# Hybrid Methods for Studies of Chemical Processes in Condensed Phase
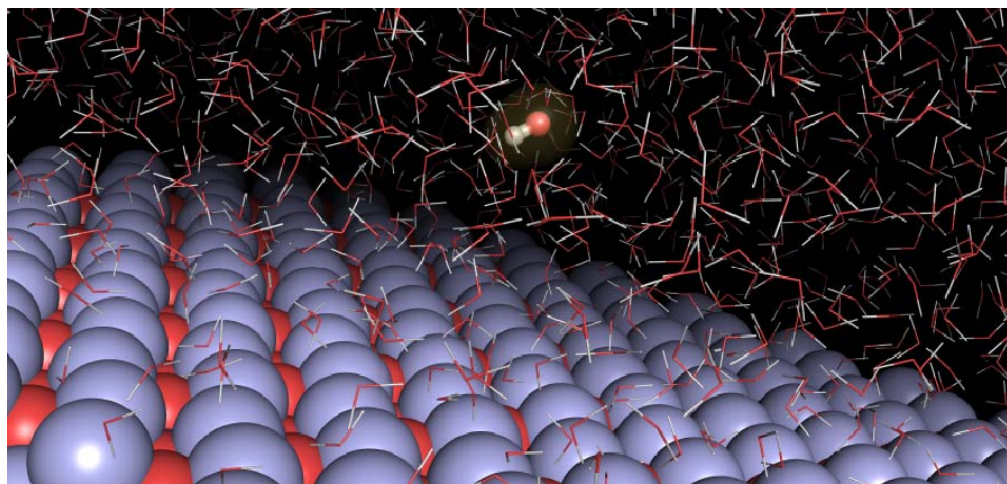
*M.Valiev, B. C. Garrett, K. Kowalski, P.D. Fan, S. M. Kathmann, G. K. Schenter, M. Dupuis*
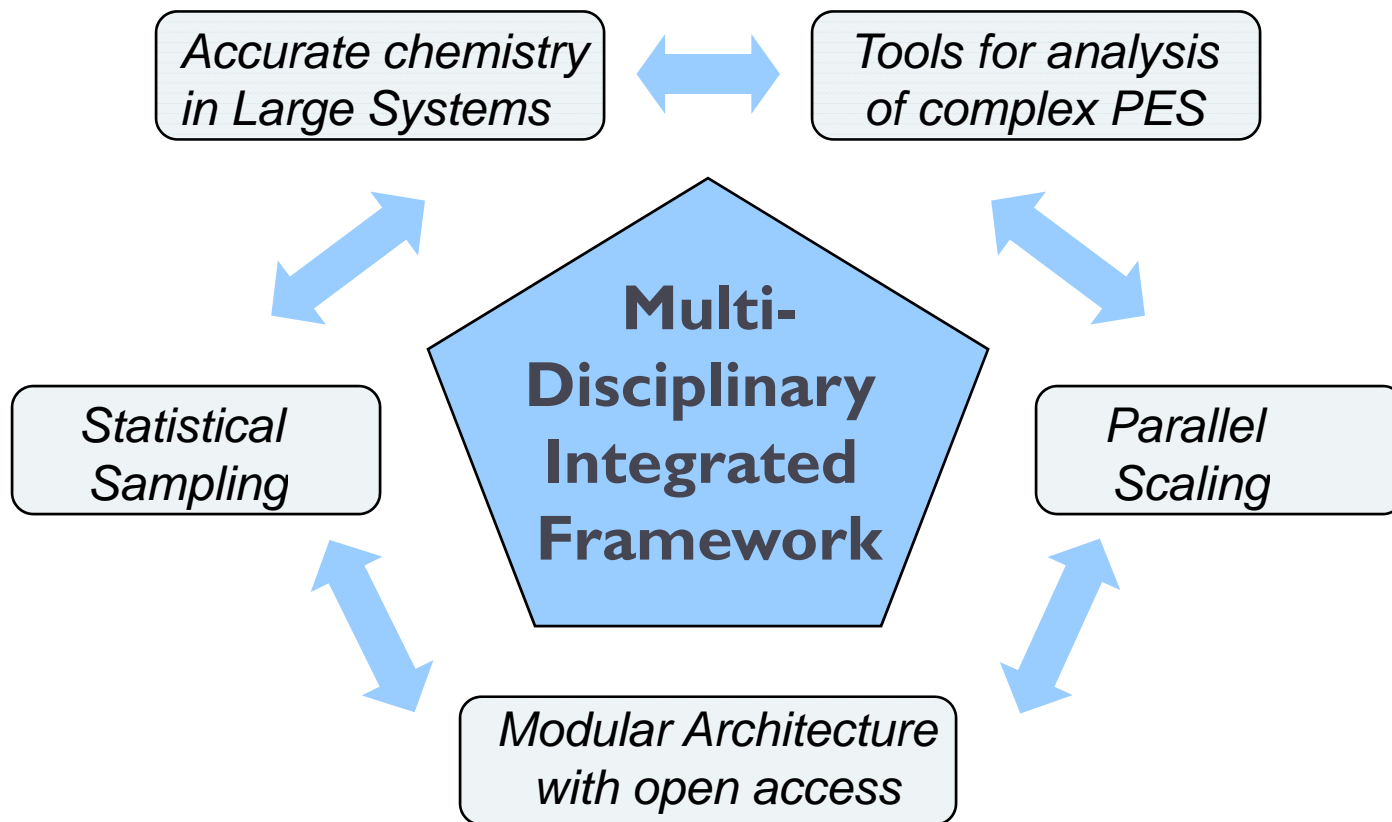
*marat.valiev@pnl.gov*

# Chemical Processes in Condensed Phase

- ▸ Chemical transformations that involve both ground and excited states
- ▸ Large heterogeneous systems ($\sim 10^3$-$10^6$ particles)
- ▸ Complex topology of potential energy surface
- ▸ Inherently non-static: temperature/pressure fluctuations
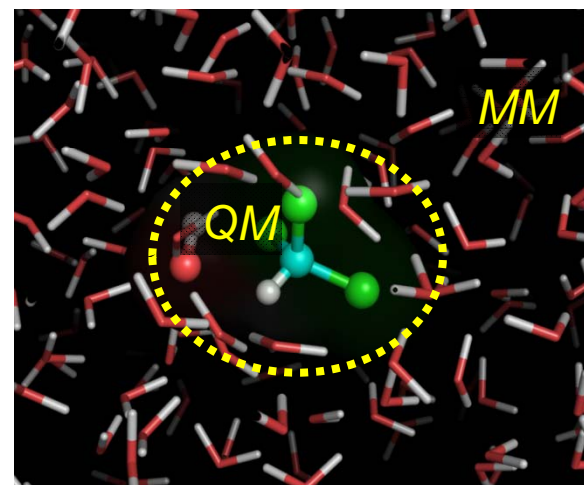
# Accurate Chemical Description

▸ Fundamental building block for the analysis of condensed phase processes.

