NWChem: Molecular Dynamics and QM/MM

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by **Battelle** Since 1965

U.S. DEPARTMENT OF **ENERGY**

Molecular Dynamics Functionality
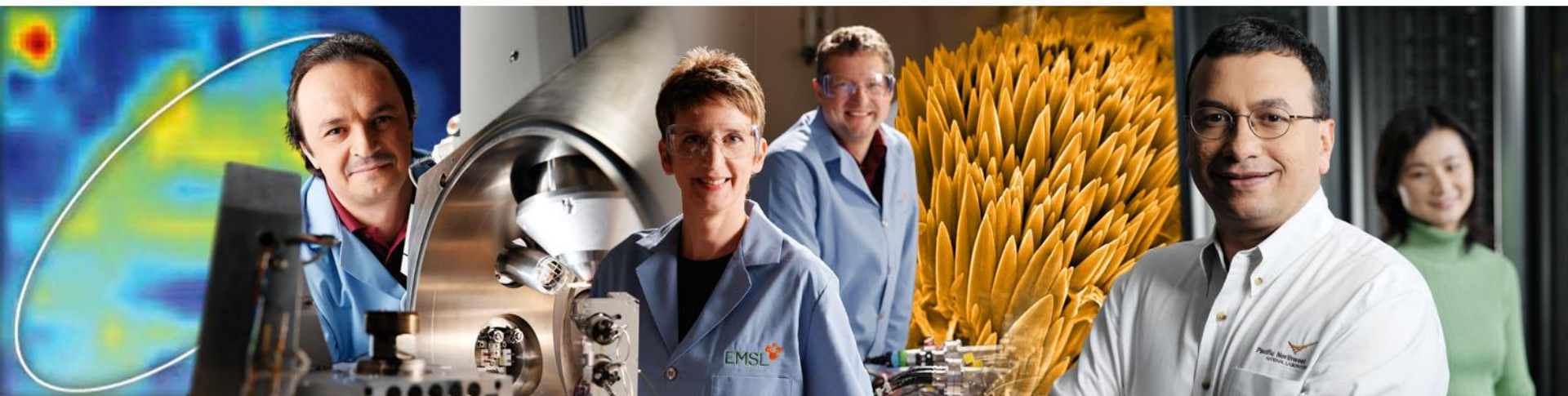
Target systems: **biomolecules**
(proteins, DNA/RNA, biomembranes)

General features: **energy evaluation (SP)**
energy minimization (EM)
molecular dynamics simulation (MD)
free energy evaluation (MSTP & MCTI)
quantum molecular dynamics (QMD)
hybrid molecular dynamics (QM/MM)

Force Fields

- **Classical empirical force fields**
  - Bond-stretching, angle-bending, torsional,
  - out-of-plane-bending, electrostatic, and
  - van der Waals non-bonded interactions

- **Electronic polarization**
  - First order
  - Self-consistent induced fields

- **Smooth particle-mesh Ewald electrostatics**
  - Effective pair potential MD only

- **Self-guided MD**

- **Integration Newton's Equations of Motion**

  (leapfrog, Browne-Clark)

- **Constant Temperature and Pressure**

  (Berendsen weak coupling)

- **Periodic Boundary Conditions**

  (minimum image convention)

- **Geometry Optimization**

  (steepest descent, conjugate gradient)

- **Twin-Range Verlet Neighbor Lists**

  (cell index method)

- **Constraints**

  (SHAKE)

Free Energy Evaluation

- **Single Step Thermodynamic Perturbation** (SSTP)

    $\Delta G = G_1 - G_0 = - RT \ln < \exp (- \Delta H/RT)>_0$ with $\Delta H = H_1 - H_0$

- **Multiple Step Thermodynamic Perturbation** (MSTP)

    $DG = S_i (G_{i+1} - G_i) = -S_i RT \ln < \exp (- \Delta H_i /RT)>_i$ with $\Delta H_i = H_{i+1} - H_i$

- **Multiconfiguration Thermodynamic Integration** (MCTI)

    $DG = S_i (G_{i+1} - G_i) = S_i < \partial H(\lambda) / \partial \lambda >_i D\lambda_i$

    Single and dual topology Hamiltonians
    Double-wide sampling
    Separation-shifted scaling (SSS)
    Potentials of mean force over multiple processors
    Statistical error correlation analysis

ffield_lvl   **ffield**   force field e.g. **amber**, **charmm**
                        lvl   **1**, **2**, **3**, **4**, **5**, **6**, **7**, **8**, or **9**

Directories supplied with NWChem are named ffield_ℓ where ℓ is one of:

s	standard parameters as published for the force field
x	extensions as published in open literature
q	contributed parameters by NWChem team
u	user contributed parameters
t	temporary
c	current

Defined in input file or ~/.nwchemrc

ffield amber
amber_1 **/software/nwchem/share/data/amber/amber_s/**
amber_2 **/software/nwchem/share/data/amber/amber_x/**
amber_3 **/home/newton/data/amber/amber_q/**
spce **/software/nwchem/share/data/solvents/spce.rst**

**system.ext**

system   user defined molecular system name
ext file type, e.g.   **pdb**   PDB file
                            top   topology file
                            seq   sequence file

system_calc.ext

system   user defined molecular system name
calc   user defined identification for the calculation identification, e.g. **em, md002,
     tiA**
ext file type, e.g.   **out**   output file
                            rst   restart file
                            qrs   energy minimized restart file
                            prp   property file
                            trj   trajectory file
                            gib   free energy file

EMSL
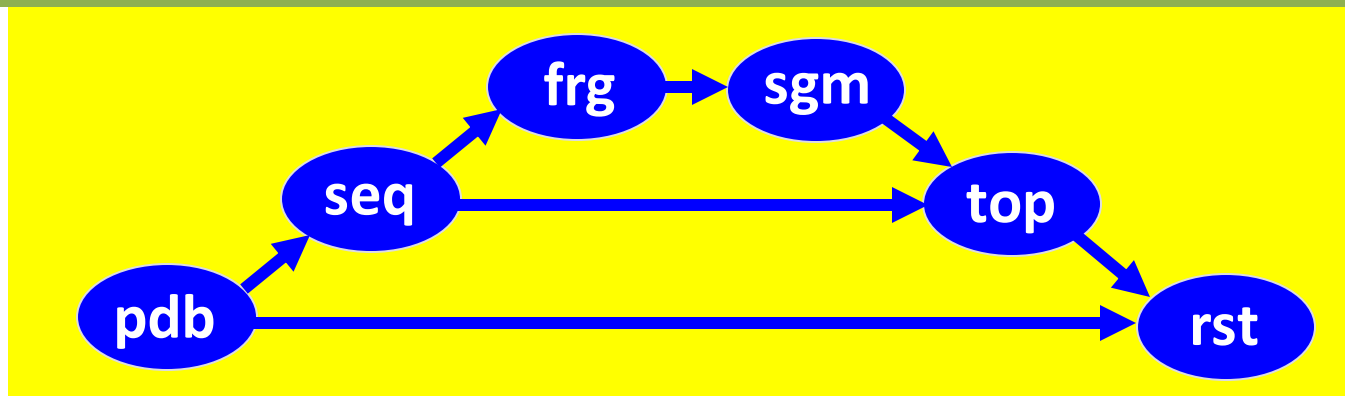
Functionality:

- Topology and restart file generation
- Coordinates from pdb formatted file or geometry input
- Fragment and segment file generation from coordinates
- Solvation
- Potential of mean force functions
- Topology modification for free energy and QM/MM calculations
- File format conversion, e.g. from rst to pdb

Requirements

- PDB format, i.e. IUPAC atom names, residue names, etc.
- Automated atom typing based on force field typing rules
- Force field parameters from par file(s)
- Partial atomic charges from frg or sgm files

Pacific Northwest
NATIONAL LABORATORY

U.S. DEPARTMENT OF
ENERGY

Proudly Operated by Battelle Since 1965

- PDB format, i.e. IUPAC atom names, residue names, etc.
- Automated atom typing based on force field typing rules
- Force field parameters from par file(s)
- Partial atomic charges from frg or sgm files

```
memory noverify heap 1 mb stack 48 mb global 24 mb
start crown
basis "ao basis" print
  H library sto-3g
  C library sto-3g
  O library sto-3g
  Na library sto-3g
end
```

Define basis sets in case partial charges need to be calculated

```
prepare
 system crown_em
 modify atom 7:Na set 3 type K
end
task prepare
```

Read coordinates from **crown.pdb**
Specify atom type change in topology
Generate topology file **crown.top**
Restart file **crown_em.rst**

MD Input Example

```
md
 system crown_md
 data 1000                                    Molecular dynamics input:
 isotherm                                     NpT ensemble, 1000 steps
 isobar
 record rest 1000 coord 100 prop 10
end
task md dynamics

task shell "cp crown_md.rst crown_ti.rst"     Copy restart file.

md
 system crown_ti
 equil 1000 data 2000 over 1000 step 0.002    Free energy simulation
 isotherm                                        MCTI: 21*(1000+2000)
 isobar                                                  NpT ensemble
 new forward 21 of 21
 print step 100 stat 1000
 record rest 1000 free 1
end
task md thermodynamics
```

```
memory heap 8 mb stack 64 mb global 24 mb
start job

analysis
  system system_calc
  reference system_calc.rst
  file crown_md?.trj 0 10          # ? Is wild card, replaced by series 0-10
  select _C? _O?                   # select subset of atoms
  essential                        # Perform essential dynamics analysis
  project 1 crown_vec1               Projection onto specified vector
  project 2 crown_vec2
end

task analyze
```

Files required: **system.top** (molecular topology)

                           system_calc.rst (coordinates)

```
memory noverify heap 10 mb stack 32 mb global 32 mb
start crown
basis "ao basis" print
  H library sto-3g
  C library sto-3g
  O library sto-3g
  Na library sto-3g
end
```

Define basis sets

```
prepare
 system crown_em
 modify atom 7:Na set 3 type K
end
task prepare
```

Read coordinates from crown.pdb
Specify atom type change in topology
Generate topology file crown.top
Restart file crown_em.rst

```
md
 system crown_em
 sd 1000
 noshake
end
task md optimize
```

Perform energy minimization

```
prepare
 read crown_em.qrs
 write crown_em.pdb
 solvate
 write crown_md.rst
end
task prepare
```

Solvate the solute
Generate crown_md.rst

```
md
 system crown_md
 data 1000
 isotherm 298.15 trelax 0.1 0.1
 isobar
 record rest 1000 scoor 100 prop 10
end
task md dynamics


task shell "cp crown_md.rst crown_ti.rst"


md
 system crown_ti
 equil 1000 data 2000 over 1000 step 0.002
 isotherm 298.15 trelax 0.1 0.1
 isobar
 new forward 21 of 21
 print step 100 stat 1000
 record rest 100 free 1
end
task md thermodynamics
```
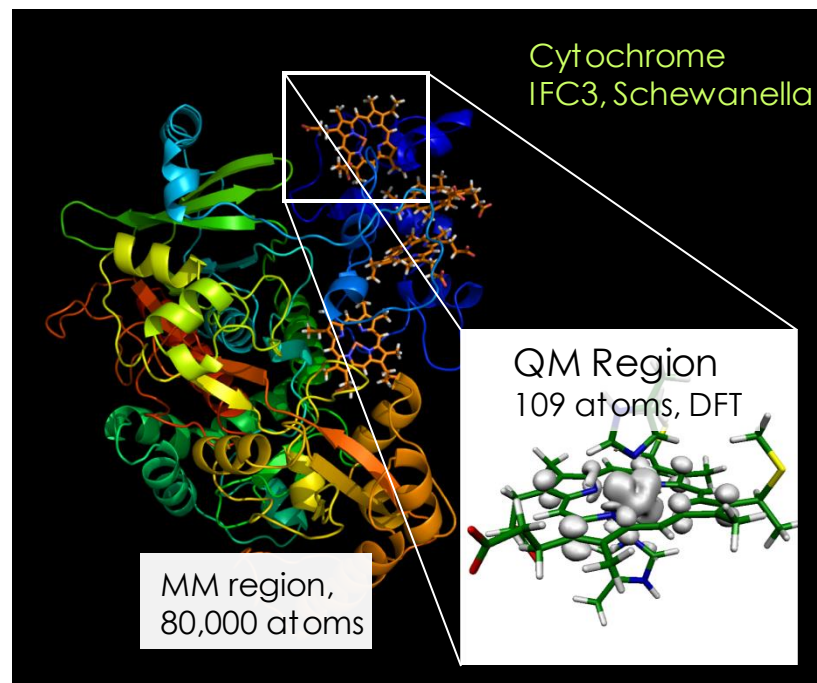
Molecular dynamics input:
   NpT ensemble, 1000 steps

Copy restart file.

