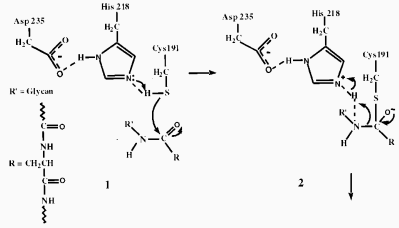
Day 2 Period 2:
Biomolecular potential energy functions

Soohaeng Yoo Willow
syoo7@iit.edu
6/29/2021

This period will consist of

1. molecular mechanics (MM),
2. quantum mechanics (QM),
3. and mixed quantum mechanics/molecular mechanics (QM/MM) force fields.
4. walk-through of a python script to estimate MM, QM, and QM/MM potential energy and gradients.



Introduction

> When our projects are modeling reactions in biomolecular systems as follows: charge transfer, electronic excitation, and chemical reactions,

> Quantum-mechanics/molecular-mechanics (QM/MM) methods becomes our choice for solving above problems.

- A QM method is used for the chemical active region (e.g. substrates and co-factors in an enzymatic reaction).
- A MM method is for the surroundings (e.g. protein and solvent).
- Senn and Thiel, *Angw. Chem. Int. Ed.* **2009**, *48*, 1198.

- The potential energy E_{sys} of the system consisting of the QM and MM regions is

$$E_{\mathrm{sys}} = E_{\mathrm{QM}} + E_{\mathrm{MM}} + E_{\mathrm{QM/MM}}$$

- E_{QM} : the potential energy of the QM region
- E_{MM} : the potential energy of the MM region
- $E_{\mathrm{QM/MM}}$: the potential energy between QM and MM regions.

Molecular Mechanics Force Field

- The **force field** refers to the *functional form* and *parameter* sets used to calculate the potential energy of a system.
- Intra-molecular potential energy:

$$
\begin{aligned}
E_{\mathrm{intra}} &= E_{\mathrm{bond}} + E_{\mathrm{angle}} + E_{\mathrm{torsion}} \\
E_{\mathrm{bond}} &= \frac{k}{2}(r - r_0)^2
\end{aligned}
$$

- Inter-molecular potential energy:

$$
\begin{aligned}
E_{\mathrm{inter}} &= E_{\mathrm{LJ}} + E_{\mathrm{Coul}}[+E_{\mathrm{pol}}] \\
E_{\mathrm{LJ}} &= \sum_{A>B} 4\epsilon_{AB}\left[\left(\frac{\sigma_{AB}}{R_{AB}}\right)^{12} - \left(\frac{\sigma_{AB}}{R_{AB}}\right)^{6}\right], \\
E_{\mathrm{Coul}} &= \sum_{A>B} \frac{q_A\, q_B}{R_{AB}}
\end{aligned}
$$

- Parameter sets represent k, r_0, ϵ_{AB}, σ_{AB}, q_A, and so on.

Molecular Mechanics Force Field Parameter Sets

- AMBER
- CHARMM
- MMFF
- OPLS
- AMOEBA

Topology

- lists of chemical bonds, angles, torsional angles, *et al*

- Using OpenMM and AMBER force fields, we will estimate the potential energy and gradients of a MM system.

	“Effective” MM	QM–based MM	QM
Wave function	No	No	$\Psi(\boldsymbol{r};\boldsymbol{R})$
Energy	$E = E(\boldsymbol{R})$	$E = E(\boldsymbol{R})$	$E = \langle\Psi\vert\hat{H}\vert\Psi\rangle$
PES[a]	“Effective”	BO[b]	BO
ZPE[c]	Implicit	PIMD[d]	PIMD
Example	AMBER, CHARMM	AMOEBA	DFT, MP2, FMO, BIM

[a] PES: Potential Energy Surface

[b] BO: Born–Oppenheimer

[c] ZPE: Zero–point energy

[d] PIMD: Path–integral molecular dynamics

Quantum Mechanics: The Born-Oppenheimer Approximation

- Assumption: the density of electrons in a molecule is obtained in the field of fixed nuclei.(from Szabo and Ostluand, **Modern Quantum Chemistry**)

$$
\begin{aligned}
\hat{H}_{\mathrm{elec}} &= -\sum_{i=1}^{N} \frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{|\boldsymbol{r}_i - \boldsymbol{R}_A|} + \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \\
E_{\mathrm{elec}} &= \langle \Psi_{\mathrm{elec}} | \hat{H}_{\mathrm{elec}} | \Psi_{\mathrm{elec}} \rangle \\
E_{\mathrm{QM}} &= E_{\mathrm{elec}} + \sum_{A=1}^{M}\sum_{B>A}^{M} \frac{1}{|\boldsymbol{R}_A - \boldsymbol{R}_B|}
\end{aligned}
$$