▸ Critically important for reliable results

**Our Approach:**

▸ Emphasis on large scale systems

▸ Flexibility in quantum mechanical description **--** from DFT to Coupled Cluster

▸ Ground and excited states capabilities

▸ Modular implementation within NWChem

▸ Efficient parallel scalability

# QM/MM approach

- ▸ It is not possible nor advisable to describe entire system ($\sim10^3$-$10^6$) quantum mechanically
- ▸ Changes in chemical structure are typically localized
- ▸ The chemical identity of bulk of the system remains unchanged



*Example: chloroform and OH- reaction in water*

## QM/MM approach:

- ▸ Chemically reactive species treated quantum-mechanically (QM region)
- ▸ Classical description for the rest of the system (MM region)

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

▸ **All QM-dependencies are in the first term**

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

*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

# QM/MM Interface in NWChem

- ▸ **Modular implementation**
  - ▸ sits on top of quantum (QM) and molecular mechanics (MM) modules
  - ▸ Generic interface driven by function calls



QM/MM Interface
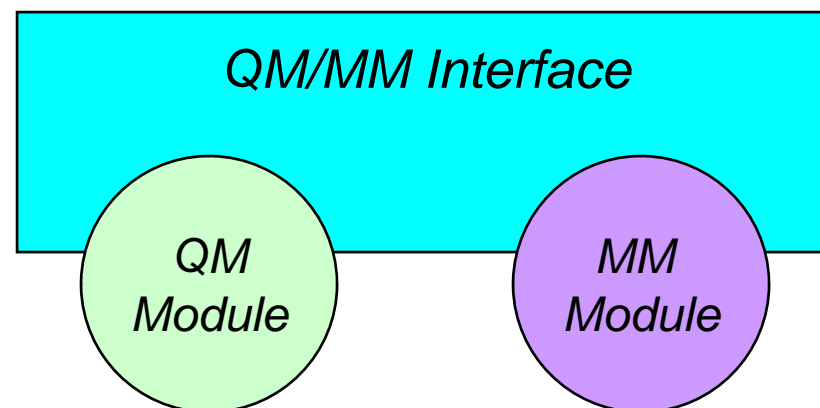
QM Module

MM Module

- ▸ **Manages data flow**
  - ▸ domain decomposition of coordinates in MM module
  - ▸ replicated geometry data in QM module
  - ▸ all data transfers happen in core

- ▸ **Easily accommodates various QM theories (DFT, MP2, CCSD, ....)**

- ▸ Supports both ground and excited state calculations

# Example: Ground and Excited States of Zn-porphyrin in solution
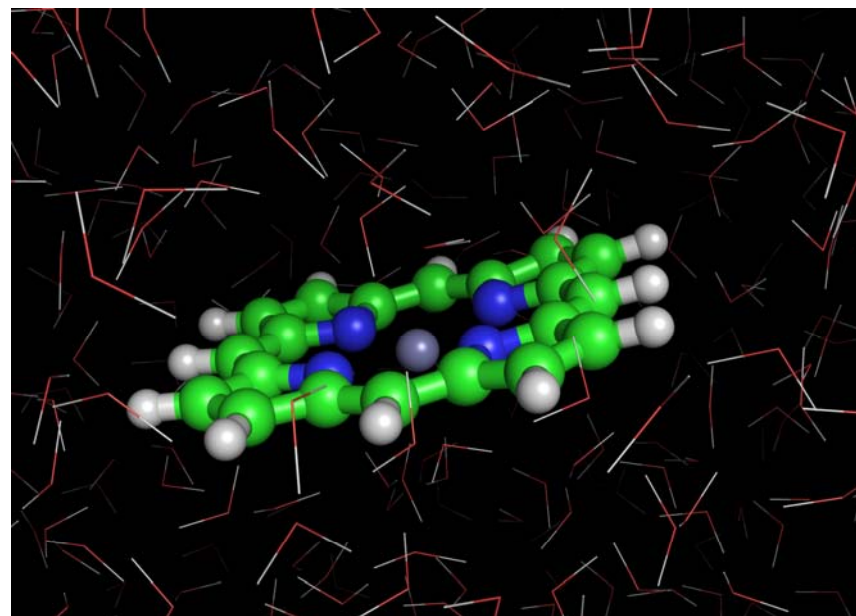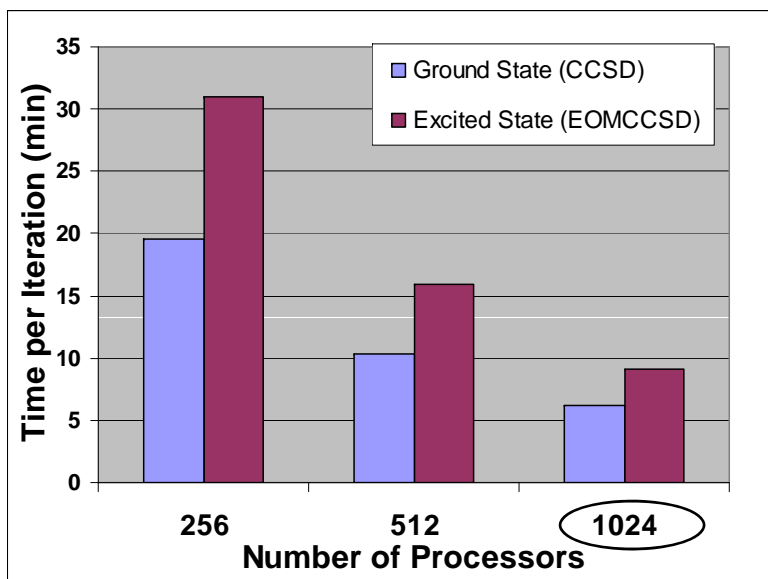


- QM region Zn-porphyrin (37 atoms) at CCSD level
- MM region – 869 SPC/E waters
- Scalable through 1024 processors

# Analysis of Potential Energy Surface

**Presence of many degrees of freedom ($10^3$-$10^6$)**

▸ Computational Expense

▸ Complexity of reaction coordinate

**Our Approach:**

▸ Takes advantage of natural scale separation (QM and MM)

▸ Provides a way to incorporate various gas phase techniques

    ▸ Optimizations, transition states

    ▸ Reaction pathways without explicit assumptions for the reaction coordinate

▶ Key Observations

  ▶ Most degrees of freedom are in MM region

  ▶ Movement of far MM atoms does not have much effect to electronic structure of QM region

▶ Can we separate optimization of QM and MM regions?
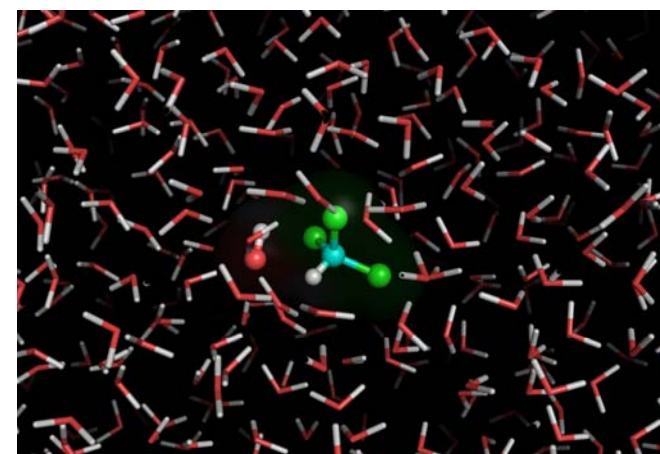
*Zhang, et al JCP 112 (8), 3483 (2000).*

- Alternating optimizations of QM and MM regions

- Efficient relaxation of MM region based on fixed effective charge model of the QM region

- Self-consistent loop ensures correct optimized geometry

Optimize QM Region
(*MM region is fixed*)

