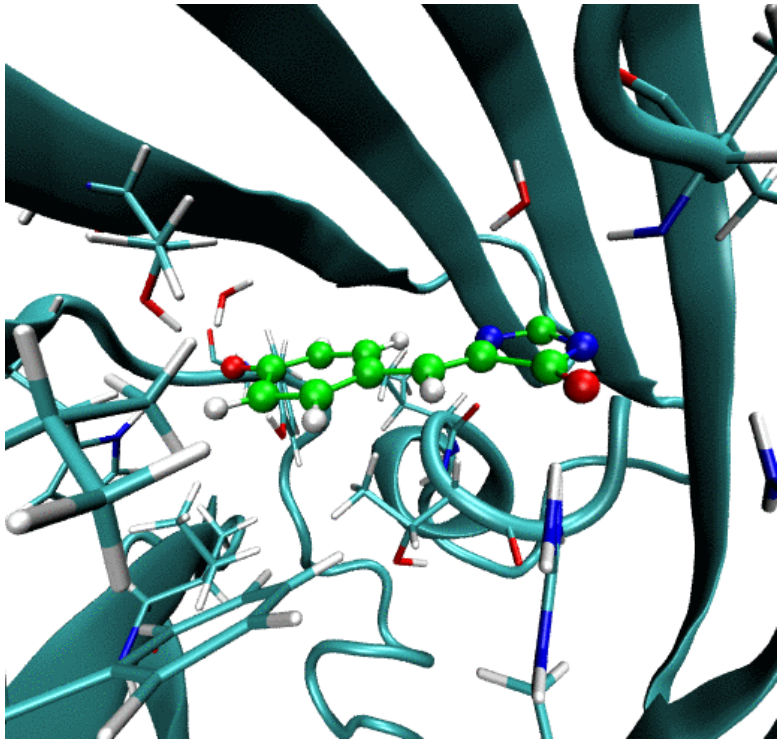
```
qmmm-active          = true
qmmm-qmgroup         = QMatoms
qmmm-qmmethod        = PBE
qmmm-qmcharge        = -1
qmmm-qmmultiplicity  = 1
```



Exercise 5: build protein system from pdb file

Do the steps (6)-(9) from the “Exercise 5”

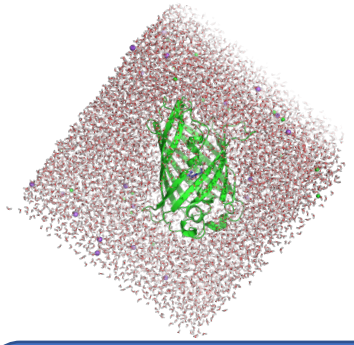


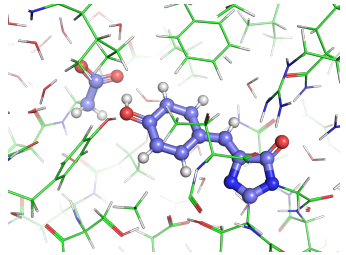


Practical: GROMACS + CP2K Part III

1. Make protein QMMM system starting from the PDB structure
2. Usage of non-standard CP2K input parameters
3. Calculation of the absorption spectra for your system