- Born's Statistical Interpretation of $\Psi_{\mathrm{elec}}(x)$: $|\Psi_{\mathrm{elec}}(x)|^2$ is a *probability density* for existence of an electron at position x.
- Using PySCF, we will estimate E_{QM} at HF, MP2, and CCSD(T).
- S. Obara and A. Saika, "Efficient recursive computation of molecular integrals over Cartesian Gaussian functions," *J. Chem. Phys.* **84**, 3963 (1986)

(QM) Fragment-Based Potential Energy

- Intra-molecular potential energy of the Ith monomer:

$$E_I - E_I^{\min} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}},$$

  where $E_I = \langle \Psi_I | \hat{H}_I | \Psi_I \rangle$ represents the potential energy of Ith monomer.

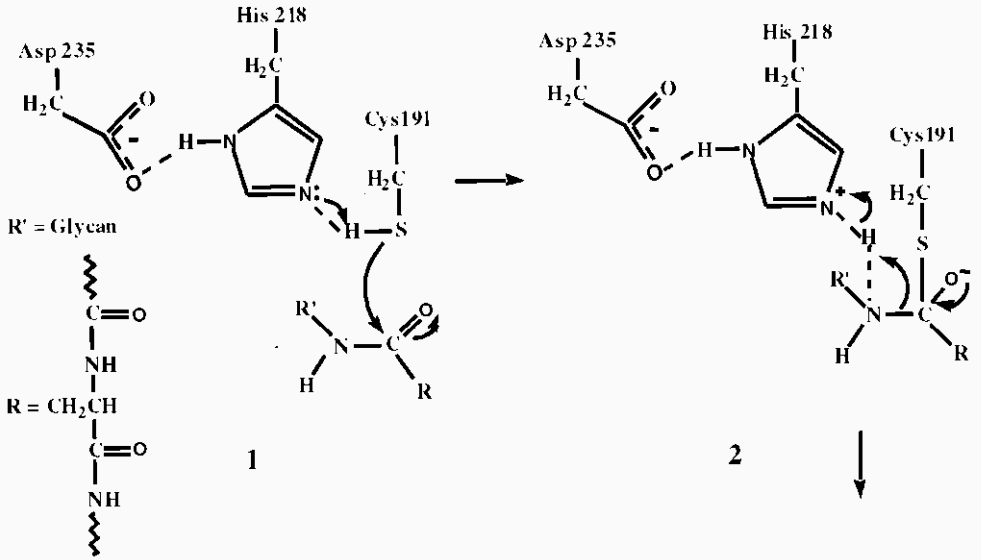
- Inter-molecular potential energy:

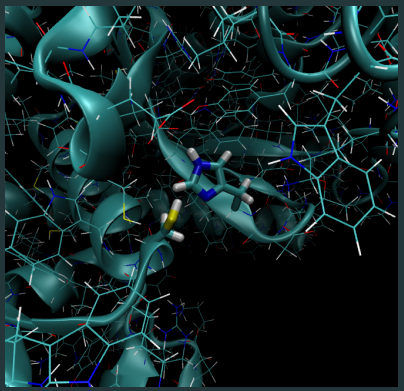
$$E_{IJ} - E_I - E_J = E_{\text{LJ}} + E_{\text{Coul}},$$

  where $E_{IJ} = \langle \Psi_{IJ} | \hat{H}_{IJ} | \Psi_{IJ} \rangle$ represents the potential energy of the dimer consisting of Ith and Jth monomers.