Generate Effective Charges
for QM region

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

$$\sum_{I} \int \frac{Z_{I} \rho(\mathbf{r}')}{\left| \mathbf{R}_{I} - \mathbf{r}' \right|} d\mathbf{r}' = \sum_{i,I} \frac{Z_{I} Q_{i}}{\left| \mathbf{R}_{I} - \mathbf{r}_{i} \right|}$$
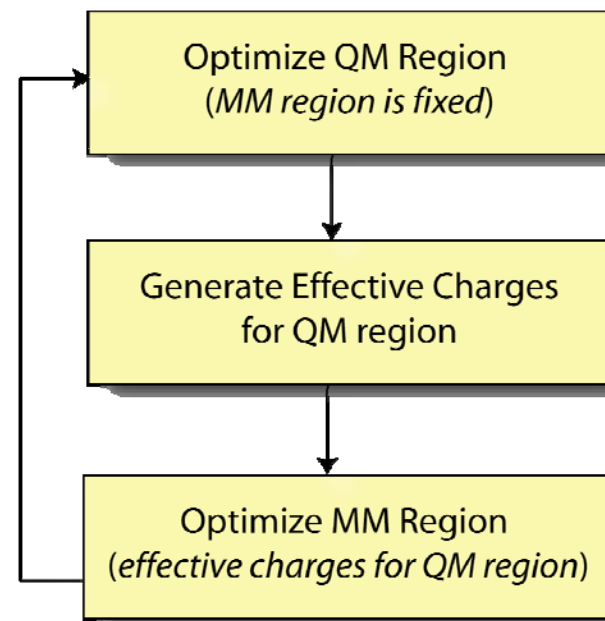
As long as the optimization cycle is convergent the frequency charges $Q_i$ updates is largely irrelevant and can be exploited for best computational efficiency

# Example: Optimization of Zn-porphyrin in solution
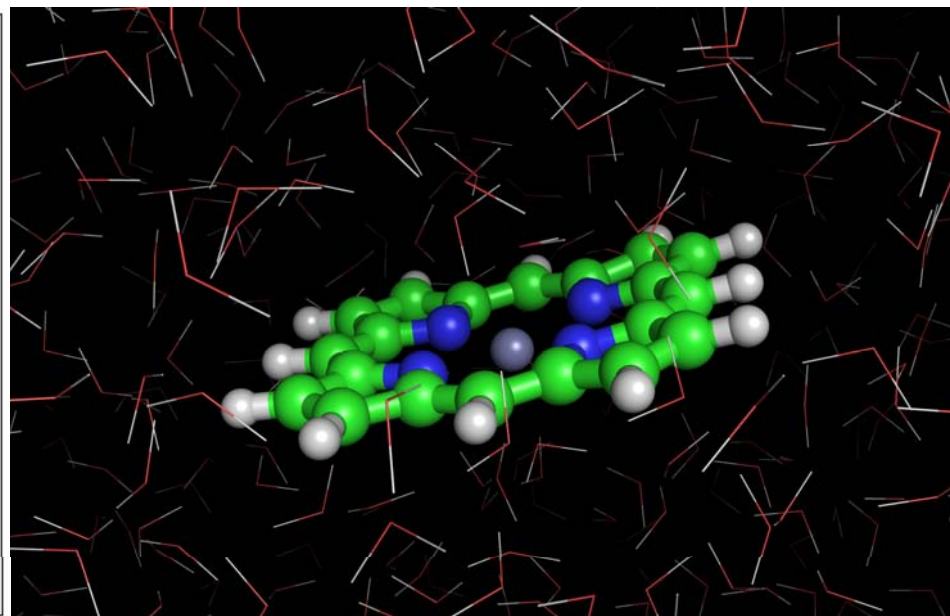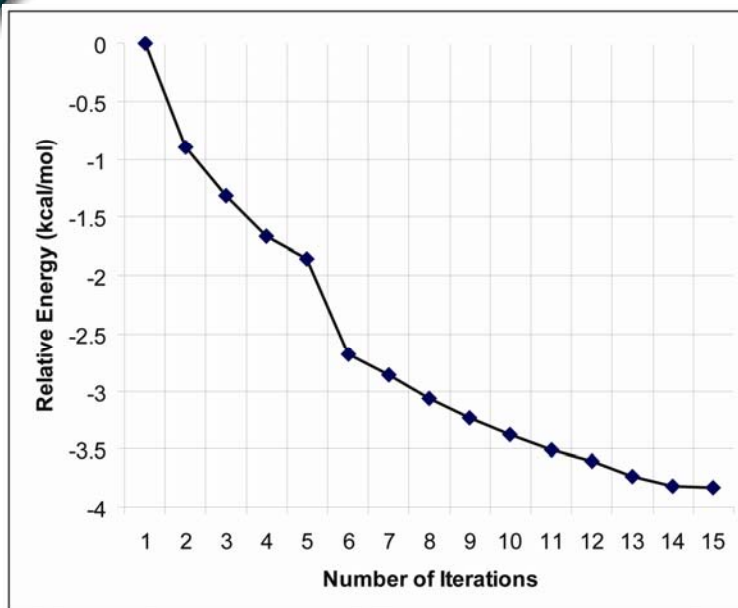


- ▸ QM region Zn-porphyrin (37 atoms) DFT/B3LYP
- ▸ MM region – 869 SPC/E waters
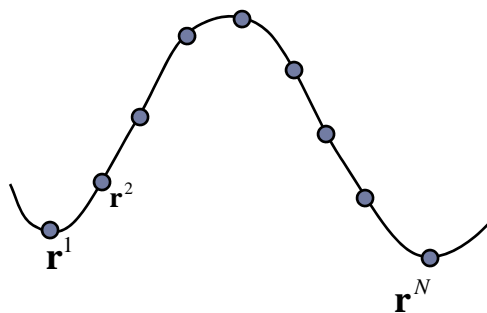- ▸ **4.5 hours on 48 processors** versus direct optimization would take ~ 2 days

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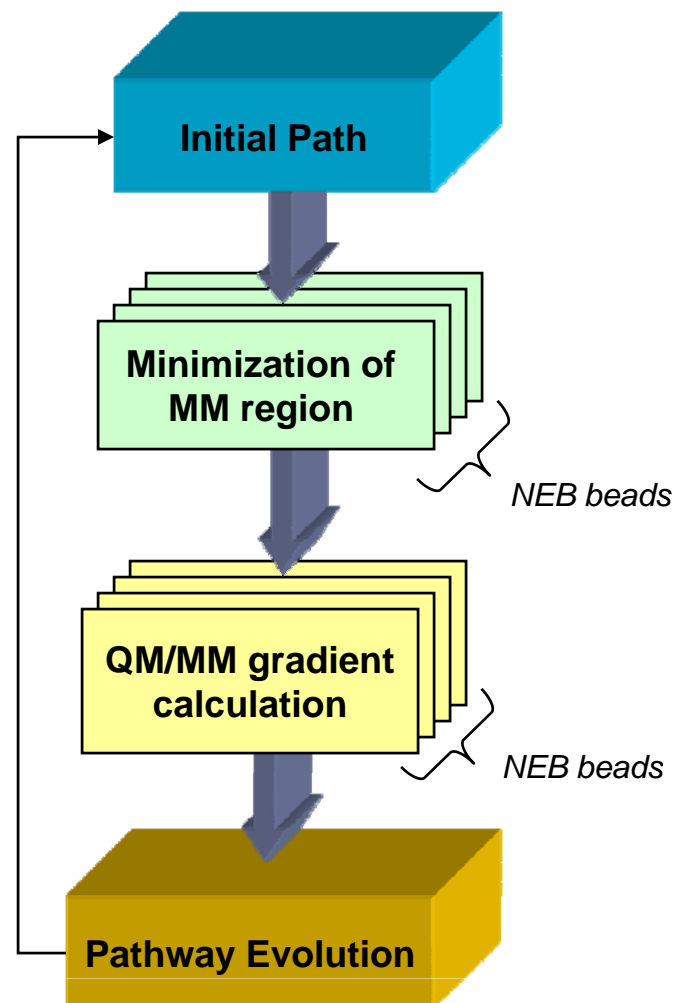
*Input File*

# QM/MM -NEB method for reaction pathways

- ▸ Pathway represented by replicas/beads of the molecular systems connected to each other through harmonic spring forces.

