GROMACS-CP2K Interface

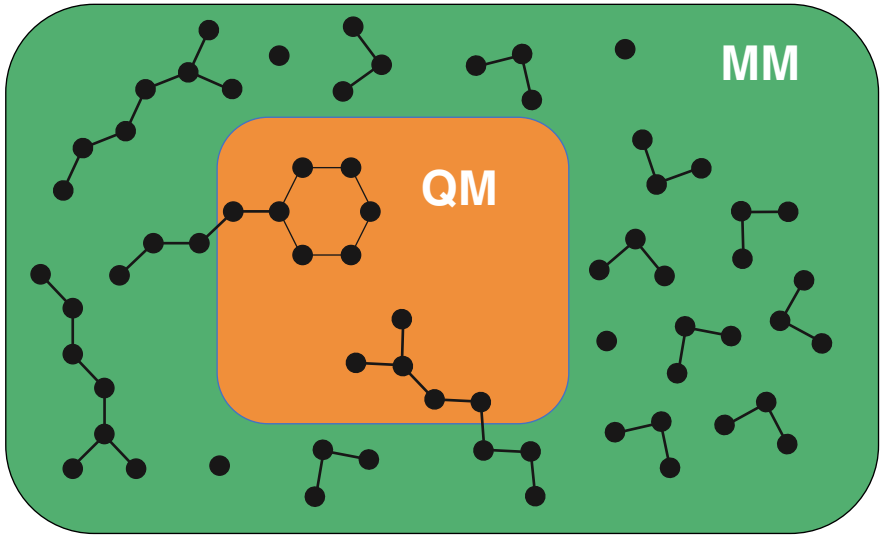
Multiscale QM/MM simulations for exploring reactions in biological systems

Dmitry Morozov

University of Jyväskylä, Finland

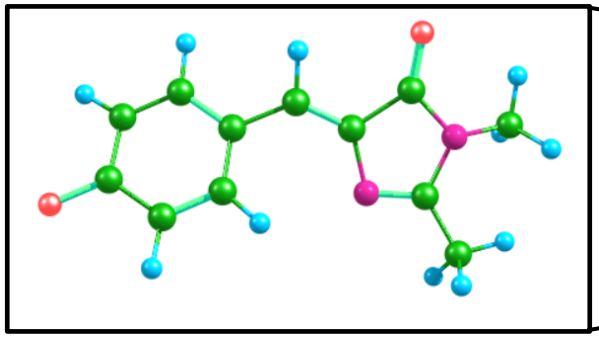
dmitry.morozov@jyu.fi

What is QM/MM and why we want to do it?



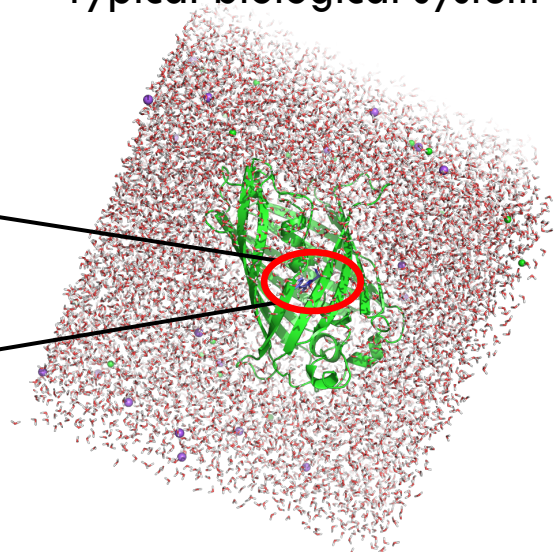
Modeling of biological systems

Typical system for QM simulations



Small number of atoms
(max of 100-200)

Typical biological system



≫10,000 atoms

Multiscale approach: QM/MM

Combining different levels of theory and resolutions

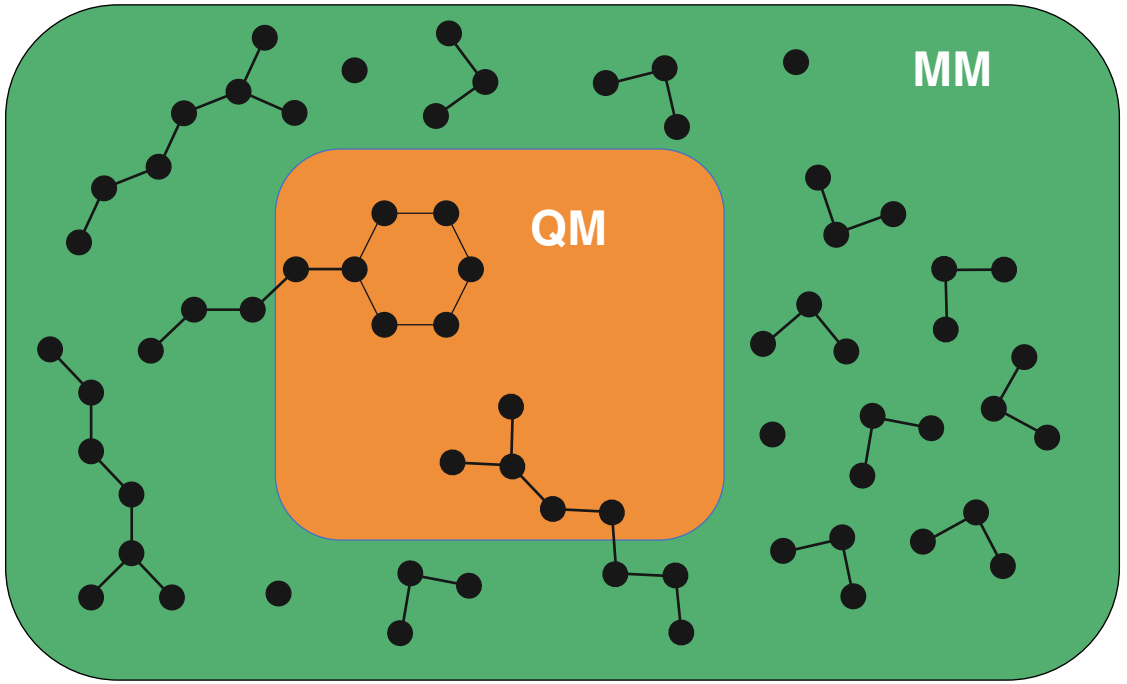
Hybrid multiscale description

A small **QM part** comprises the **chemically/photophysically active region** treated by computationally demanding electronic structure methods.

QM/MM Interface

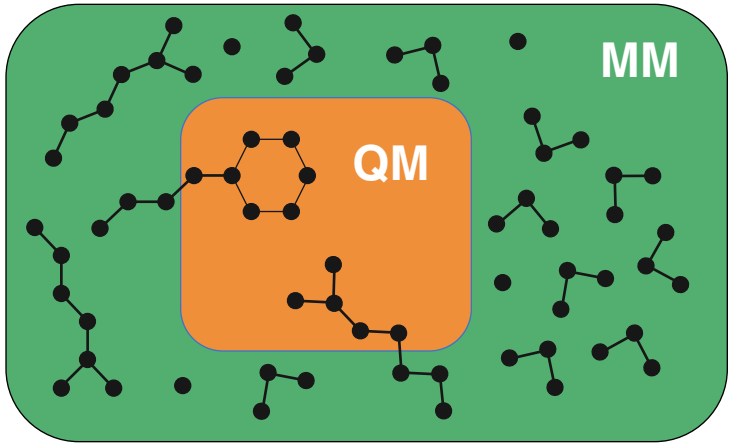
The remainder **MM part** is **described efficiently** at a lower level of theory by classical force fields.