Exercise 6: using user-defined QM input file





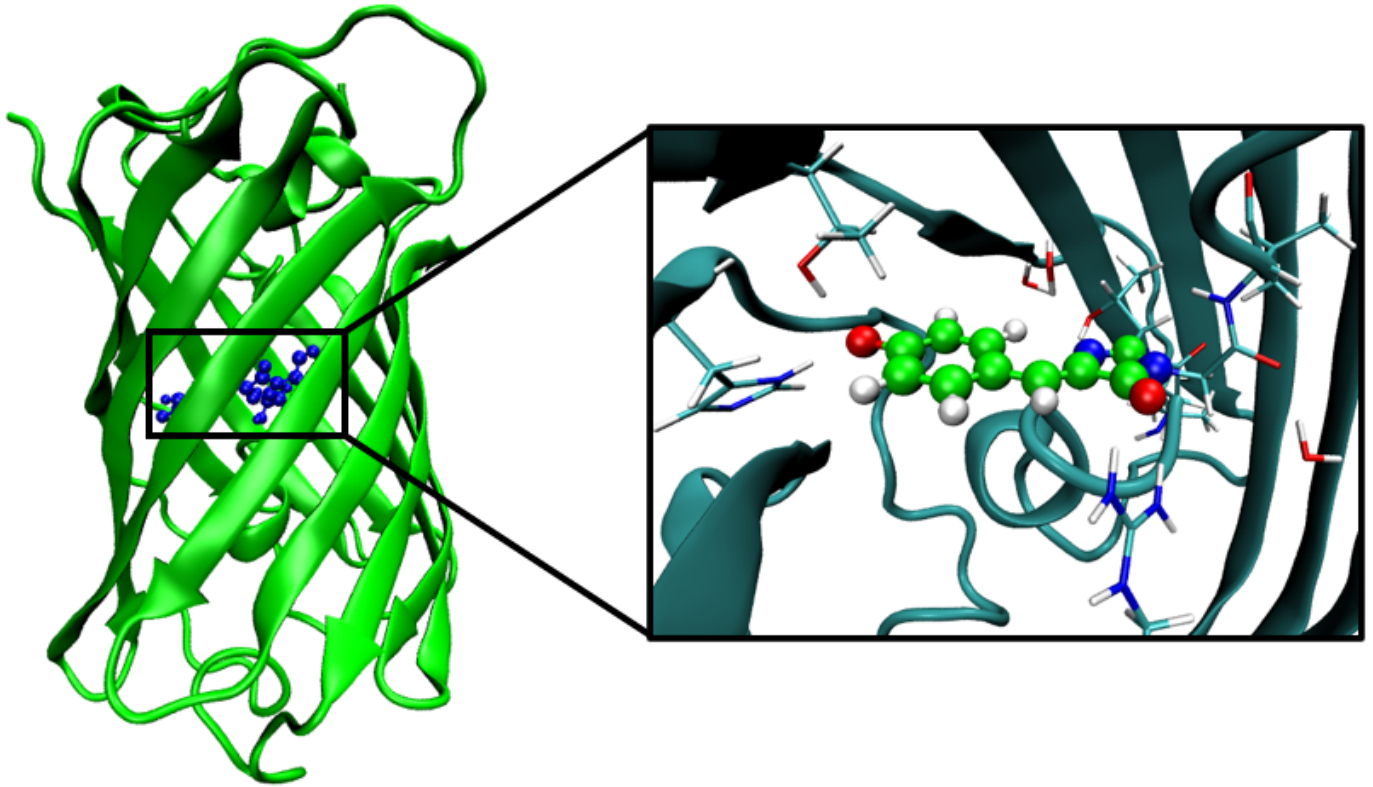
Classical MD Simulations

- Index file
- .mdp parameters for QM/MM
- Your own CP2K input

QM/MM Calculations

```
qmmm-active          = true      ; Activates QMMM MdModule
qmmm-qmgroup         = QMatoms   ; Index group with QM atoms
qmmm-qmmethod        = INPUT     ; Provide your own input for CP2K
qmmm-qminputfile     = cp2k.inp  ; Input file to be used
```

Exercise 6: using user-defined QM input file



Objective:
Simulate UV/Vis absorption spectra of EGFP protein

System:
QM part - Chromophore
QM method - PBE/DZVP-MOLOPT-GTH
TDDFT – for excitation energies
MM Frocefield - Amber03

Do the steps (1)-(5) from the "Exercise 6"