- Using the second-order many-body expansion, the potential energy of the system becomes

$$E_{\text{MBE(2)}} = \sum_I \{E_I - E_I^{\min}\} + \sum_{I>J}^{R_{IJ}<R_{\text{cut}}} \{E_{IJ} - E_I - E_J\}$$

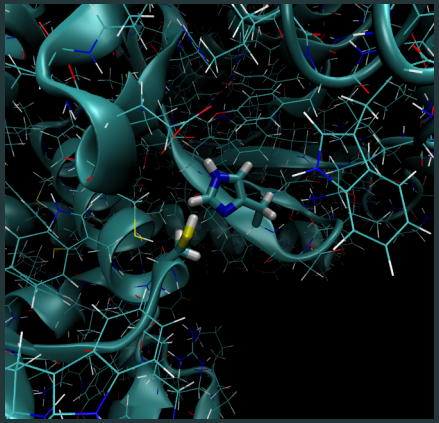

Asp 235
His 218
Cys191
R' = Glycan
R = CH2CH
1
2



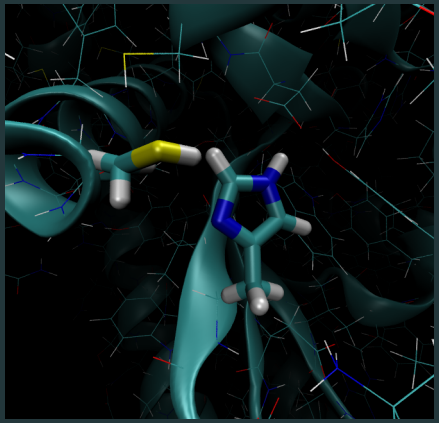
QM/MM methods for biomolecular systems

- Problem: There are covalent linkages between QM atoms and MM atoms.
- Consider that the QM region is connected to the MM region via one covalent bond: CB-CA.
- The ‘CB’ atom belongs to the QM region and the ‘CA’ atom to the MM region.
- The link-atom (L) method: the free valency is capped by an additional atom (H): CB-CA –> CB-H.

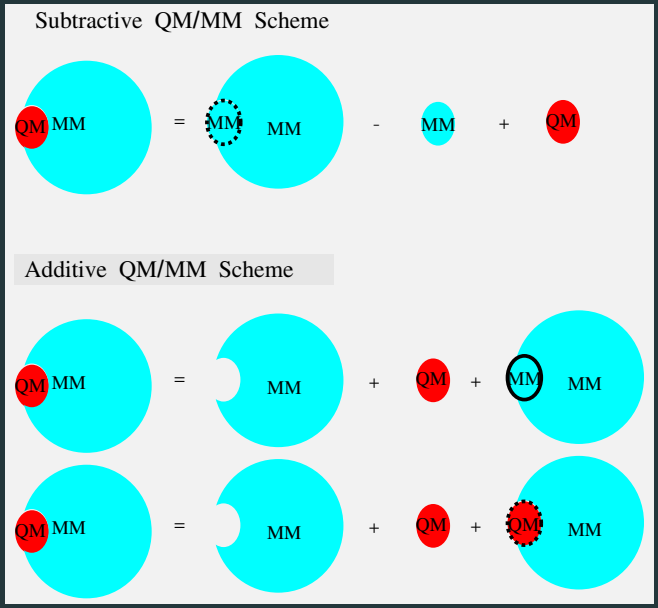
Before adding link-atoms.



After adding link-atoms.



QM/MM Schemes



Subtractive QM/MM Scheme: ONION method in Gaussian

- $$E_{\mathrm{sys}} = E_{\mathrm{MM}}(L+R) - E_{\mathrm{MM}}(L) + E_{\mathrm{QM}}(L)$$
- L is a ligand in the QM region and R is a receptor (protein) in the MM region

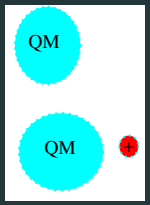
Additive QM/MM Scheme

- $$E_{\mathrm{sys}} = E_{\mathrm{MM}}(R) + E_{\mathrm{QM}}(L) + E_{\mathrm{QM/MM}}$$
- There is an explicit QM-MM coupling term $E_{\mathrm{QM/MM}}$.
- $E_{\mathrm{QM/MM}}$ collects the interaction between QM and MM regions.

$$E_{\mathrm{QM/MM}} = E^{\mathrm{bonded}}_{\mathrm{QM/MM}} + E^{\mathrm{vdW}}_{\mathrm{QM/MM}} + E^{\mathrm{elec}}_{\mathrm{QM/MM}}$$

Nonbonded and Bonded QM-MM interactions

- Nonbonded QM-MM interactions ($E^{\mathrm{vdW}}_{\mathrm{QM/MM}}$): use Lennard-Jones potential.
- Bonded QM-MM interactions ($E^{\mathrm{bonded}}_{\mathrm{QM/MM}}$): (bond stretching, angle bendding, torsional, etc) use the standard MM parameter set.