The system is separated into two parts:



Nobel Prize Chemistry 2013: Karplus, Levitt, Warshel

Interactions in QM/MM



Hybrid QM/MM scheme:

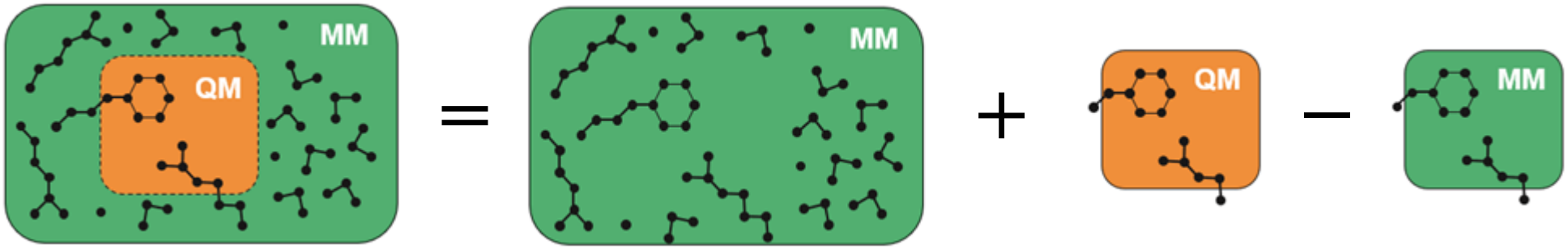
- Small part of the system at quantum (QM) level
- Rest of the system at forcefield (MM) level

The hybrid QM/MM **potential energy** contains three types of interactions:

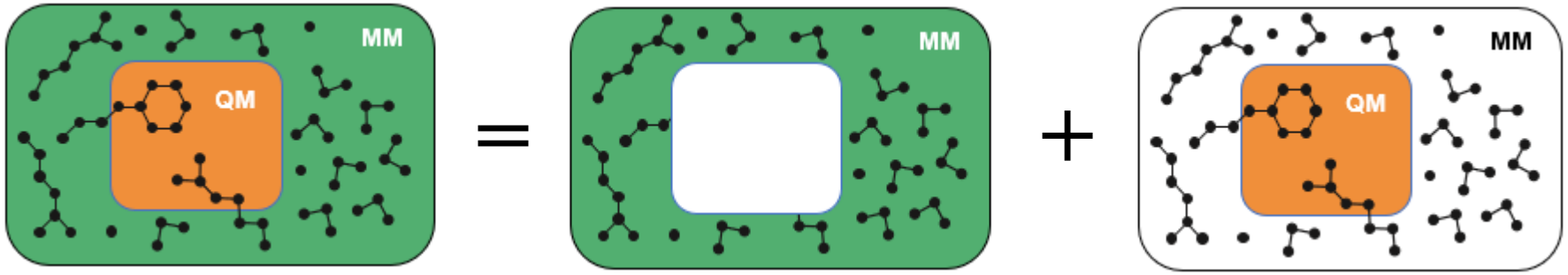
- Interactions between particles in the QM region (straightforward description)
- Interactions between atoms in the MM region (straightforward description)
- Interactions between QM and MM particles

QM and MM Coupling Approaches

- **Mechanical embedding: no influence of MM charges on the QM part,** i.e QM calculation is gas-phase-like without additional potential due to the MM atoms.



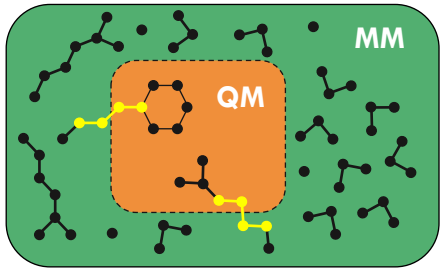
- **Electrostatic embedding: QM polarization due to the MM part included.**



- **Polarized embedding: MM polarization due to the QM part included as well** (non-self consistently or fully self-consistently).

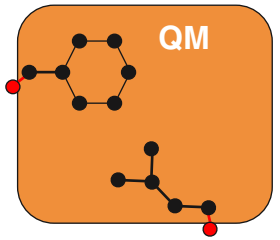
Bonded interactions

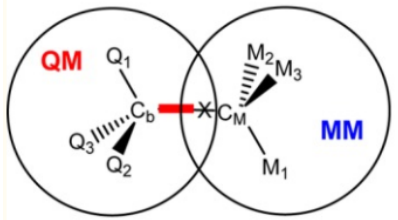
Introduced when covalent bonds connecting the QM and the MM regions are cut:



In addition, in this case care has to be taken when evaluating the QM wave function. Some methods to deal with dangling valences:

- **Monovalent capping Link-atoms:** usually hydrogen atoms at an appropriate position along the bond vector
- **Capping pseudopotentials:** linking atom with scaled down pseudopotential and the required valence charge, which requires to constrain the bond distance appropriately.
- **Generalized hybrid orbitals:** hybrid orbitals are placed on the boundary atom and one of them participates in the SCF calculation for the WF of the QM region.





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**

GROMACS FAST. FLEXIBLE. FREE.

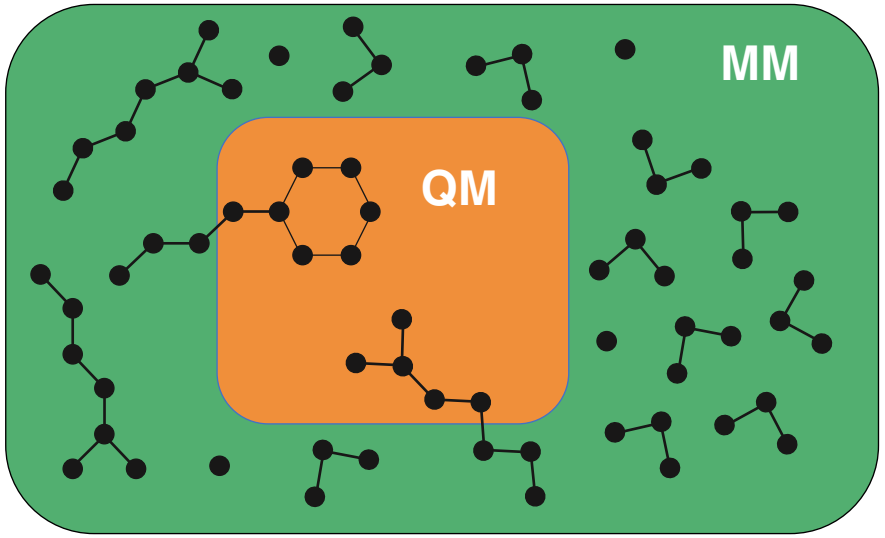
Forcefield: Classical MM-MM interactions both bonded and non-bonded (PME)