Free energy simulation input:
      MCTI: 21*(1000+2000)
            NpT ensemble

Pacific Northwest
NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

Proudly Operated by Battelle Since 1965

QM/MM approach

- Combines two different descriptions – quantum-mechanical and classical

- The level theory **changes** based on a particular region

- Reactive regions – quantum mechanical description (QM)

- Regions where no chemical changes occur (or important) are treated a the classical molecular mechanics level (MM)

Cytochrome IFC3, Schewanella

QM Region
109 atoms, DFT

MM region,
80,000 atoms

Pacific Northwest
NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

Proudly Operated by Battelle Since 1965

$$E_{qm/mm}(\mathbf{r}, \mathbf{R}; \psi) = E_{qm}(\mathbf{r}, \mathbf{R}; \psi) + E_{mm}(\mathbf{r}, \mathbf{R})$$

▸ All QM-dependencies are in the first term

$$E_{qm}[\mathbf{r}, \mathbf{R}; \psi] = E_{qm}^{int}[\mathbf{r}, \mathbf{R}; \psi] + E_{qm}^{ext}[\mathbf{r}, \mathbf{R}; \rho]$$

Internal QM energy
(theory dependent)

Coulomb interactions
with MM atoms
$$\sum_I \int \frac{Z_I \rho(\mathbf{r}')}{|R_I - \mathbf{r}'|} d\mathbf{r}'$$

▸ All other classical terms are in the second term

  ▸ Bonded, angle, dihedral

  ▸ Coulomb interactions

  ▸ Vdw interactions

Pacific Northwest
NATIONAL LABORATORY

U.S. DEPARTMENT OF
ENERGY

Proudly Operated by **Battelle** Since 1965

QM/MM Interface in NWChem

$$H\psi = E\psi$$

QM/MM

MM

QM
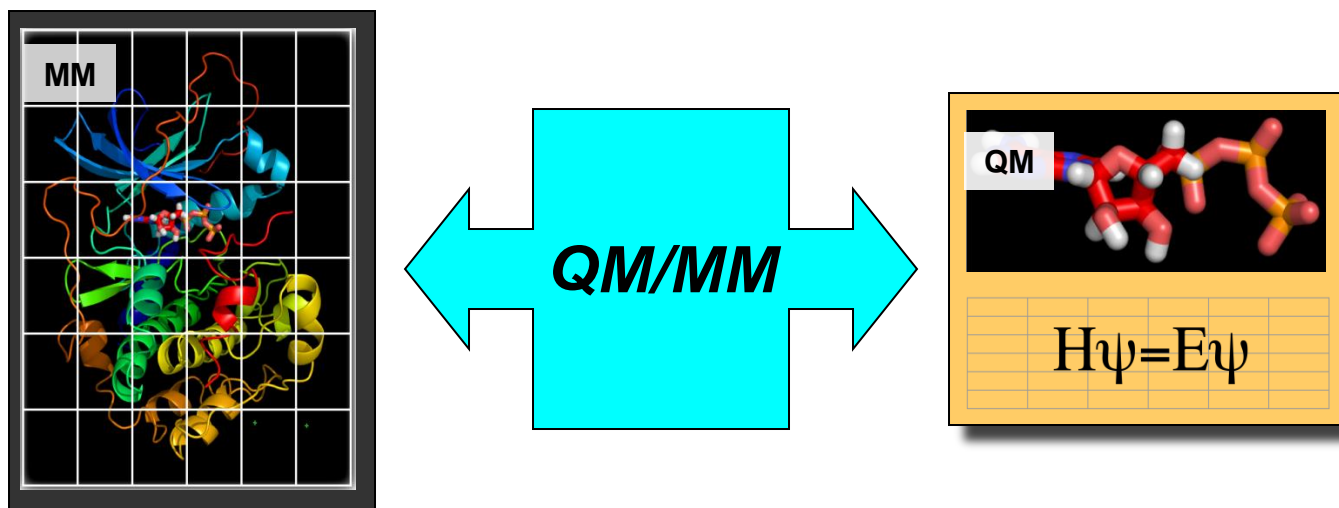
- ■ Modular implementation
  - ◆ sits on top of molecular mechanics (MM) and quantum (QM) modules
  - ◆ Generic interface driven by function calls
- ■ Manages data flow
  - ◆ domain decomposition of coordinates in MM module
  - ◆ replicated geometry data in QM module
  - ◆ all data transfers happen in core
- ■ Dispatches high level operations (e.g. optimization)

Pacific Northwest
NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

Proudly Operated by **Battelle** *Since 1965*

- Ground and excited state properties
- Structural optimization
- Reaction Pathway Calculations
- Dynamical Simulations
- Statistical Sampling (free energies)

Pacific Northwest
NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

Proudly Operated by **Battelle** *Since 1965*

Example of QM/MM Single Point Calculation

```
start capk
charge 2

prepare
 source capk.pdb
 modify segment 358 quantum
 …
 ignore
 update lists
 write capk.rst
end
task prepare

md
 system capk
 cutoff 1.5
end

basis
  *   library "Ahlrichs pVDZ"
end

qmmm
 link_atoms hydrogen
end

task qmmm dft energy
```

QM region definition
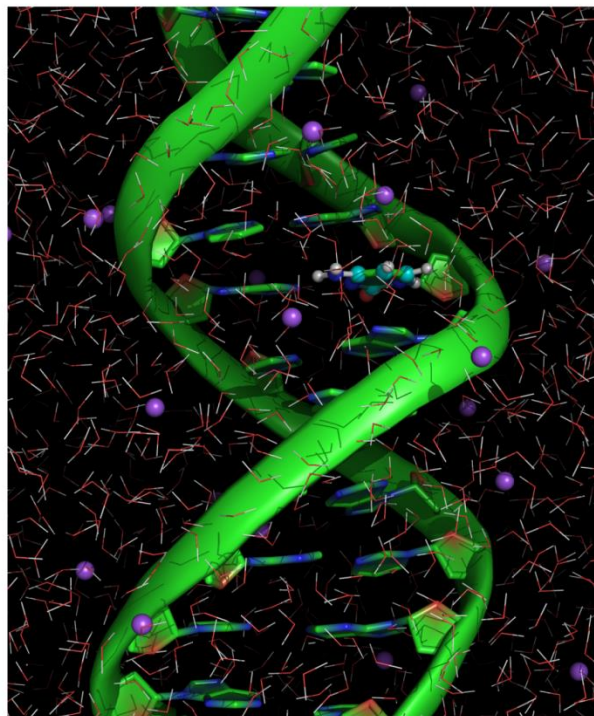
Hydrogen Link Atoms

Commence QM/MM DFT Calculation

Active Site of cAPK kinase

75 atoms, 817 basis functions

Scaling of QM/MM-DFT calculation

Time (s) vs Processors (8, 16, 32, 64, 96, 128)
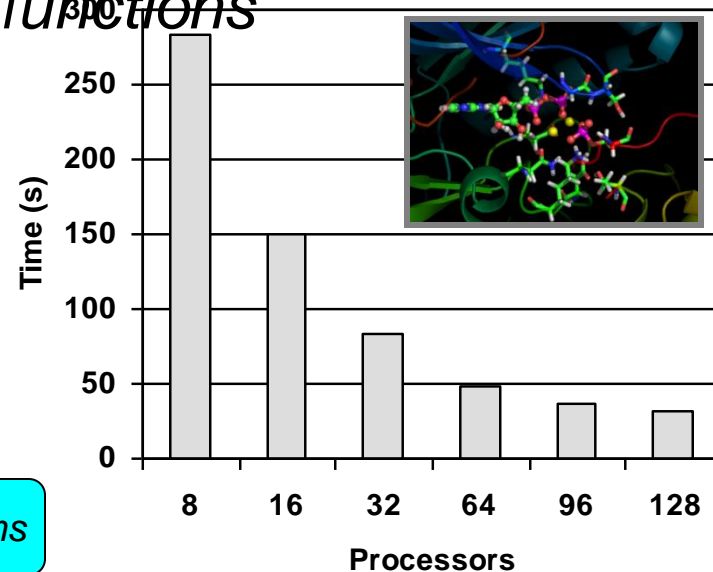
Valiev, M., et al. JPC B, 2007. 111(47): p. 13455-13464.