Electrostatic interaction between QM and MM regions

> Electron density (highlighted with blue color) in the QM region is perturbed when the MM charges are provided.

> This perturbation increases as the MM charges are closer to the QM region.

Electrostatic QM-MM interaction: $E^{\text{elec}}_{\text{QM/MM}}$

1. mechanical embedding: use the MM-MM electrostatics.

$$E^{\text{elec}}_{\text{QM/MM}} = E^{\text{Coul}}_{\text{QM/MM}} = \sum_{A \in \text{QM}} \sum_{B \in \text{MM}} \frac{q_A q_B}{|\boldsymbol{R}_A - \boldsymbol{R}_B|}$$

2. electrostatic embedding: perform the QM calculation as follows:

$$\begin{aligned} E^{\text{elec}}_{\text{QM/MM}} &= \langle \Psi_{I:Q_I} | \hat{H}_{I:Q_I} | \Psi_{I:Q_I} \rangle - \langle \Psi_I | \hat{H}_I | \Psi_I \rangle \\ &= E^{\text{Coul}}_{I:Q_I} + E^{\text{pol}}_{I:Q_I} \end{aligned}$$

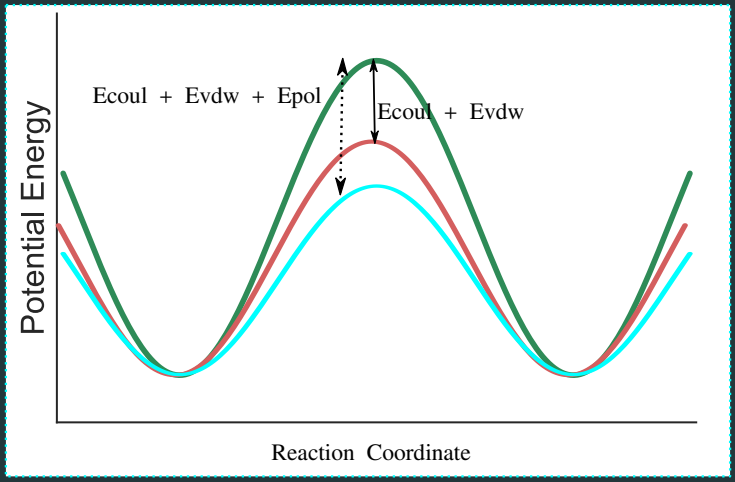
- The electronic structure of the QM region depends on the charge distribution of the MM region and is polarized by it.
- Special care: the MM charges are placed in immediate proximity to the QM electron density at the QM-MM boundary, which give overpolarization.

Polarization Energy: $E^{\mathrm{pol}}_{I:Q_I}$ **or** $\Xi^{\mathrm{pol}}_{I:Q_I}$

$$
\begin{aligned}
\Xi^{\mathrm{pol}}_{I:Q_I} &= \langle \Psi_{I:Q_I} | \hat{H}_{I:Q_I} | \Psi_{I:Q_I} \rangle - \langle \Psi_I | \hat{H}_{I:Q_I} | \Psi_I \rangle \\
&= \langle \Psi_{I:Q_I} | \hat{H}_{I:Q_I} | \Psi_{I:Q_I} \rangle - \langle \Psi_I | \hat{H}_I | \Psi_I \rangle - \langle \Psi_I | \hat{H}_{[I:Q_I]} | \Psi_I \rangle \\
&= \langle \Psi_{I:Q_I} | \hat{H}_{I:Q_I} | \Psi_{I:Q_I} \rangle - E_I - E^{\mathrm{Coul}}_{I:Q_I} \\
\hat{H}_{I:Q_I} &= \hat{H}_I + \hat{H}_{[I:Q_I]} \\
\hat{H}_{[I:Q_I]} &= \sum_{A \in I} \sum_{B \in Q_I} \frac{Z_A q_B}{|\boldsymbol{R}_A - \boldsymbol{R}_B|} - \sum_{i \in I} \sum_{B \in Q_I} \frac{q_B}{|\boldsymbol{r}_i - \boldsymbol{R}_B|}
\end{aligned}
$$

- I represents a molecule in the QM region
- Q_I represents a set of MM charges.