+

Integration: Classical MD using fully periodic QM/MM forces

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

Forcefield Quickstep GEEP



Forcefield (MM) - GROMACS

- Force field description of MM region

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

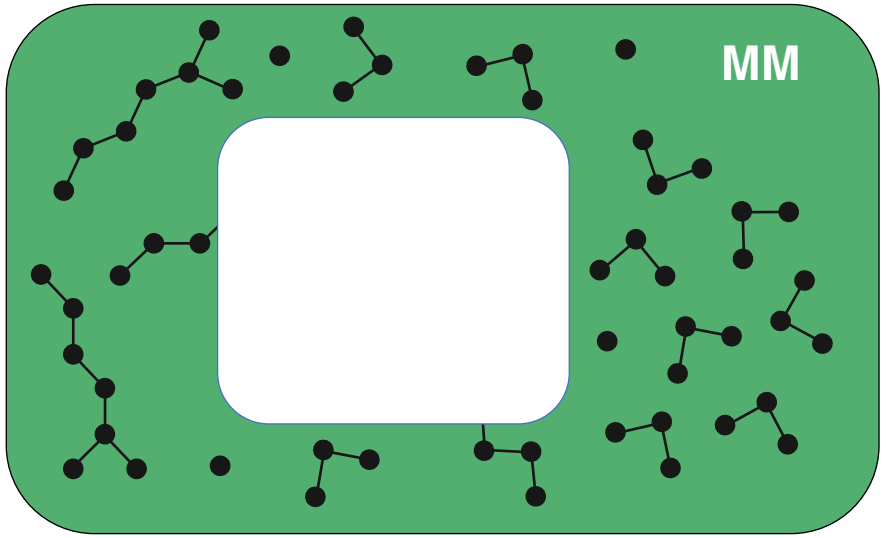
Forcefield Quickstep GEEP

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

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

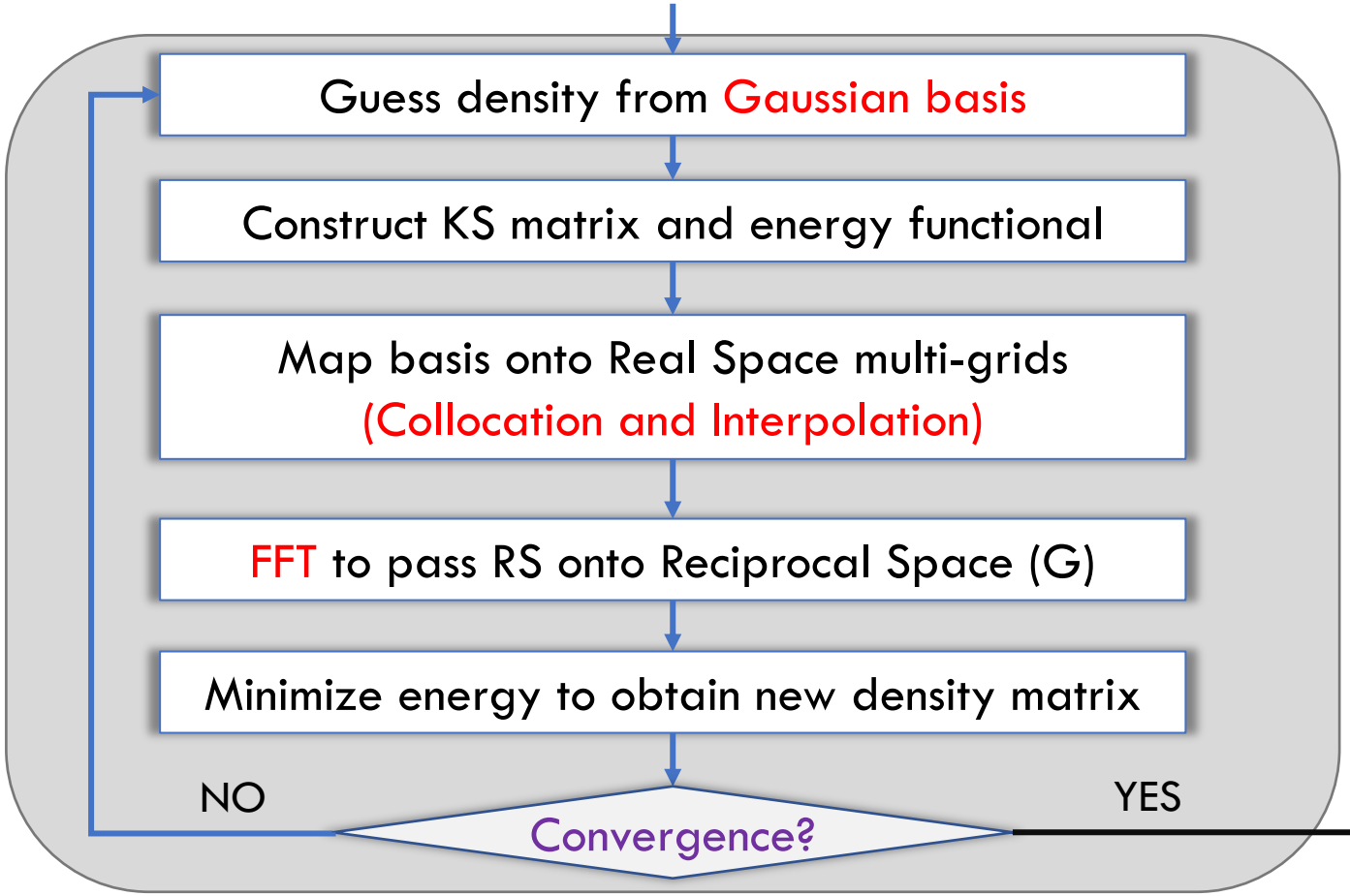
$$+ \sum_{torsions} \frac{1}{2} k_\phi [1 + \cos(n\phi - \phi_0)]$$

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



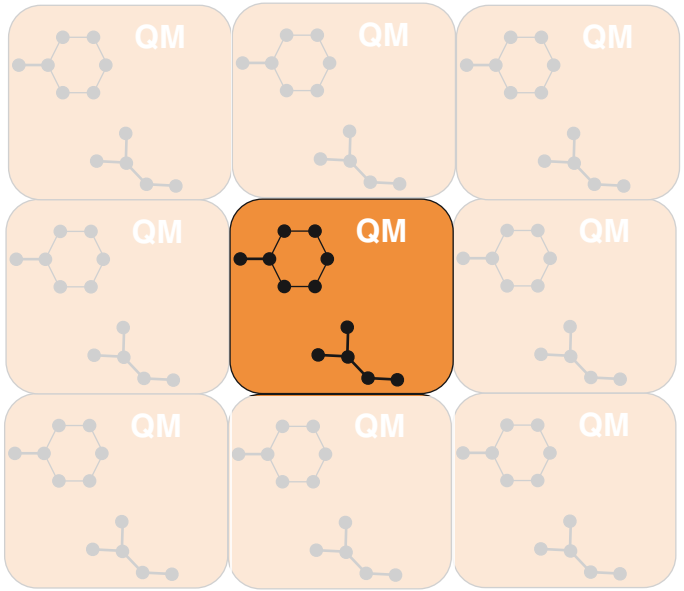
Quickstep (QM) - CP2K

QM region as CP2K input



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

Forcefield Quickstep GEEP



Energy, Forces and other properties

GEEP for QM/MM Coupling - CP2K

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

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

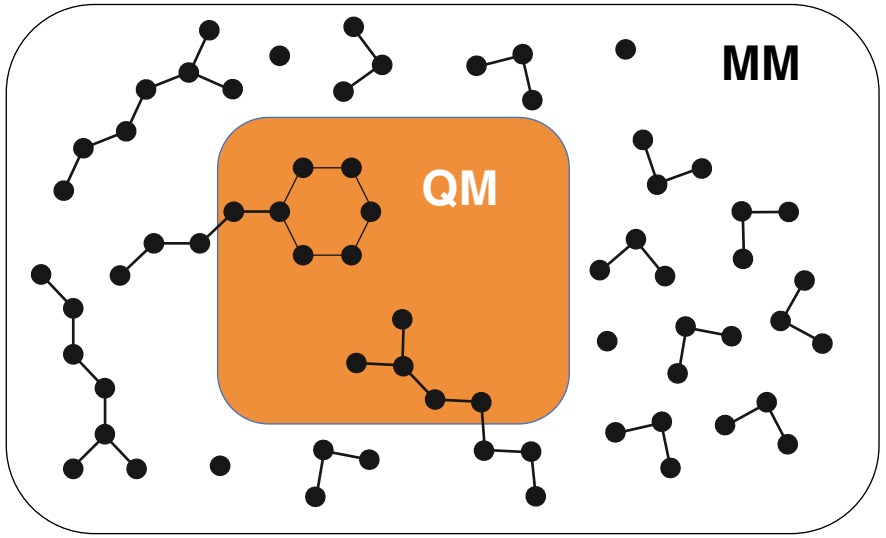
Forcefield Quickstep GEEP

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

$$v_I^{smear}(|\mathbf{r} - \mathbf{R}_I|) = \frac{\mathrm{Erf}\left(\frac{|\mathbf{r} - \mathbf{R}_I|}{\mathbf{r}_{c,I}}\right)}{|\mathbf{r} - \mathbf{R}_I|} = \sum_{N_g} A_g e^{-\left(|\mathbf{r} - \mathbf{R}_I|/G_g\right)^2} + R_{low}(|\mathbf{r} - \mathbf{R}_I|)$$