Pacific Northwest
NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

Proudly Operated by Battelle Since 1965

Excitation Spectrum Time Series
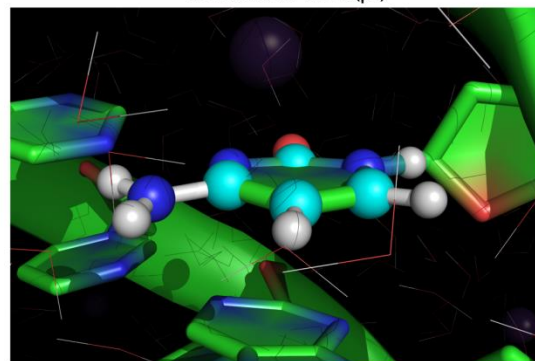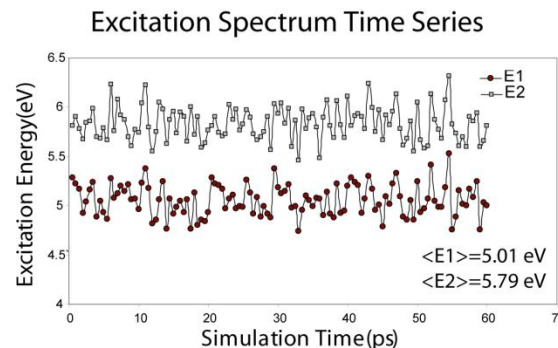
- QM/MM coupled-cluster CR-EOMCCSD(T) calculations of two lowest excited states
- Protein environment has a significant influence on the excitation leading to a 0.4 eV stabilization of the $\pi\pi^*$ excited state compared to gas phase
- M. Valiev, K. Kowalski, JCP, 125(21), (2006)

QM/MM Optimization

- Large system sizes (10^3-10^6) makes direct optimizations impractical even with QM/MM approximation

- Key observations

  - Most of degrees of freedom are in MM region

  - The structure of far away MM regions has a little effect on the structure of QM region

  - Small displacements of MM atoms affect little the electronic structure of QM region.
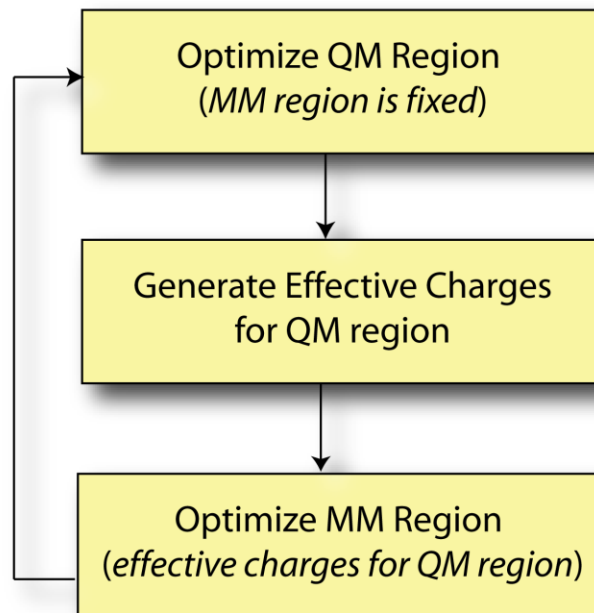
- Decouple optimization of QM and MM regions

Optimization Algorithm

1. QM Region Optimization

    a. MM region is fixed
    b. Typically 20-30 steps
    c. Requires solution of Schrödinger Equation
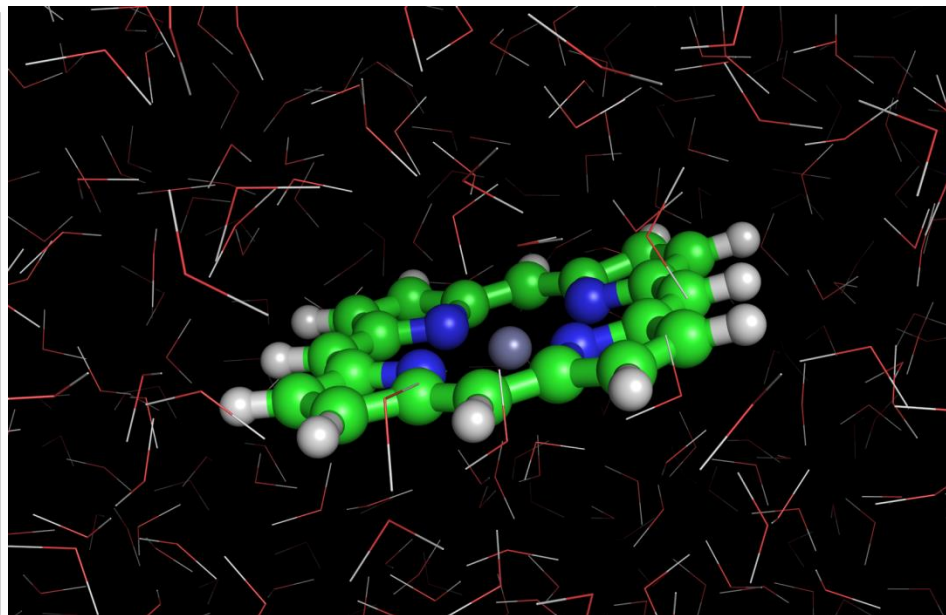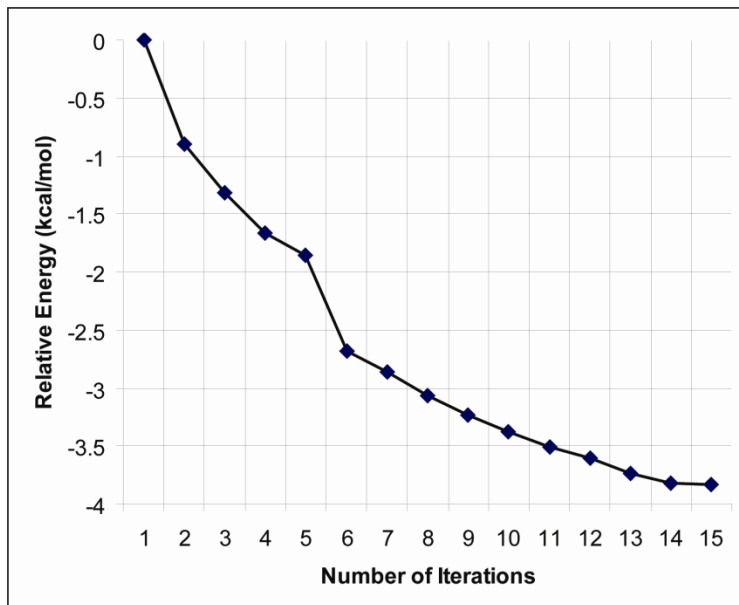
2. MM Region Optimization