- ▸ Reaction is assumed to be driven by QM region



$$E^{NB} = \sum_{n=1}^{N} E\left[\mathbf{r}^n, \mathbf{R}; \psi\right] + \sum_{n=1}^{Nb} k\left|\mathbf{r}^n - \mathbf{r}^{n-1}\right|^2$$

**Initial Path**

**Minimization of MM region**

*NEB beads*

**QM/MM gradient calculation**

*NEB beads*

**Pathway Evolution**

*Valiev, Garrett, et al, J. Chem. Phys.* **127**, *051102 (2007)*

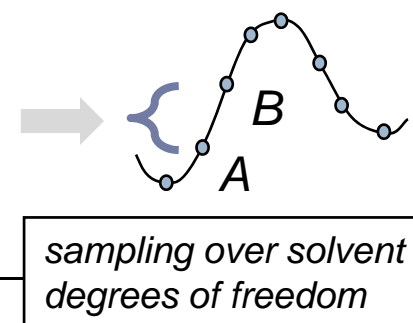# Statistical Sampling

▸ Critical for proper description of finite temperature/pressure fluctuations

▸ Enormous computational challenge for QM-based methods (40,000 samples or more)

## Our Approach:

▸ Combines several layers of different QM/MM representations

▸ Potential of Mean Force (PMF) at the coupled cluster level of theory

▸ Thermal averages of various observables

# Free Energy Profiles for Reactions

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

sampling over solvent degrees of freedom

- ▸ Reaction profiles in condensed phase should be described in terms of free energies or PMF's (not the total energy)

- ▸ Two key requirements:
    - ▸ Sufficient statistical averaging ($10^4$ - $10^5$ samples)
    - ▸ Accurate quantum mechanical description
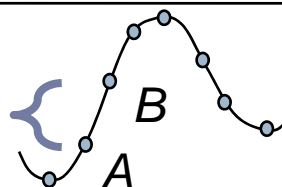
- ▸ Can we have both?

# QM/MM Representations

▸ Different flavors of QM/MM representations

*Computational Efficiency* →

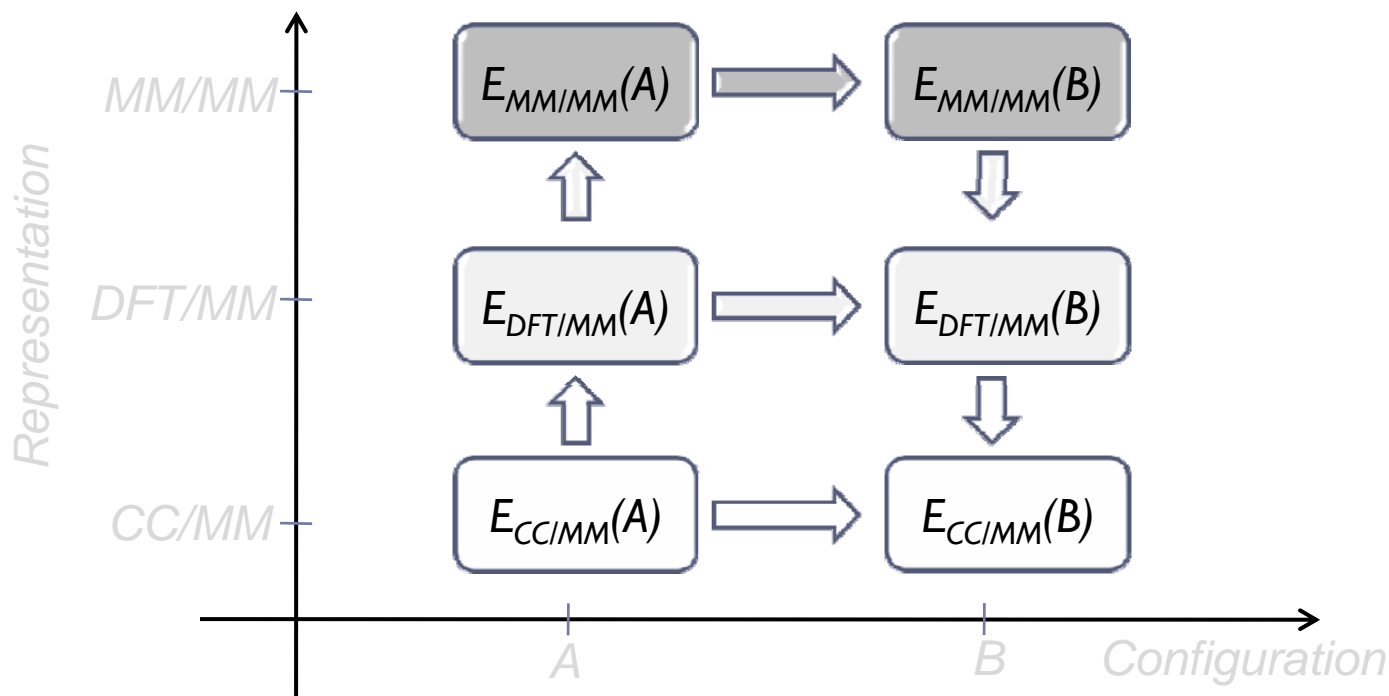| CC/MM | DFT/MM | MM/MM |
|---|---|---|

← *Complexity and Accuracy*
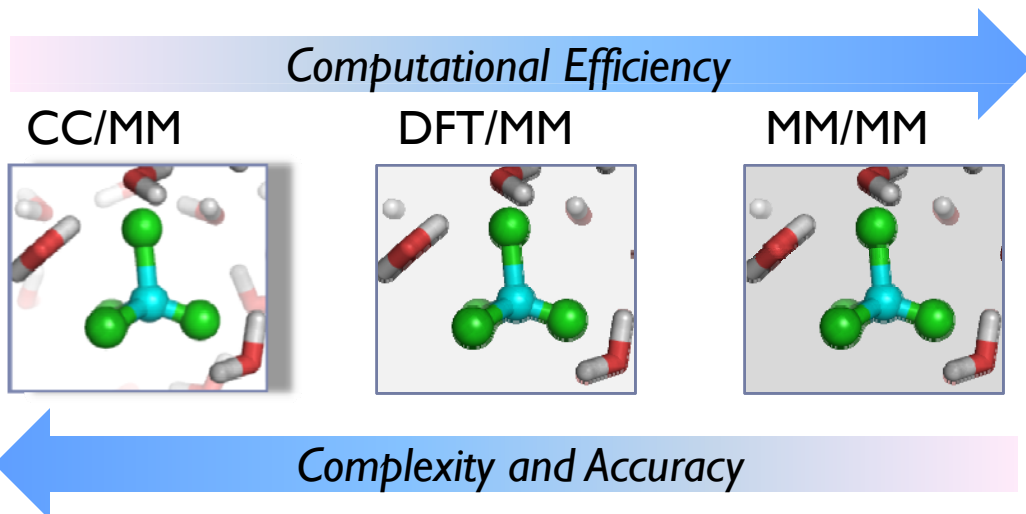
▸ Example of MM/MM representation - QM atoms are replaced by effective point charges $Q_i$ reproducing correct field