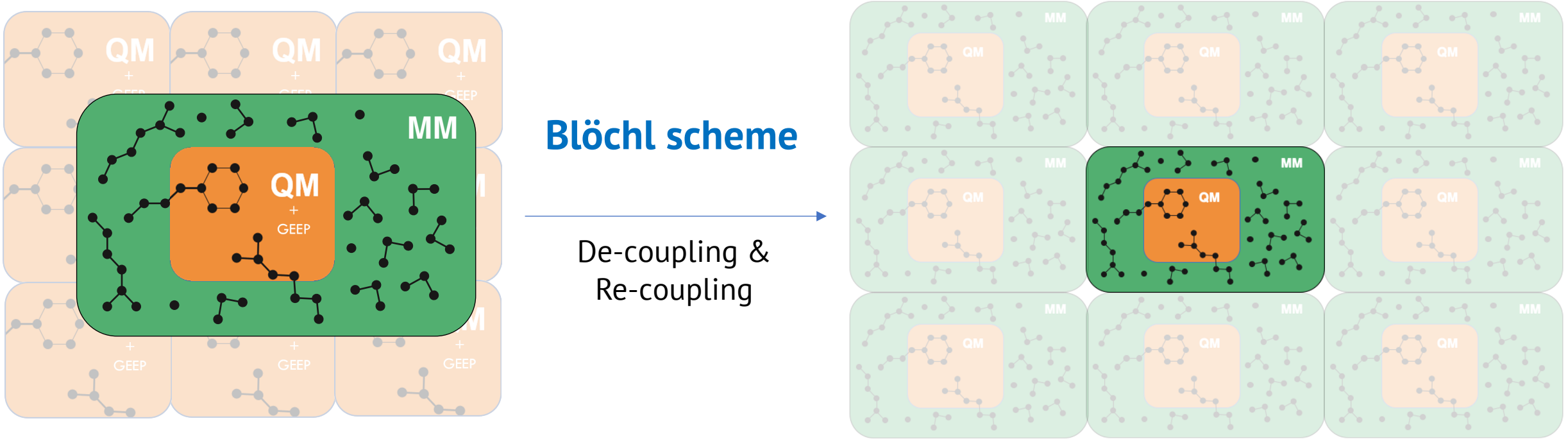
+

Real Space multi-grid approach

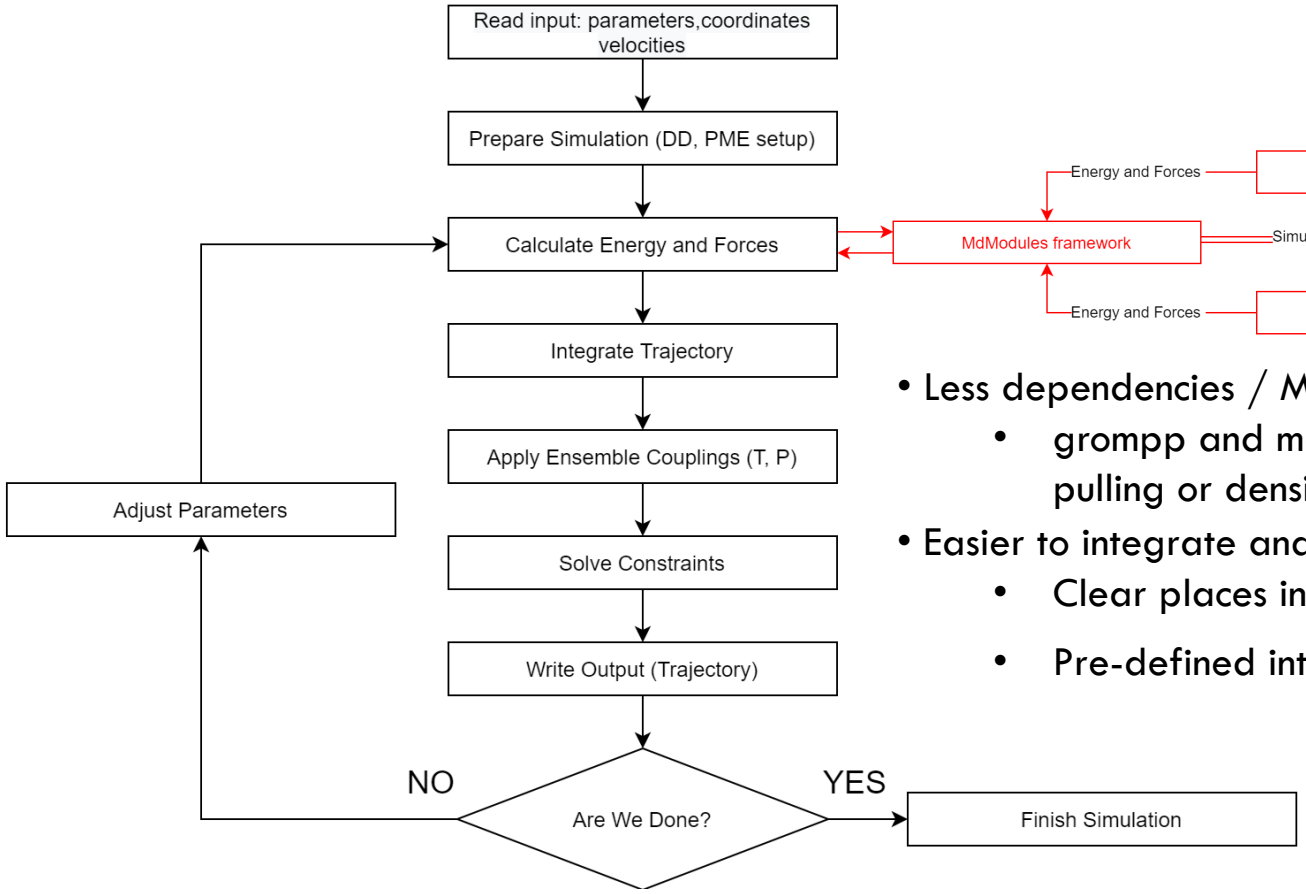


Fully periodic QM/MM

- 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



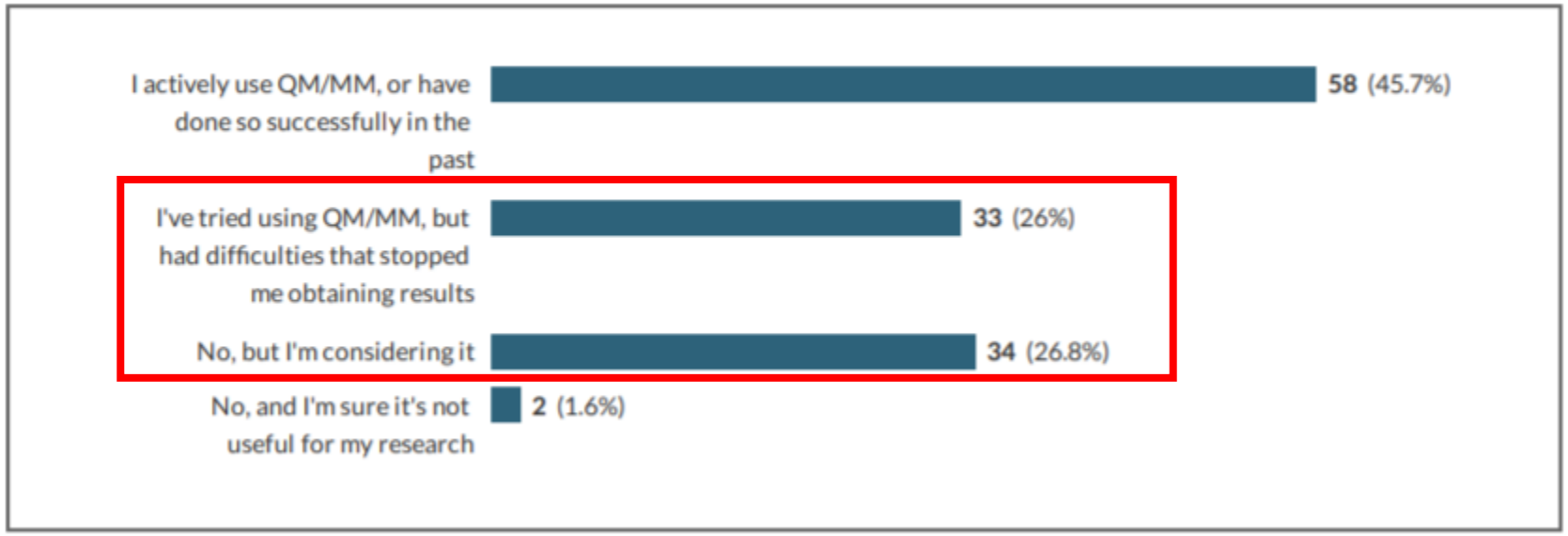
Simulation flow – Gromacs (MdModules)



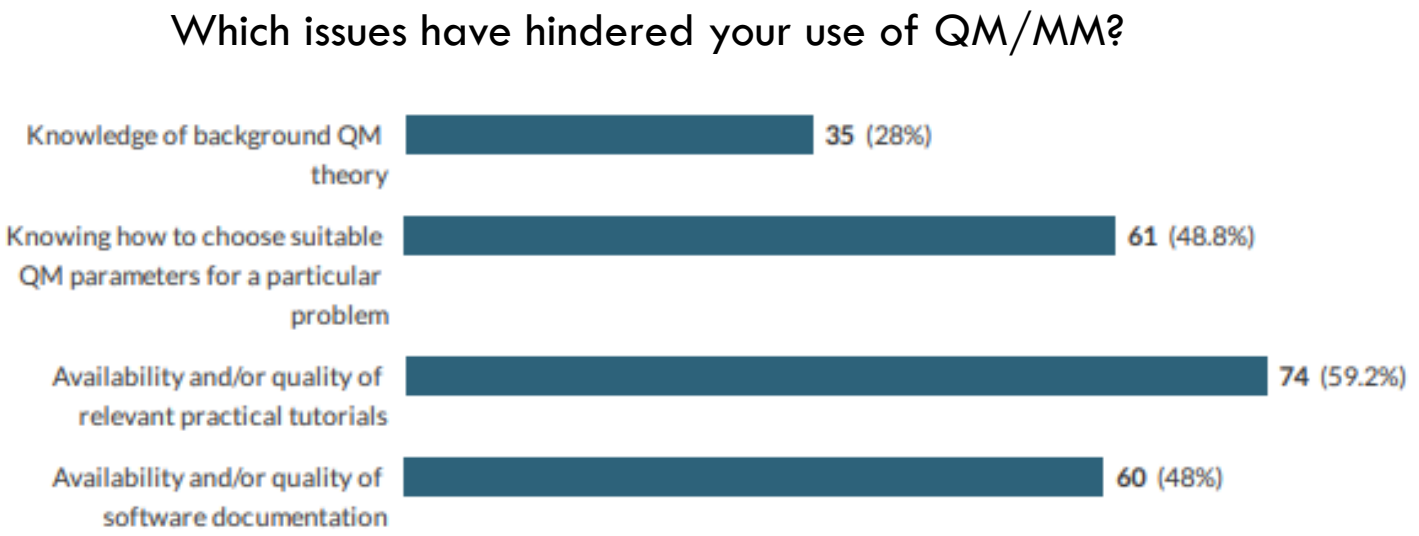
- Less dependencies / More encapsulation
  - grompp and mdrun does not need to know about QM/MM, pulling or density-guided simulations
- Easier to integrate and maintain
  - Clear places interaction of new code with GROMACS
  - Pre-defined interaction pattern with rest of code

Why QM/MM is difficult to use?