    a. QM region is fixed
    b. 1000-3000 steps
    c. QM region is represented either as effective point charges or static electron density distribution

3. Repeat the cycle until convergence

Optimize QM Region
(*MM region is fixed*)

Generate Effective Charges for QM region

Optimize MM Region
(*effective charges for QM region*)

Pacific Northwest
NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

Proudly Operated by **Battelle** *Since 1965*

EMSL

- QM region Zn-porphyrin (37 atoms) DFT/B3LYP

- MM region – 869 SPC/E waters

- **4.5 hours on 48 processors** versus direct optimization would take ~ 2 days

```
….
qmmm
region   qm solvent
maxiter  10 3000
method   lbfgs  sd
ncycles 20
density espfit
end
task qmmm dft optimize
```
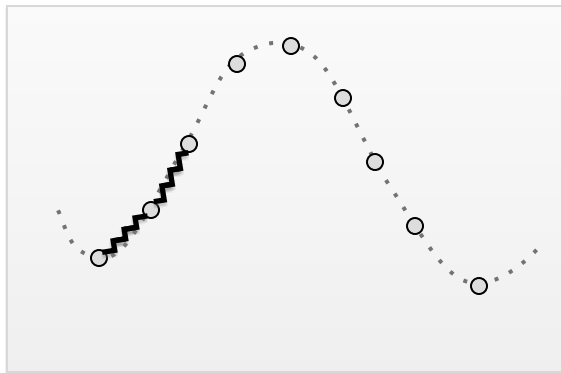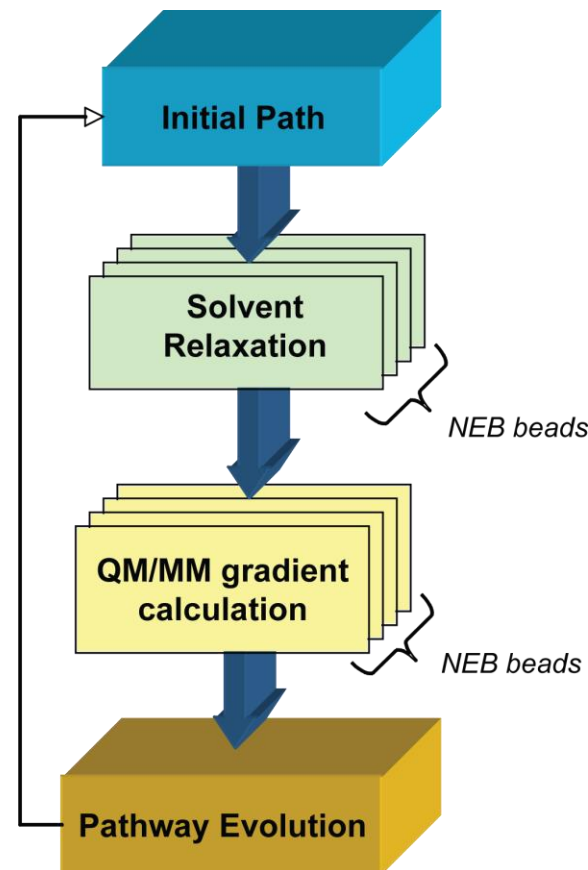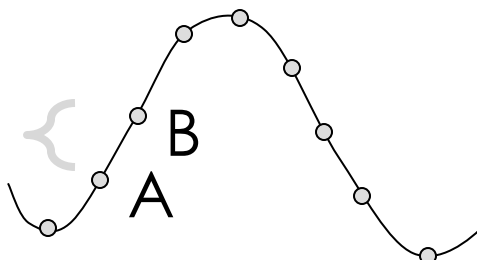
Input File

Pacific Northwest
NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

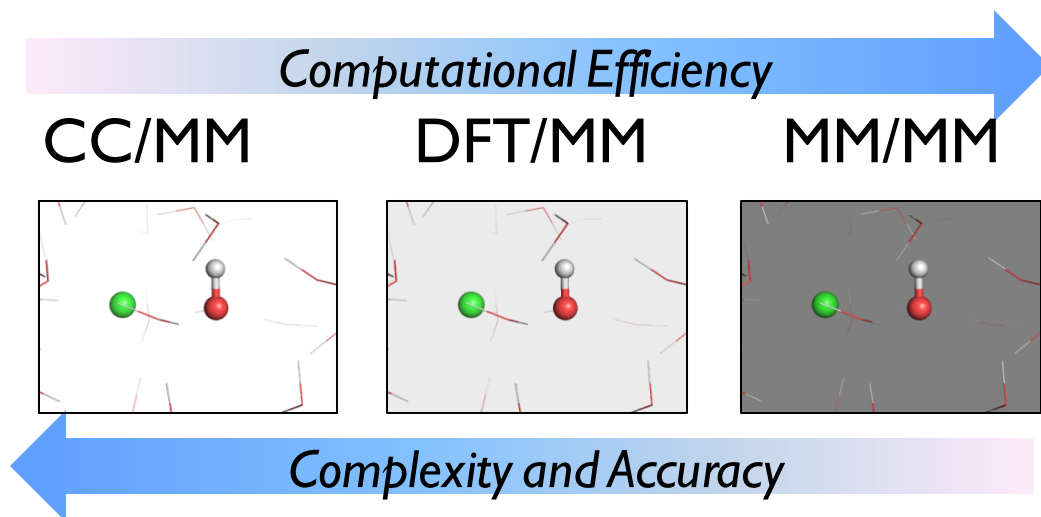Proudly Operated by **Battelle** Since 1965