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

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



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

*Valiev, et al, J. Chem. Phys.* **127**, *051102 (2007)*

# Calculation of MM/MM free energy

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

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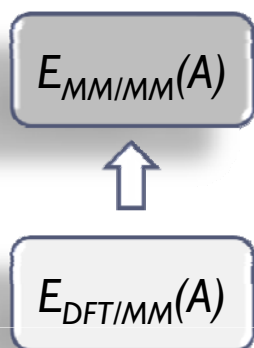
▸ Transformation between A and B configurations

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

▸ 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

$E_{MM/MM}(A)$

$E_{DFT/MM}(A)$

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

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

‣ "**V**ertical" change of transformation (fixed QM region)

‣ MM representation is closely tailored to DFT by point charge fitting

‣ Can utilize free energy perturbation approach by resampling MM/MM trajectory (*Wood et al, JPC A, 2002*)

$E_{DFT/MM}(A)$

$$\Delta W_{AA}^{CC \rightarrow DFT} = -\frac{1}{\beta} \ln \left\langle e^{-\beta \left( E_{CC/MM}(\mathbf{r}_A, \mathbf{R}; \Psi_A) - E_{DFT/MM}(\mathbf{r}_A, \mathbf{R}; \psi_A) \right)} \right\rangle_{DFT/MM}$$

$E_{CC/MM}(A)$

*Do we really need to do statistical average here ?*

- *Coupling between QM and MM region depends only the density not the wavefunction*
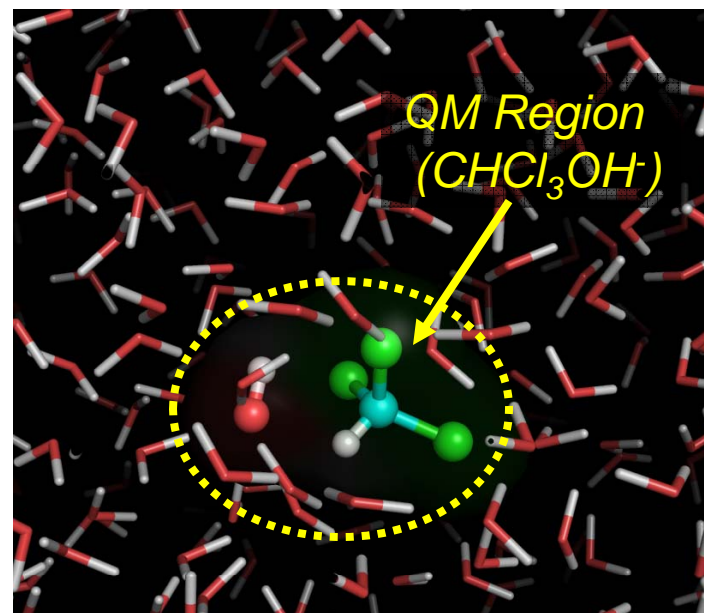
$$\sum_I \int \frac{Z_I \rho(\mathbf{r}')}{|\mathbf{R}_I - \mathbf{r}'|} d\mathbf{r}'$$

- *DFT density (at least formally) should be pretty close to CC density*

$$\Delta W_{AA}^{CC \rightarrow DFT} \approx E_{CC/MM}(\mathbf{r}_A, \mathbf{R}; \Psi_A) - E_{DFT/MM}(\mathbf{r}_A, \mathbf{R}; \psi_A)$$
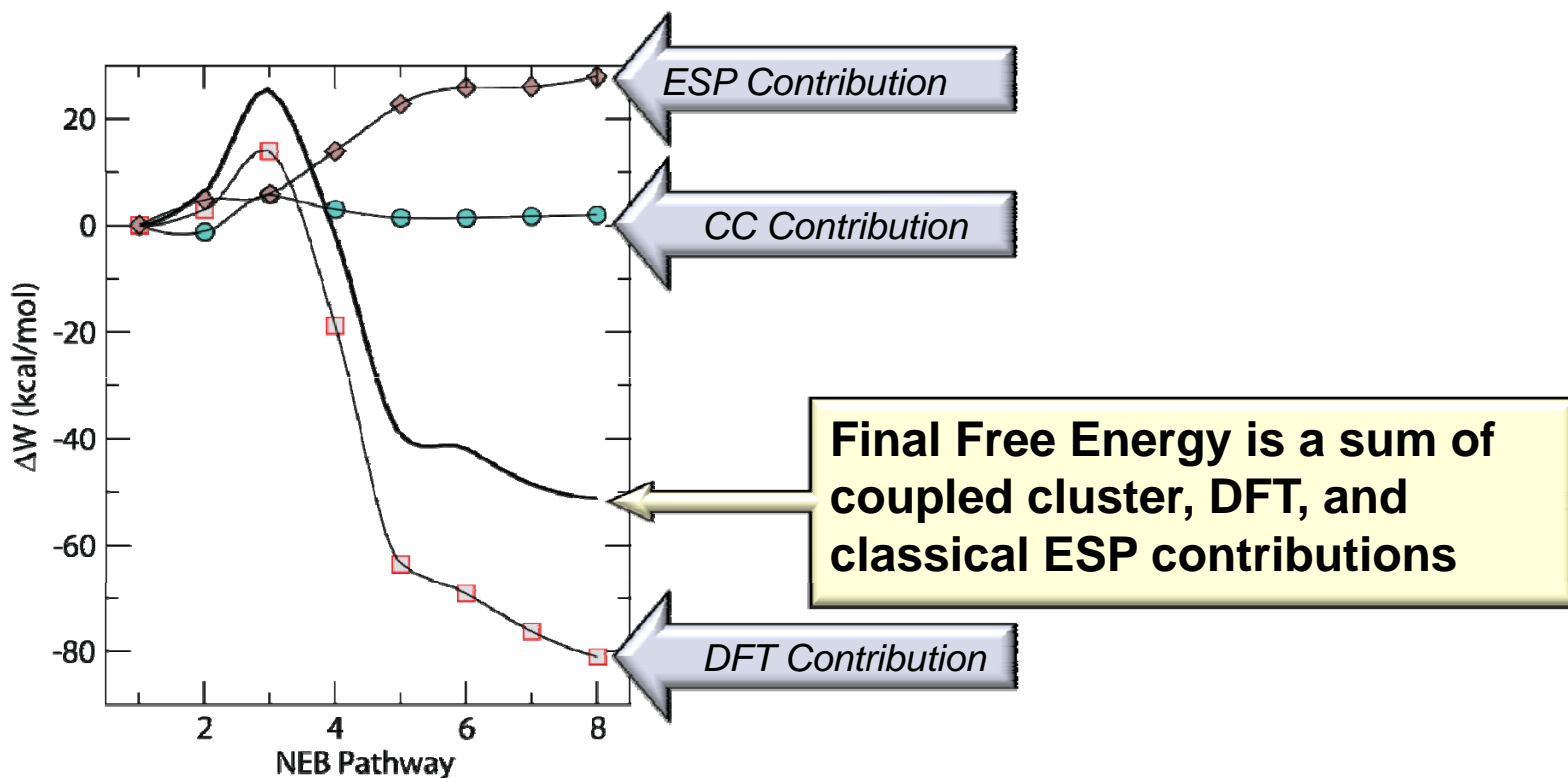
# CHCl$_3$+OH$^-$ reaction