Do you use QM/MM simulations in your research?



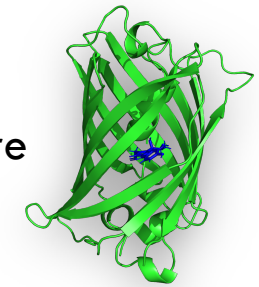
Why QM/MM is difficult to use?

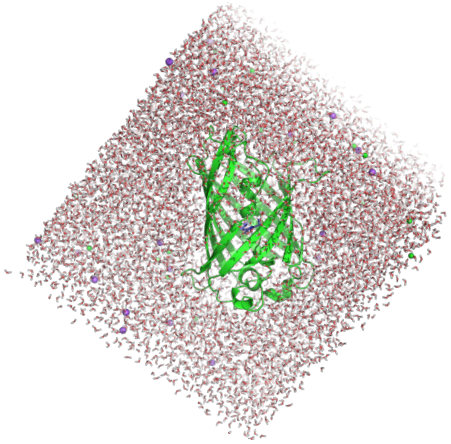


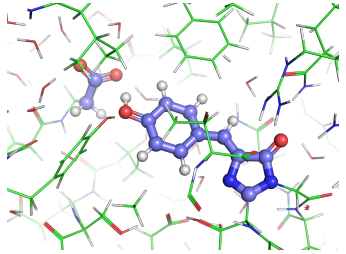
Make it simpler to use!

Typical workflow for biochemical QM/MM modelling

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
- …

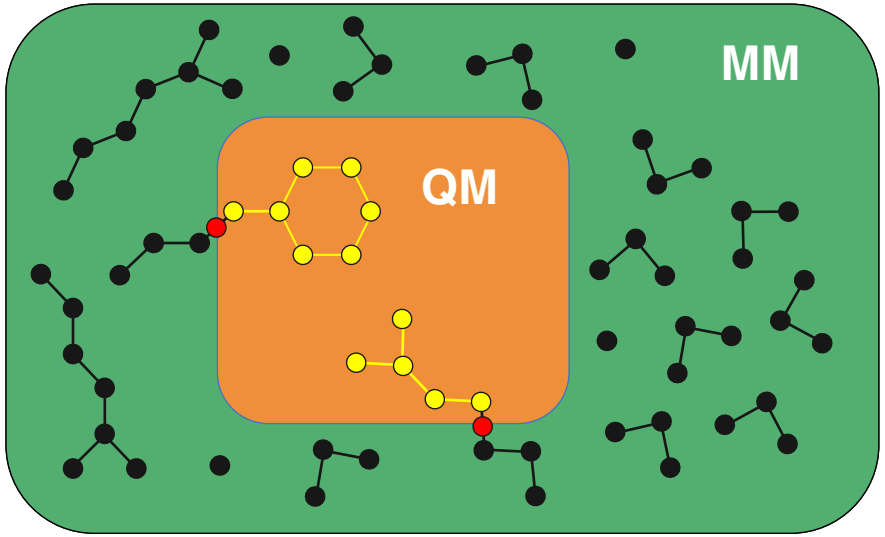
- Index file
- .mdp parameters for QM/MM

- Keep it flexible for advanced users

QM/MM Calculations

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

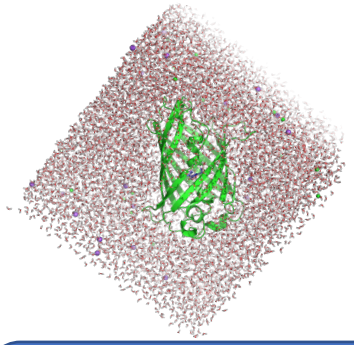
Topology processing and features

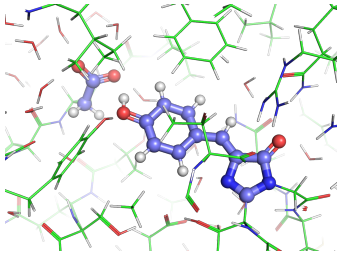


1. Remove partial charges from QM atoms
2. Remove LJ interactions between QM-QM atoms (keep with QM-MM)
3. Remove Bonds (which connects 2 QM atoms)
4. Remove Angles, which contains 2 or more QM atoms
5. Remove Dihedrals, which contains 3 or more QM atoms
6. Cleanup water constraints (SETTLE) from QM waters
7. Generate CP2K Input (including Link atoms)

- Sample 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

Easy to use: simple QM/MM setup

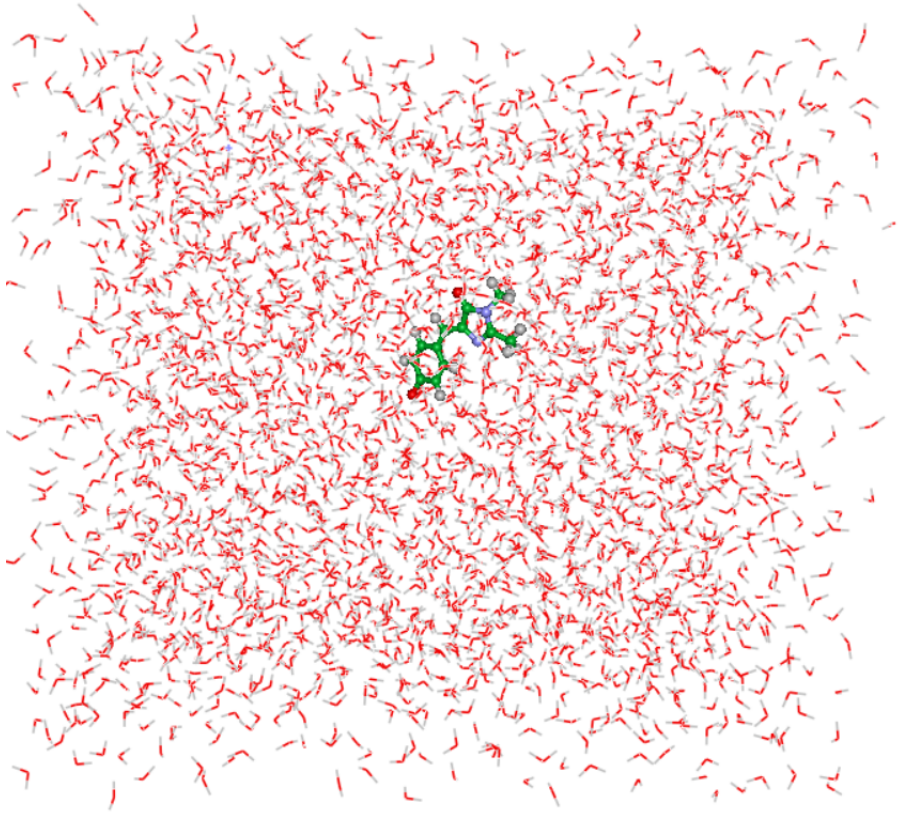




Classical MD Simulations → - Index file - .mdp parameters for QM/MM → - topol.tpr - topol.inp (CP2K Input) - topol.pdb (QM-MM split and point charges) → **QM/MM Calculations**

```
qmmm-active          = true      ; Activates QMMM MdModule
qmmm-qmgroup         = QMatoms   ; Index group with QM atoms
qmmm-qmmethod        = PBE       ; Pre-set parameters for DFT (PBE/DZVP)
qmmm-qmcharge        = 0         ; Total Charge of QM system
qmmm-qmmultiplicity  = 1         ; Spin-state of QM system
```

Example system



QM subsystem : HBDI (GFP Chromophore)