Potential Surface Surface in a Chemical Reaction

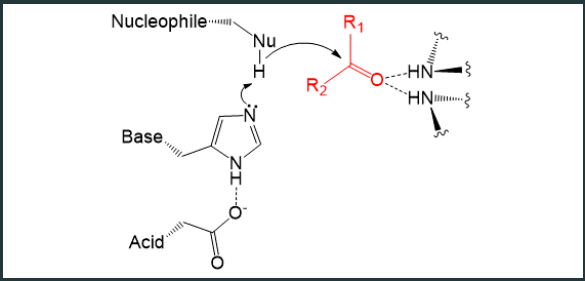


- Green line: QM
- Red line: QM/MM with the MM interaction between QM and MM regions
- Blue line: QM/MM with the polarization energy.

Questions? or Exercise

Catalytic triad

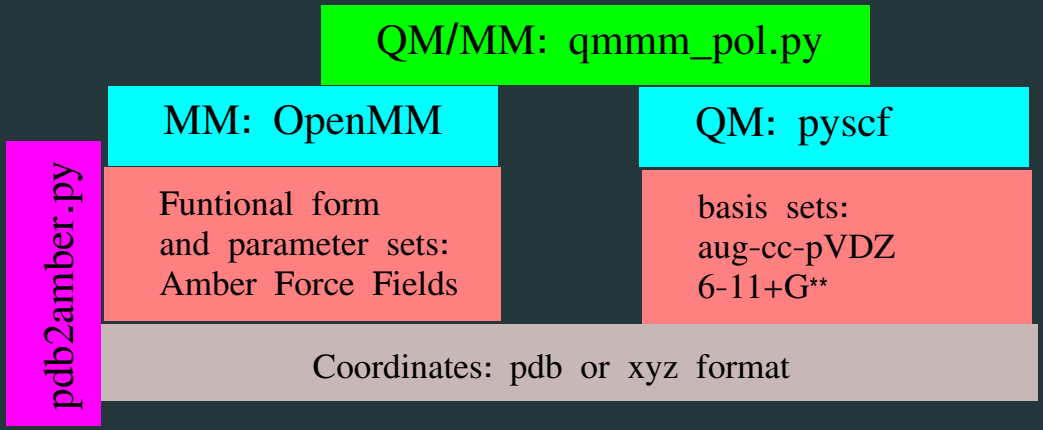
- An Acid-Base-Nucleophile triad is a motif.
- Acid residues: aspartate, glutamate
- Base residue: histidine
- Nucleophile residues: serine (-OH) or cysteine (-SH)



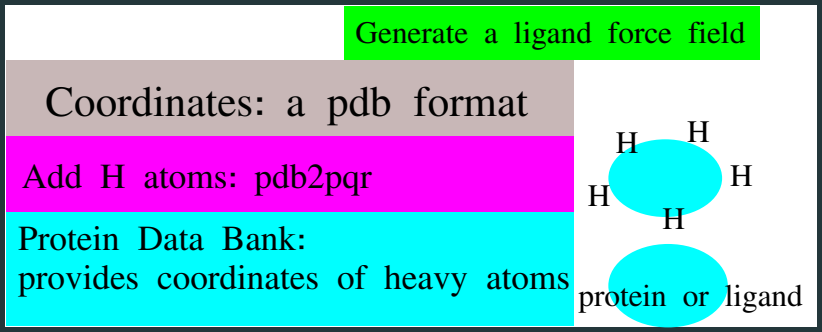
Tobacco etch virus (TEV) protease

- J. Biol. Chem. Vol. 277 pp.50564 (2002).
- The catalytic triad residues His:46, ASP:81, and CYS:151
- The cleavage site: the peptide bond between GLN (Q) and SER (S).
- PDB: 1LVM and 1LVB