- **QM region - CCl$_4$**
  - DFT/B3LYP
  - CCSD(T)
- **MM regions water**
  - 30 A cubic box
  - SPC/E model



QM Region (CHCl$_3$OH$^-$)

- Reaction Pathway was determined using NEB QM/MM approach
- 8-96 cpu's about a week worth of work

*Valiev, et al, J. Chem. Phys. **127**, 051102 (2007)*

Pacific Northwest National Laboratory
Operated by Battelle for the U.S. Department of Energy

# Free Energy Reaction Profile



ESP Contribution

CC Contribution

**Final Free Energy is a sum of coupled cluster, DFT, and classical ESP contributions**

DFT Contribution

- *CC/MM approach gives a free energy reaction barrier of 29.3 kcal/mol and an overall reaction free energy of -46.7 kcal/mol.*
- *DFT/MM treatment of the same process underestimates the reaction barrier (24.6 kcal/mol) but agrees well on the reaction free energy (-47 kcal/mol).*

*Valiev, Garrett, et al, J. Chem. Phys. **127**, 051102 (2007)*

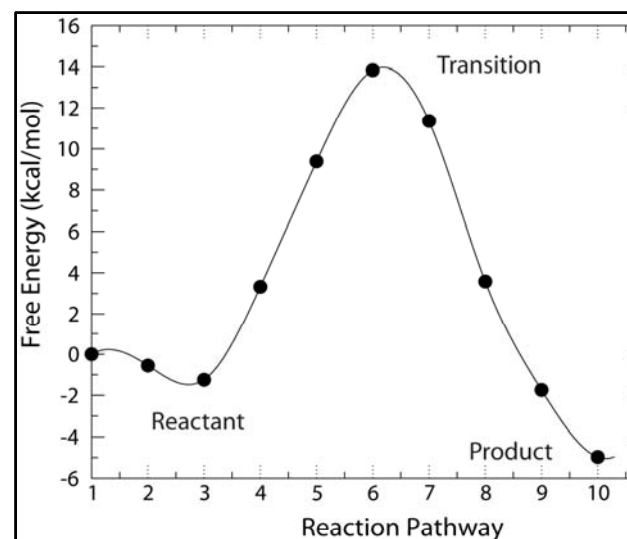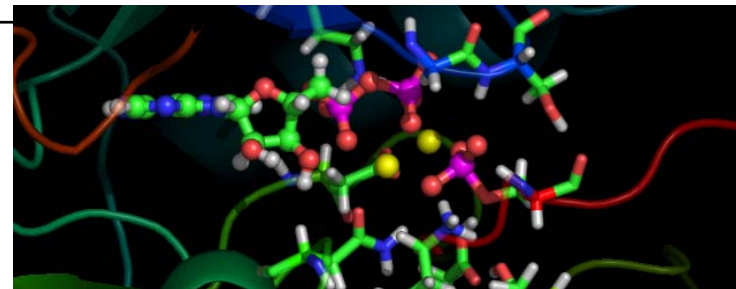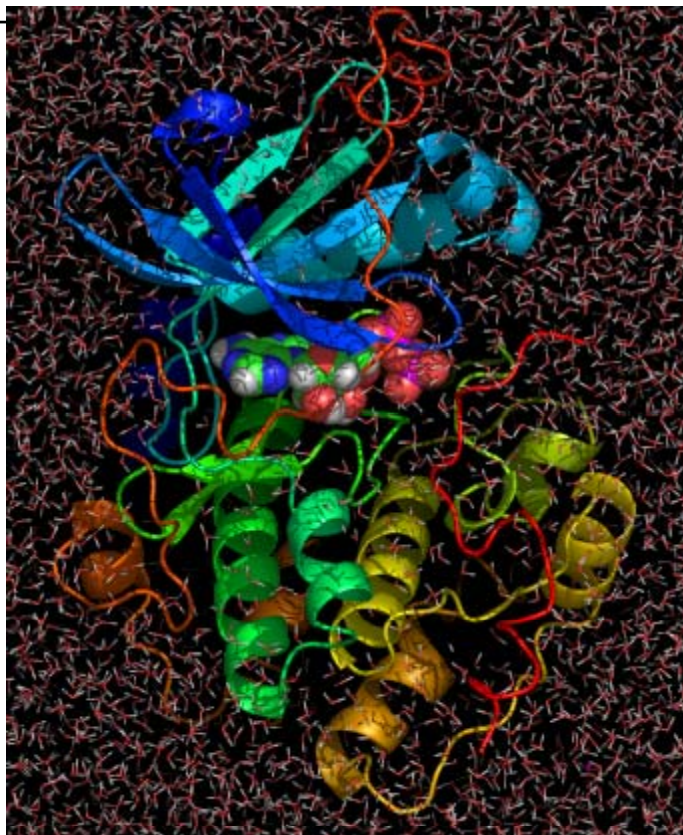$$CCl_4 + e^- \rightarrow CCl_3^{\bullet} + Cl^-$$
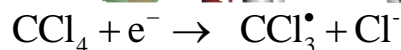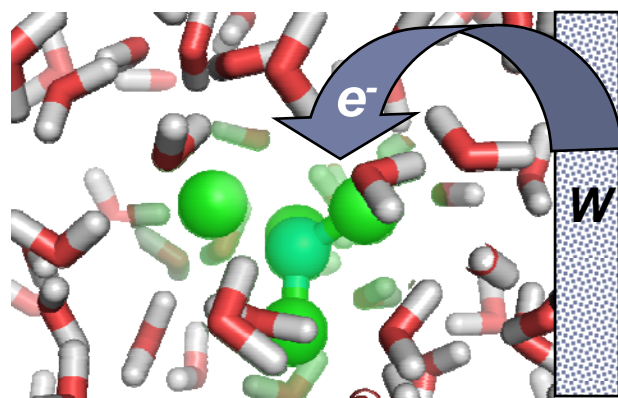


- ‣ Concerted electron transfer bond breaking mechanism
- ‣ Free energy calculation of dissociation profiles and redox potential using CCSD(T)/MM and DFT/MM descriptions
- ‣ CCSD(T) are in good agreement with experimental data for oxide-free iron cathodes
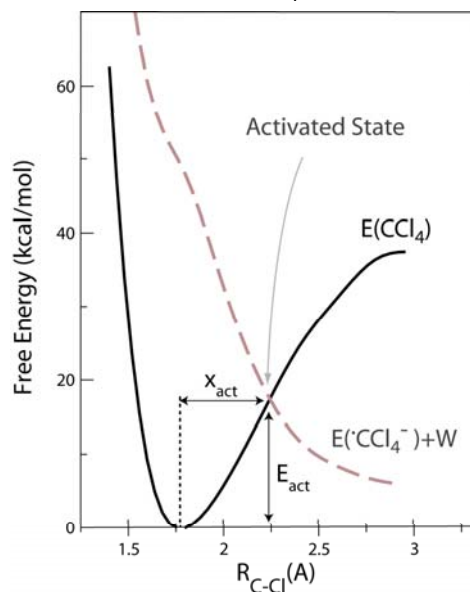
*Free Energy Activation Barriers at W=-1.2V*