MM subsystem : 3237 TIP3P waters and Na^+ Ion

QM charge: -1

QM multiplicity: 1

Functional: PBE

Basis: DZVP-MOLOPT

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

CP2K: Generated Input File

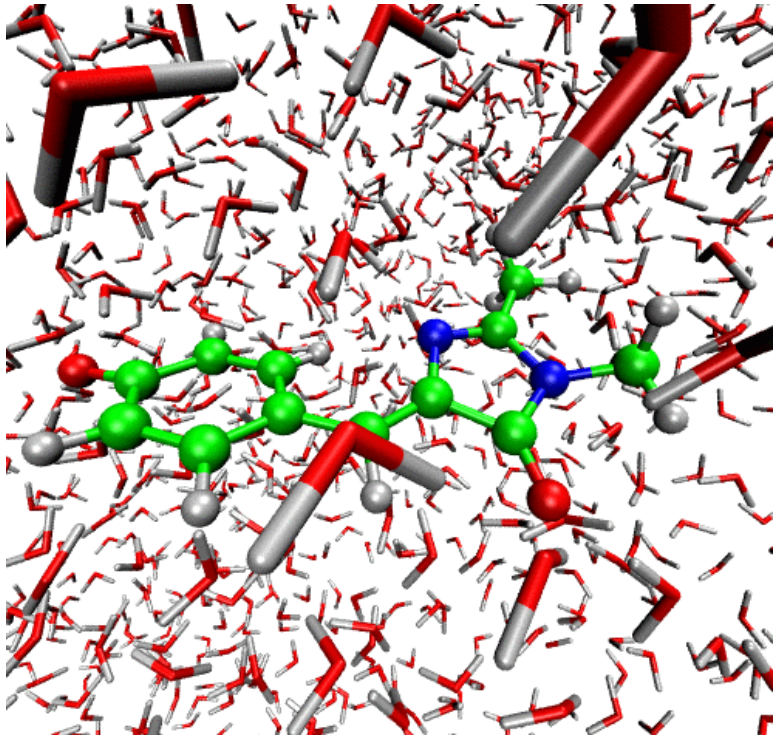
```
&GLOBAL
   PRINT_LEVEL LOW
   PROJECT GROMACS
   RUN_TYPE ENERGY_FORCE
&END GLOBAL
&FORCE_EVAL                    ! parameters for force evaluation
  METHOD QMMM                  ! method employed e.g. QMMM (Quickstep + GEEP)
  &DFT                         ! Grids setup, SCF and Functional parameters
    .... contents of DFT section
  &END DFT
  &QMMM                        ! Set up for QM box, GEEP, Blöchl scheme, Link atoms
    .... contents of QMMM section
  &END QMMM
  &MM                          ! Needed to switch off MM-MM interactions
    .... contents of MM section
  &END MM
  &SUBSYS                      ! Coordinates, atom kinds (Gaussian basis sets)
    .... contents of SUBSYS section
  &SUBSYS
&END FORCE_EVAL
```

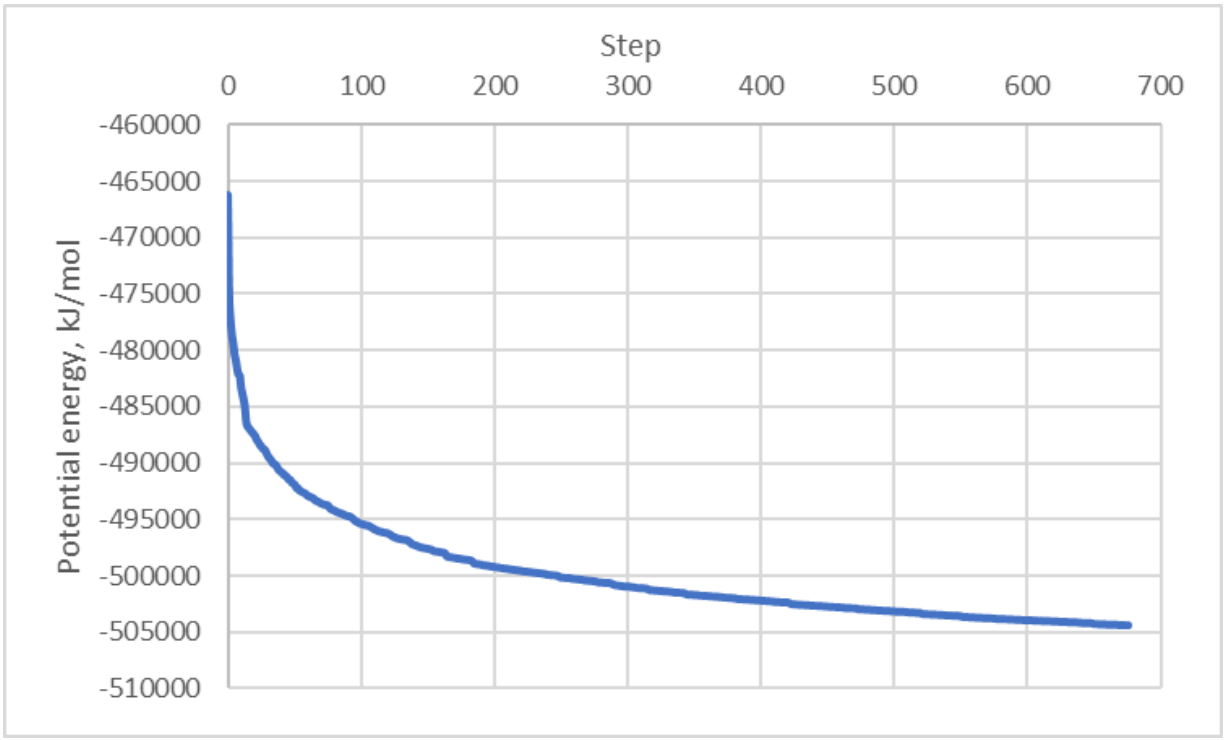
CP2K: Point Charges

```
....
ATOM     24  H   QM     1      22.990  26.200  25.510  1.00  0.00        H   0.000000
ATOM     25  C   QM     1      24.150  24.930  26.660  1.00  0.00        C   0.000000
ATOM     26  O   QM     1      23.620  23.820  26.190  1.00  0.00        O   0.000000
ATOM     27  O   MM     2      23.740  30.770  14.650  1.00  0.00        O  -0.820000
ATOM     28  H   MM     2      24.150  29.860  14.630  1.00  0.00        H   0.410000
ATOM     29  H   MM     2      24.070  31.260  15.460  1.00  0.00        H   0.410000
ATOM     30  O   MM     2      16.410  39.470  20.500  1.00  0.00        O  -0.820000
ATOM     31  H   MM     2      17.320  39.650  20.870  1.00  0.00        H   0.410000
ATOM     32  H   MM     2      16.140  40.230  19.900  1.00  0.00        H   0.410000
ATOM     33  O   MM     2      16.020  28.300  22.750  1.00  0.00        O  -0.820000
ATOM     34  H   MM     2      16.010  28.580  23.700  1.00  0.00        H   0.410000
ATOM     35  H   MM     2      15.130  27.920  22.490  1.00  0.00        H   0.410000
....
```

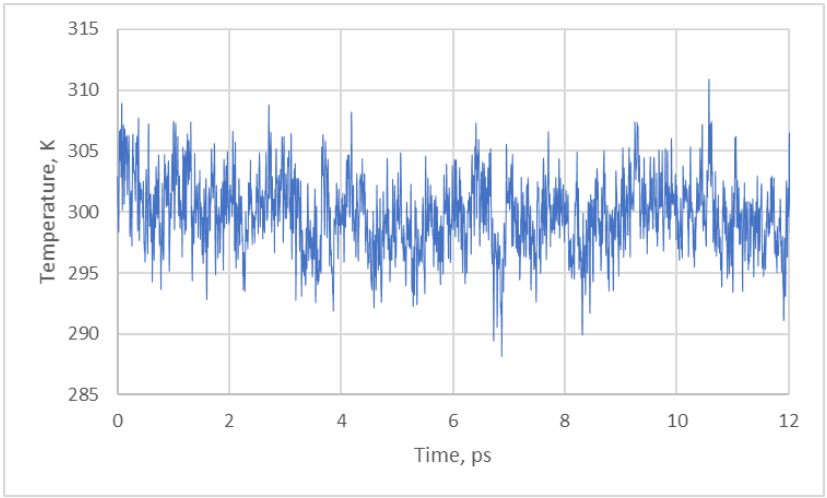
TIP3P water

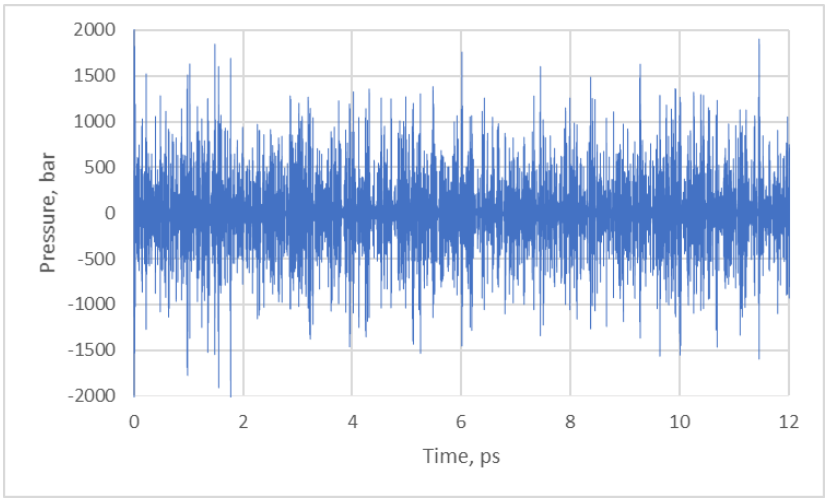
Energy minimization





Dynamics with Pressure and Temperature coupling (NPT)