- Nudged Elastic Band Method

- Pathway approximated discrete set of intermediate structures

- Beads represent different snapshots of reactive QM region along the pathway

- Forces on beads are calculated at the relaxed solvent configuration

Initial Path

Solvent Relaxation

NEB beads

QM/MM gradient calculation

NEB beads

Pathway Evolution

$$\Delta W_{AB} = -\frac{1}{\beta} \ln \left\langle e^{-\beta(E(A)-E(B))} \right\rangle_A$$
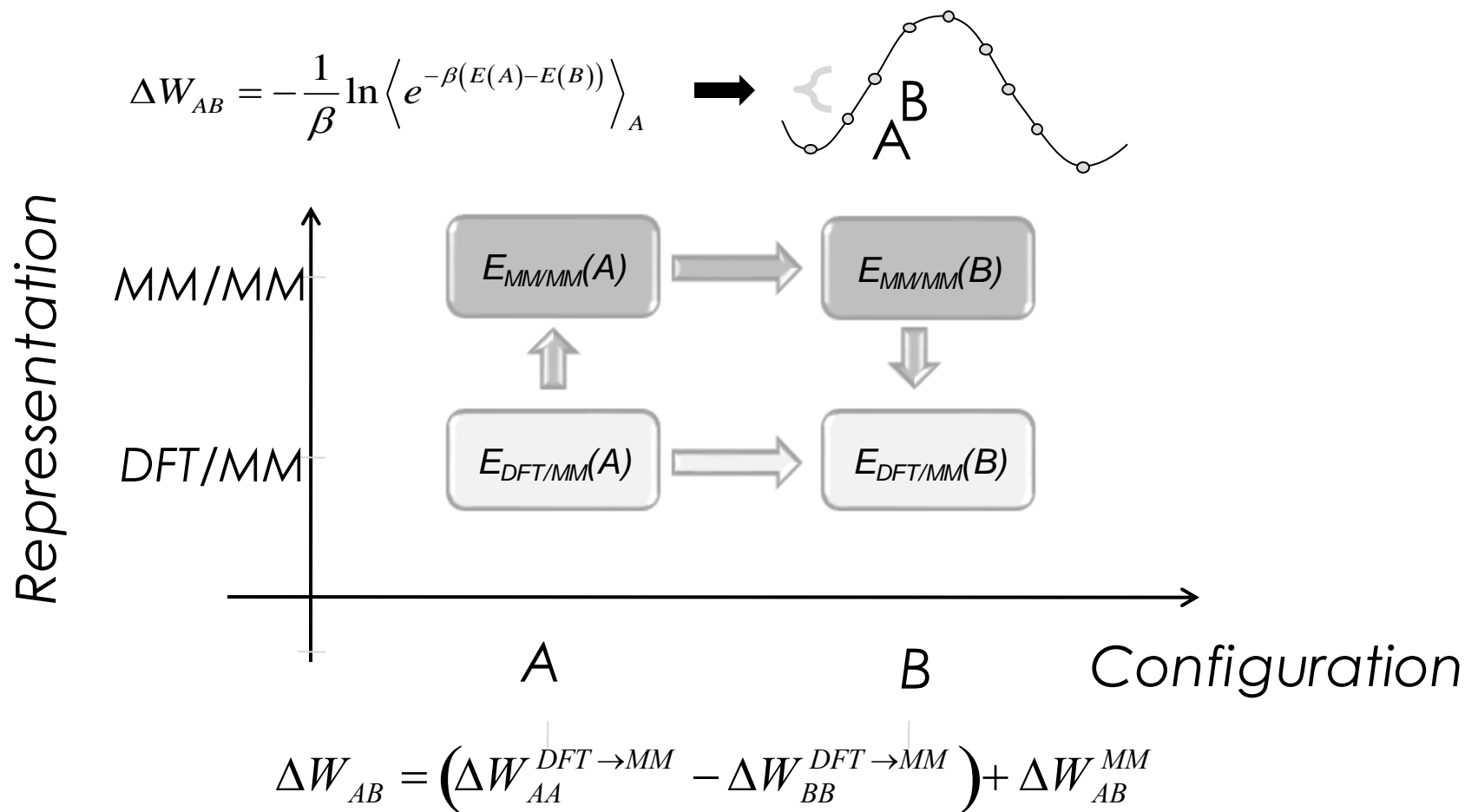
B
A

- Accurate quantum mechanical description is a major challenge especially for high level methods ($10^4 - 10^5$ energy evaluations)

- The solution

  - ◆ Introduce intermediate less expensive representation(s)
  - ◆ Redistribute sampling using thermodynamic cycles

Pacific Northwest
NATIONAL LABORATORY

U.S. DEPARTMENT OF
ENERGY

Proudly Operated by Battelle Since 1965

■ Different QM/MM representations

Computational Efficiency

CC/MM   DFT/MM   MM/MM
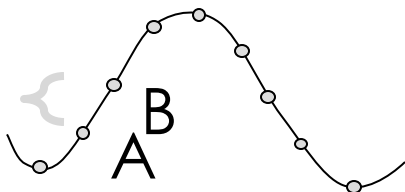
Complexity and Accuracy

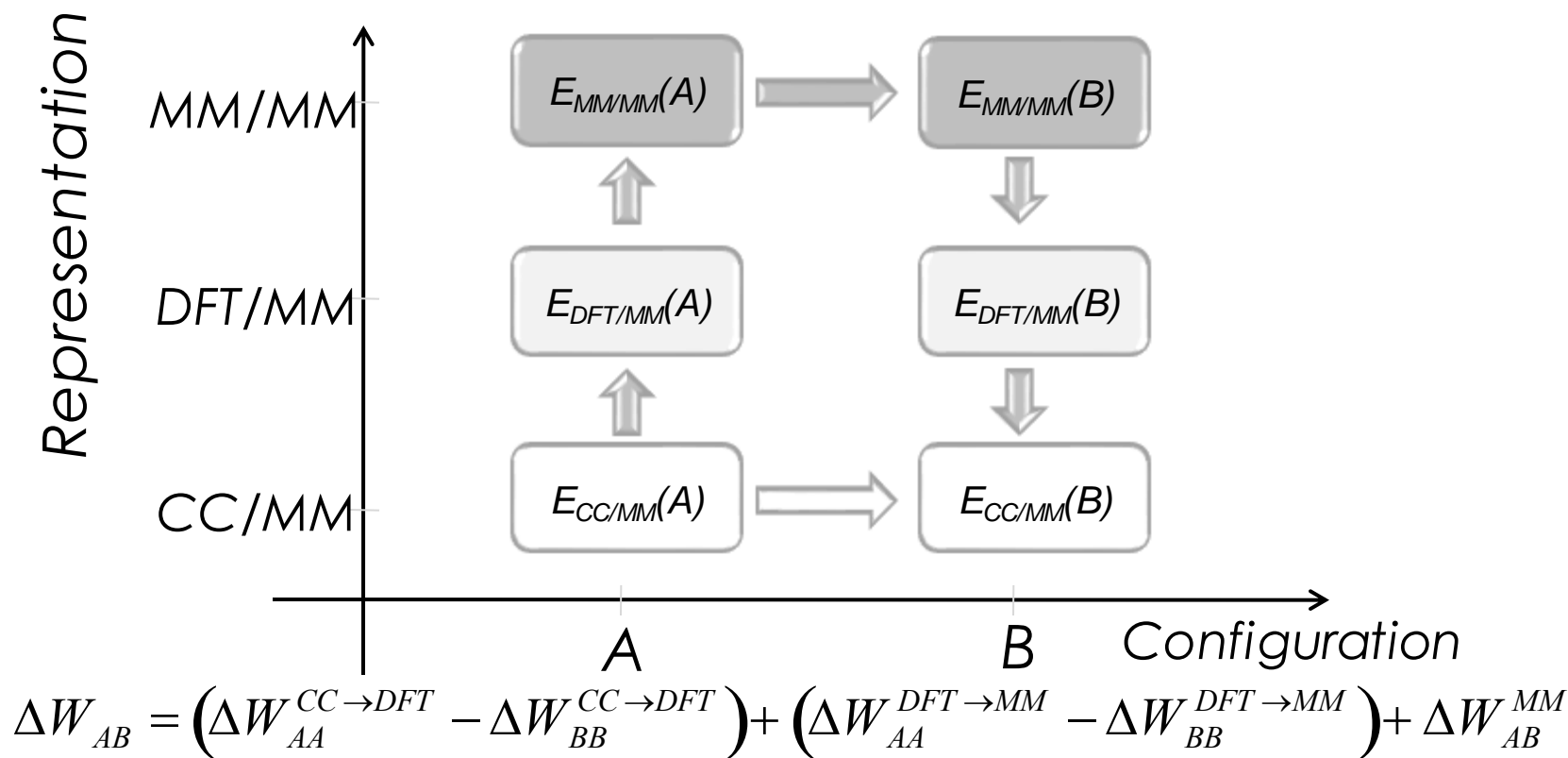■ Example of MM/MM representation - QM atoms are replaced by effective point charges Q_i reproducing correct field