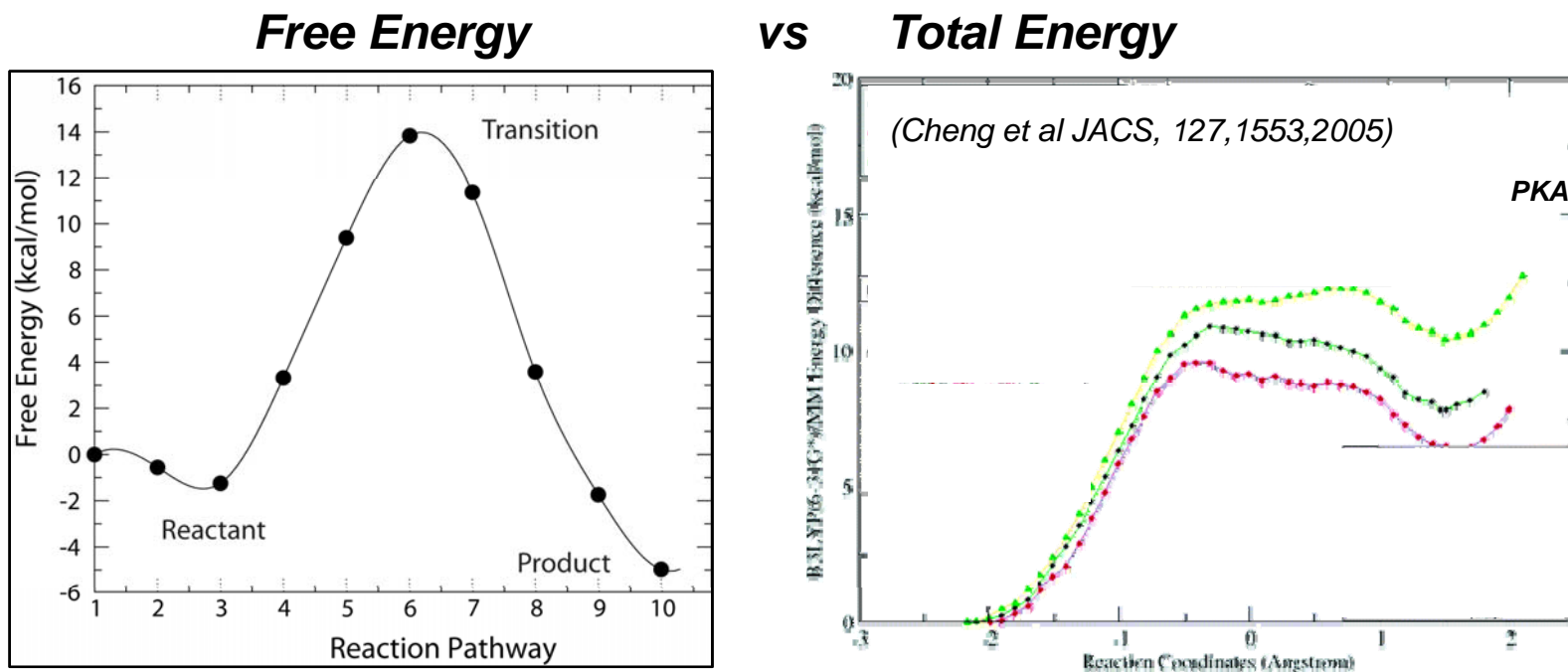|  | CCSD(T)/MM | DFT/MM | Exp |
|---|---|---|---|
| $E_{act}$ | 8.1 kcal/mol | 4.3 kcal/mol | 8.0 kcal/mol |

- *Catalyze the transfer of the γ-phosphoryl group of ATP
- *Determination of the reaction pathway using NEB QM/MM approach
- *Free energy reaction profile at DFT/MM level (75 QM atoms)

# Free Energy is Important!

### Free Energy  vs  Total Energy



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

PKA

- *Free energy calculations provide a definite improvement over total energy methods*
- *Free reaction barrier of 15 kcal/mol versus experiment 12.8 kcal/mol*
- *We observe exothermic reaction (3 kcal/mol) in good agreement with experiment estimates (4-5 kcal/mol)*

# Excited States of Liquid Water

▸ Optimized at DFT/MM level
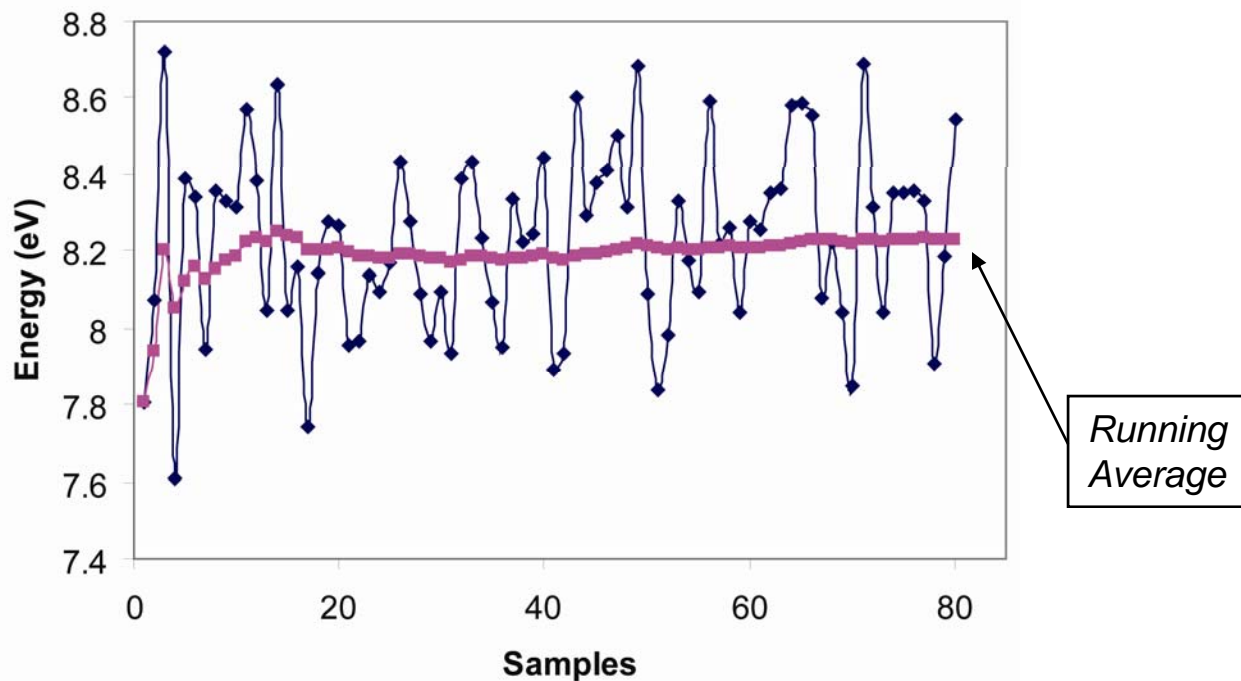
▸ Significant deviations between TDDFT and CCSD results

|  | TDDFT | CCSD |
|---|---|---|
| Gas | 6.85 | 7.46 |
|  | 8.24 | 9.22 |
| Solvated | 7.99 | 8.67 |
|  | 9.06 | 10.49 |