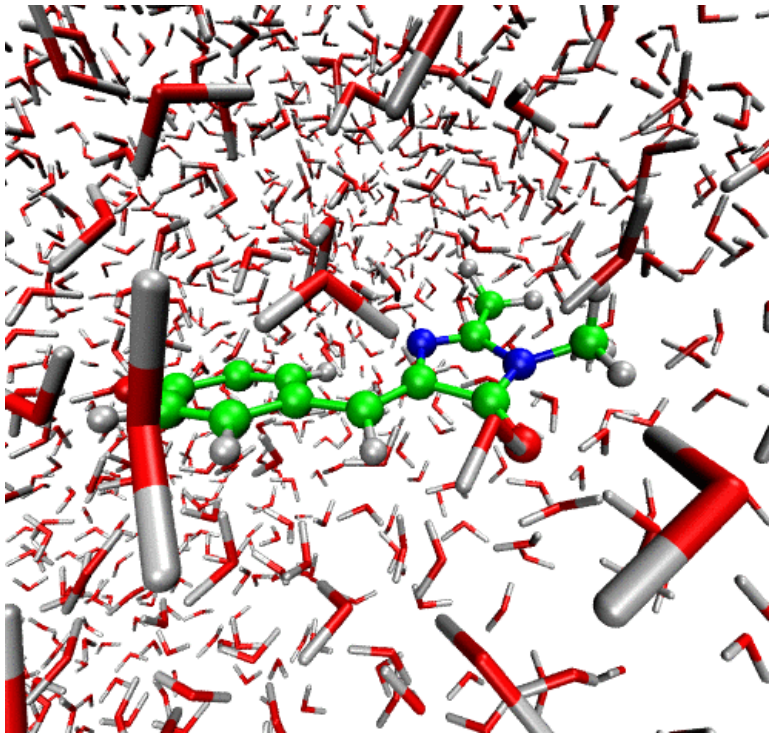




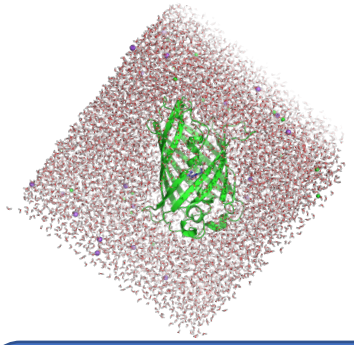
Thermostat: V-rescale 300K, 0.1 ps

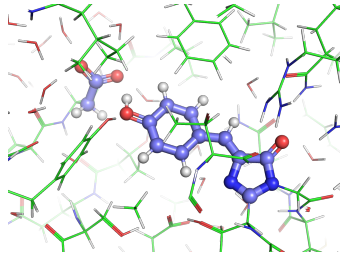
Barostat: Berendsen 1 bar, 0.1 ps

Dynamics: 1 fs time-step, approximately 10 ps/day



Flexible input: advanced QM/MM setup





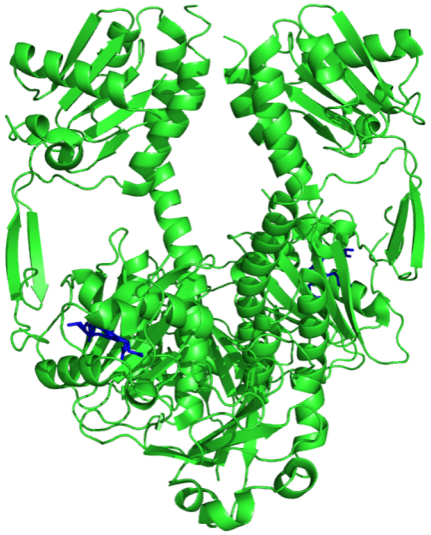
Classical MD Simulations

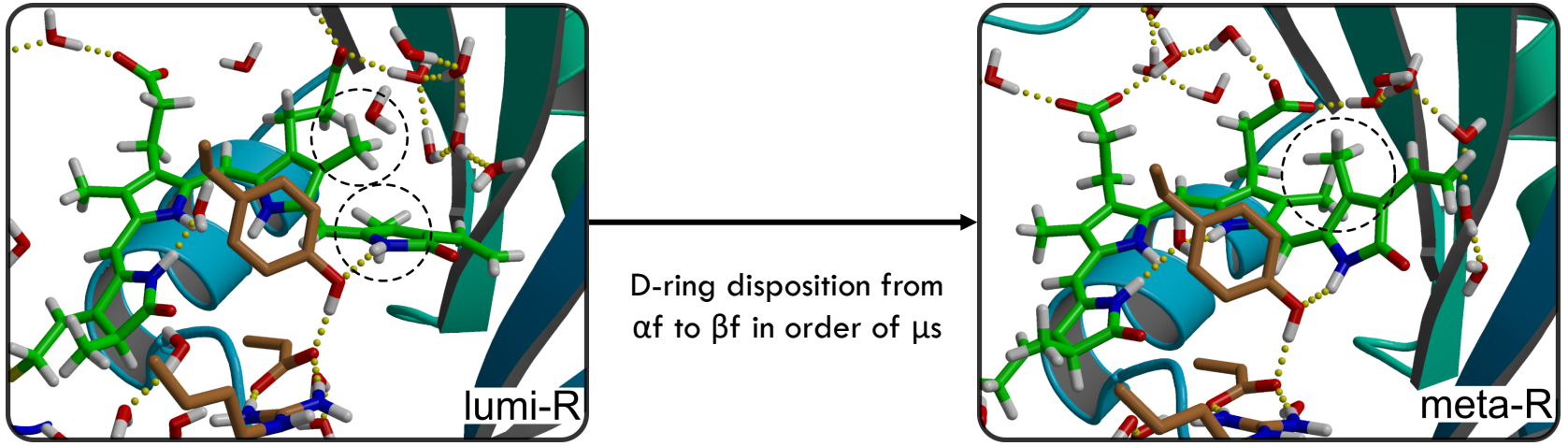
- Index file
- .mdp parameters for QM/MM
- Your own CP2K input