Exercise 6: using user-defined QM input file

egfp-qmmm-spec.inp

```
&FORCE_EVAL
...
  &DFT
  ...
  &END DFT
  &PROPERTIES                          ! Request additional properties to be calculated after SCF
    &TDDFPT                            ! TDDFT excitations
      NSTATES      5                   ! Number of excited states to calculate
      MAX_ITER     10                  ! Maximum Davidson diagonalization Iterations to be performed
      CONVERGENCE [eV] 1.0e-3          ! Convergence of energies in eV
    &END TDDFPT
  &END PROPERTIES
...
&END FORCE_EVAL
```

Exercise 6: using user-defined QM input file

less md-qmmm-spec.mdp

```
; CP2K QMMM parameters
qmmm-active             = true    ; Activate QMMM MdModule
qmmm-qmgroup            = Qmatoms ; Index group of QM atoms
qmmm-qmmethod           = INPUT   ; Method to use
qmmm-qminputfile        = egfp-qmmm-spec.inp ; external input file
```

Practical: GROMACS + CP2K Part III

1. Make protein QMMM system starting from the PDB structure
2. Usage of non-standard CP2K input parameters
3. Calculation of the absorption spectra for your system

Exercise 6: TDDFT excitations

```
>> less egfp-qmmm-spec.out
```

Results of TDDFT calculation will look like that:

```
 R-TDDFPT states of multiplicity 1

          State    Excitation        Transition dipole (a.u.)        Oscillator
          number   energy (eV)       x           y           z     strength (a.u.)
          ------------------------------------------------------------------------
 TDDFPT|      1       2.00058  -3.5991E-02 -5.4149E-02 -7.9349E-03   2.10286E-04
 TDDFPT|      2       3.08318   1.3797E+00 -1.7284E-01  6.5479E-01   1.78424E-01
 TDDFPT|      3       3.22153   2.4009E+00 -9.8621E-01  1.1151E+00   6.29837E-01
 TDDFPT|      4       3.54032  -4.8474E-01 -1.9293E-01 -9.7242E-02   2.44295E-02
 TDDFPT|      5       3.55772  -5.5083E-01  3.7988E-01 -2.2543E-01   4.34543E-02
```

We can gather that information over the trajectory:

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

Do the step (7) from the “Exercise 6

Exercise 6: convolving the spectra

>> less excitations

E_i, eV	f, a.u.
2.00058	2.10286E-04
3.08318	1.78424E-01
3.22153	6.29837E-01
3.54032	2.44295E-02
3.55772	4.34543E-02
2.04421	2.02664E-04
3.10716	1.86226E-01
3.24825	6.34929E-01
3.56783	3.07195E-02
3.59003	4.00265E-02
2.13146	2.12315E-04
3.12118	1.43032E-01
3.30321	7.14998E-01
3.61706	5.70336E-02
3.67955	2.27674E-02

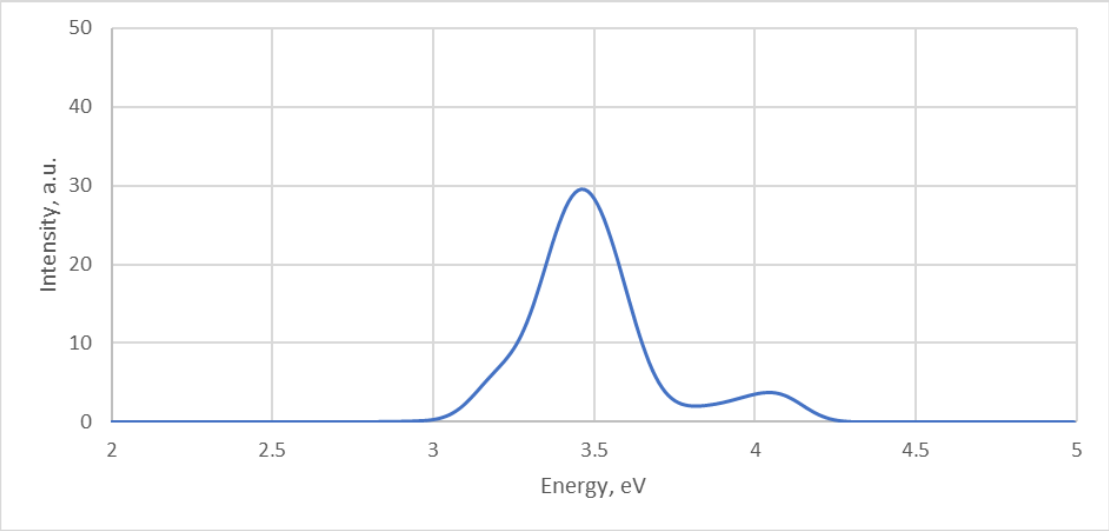
$$I(E) = \sum_{i}^{N} f * e^{-(E-E_i)^2/\sigma^2}$$