▸ Experimental number for 1st excited state 8.2 eV
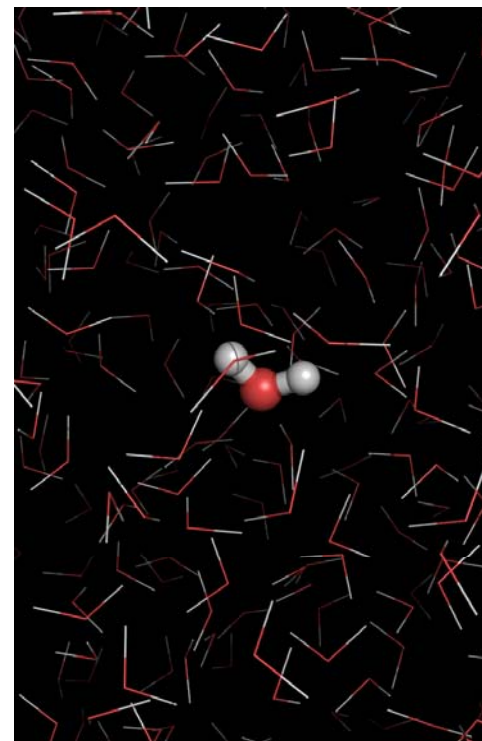


*QM water embedded In 30A box of classical SPC/E waters*

# Thermal Averaging



- 40 ps fully classical SPC/E water dynamical simulation
- Resampled at 0.5 ps rate with CCSD/MM excited state calculations
- The average excitation energy is 8.2 eV in excellent agreement with experiment

# Example Application: DNA photostability

The irradiation of DNA by UV light causes an electronic excitation of its base chromophores.

Depending on the nature of the excitation this can lead to DNA damage.

We have calculated finite temperature averages of excited state of cytosine base in DNA.

Our results indicate that protein environment plays a protective role by stabilizing "good" $\pi\pi^*$ excited state



Cytosine
CR-CCSD(T)

12-mer DNA fragment
5760 water molecules
Amber force field
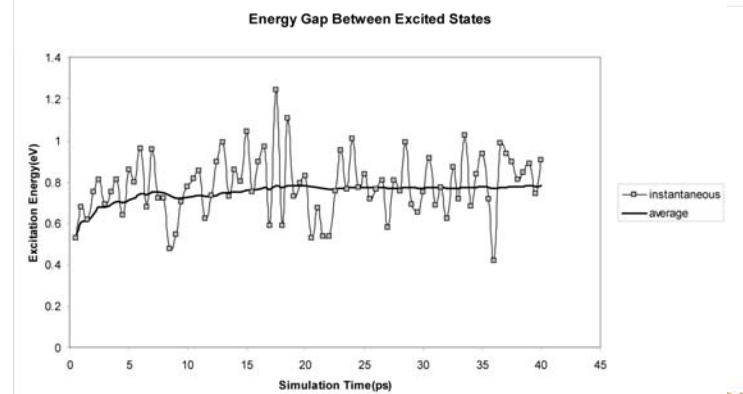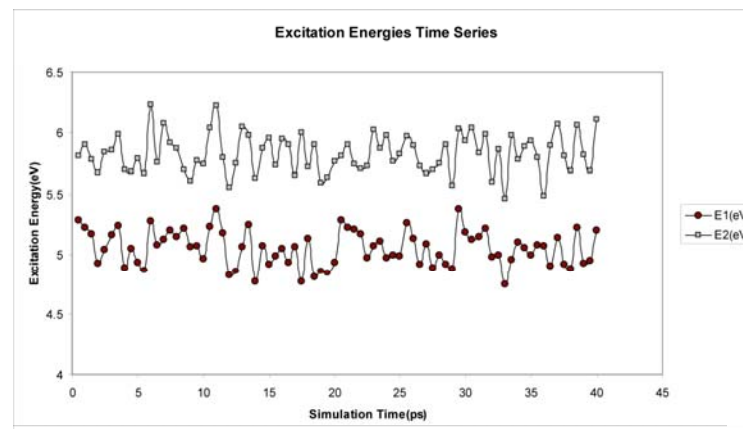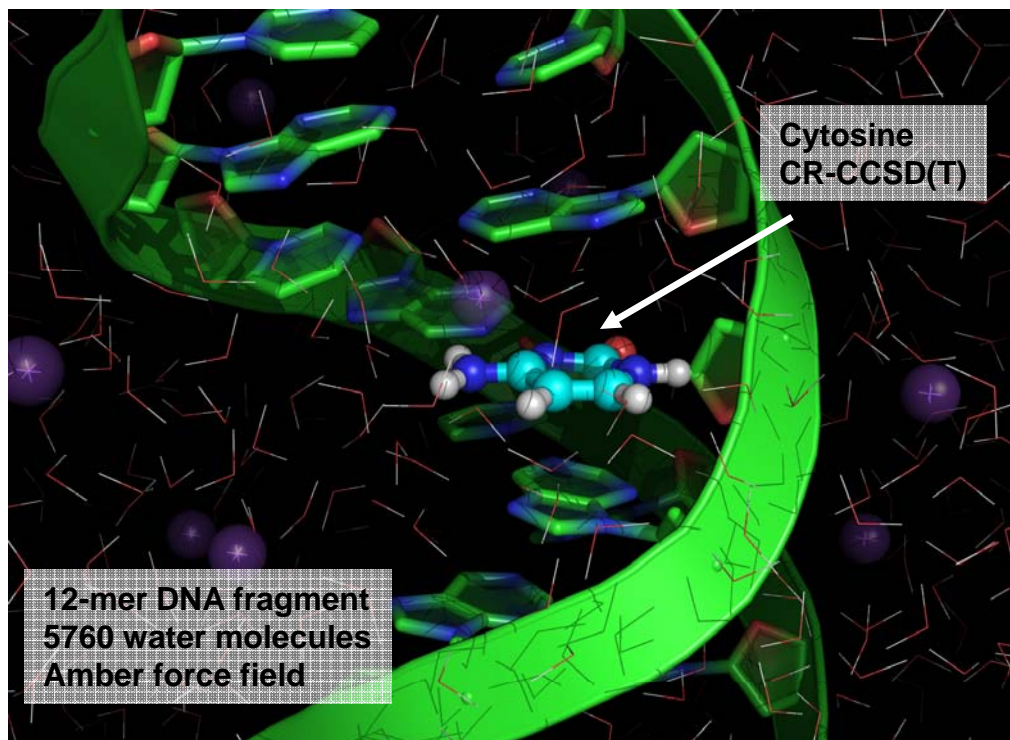
# Summary of results

- Integrated framework for the analysis of chemical processes in condensed phase
  - Accurate chemistry up to coupled cluster level of accuracy within QM/MM framework
  - Efficient analysis of potential energy surface (critical points, reaction pathways)
  - Efficient statistics (free energy activation barriers, thermal averages)
  - Excellent parallel scalability
- Production level code implemented in NWChem
  - Free source code availability
  - Parallel executables for major computing platforms

- Moving beyond QM/MM
  - More generalized hybrid models (e.g. DFT/CC/MM)
  - Fragment based descriptions
  - Dynamical QM regions
- Discovery-type approaches to locating possible reaction channels
- Reaction pathways on PMF
- Excited States Processes
  - Reaction pathways
  - Free energies
  - Non-adiabatic dynamics
- **Peta-scale computer architectures**