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
```

Protein simulations: Phytochrome





Objective:

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

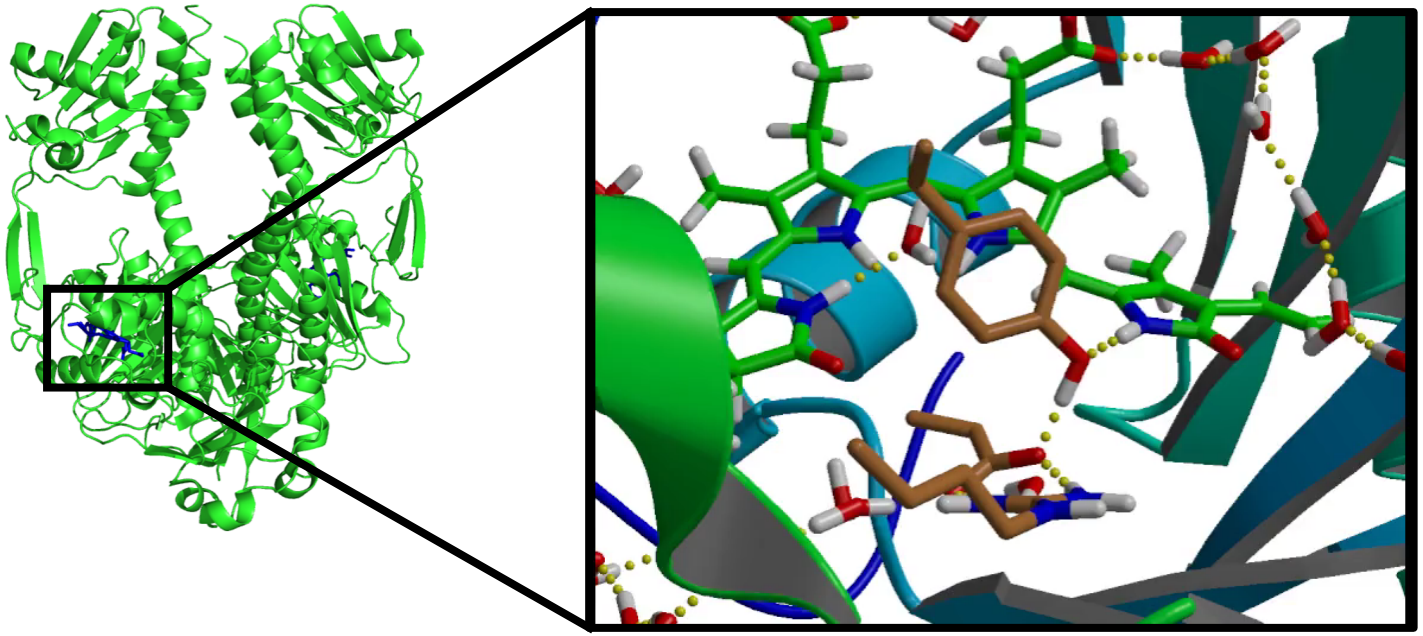
Free energy with Umbrella sampling simulations

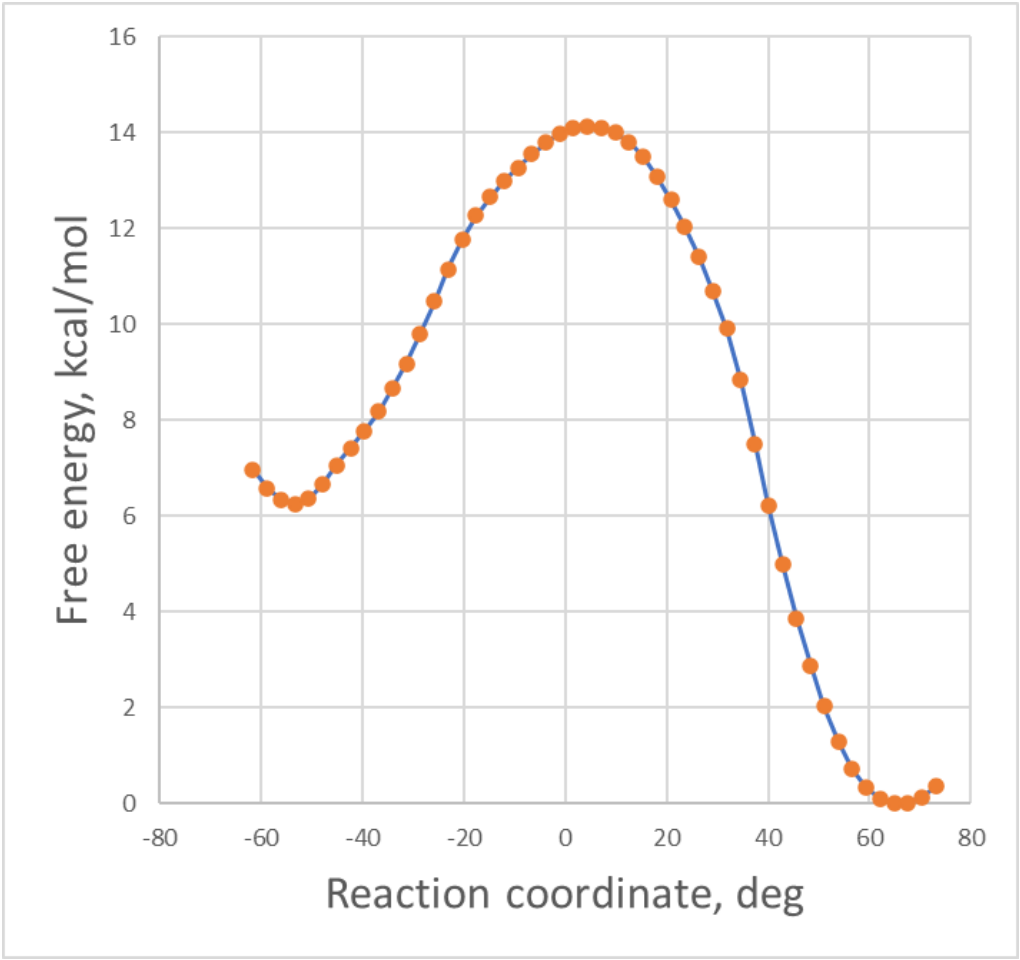
QM part - Chromophore

QM method - PBE/DZVP-MOLOPT-GTH

MM Forcefield - Amber03

Protein simulations: umbrella sampling





CP2K: Advanced Input Parameters

For further information:

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

1) Best practice guides
2) Explanations on parameters and combinations
3) Simulations of biological systems with CP2K
4) Understanding of CP2K output
5) Troubleshooting

Acknowledgments

Prof. Gerrit Groenhof (JYU)

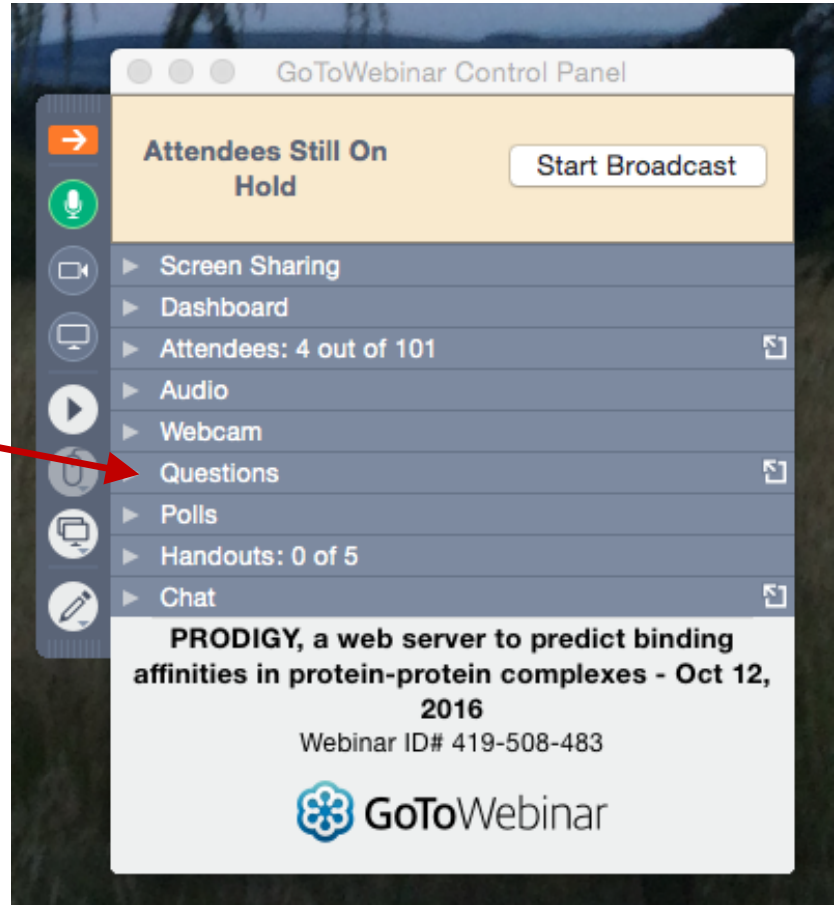
Dr. Christian Blau (KTH, Gromacs Team)

M.Sc. Vaibhav Modi (JYU)
Dr. Arno Proeme (UEDIN)
Dr. Holly Judge (UEDIN)
Dr. Emiliano Ippoliti (Jülich)

Audience Q&A session

- Please use the Questions function in GoToWebinar application
  - If you *don't have audio*, please mention that in the question.

- Any other questions or points to discuss after the live webinar? Join the discussions at http://ask.bioexcel.eu.



Next webinar occasions

See https://bioexcel.eu/

18 January 2021

Student Webinar: Winter School 2020 Edition

Starting with February 2021

Spring BioExcel Webinars

- Paul Bauer/Berk Hess – Feature of GROMACS 2021
- Magnus Lundborg/Berk Hess -Solvation free energy and Advance Weight Histogram
- Joe Jordan - NB-library

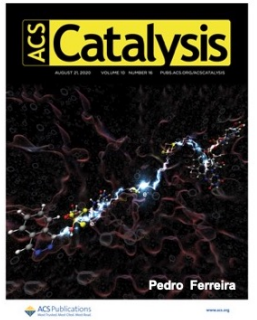
QM/MM Best Practice workshop webinar: Studies on enzyme-catalysed reactions

Prof. Maria João Ramos

University of Porto

10/12/2020

https://bioexcel.eu/qmmm-best-practice-workshop-4



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.