σ – parameter defining gaussians half-width
For example 0.1 eV in that case

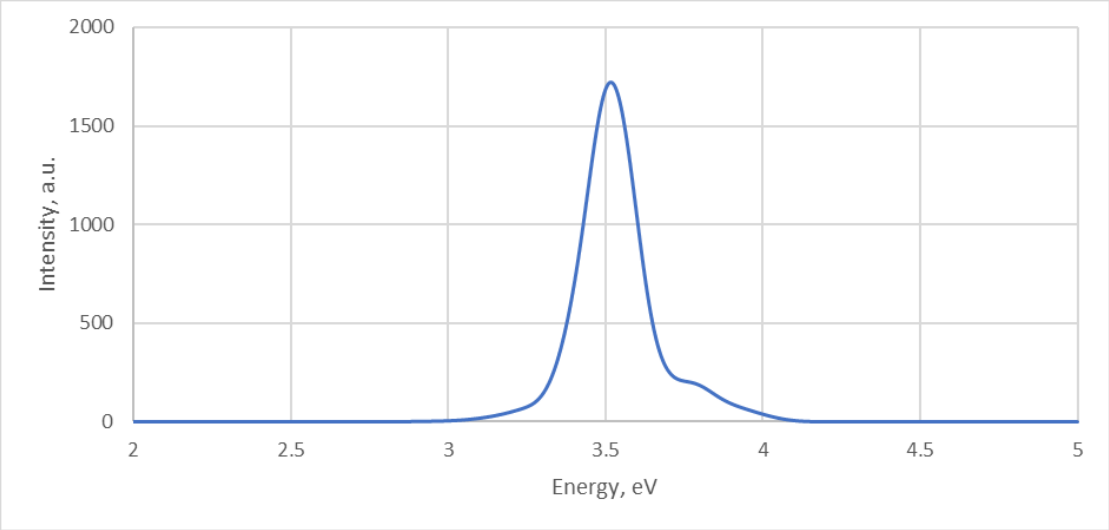
Do the step (8) from the “Exercise 6

Exercise 6: Results

After 100fs sampling



After 3ps sampling



Further information

1) CP2K parameters and best practices:

   https://docs.bioexcel.eu/qmmm_bpg/en/main/

2) Best practices in QM/MM webinar series:

   https://bioexcel.eu/events/virtual-workshop-best-practices-in-qm-mm-simulation-of-biomolecular-systems/

3) Bioexcel YouTube channel:

   https://www.youtube.com/c/BioExcelCoE/videos

BioExcel Partners

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

NORMAN CONSULTING

EMBL-EBI

KTH VETENSKAP OCH KONST

THE UNIVERSITY OF EDINBURGH

JÜLICH FORSCHUNGSZENTRUM

IRB BARCELONA
INSTITUTO DE INVESTIGACIÓN BIOMÉDICA

Utrecht University

acrosslimits

Ian Harrow Consulting

MAX-PLANCK-GESELLSCHAFT

MANCHESTER 1824

UNIVERSITY OF JYVÄSKYLA

NBD NOSTRUM BIODISCOVERY

Horizon 2020
European Union Funding
for Research & Innovation

BioExcel is funded by the European Union Horizon 2020 program under grant agreements 675728 and 823830.