$$E_{qm} = \sum_{i,I} \frac{Z_I Q_i}{|\mathbf{R}_I - r_i|} \quad \Longleftarrow \quad \sum_I \int \frac{Z_I \rho(\mathbf{r}')}{|\mathbf{R}_I - \mathbf{r}'|} d\mathbf{r}' = \sum_{i,I} \frac{Z_I Q_i}{|\mathbf{R}_I - \mathbf{r}_i|}$$

Pacific Northwest
NATIONAL LABORATORY

U.S. DEPARTMENT OF
ENERGY

Proudly Operated by Battelle Since 1965

EMSL

$$\Delta W_{AB} = -\frac{1}{\beta} \ln \left\langle e^{-\beta(E(A)-E(B))} \right\rangle_A$$

Representation

MM/MM — $E_{MM/MM}(A)$ ⟹ $E_{MM/MM}(B)$

DFT/MM — $E_{DFT/MM}(A)$ ⟹ $E_{DFT/MM}(B)$

A   B   Configuration

$$\Delta W_{AB} = \left(\Delta W_{AA}^{DFT \to MM} - \Delta W_{BB}^{DFT \to MM} \right) + \Delta W_{AB}^{MM}$$

*Valiev et al JCP **127**, 051102 (2007)*

Pacific Northwest
NATIONAL LABORATORY

U.S. DEPARTMENT OF
ENERGY

Proudly Operated by *Battelle* Since 1965

$$\Delta W_{AB} = -\frac{1}{\beta} \ln \left\langle e^{-\beta(E(A)-E(B))} \right\rangle_A$$

Representation

MM/MM — $E_{MM/MM}(A)$ → $E_{MM/MM}(B)$

DFT/MM — $E_{DFT/MM}(A)$ $E_{DFT/MM}(B)$

CC/MM — $E_{CC/MM}(A)$ → $E_{CC/MM}(B)$

A   B   Configuration

$$\Delta W_{AB} = \left(\Delta W_{AA}^{CC \rightarrow DFT} - \Delta W_{BB}^{CC \rightarrow DFT} \right) + \left(\Delta W_{AA}^{DFT \rightarrow MM} - \Delta W_{BB}^{DFT \rightarrow MM} \right) + \Delta W_{AB}^{MM}$$

Pacific Northwest
NATIONAL LABORATORY

U.S. DEPARTMENT OF
ENERGY

Proudly Operated by Battelle Since 1965

$$E_{MM/MM}(A) \longrightarrow E_{MM/MM}(B)$$

- Can use any of the methods developed for classical free energy calculations
- Transformation between A and B configurations

$$\mathbf{r}_\lambda = (1-\lambda)\mathbf{r}_A + \lambda\mathbf{r}_B$$

$$Q_\lambda = (1-\lambda)Q_A + \lambda Q_B$$

- Free Energy Perturbation

$$\Delta W_{AB}^{ESP} = -\sum_i \frac{1}{\beta} \ln\left\langle e^{-\beta\Delta E_{\lambda_i \to \lambda_{i+1}}^{ESP}} \right\rangle_{\lambda_i}$$

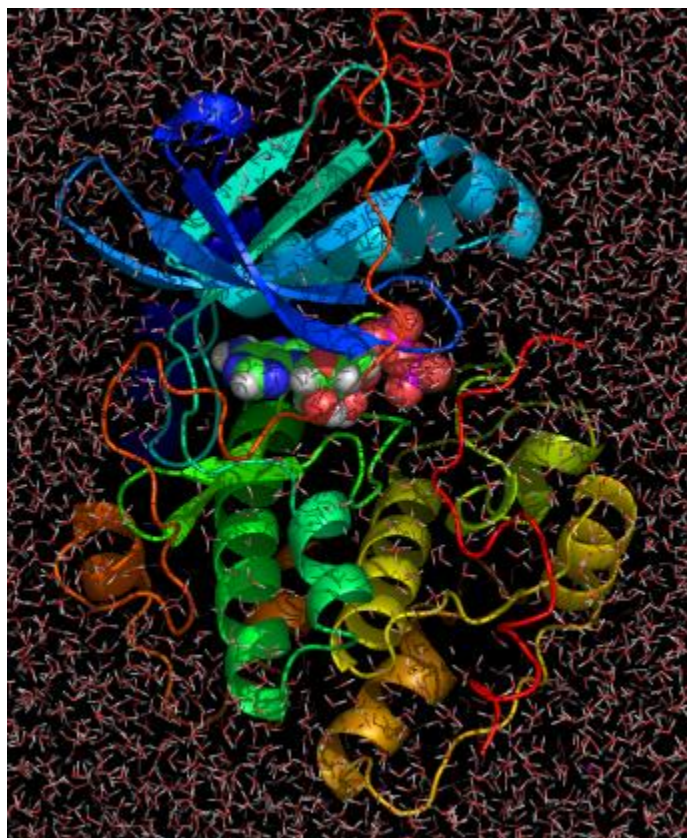# Calculation of DFT->MM free energy

$$E_{MM/MM}(A)$$

$$E_{DFT/MM}(A)$$

$$\Delta W_{AA}^{DFT \to MM} = -\frac{1}{\beta} \ln \left\langle e^{-\beta \left(\Delta E_{AA}^{DFT \to MM} \right)} \right\rangle_{MM/MM}$$

$$\boxed{\Delta E_{AA}^{DFT \to MM} = E_{DFT/MM}(\mathbf{r}_A, \mathbf{R}; \psi_A) - E_{MM/MM}(\mathbf{r}_A, \mathbf{R}; Q_A)}$$