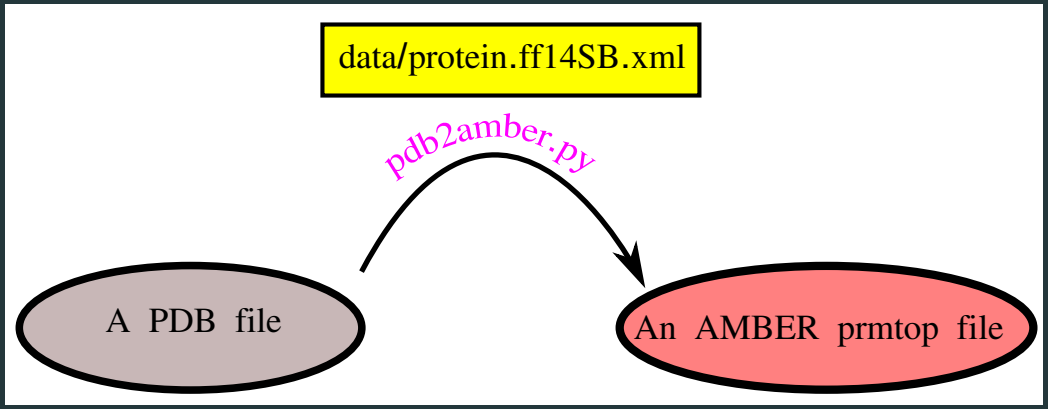
A Bird's Eye View



A Bird's Eye View



A Bird's Eye View



Prerequisites: Install the dependencies

- python 3.7
- (MM) Install OpenMM

```
$ conda install -c conda-forge openmm
or
$ conda install -c omnia openmm
If CUDA is supported, you can install openmm as follows:
$ conda install -c omnia/label/cuda101 openmm
```

- (Adding H atoms) Install pdb2pqr

```
pip install pdb2pqr
```

- (QM) Install pyscf

```
pip install pyscf
```

- (IO of binary data) install h5py

```
pip install h5py==2.10.0
```

- (GeomOpt) Install pyberny

```
pip install pyberny
```

Prerequisites: Exercise

- https://github.com/swillow/pdb2amber
- https://github.com/swillow/modelingworkshop

MM_Step 2. Add H atoms to a protein

- using pdb2pqr: pip install pdb2pqr

```
pdb2pqr30 --ff AMBER --with-ph=7 --nodebump
--ffout=AMBER abc.pdb abc.pqr
```

- from Documentation of the PDB2PQR software The use of continuum solvation methods such as APBS requires accurate and complete structural data as well as force field parameters such as atomic charges and radii. Unfortunately, the limiting step in continuum electrostatics calculations is often the addition of missing atomic coordinates to molecular structures from the Protein Data Bank and the assignment of parameters to these structures. To address this problem, we have developed PDB2PQR.

Generate Ligand Force Fields: step 1

1. Using the UCSF Chimera, open 'pdb' file.
2. Add Hydrogen atoms:

   'Tools'–>'Structure Editing'–>'AddH'

3. check the molecular structure and identify the charge of the ligand
4. Add Charges:

   'Tools'–>'Structure Editing'–>'Add Charge'

5. save it as a 'mol2' file

   'File'–>'Save Mol2'

Generate Ligand Force Fields: step 2

1. option Using an ambertools, rewrite 'mol2' for the ambertools:

```
antechamber -i ligand.mol2 -fi mol2 -o ligand_amber.mol2 -fo mol2 -rn LIG
```

[In this step, the atom types are reassigned for General Amber Force Field (gaff)]

2. generate frcmod

```
parmchk2 -i ligand_amber.mol2 -o ligand_amber.frcmod -f mol2
```

3. generate the parameter and topology file using a General Amber Force Field

```
tleap -s -f ligand.in
```

ligand.in

```
source leaprc.gaff2
mol = loadmol2 ligand_amber.mol2
loadamberparams ligand_amber.frcmod
saveamberparm mol ligand.prmtop ligand.inpcrd
quit
```

Code

```
from pyscf import gto, scf
mol = gto.M(atom='H 0 0 0; H 0 0 1.2', basis='ccpvdz')
mf = scf.RHF(mol)
mf.kernel()
```

```
python run_qmmm.py -i input.json
```

input.json

```
{
    "theory": "qm",
    "job": "ener", # gopt
    "qm": {
        "method": "scf",
        "basis": "6-31gs",
        "charge": 0,
        "fname_geom": "qm_step4.xyz",
        "esp": true,
        "esp_opts": {"resp": true,"resp_hfree": true}
    }

}
```

Sample Code (OpenMM >= 7.5)

```
from openmm.app import *
from openmm import *
from openmm.unit import *
from sys import stdout
```

Sample Code (OpenMM <= 7.4.2)

```
from simtk.openmm.app import *
from simtk.openmm import *
from simtk.unit import *
from sys import stdout
```