- "**V**ertical" change of transformation (fixed QM region)
- MM representation is closely tailored to DFT by point charge fitting
- Approximate by the energy difference
- Can utilize free energy perturbation approach by resampling MM/MM trajectory
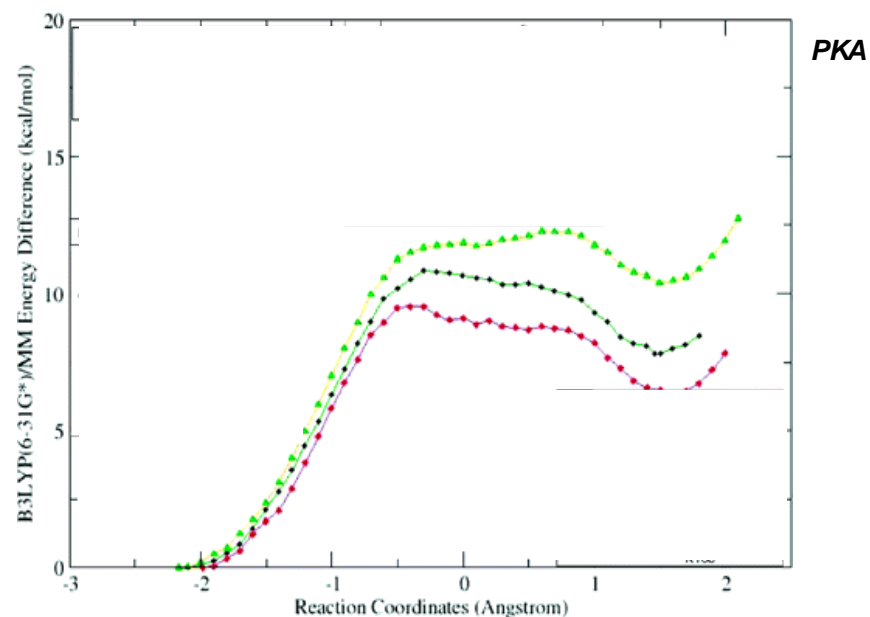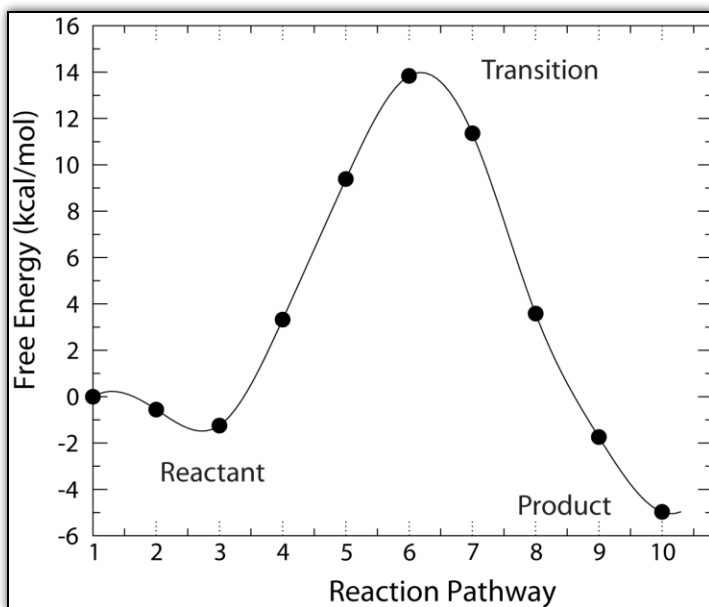
~15 kcal/mol reaction barrier
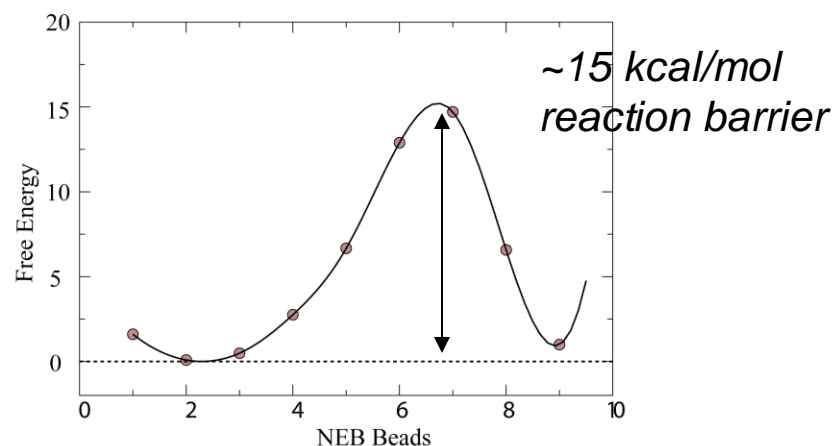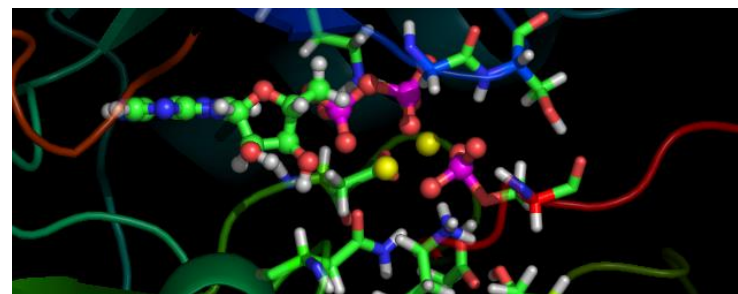
- cAPK protein kinase catalyze the transfer of the g-phosphoryl
- Determination of the reaction pathway using NEB QM/MM ap
- Calculation of free energy using effective charge approximatio
- Valiev et al J. Phys Chem B. 111(47):13455-64. (2007)

(Cheng et al JACS, 127,1553,2005)

PKA

Questions?

www.**emsl**.pnl.gov

Pacific Northwest
NATIONAL LABORATORY

U.S. DEPARTMENT OF
ENERGY

Proudly Operated by **Battelle** Since 1965